768

A STUDY OF TRACKING PERFORMANCE DURING AUDITORY DISTRACTION

by

ARLEN AARON MICHAELS

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

Department of Psychology
University of Stirling
January 1987

Stirling Scotland

5/87

ABSTRACT

The distraction of attention by irrelevant information is normally considered to have detrimental effects on human performance, but published accounts have provided only limited evidence. This study examined the effects of auditory distraction on the execution of a visual compensatory tracking task. It was hoped that a sensorimotor task demanding much information-processing capacity would create appropriate conditions for the detection of distraction effects.

In order to establish a suitable tracking requirement, an adaptive task was designed. Its purpose was to impose a heavy mental workload without the task becoming uncontrollable. To achieve this, the tracking forcing function bandwidth was adjusted by computer throughout the task in response to measured changes in the operator's error. Observation confirmed that the procedure provided a substantial workload throughout practice. The adaptive task was used in subsequent experiments as a training aid and as a method for determining the optimum tracking load for each subject.

Three experiments compared tracking performance during auditory distraction to performance in nondistracted conditions. Random presentations of moderate intensity (82 dBA) bursts of white noise resulted in increased RMS error and increased tracking response delay during the few seconds following noise onset. The effects were highly localised and were very shortlived, surviving only a few exposures. The magnitude of effect was also influenced by the neighbouring forcing function dynamics. Distraction effects were greatly reduced if noise presentation followed a predictable periodic schedule. When randomly presented verbal messages were used as distractors instead of noise, the increase in tracking error was not so localised and subjects did not adapt to distraction so quickly.

Because periodicity had so noticeably promoted adaptation a final experiment investigated the formation of temporal expectancy during tracking. Subjects were asked to predict when intermittent auditory signals would occur. It was found that subjects anticipated temporal events readily, probably using a rational strategy, and that they formed expectancies in very similar fashion whether or not they were tracking. The relative insensitivity of expectancy formation to a simultaneous tracking load (and vice versa) was in accord with the broad suggestion that temporal expectancy could participate in the adaptation process.

This study corroborated reports that distraction effects are generally variable, task-dependent, and of short duration. It is proposed that auditory distractors affect tracking performance in a direct way by commanding selection and analysis, and sometimes indirectly by altering arousal. Even so, tracking is probably not a very susceptible task. The explanations for this suggest that tracking tolerates discontinuities in information sampling and in motor processing rather better than has often been supposed, and that this-- in combination with the often very brief timescale of the distraction effects-- affords tracking some immunity from distraction. Also, it seems likely that expectancies of various sorts, including temporal expectancy, mediate the effects of distraction. Following this interpretation, distractibility should be seen as a contributor to normal attentional functioning, because it can provide information to assist in the formation of predictive internal models to guide selection.

Passengers must not speak to the driver or distract his attention without good cause while the vehicle is moving.

- public notice displayed in Scottish buses (1977)

Noise is the most impertinent of all forms of interruption. It is not only an interruption, but also a disruption of thought. Of course, when there is nothing to interrupt, noise will not be so particularly painful. Occasionally it happens that some slight but constant noise continues to bother and distract me for a time before I become distinctly conscious of it. All I feel is a steady increase in the labor of thinking-- just as though I were trying to walk with a weight on my foot.

- Arthur Schopenhauer, *Parerga und Paralipomena* (1851)

I said it very loud and clear;
I went and shouted in his ear.
But he was very stiff and proud;
He said "You needn't shout so loud!"

- Lewis Carroll, *Through the Looking Glass* (1872).

ACKNOWLEDGEMENTS

This work was largely supported by a Postgraduate Scholarship awarded by the National Research Council of Canada.

Neville Moray has been my teacher, colleague, and friend for many years. Neville introduced me to human factors psychology and inspired this research. I thank him for his supervision, for his enthusiasm, and for the intellectual freedom he encouraged.

Bill Phillips viewed the work from the perspective of a cognitive psychologist and his thoughtful suggestions were invariably helpful. I am very grateful to him for his constant support and guidance.

I also thank Ranald Macdonald, who patiently read many early drafts, provided invaluable advice on statistical approaches, and skilfully steered me through many difficulties.

The computer instrumentation used in the experiments required much engineering support and I am grateful to many members of the Psychology Department's technical staff. In particular I am indebted to Angus Annan. I would also like to single out Allan Hutton, Bob Lavery, Jim Nimmo, John Russell, and Bruce Sutherland for my appreciation.

Thanks also to Tony Hewitt, director of the university's computer unit, who gave his support at crucial times. The computer-generated figures reproduced throughout the thesis were nearly all drawn using a program developed by Richard Bland.

Finally, my deepest thanks go to my wife Jan and my children Sean and Robin. Without their love and help and patience I would not have been able to carry out this work. Jan assisted in many, many ways. The children, of course, taught me a good deal about distractibility. Jan and the children have really been partners in the endeavour. To them I dedicate this thesis.

CONTENTS

Chapter 1 The investigation of distractibility 1
Introduction 2
Early distractibility research 5
Auditory distraction during discrete-response sensorimotor tasks 10
Distractibility, capacity, and mental workload 15
Tracking 17
Auditory distractibility during tracking 21
The special problem of noise stress 27
The present research 31

Chapter 2 The adaptive tracking task 33
Introduction 34
Adaptive tracking task: General Procedure 47
Tracking Experiment 1 51
Tracking Experiment 2 58
Tracking Experiment 3 74
Adaptive tracking: Conclusions 91

Chapter 3 Effects of auditory distraction on tracking 93
Introduction 94
Distraction Experiment 1 99
Distraction Experiment 2 138
Distraction Experiment 3 156
Auditory distraction: Conclusions 179

Chapter 4 Temporal expectancy during tracking 182
Introduction 182
Expectancy Experiment 189
Temporal expectancy: Conclusions 235

Chapter 5 Final discussion 237
The attentional demands of auditory distraction 239
The attentional demands of tracking 250
The problem of distractibility measurement 256
Distractibility, internal models, and attention 265

Appendix A Questionnaire used in Expectancy Experiment 271

Appendix B Toward a model for temporal expectancy 272

References 287

LIST OF FIGURES

CHAPTER 1

1-1 Compensatory tracking as a single-loop manual control system. (p19)

CHAPTER 2

2-1 Adaptive algorithm for the tracking task. (p39)

2-2 Characteristics of tracking performance. (p40)

2-3 Calculation of the bandwidth increment in the adaptive task. (p45)

2-4 Zero-order compensatory tracking system. (p48)

2-5 Tracking Experiment 1. Effect of practice on the forcing function bandwidth. (p53)

2-6 Tracking Experiment 1. Effect of practice on the normalised RMS error. (p56)

2-7 Tracking Experiment 2. The subjective difficulty rating scale. (p60)

2-8 Tracking Experiment 2. Forcing function bandwidth. (p61)

2-9 Tracking Experiment 2. Tracking normalised RMS error. (p63)

2-10 Tracking Experiment 2. Subjective difficulty ratings. (p69)

2-11 Tracking Experiment 3. Forcing function bandwidth. (p78)

2-12 Tracking Experiment 3. Tracking normalised RMS error. (p79)

2-13 Tracking Experiment 3. Subjective difficulty ratings. (p83)

2-14 Tracking Experiment 3. Tap level in the secondary task. (p86)

CHAPTER 3

3-1 Distraction Experiment 1 (Random noise). Intersignal intervals. (p102)

3-2 Distraction Experiment 1 (Random noise). Tracking performance during the experiment. (p106)

3-3 Distraction Experiment 1 (Random noise). Tracking RMS Error: Distraction compared to no distraction. (p112)

3-4 Distraction Experiment 1 (Random noise). Difference between Distracted and Nondistracted RMS Error: stability within the trial. (p116)

3-5 Distraction Experiment 1 (Random noise). RMS Error during the trial: effect of increasing exposure to distraction. (p118)

3-6 Distraction Experiment 1 (Random noise). RMS Error during the trial: dependency on track characteristics during distraction. (p119)

3-7 Distraction Experiment 1 (Random noise). Tracking records for Subjects CM and AI, showing Distracted and Nondistracted joystick response. (pp121,122)

3-8 Distraction Experiment 1 (Random noise). Joystick movement size: First four trials only. (p124)

3-9 Distraction Experiment 1 (Random noise). Tracking response delay: Distraction compared to no distraction. (p129)

3-10 Distraction Experiment 1 (Random noise). Difference between Distracted and Nondistracted response delay: stability within the trial. (p132)

3-11 Distraction Experiment 1 (Random noise). Response delay during the trial: effect of increasing exposure to distraction. (p133)

3-12 Distraction Experiment 2 (Periodic noise). Tracking performance during the experiment. (p140)

3-13 Distraction Experiment 2 (Periodic noise). Tracking RMS Error: Distraction compared to no distraction. (p142)

3-14 Distraction Experiment 2 (Periodic noise). Tracking RMS Error: First six trials only. (p143)

3-15 Distraction Experiment 2 (Periodic noise). Joystick movement size: First four trials only. (p147)

3-16 Distraction Experiment 2 (Periodic noise). Tracking response delay: Distraction compared to no distraction. (p149)

3-17 Distraction Experiment 3 (Verbal distractors). Intersignal intervals. (p158)

3-18 Distraction Experiment 3 (Verbal distractors). Tracking performance during the experiment. (p161)

3-19 Distraction Experiment 3 (Verbal distractors). Tracking RMS Error: Distraction compared to no distraction. (p165)

3-20 Distraction Experiment 3 (Verbal distractors). Distracted and Nondistracted RMS Error: dependency on track characteristics. (p168)

3-21 Distraction Experiment 3 (Verbal distractors). Tracking response delay: Distraction compared to no distraction. (p170)

CHAPTER 4

4-1 Expectancy Experiment. Intersignal intervals. (p193)

4-2 Expectancy Experiment. Comparison of tracking performance (with prediction task) to performance in Tracking Experiments 1 and 2 (tracking alone). (p197)

4-3 Expectancy Experiment. Distributions of prediction intervals. (p200)

4-4 Expectancy Experiment. "Aligned" distributions of prediction intervals. (p202)

4-5 Expectancy Experiment. Distributions of prediction intervals with correction for trial elimination effect. (p208)

4-6 Expectancy Experiment. Autocorrelation functions of prediction intervals. (p217)

4-7 Expectancy Experiment. Crosscorrelation of prediction interval with previous intersignal intervals. (p220)

4-8 Expectancy Experiment. Distributions of reaction responses. (p224)

CHAPTER 5

5-1 Functions serving attentional selection. (p270)

APPENDIX B

B-1 Expectancy Experiment. Correlation functions showing optimum exponential decay model for expectancy. (p283)

LIST OF TABLES

CHAPTER 2

2-1 Tracking Experiment 2. Effect of practice on the RMS error. (p65)

2-2 Tracking Experiment 2. ANOVA summary: effects of bandwidth and practice on the average normalised RMS error. (p66)

2-3 Tracking Experiment 2. ANOVA summary: effects of bandwidth and practice on the subjective difficulty rating. (p70;

2-4 Tracking Experiment 3. ANOVA summary: effects of bandwidth and practice on the average normalised RMS error. (p81)

2-5 Tracking Experiment 3. ANOVA summary: effects of bandwidth and practice on the subjective difficulty rating. (p84)

2-6 Tracking Experiment 3. ANOVA summary: effects of bandwidth and practice on the mean tap level. (p88)

2-7 Tracking Experiment 3. ANOVA summary: effects of bandwidth (Adaptive and Matched only) and practice on the mean tap level. p(89)

CHAPTER 3

3-1 Distraction Experiment 1 (Random noise). ANOVA summary: effects on tracking RMS Error. (p113)

3-2 Distraction Experiment 1 (Random noise). ANOVA summary: effects on tracking RMS Error, first six trials only. (p115)

3-3 Distraction Experiment 1 (Random noise). ANOVA summary: effects on joystick movement size, first four trials only. (p125)

3-4 Distraction Experiment 1 (Random noise). ANOVA summary: effects on tracking response delay, first eight trials only. (p130)

3-5 Distraction Experiment 2 (Periodic noise). ANOVA summary: effects on tracking RMS Error, first six trials only. (p146)

3-6 Distraction Experiment 2 (Periodic noise). ANOVA summary: effects on tracking response delay, first twelve trials only. (p150)

3-7 Distraction Experiment 3 (Verbal distraction). Texts of messages used as distractors. (160)

3-8 Distraction Experiment 3 (Verbal distraction). ANOVA summary: effects on tracking RMS Error. (p166)

3-9 Distraction Experiment 3 (Verbal distraction). ANOVA summary: effects on tracking response delay. (p172)

CHAPTER 4

4-1 Expectancy Experiment. Effect of ISI distribution on tracking. (p198)

4-2 Expectancy Experiment. Prediction interval characteristics. (p204)

4-3 Expectancy Experiment. Mode of the prediction interval distribution. (p211)

4-4 Expectancy Experiment. Autocorrelation of prediction intervals, showing Lag 1 only. Results for individual subjects. (p218)

4-5 Expectancy Experiment. Crosscorrelation of prediction interval with previous intersignal intervals: comparison of gaussian and uniform results. (p222)

4-6 Expectancy Experiment. Reaction time characteristics. (p225)

4-7 Expectancy Experiment. Correlation of reaction time with concurrent intersignal interval. (p227)

4-8 Expectancy Experiment. Results of questionnaire (partial). (p229)

APPENDIX B

B-1 Expectancy Experiment. ANOVA Summary. Effects of loading and weighted-average size on the optimum correlation of prediction with $I_{Nweighted}$. (p285)

B-2 Expectancy Experiment. Values of c, the decay constant, to yield optimum correlation of prediction with $I_{Nweighted}$. (p286)

CHAPTER 1: THE INVESTIGATION OF DISTRACTIBILITY

Overview

There has been comparatively little research describing true distraction effects on information processing skills. Traditionally, distractibility has been seen as an involuntary diversion of attention, often resulting in errors or other deficiencies in performance. A review of the effects of (mainly auditory) distraction on (mainly visuomotor) tasks revealed that while some studies identified distraction effects, many investigations could not. When there was a distraction effect, it typically impaired performance, but there have also been exceptional observations of performance enhancement. Effects have tended to diminish rapidly during prolonged exposure to distractors. Loud noise has been widely used as a distractor in experimental studies, but it may not be ideal because it can have other stressful effects. The tasks most likely to be affected by distraction seemed to require continuous information processing of a complex sort, especially with some form of time pressure. Tasks demanding much processing capacity, such as tracking tasks, would appear to be ideally suited to the investigation of distraction effects.

Introduction

The fruitful renaissance in attention research which began in the 1950s was fueled by at least three events: a war, which had exposed the significance of human factors in complex systems; the development of cybernetics and information theory, which offered powerful new ideas to experimental psychology; and the invention of the tape recorder, which provided the tool necessary for testing many of the theories about attention. Most of the research concentrated on the processes of selective attention and the dramatic failures of selection, including the lapses found in vigilance performance. While a large body of theory and experiment accumulated, well-summarised in reviews by Kahneman (1973) and Moray (1969), the work generally concentrated on the phenomena of *voluntary* selection and tended to pass over the *involuntary* aspects of attentional functioning. It was probably enthusiasm for the information-processing approach which caused the neglect of involuntary attention processes such as distractibility. Information-processing interpretations of performance stressed the active, strategic capabilities of cognitive functions, and during nearly three decades of experiment there was little interest shown in the seemingly nonstrategic, unintentional, and unwanted diversions of processing resources caused by distraction.

Despite this neglect, it is increasingly obvious that distractibility is an important phenomenon which warrants serious investigation. It clearly contributes to the ever-growing problems associated with the human supervisory control of complex technological systems such as nuclear power plants, especially by interfering with the rapid diagnosis of system faults (well represented in Rasmussen and Rouse, 1981). Indeed, the American Federal Aviation Administration has become so concerned about the effects of pilot distractibility on air safety that a "sterile cockpit" regulation is now being enforced which limits nonessential conversation and activity on the flight deck of a commercial aircraft during critical stages of flight (Wiener, 1985). With so little known about distractibility, it is very difficult to take more specific countermeasures. Neither the effects of distractibility nor the situations susceptible to it have been convincingly established.

Distractors are most appropriately defined as irrelevant extraneous events which attract attention to themselves and away from the observer's chosen task. It is important that the distractors achieve their effect by redirecting attention: events should not be described as distractors if they directly impede the ongoing task. In most experiments, this requirement has been partially solved by directing the task and the distractors to different sense modalities. Furthermore, if the concept of distraction is to be useful it must contain the implication of task-irrelevance; that is, true distractors must not appear to bring anything relevant to the task environment such that they would actually warrant having attention allocated to them.

In practice, these have been difficult conditions to fulfil. Within the enormous literature on selective attention, timesharing, and the limitations of processing capacity there is a substantial body of research describing dual-task performance. Nearly all of this research is unhelpful in seeking true distraction effects because in dual-task experiments neither task is in any sense extraneous to the subject's dedicated attention. Nonetheless, interference from a secondary task is sometimes described as "distraction" (see, for example, Kalsbeek, 1964) even though the subject would be expected to attend to both primary and secondary demands. Other studies supposedly imposing distraction have compromised their validity by interfering excessively with the subject's task, as in an experiment by Howell and Briggs (1959) in which the visual "distraction" added to a tracking display was the random movement of some of the display elements-- thereby directly interfering with the subject's primary source of information. A more subtle misjudgement created a similarly doubtful interpretation of distraction in a visual search experiment by Holahan, Culler, and Wilcox (1978). In this account, nontarget elements that had been added to the search display were described as "distractors"; but it is certainly arguable that these extra stimuli transformed the search conditions into a completely different task. The last two examples illustrate the difficulty of imposing irrelevant distraction in a legitimately noninvasive way using the same sense modality as the imperative task. Johnson and Cole (1976) did attempt to accomplish this visually by projecting transparencies of interesting but task-irrelevant images onto the same display screen as the stimuli for a choice reaction task, but even here there were acknowledged difficulties controlling interactions of image luminance contrast, so the method proved problematic and far from ideal. For reasons such as these, many studies concerned ostensibly with distraction effects must be interpreted with considerable care.

Early distractibility research

The investigation of distractibility has not always been neglected. Early in this century distractibility was examined alongside the other perplexing, but recognisably important, properties of mental life. The concept had undoubted intuitive appeal and was taken for granted as an inescapable factor in perception and learning (James, 1890). It was assumed that distraction interfered with attentiveness and impaired performance. Experiments were carried out to identify the most distracting features of stimuli and to measure their effects, particularly on higher intellectual functions. Woodworth and Schlosberg (1954) have summarised the early research into the fluctuations of attention. To begin with, variations in attending were found to be affected by:

-absolute stimulus factors, such as brightness;

-relative stimulus factors such as colour and size;

-the interest inherent in the stimulus object, which influences the observer's "satiation" (a somewhat recursive construct representing the observer's inclination to dwell on any one stimulus)

This was casting a very wide net indeed. Moreover, even the earliest investigators recognised that distractors produced "paradoxical" outcomes (Myers, 1909, p.323). Equivocal effects on performance were associated with such within-subject variability and effects tended to disappear early in the experiment (for example: Cassel and Dallenbach, 1918; Harmon, 1933; Pollock and Bartlett, 1932). Reviewing the early work, Woodworth and Schlosberg (1954) bravely tried to list the experimentally demonstrated effects of distraction. They concluded that attempts to measure adverse effects of distraction on higher mental processes (such as reasoning and problem-solving) exposed effects only rarely, and even then impairment would be slight. They suggested that motor performance might exhibit increased muscular effort, as demonstrated by pressing a response key more vigorously, when distraction stress occurred. It was generally accepted that a distractor would exert its effect only for a brief time. Substantial variability in the effect of a given distractor could be expected among subjects and even within the same task by a single subject. When a period of noise acted as a distractor, the offset of the noise was often found to serve as effectively as a distractor as the onset; again the effect was temporary. Woodworth and Schlosberg agreed that there were clearly profound *adaptation* processes which counteracted the continuing influence of distraction, so that effects would diminish rapidly with repeated exposure. In addition to this adaptation (sometimes referred to as *habituation*) it seemed that the subject's experience, motivation, and expectation could be manipulated to cause differences in the distraction effect, and it was suspected that emotional responses (anger at being distracted, for example) would also participate in the final outcome.

In the relatively small number of experimental studies of distractibility, all of these themes have recurred. That is, the effects of distractors are difficult to detect, variable, and often exhibit adaptation. Distraction experiments which have shown effects have usually identified short-lived performance decrements. But there have also been puzzling improvements in performance during distraction: for example, Kryter (1950) described the results of a wartime study which suggested that artillery aiming skill improved in very noisy conditions; more recent examples of performance enhancement have appeared in a visually distracted selection task by Johnson and Cole (1976), and in a visual tracking task conducted by Gawron (1982) in the presence of moderate noise. Throughout the history of the topic, some investigators have suspected that when distractors are present, performance deficits may alternate with cycles of compensatory improvement, as if the victims of distraction redouble their efforts following impairment (Cassel and Dallenbach, 1918; Fisher, 1984b; Harmon, 1933; Kryter, 1970; Morgan, 1916).

To some extent, the problem with the earliest studies was that they tried to understand the impact of distraction on complex cognitive processes which are themselves imperfectly understood, such as intellectual activity. These efforts did not often succeed even when the experimenters went to great lengths. The flavour of these vigorous investigations is well captured in the distraction study described by Hovey (1928). Hovey asked college students to complete the American Army Alpha intelligence test under conditions of elaborate distraction. The test took about twenty minutes. During this time the subjects endured the intermittent distraction created by seven different electric bells ringing in different locations, five buzzers, two organ pipes, three whistles, a large circular metal saw that was struck with wooden and iron beaters, a crackling electric spark gap, a phonograph playing music, and a powerful spotlight flashing about the room (but not directly into the subjects' eyes). This astonishing commotion was enhanced by the experimenter's four accomplices who marched about wearing peculiar clothes, rolled kegs of nails past the subjects, and manipulated strange equipment; meanwhile a famous photographer took pictures. The extraordinary conditions were fatiguing and stressful to the subjects (even sufficient, according to Hovey, "to cause crying spells afterwards, in the cases of two girls" p.590) but seemingly had little effect on performance. Compared to a previously-matched control group who worked in undisturbed conditions, the distracted group suffered what Hovey judged to be a negligible decrement (it was less than 3%) in their test scores and he concluded that distraction was not a serious handicap in such a task. Interestingly, rather similar results were reported some years later in a more conventional experiment by K R Smith (1951). Smith used the Minnesota mental test battery and a much less theatrical form of distraction, namely bursts of very loud (100 dB) white noise. He found very slightly diminished accuracy on only one of the three subtests, and concluded that the effects of distraction were of no practical significance in this task. Harmon (1933) conducted a careful long-term examination of the effects of noise on a mental arithmetic task (adding three-digit numbers) and, simultaneously, on metabolic rate (oxygen absorption and heart rate). The distractors were recordings of either office sounds (50-65 dB) or street noises (65-75 dB). Although Harmon observed minor changes in the number of arithmetic problems solved and the accuracy of the solutions, the changes were compromised by much day-to-day variability. He concluded that the noises had no reliable effect on performance in the arithmetic task, although he did propose that the distraction had, for a time, prompted an increase in oxygen absorption and pulse rate. Harmon's well-controlled study was one of the first to document convincingly the high level of within-subject variability and Harmon stressed the urgent need to "fractionate" the data in order to discern adaptation effects.

Experiments such as these leave the general impression that global effects of distraction, particularly on intellectual functions, cannot be measured easily. Even when subjects are consciously irritated by the distractors they seem to be able to maintain adequate performance levels overall, perhaps by compensating for the effects during periods of non-distraction. Certainly subjective feelings of discomfort are not associated in any clear way with performance deficits: the subjective annoyance of noise is increased by making the sounds unexpected, intermittent, or loud but those noises which produce the greatest irritation are not necessarily the most distracting as measured by performance (Kryter, 1950; Pollock and Bartlett, 1932; Plutchik, 1959).

If gross or high-level effects of distraction are difficult to detect, a more satisfactory approach would be to employ a small number of well-defined distracting stimuli and undertake to detect brief changes in performance just following the distractor. However, the requirement for a relatively small number of measurements (to prevent adaptation) of what are perhaps very shortlived effects creates real statistical and methodological problems. As Moray (1976, 1979b) has pointed out, most measures of performance are long-term time averages which are not sensitive to short-term transients. Moreover, the need to sustain the irrelevance of the distractors and to avoid direct obstruction of the task imposes additional constraints on the sorts of experiments which are possible. In the event, most investigators of distractibility have resorted to studies of sensorimotor tasks. The tasks have primarily depended on vision, with auditory distractors to preclude directly invasive (and therefore inadmissible) effects such as visual masking. The auditory distractors have nearly always used broadband white noise, presumably because it largely bypasses the more variable properties of subjects' hearing ability (Miller, 1948) while upholding the principle of distractor irrelevance.

Auditory distraction during discrete-reponse sensorimotor tasks

Even very early descriptions of distractibility, citing experiments such as those by Cassel and Dallenbach (1918), suggested that the speed of simple reaction time was not affected in any systematic way by the presence of irrelevant background noise. Consequently, effects were sought in more complex motor skills, such as the placing and removal of small mechanical parts (Pollock and Bartlett, 1932). In these early experiments it was confirmed that when noises were presented as distractors, complicated motor tasks were more likely to suffer some loss of efficiency (demonstrated, for example, by fewer or slower responses) than were simple repetitive tasks. The appropriately demanding tasks were deemed to be those requiring strategies, rapid decisions under time pressure, and relatively unpatterned movements unlikely to become automatised. Even so, the impairment was likely to affect only some subjects, mainly during the earliest exposures, and would probably be very slight. Indeed, the performance of some subjects seemed actually to benefit from the presence of moderate background noise and the objective evidence of this agreed with their introspective reports.

Taking the suggestion of task complexity still further, Boggs and Simon (1968) asked subjects to carry out two tasks at the same time, sometimes exposing them to randomly scheduled half-second bursts of loud (92 dB) bandsaw noise. The primary task was a four-choice visual reaction time test. As the secondary task, the subjects listened to a series of spoken digits and called out whenever they detected an odd-even-odd digit sequence. The noisebursts were carefully timed not to coincide with the digit message. The subjects performed the primary choice task well (making no errors during the test conditions) and the noises did not have any effect on the average reaction time. What did suffer was the subsidiary digit detection performance, which contained more errors during noisy conditions. This deterioration in the secondary task performance conforms to the normal pattern in dual-task studies; namely, the low-priority task tends to suffer most when processing resources are overstretched (Rolfe, 1973). Notably, Boggs and Simon found that the errors in the secondary task were even more numerous if the primary choice task was made more difficult by introducing spatial incompatibility between the stimulus lights and their response keys. This observation certainly suggested that distraction effects were likely to be more severe during more demanding tasks. While Boggs and Simon had reported that the overall reaction time was not significantly affected by the noisy conditions, they did not comment on the effects of the noisebursts on individual reactions (such as the responses which immediately followed noise onset). The usefulness of a closer analysis of distraction effects became more apparent later.

A very tight coupling of distractor presentation with distractor effect was illustrated clearly in an experiment by Fisher (1972). She presented irrelevant two-second bursts of 80 dB white noise during a serial response task and looked for extremely localised effects. In this task, an array of five signal lamps lit in random sequence. Whenever a lamp turned on, the subject had to strike its response contact as quickly as possible, whereupon the next lamp in the sequence would appear. The presentation of a signal was triggered by the completion of the response to the previous signal. The response latencies and any response errors were recorded. The twelve subjects served as their own controls, with a counterbalanced design that gave the task to each subject under both quiet and noisy conditions.

Looking first at overall performance, Fisher was unable to confirm any reliable change in the response latency mean or variance or the response errors under noisy conditions. (This bears a similarity to the absence of a general reaction time effect in the Boggs and Simon (1968) report.) She next compared the data from those responses which occurred while noisebursts were in progress to the results recorded during the silences between bursts, but again none of the differences were significant. Fisher discovered that she had to analyse the distractor trials one by one. She found that the first response immediately following a noiseburst, but only the first response, sometimes suffered from a prolonged latency. The effect followed noise onset but not offset (but Fisher noted that since the bursts were of constant duration the offset was implicitly predictable). Thus the duration of distractor influence appeared to be quite short in this experiment, certainly lasting no longer than one second in each instance. Furthermore, not all distractors were effective. This was shown by the response latency distribution. While the average duration of the first post-burst latency was certainly significantly longer, examination of the histogram of these responses suggested that only a proportion of such responses had actually been affected. The effect looked as if it resulted from additions to the long latency tail of the response distribution, not from a general shift of the distribution toward longer latencies.

The close comparison of pre-burst versus post-burst response latencies was the only rewarding analysis Fisher could identify. The necessity for such fine-grained analysis of response delays was underlined by

Conrad's (1973) failure to detect any effects of auditory distraction on performance in another serial response task. Conrad used louder noise (93 dBA) and tried periodic and aperiodic presentations of two-second bursts as well as continuous noise. The task was to press a button defined by a previously-displayed digit code, and occasionally the code had to be remembered from the previous trial so some short-term memory load was imposed. Before the experiment began, the subjects answered a questionnaire designed to rate their susceptibility to distraction, and they were sorted into low- and high-sensitivity groups. Conrad took the trouble to record several psychophysiological variables, but his behavioural measures did not include response latencies, only selection errors. Although he noticed some significant changes in blood volume, Conrad did not find any effects on the error scores under any conditions of noise or rated sensitivity.

The effects noted by Fisher (1972) proved to be very task-sensitive. In 1973, Fisher reported further results of serial response experiments in which manipulations of stimulus predictability and stimulus-response compatibility altered the magnitude of the distraction effect. One result was completely contrary to expectation. By attempting to make the task more difficult by adding spatial disparity between the signal lamps and their response contacts (as Boggs and Simon, 1968, had done), the distraction effect of the noise was reduced (which was opposite to the Boggs and Simon outcome and the usual theoretical prediction). Fisher (1973) could not easily account for this loss of effect and put forward a tentative arousal/effort hypothesis to explain the anomaly. This sensitivity to task parameters in a carefully conducted study echoes the historical warning that distraction effects may be extremely variable.

Distractibility, capacity, and mental workload

The phenomenon of distractibility implies a limited capacity restriction on attention. Distraction effects seem to be evidence of irregularities in the allocation of processing resources, resulting from constraints which may be structural or functional or both. Several decades of research on human performance have shown that processing capacity is a flexible resource and that in many cases human information transmission cannot be adequately explained by a straightforward single-channel model (Kahneman, 1973). Indeed, under some circumstances processing capacity may seem not to be limited: it is now accepted that limits can be very task-specific and can be affected by practice (Johannsen, Moray, Pew, Rasmussen, Sanders, and Wickens, 1979). There are always times when the analysis of new information takes precedence over the prevailing attentional focus. This high-priority switching is typified by the orienting reflex toward stimuli which are biologically significant or highly novel (Berlyne, 1960; Lynn, 1966). But attractions such as novelty may not ensure measurable distractibility, as the elaborate experiment by Hovey (1928), described earlier, so clearly illustrated. Experiments which have demonstrated distraction effects have often stressed the need for "difficult" tasks (Eschenbrenner, 1971; Finkelman and Glass, 1970; Hack, Robinson, and Lathrop, 1965; Plutchik, 1961; Pollock and Bartlett, 1932). Johnson and Cole (1976) described explicitly how they had to make their task more difficult before distraction effects could be detected. The emphasis on difficulty implies an increase in the mental workload imposed by the task. Mental workload is a convenient, even indispensable, term to denote in a general way the attributes of cognitive effort; but it eludes satisfactory definition because it subsumes difficult correlates such as capacity limitations and subjective experience. While mental workload is not entirely determined by spare capacity, it is clearly related to capacity (Sanders, 1979). By increasing the mental workload in a task, it is assumed that less capacity is available for processing information outside the task. Support for this assumption comes largely from studies of secondary task performance, in which changes in the performance of a subsidiary task are taken as evidence of changes in spare processing capacity (Pew, 1979; Rolfe, 1973; Wickens, 1979). The rationale of distraction experiments depends on two assumptions: that by increasing mental workload, capacity will be less freely available to deal with distractors; and that there are tasks which will be appropriately sensitive to variations in the processing resources made available to them.

Tracking

The conclusion that complex and continuous processing of information would be most likely to expose temporary effects seems to have prompted investigators to employ tracking tasks in distractibility research. Tracking generally refers to a class of sensorimotor skills in which the operator (subject) attempts to control the state of a system which is undergoing change due to environmental disturbance. The favourite laboratory tasks involve vehicle control simulations or visual displays which respond to instability in an electronic circuit. Contemporary interest in tracking stems not only from its practical relevance in tasks such as vehicular control, but also because it provides a convenient laboratory paradigm for skills requiring continuous information processing.

In any tracking task, the operator monitors the effects of a disturbance on a physical system and attempts to control the effects. The properties of the complete man-machine system depend both on the characteristics of the physical components (the *plant*) and on the properties of the human controller. Much of the science of control engineering is concerned with identifying the relationship between plant input and output and this is usually embodied in a mathematical transfer function. In principle, the same technique can be applied to the man, but in practice this is more difficult because the man's behaviour is never as deterministic as the plant's. In spite of this the methods of control engineering have proven to be immensely powerful and useful tools for understanding man-machine systems, particularly in the area of manual control skills (and have been given superb treatment in a book by Kelley, 1968). The concepts of control engineering are well suited to descriptions of tracking behaviour, and some of these ideas appear in the following short summary of visual tracking displays and their associated controls.

It is usual to distinguish between two types of tracking task. In *compensatory* visual tracking, the operator's display indicates only the error in the system which he is trying to reduce and he must use his control to compensate for this error. In *pursuit* tracking, the operator views both the target state of the system and the present state of the control; he must try to make them match. Compensatory tracking is generally much more difficult, mainly because it is harder for the tracker to anticipate the behaviour of the disturbance (Poulton, 1974). Compensatory tasks are also of considerable practical significance in aircraft control; consequently they have figured more prominently in tracking research. The compensatory tracking task may be schematically represented by the single-loop manual control system of Figure 1-1. This shows that the operator acts on the *error*, defined as the difference between the required system output and the actual system output. The error is visually sampled and it is processed somewhat imperfectly by the operator, whose motor output adjusts the control signal. If the forcing function is a random disturbance, its rate of change has a major effect on tracking performance. Humans cannot competently track signal rates much higher than 1 Hz in compensatory tasks (Pew, 1974). Tracking performance depends not just upon the important effects of

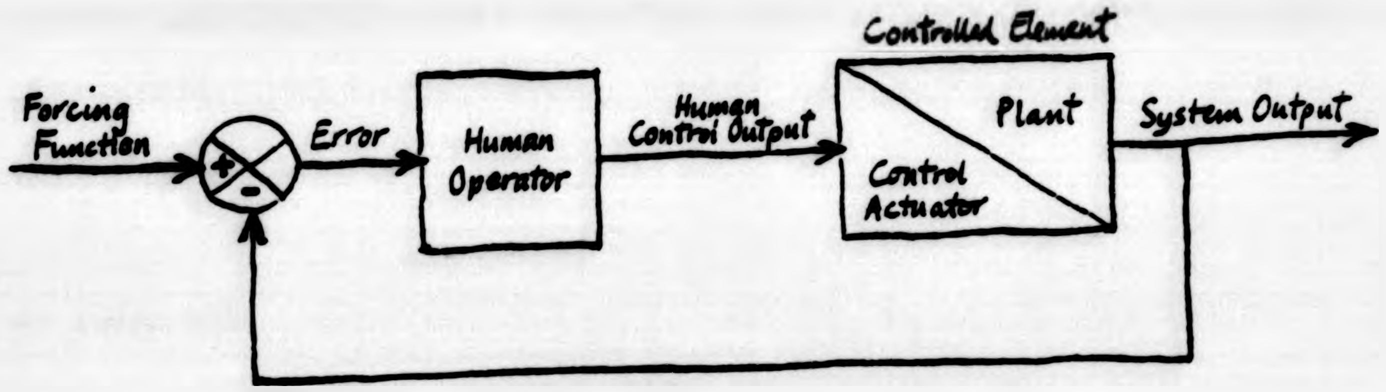

FIGURE 1-1 Compensatory tracking as a single-loop manual control system.

the forcing function and the properties of the plant, but also upon the characteristics of the human operator and the dynamics of the control actuator.

Models of the human operator generally describe the human transfer function properties in terms of response delay, gain, and noise (Sheridan and Ferrell, 1974). The operator's control movements will tend to lag behind the disturbance. This response delay will tend to diminish as the operator gains skill. The gain factor describes how large the corrective output is in response to a given size error. Noise is a reference to the variability in the operator's information processing which cannot be described adequately by the delay and gain components. The models, then, typically assign nonlinear effects to a noise term.

In laboratory tasks the control actuator is usually a joystick which moves along one axis (normally from left to right) but which in some systems may move in two axes or even three (left/right, forward/backward, up/down). The effect of joystick movement depends on the system's *order of control*. If the position of the joystick directly alters the display position, this is described as a zero-order control law. If the outcome is dependent on the velocity of the joystick movement, the control is first-order. Each order of control is a higher derivative of simple position control, and in human terms becomes more difficult to use. Tracking systems rarely attempt to go beyond control based on an acceleration of the joystick (second-order control) because of its difficulty.

In tracking tasks there are numerous methods for measuring the quality of the operator's performance (Kelley, 1969a; Poulton, 1974). The most obvious measure, the mean value of the error, is rarely useful because if the tracker is executing the task properly, the errors fluctuate between positive and negative values and will average to zero. It is therefore much more helpful to score the mean absolute error, mean squared error, or-- most commonly-- the root mean squared (RMS) error. Assuming that the error has a mean of zero, the mean squared error corresponds to the variance of the error. It is important to recognise that the operator's response delay can have a substantial effect on an error measure. For example, even if the operator could copy the track exactly, his processing delay causes his control ouput to lag behind the forcing function and this misalignment of the signals ensures a nonzero error. For this reason, a lagged crosscorrelation analysis of the signals may sometimes be useful to identify response delay effects.

Auditory distractibility during tracking

Typically, experiments to demonstrate distraction effects on tracking seek to identify irregularities in some measure of the tracking error. An early example of such an experiment was described by Plutchik (1961), who asked his subjects to track a compensatory display on an oscilloscope screen. The subjects manipulated a zero-order joystick control to minimise the deflections of the display caused by a 0.4 Hz sinewave signal disturbance. The tracking performance score was a modified absolute error measure, and it was based on two minutes' worth of data. Although he used extremely loud pulsed auditory distractors (usually 115 dB toneburst at 2500 Hz, not white noise) Plutchik could not detect any reliable effect of noise on the mean tracking error, nor on the error variance. The same subjects also carried out a mirror tracing task under similar conditions of distraction. In the mirror task there was no significant effect of noise on the time needed to complete the trace, nor any change in the duration of errors, but the error times were significantly more variable when distracted. Plutchik's (1961) tracking results have been widely cited (for example, in Eschenbrenner, 1971) as an example of the insensitivity of visual tracking performance to auditory distraction, yet it is clear from the description of the apparatus that the behaviour of the tracking system was necessarily rhythmic and predictable, and did not impose a complex task by normal standards.

Although Hack *et al.* (1965) did not comment on this shortcoming of Plutchick's (1961) experiment, they did suspect that the performance measure used in that study was not sufficiently sensitive. Hack *et al.* attempted to distract subjects during compensatory tracking by presenting rather softer (60 dB) brief pulses of white noise. Their idiosyncratic measure of tracking proficiency was based on the average slope of the error trend; that is, it was a composite measure combining information about the error amplitude and also its instability. The scores were calculated during 10 one-minute tracking trials, with one group of subjects encountering the noise distraction and another group providing the silent control data. This procedure did not reveal an overall performance change in noise, but average performance during (only) the first five noise trials was significantly worse. It appeared, then, that habituation to distraction had been extremely rapid. Close inspection of Hack *et al.*'s plotted data suggests that even within the earliest five trials the size of the distraction effect was quite variable, but the authors did not remark on this. The tracking task studied by Hack *et al.* used a display rather similar to that of Plutchik (1961) but the task seems to have been considerably more complex. The track disturbance was a mix of three sinusoids and so the display movement was more complicated and more difficult for the tracker to predict (see Poulton, 1974 for a comparison of single- and mixed-sinusoid tracks). Also, the display and the control joystick moved in two axes. Thus, in spite of the reduced loudness of the distractors used in this experiment, brief disturbances were observed.

Since Hack *et al.* did not present any results using average error scores like those of Plutchik (1961), it is not apparent whether their scoring method was more sensitive. Other factors could have contributed to the observation of distraction effects; especially the reduction of the averaging time for the performance measure from two minutes to one minute, and the use of a more complex tracking task. Variations among all these factors have no doubt contributed to the confused picture of distraction effects in the literature. Thus

Theologus, Wheaton, and Fleishman (1974) failed to find distraction effects on error in their study of compensatory tracking performance. Compared to Hack *et al.* (1965), their distractors seemed to carry more potential for disruption because they were louder (85 dB), were of varying duration, and arrived more unpredictably. The tracking task again used an oscilloscope display and moved in two axes; moreover, the joystick was velocity-sensitive (first-order control) rather than position-sensitive (zero-order) as it had been in other studies, and people find higher orders-of-control more difficult. However, Theologus *et al.* did not use a random track disturbance but a sinusoid, and their performance estimate was mean integrated error measured over the comparatively long span of five minutes.

In contrast, Eschenbrenner (1971) did report a performance decrement caused by intermittent auditory distraction during a compensatory task conducted in a space flight simulator. This study seems to have been the most thorough examination of distracted tracking published. The apparatus was a space vehicle optical tracking system, comprising a two-axis compensatory controller for maintaining image stability in a telescope finder. The performance measure was the total length of time the operator was able to keep the image jitter from exceeding a criterion level during each 40-second simulated orbit of the spacecraft. The distractors were bursts of white noise of 50, 70, or 90 dB intensity. The noise was scheduled to be periodic distraction (on for 2 seconds, off for 2 seconds), aperiodic (on for 2 seconds, off for a random time between 0.5 to 3.5 seconds), or continuous. The different schedules were given to separate groups of subjects, with a control group receiving no distraction. Within the distracted groups, each subject experienced all loudness levels, counterbalanced. The subjects were trained for two days. The test trials took a further three days, with a different loudness level each day. Eschenbrenner found that all the distracted groups performed worse than the control group, using time below criterion as his performance measure. As the intensity of distraction increased, so did the deterioration of performance. The aperiodic schedule resulted in significantly worse performance than either the periodic or continuous regime, but the difference between the periodic and continuous presentation was not significant. Eschenbrenner stated that these effects held up consistently across the distraction trials in each session. He supported this claim with summary statistics, but since no raw data were presented it is impossible to say whether there were, in spite of the lack of statistical evidence, any weak trends suggestive of habituation. Eschenbrenner attributed the absence of habituation to the relatively small number of distraction trials (20 per session) and the fact that testing had been distributed over several days.

Finkelman and Glass (1970) combined tracking with a memory task to investigate distractibility. They used a compensatory tracking display driven by a step function, with first-order control dynamics, as the high-priority primary task in a dual-task experiment. The secondary task was a delayed digit recall test. The distractors were intervals of 80 dB white noise which were either "predictable" (9-second bursts separated by 3-second silences) or "unpredictable" (bursts from 1 to 9 seconds, separated by silences of 1 to 3 seconds, randomly combined). Finkelman and Glass scored tracking performance using only time-on-target (TOT) as the critical measurement. They reported that tracking TOT was unaffected by the presence of noise. The use of TOT to evaluate tracking performance was unfortunate: it is now recognised as a crude measure which wastes information concerning the size of the operator's error (Poulton, 1974). Nonetheless, digit recall errors in the subsidiary task did increase significantly, provided that the noise scheduling was unpredictable. The two tasks were also tested separately. When taken out of the dual-task paradigm, there was no significant distraction effect on either recall errors or TOT, supporting Finkelman and Glass's claim that it was necessary to take up much of the operator's processing capacity before the noises could be seen to affect performance. Finkelman and Glass also made the interesting speculation that predictable noises were less distracting because less capacity needed to be devoted to the job of anticipating their arrival.

It is striking that all of the tracking studies identified during this review sought evidence of distraction effects in measures related solely to the precision of error correction. None looked for distractor-induced changes in the operator's response delay, for example.

The special problem of noise stress

Because so many distraction experiments have used moderate to intense white noise as distractors, they overlap the vast number of studies concerned with the effects of noise on human performance. This must certainly account for some of the confusing and contradictory findings in the distractibility literature, because the general effects of noise are far from simple. Cohen and Weinstein (1983) stated succinctly that "the noise-performance literature is complex, often at least seemingly inconsistent, and subject to a number of different interpretations" (p 47). Nonetheless, there are two properties of noise stress which certainly contribute to the distractibility variations. The intensity of noise and its intermittency both influence performance in ways which are not simple or consistent.

Many reviews of noise research (Broadbent, 1971; Broadbent, 1979; Burns, 1968; Cohen, 1969; Kryter, 1950; Kryter, 1970; Poulton, 1978; Poulton, 1979; Rossi, 1983) have pointed out that there seems to be a distinction between the effects of noise which are due to incidents of conscious irritation or distraction, and effects which are stressful in a more diffuse way. The intensity of the noise has a good deal to do with this difference. Broadbent (1971) identified noise above 95 dB as being stress-related and generally harmful to performance, but concluded that below 95 dB the interference effects of noise (especially in intermittent bursts) were more equivocal: sometimes impairing and sometimes facilitating performance. Even this distinction had to be qualified, because it was recognised that the sudden introduction of any noise regardless of intensity could have an initial transitory effect which could differ in quality and quantity from the longer-term effects on performance. This subdivision of effects according to loudness seemed tidy, but proved to be premature. In a later survey Broadbent (1983) revised the previous 95 dB minimum for stress effects, pointing out that behavioural consequences were now being demonstrated with continuous noise at 80 to 85 dB, especially in tasks which incorporated time pressure. Note that the stressful effects emphasized here are considered to operate over and above the temporary influences of distraction. Thus it seems likely that at more moderate as well as at very high intensities, noise stress and distractibility effects interact. This would perhaps account for the frequent observation that performance in noise is highly variable within-task (Broadbent, 1979; Cohen and Weinstein, 1983) and that performance may fluctuate so readily between deterioration and improvement that no change is measurable overall (Cohen, 1969).

It is therefore understandable that noise effects often depend on whether the noise is continuous or intermittent. If it is likely that increased loudness exacerbates stress, increased intermittency probably enhances distractibility (which possibly has its own impact on stress). The compound effect may be difficult to predict, prompting Poulton (1979) to caution his colleagues not to employ intermittent sounds at all when studying noise stress. The relationship between noise stress and performance has proved too complex to be encompassed by a simple theory, especially when moderate, rather than extremely loud, noise is discussed. The most popular hypotheses have invoked arousal (Broadbent, 1971) as an intervening variable, but experience has shown that to exploit the construct usefully arousal has to be treated as a much more complex multidimensional concept than originally proposed (Broadbent, 1993). It has usually been found that the optimum arousal level for a difficult task tends to be lower than for an easy task; but as Fisher (1984a) pointed out, this description has little utility without knowledge of the base level of arousal the subject brings to the task. Uncertainty about arousal's multidimensional makeup and its measurement may be only part of the problem in estimating the effects of noise: Fisher argued persuasively that some stressors could give the subject "worry work" (such as concern for the consequences of failure) which interacts with and compounds the mental workload already imposed by the task. Furthermore, there are genuine limits on the explanatory power of the term arousal when it is interpreted as some sort of unitary property governed by a physiological state, especially if it is viewed as a simple (quantitative) resource driving behaviour. Hockey (1984) prefers to describe arousal as a kind of tuning process which corresponds to a (mainly qualitative) patterning of cognitive resources. The final pattern of resources can be affected in quite complex ways by different combinations of stressors. This provision is necessary because the experimental literature contains much evidence that stressors do not always interact by simple additive or subtractive combinations of individual effects. In Hockey's formulation, performance is determined by the overall selection of resource states: this is a multicausal outcome which cannot be sensibly sustained by a unidimensional interpretation of arousal.

Thus it is not possible to define a straightforward rule such as "stress increases arousal which impairs performance." For example, target detection in vigilance tasks seems to benefit from moderate varied noise but suffers from loud noise (Broadbent, 1979). Also, Gawron (1982) found that in a single axis compensatory tracking task, continuous noise at 55, 70, and 85 dB actually facilitated tracking performance. Although the effect was small the tracking error decreased reliably as the sound intensity (and presumably the subjects' arousal) increased. In some situations, noise may act in very task-dependent ways to bias information processing strategies such as the subject's pattern of attentional selection (Hartley, 1981; Hockey, 1970a; Hockey, 1970b). Such effects may occur without being seen to either harm or enhance overall performance (A P Smith, 1985; Woodhead, 1966).

To this point, only a little has been said about the sensitivity of different tasks to noise stress. The general view is that noise is more likely to impair performance when there are extreme demands on the subject. The load on the subject may be imposed by set, time pressure (that is, by reducing the amount of time available to make decisions), and memory demands. A good cross-section of such factors is discussed in Rossi (1983). These characteristics are much the same as the task properties which have traditionally been seen to aggravate distractibility.

The present research

The implications of the distractibility literature are that auditory distraction effects are more likely to occur when the distractors are infrequent, unpredictable, novel, and loud. However, if the distractors are excessively loud, they may produce stress-induced effects in addition to, or in place of, distraction effects. The availability of spare capacity probably affects the measurability of the effects, and tasks imposing a high mental workload would seem most likely to suffer impairment during distraction. It is to be expected that sensitive performance measures which are responsive to very short transients are necessary. The most probable sensorimotor effects would include prolonged response latencies and elevated tracking error.

The research described in this paper explored three issues. The first concern was to develop a compensatory tracking task which would provide rapid training while imposing a large mental workload. This was achieved by making the difficulty of the task adaptive in response to the subject's demonstrated ability. It was expected that a task such as this would promote conditions sensitive to auditory distraction. The tracking requirement was intended to demand continuous sensorimotor processing; the adaptive technique was to help to ensure that each subject was operating close to the limit of skill.

The second, and primary, goal of the research was to investigate the effects of auditory distraction on tracking. Using noise distractors, the importance of unpredictable distractor scheduling was examined by comparing the effects of random and periodic presentation rates. The tracking results were analysed in detail, with additional measures besides error coming under consideration. This provided a far more comprehensive picture of distraction effects on compensatory tracking than has previously been available. In consideration of the widespread use of loud noise as a stressor, some observations were also made on the impact of verbal distractors in order to gain some impression of the relative efficacy of verbal material compared to noise.

The final experimental section was prompted by the observations which have stressed the importance of distractor regularity in promoting adaptation to the distractors. An experiment explored the hypothesis that a form of temporal expectancy could contribute to subjects' adaptation to distraction. For expectancy to assist in the adaptation to distraction during tracking, it would have to be relatively insensitive to the load imposed by tracking. For this reason an examination was made of the ability to anticipate the arrival time of uncertain events, and of the effects of the tracking task load on predictive strategies. The aim was to determine whether temporal expectancy would be restricted or eliminated by tracking at high workload levels.

CHAPTER 2: THE ADAPTIVE TRACKING TASK

Overview

An adaptive tracking task was devised in which the task demands changed in a rational way in response to the subject's performance. As the operator gained skill, the track characteristics were adjusted to increase the difficulty of the task and so promote greater skill. When performance faltered, the task was made easiar until performance recovered. The adaptive task was developed as a tool to assist in later investigations of distractibility. The task was designed to load the subject's processing capacity as much as possible without the subject losing control.

Adaptive tracking was compared experimentally to constant-demand nonadaptive tracking tasks. It was demonstrated that the adaptive procedure imposed a difficult tracking requirement. Under adaptive conditions, tracking performance contained large average errors and subjects rated the task as difficult throughout practice. When nonadaptive tracking tasks were used, the level of error was determined by the initial complexity of the track and practice appeared to have little effect on the quality of performance. An objective measure of mental workload during tracking was obtained from the outcome of a simultaneous secondary task. A finger-tapping task was added alongside the primary tracking requirement, and the tap interval irregularity was taken as an index of workload. The tapping experiment demonstrated that the adaptive tracking procedure kept the subject heavily loaded.

It was concluded that this type of adaptive task should create suitable workload conditions for the measurement of the effects of auditory distraction.

Introduction

In the typical compensatory visual tracking task the experimenter sends a noisy electrical signal, the *forcing function*, to a tracking system incorporating an oscilloscope display. The display shows a bright line moving back and forth across the screen. The operator is asked to try to keep the line in a fixed target position, and is provided with a joystick with which he can exert control over the line's position. The movements of the line result from the combined effects of the forcing function and the operator's joystick movements. The operator serves as part of the control system. The deflection of the line from the target position is a measure of the difference, or error, between the forcing function disturbance and the control system output. In compensatory tracking the operator does not view the movement of the track directly nor the signal from the joystick. Instead the bright line deflection indicates the error between them. Appropriate movements of the joystick compensate for the movements of the track caused by the forcing function and reduce the error. If the joystick signal exactly matches the forcing function, the error is zero and the displayed line coincides with the target position. Otherwise, the line's movement away from the target position illustrates clearly to the operator how much compensation is required from the joystick.

In most tracking tasks the disturbances of the track are by intent not very predictable. The forcing function enters the system as a time-varying voltage. With Gaussian random noise sources the forcing function may be thought of as a mixture of high and low frequency signals. The more high frequencies incorporated, the more erratic the display's behaviour appears, and the more difficult it is to track. The actual high frequency content can be adjusted by passing the forcing signal through an electronic low-pass filter circuit before it enters the tracking system. This provides a straightforward way of altering the difficulty of the tracking task: by adjusting the filter's cutoff point up or down the contribution of high frequency noise may be increased or reduced. Moreover, a computer can be used to measure how well the subject is controlling the display and the computer can adjust the filter setting if it becomes necessary to change the difficulty of the task.

Using this computer control technique, the earliest work in the research programme sought to establish an *adaptive* compensatory tracking task which would serve as a sound methodological tool in the later experiments. The basic idea behind the task was that it would adapt appropriately to the operator's demonstrated ability. As the subject displayed greater skill, the tracking demands would temper the improvement in performance by becoming more strenuous, and if the subject could not maintain his level of performance the tracking requirements would ease before the subject became discouraged.

Pew (1974) and Poulton (1974) have provided good general reviews of human tracking behaviour. In general terms the usefulness of making the task difficulty dependent on performance is easy to appreciate. Conventional tracking skill is known to be sensitive to time-on-task, and extended practice is usually needed to acquire competence. Substantial individual differences are commonly observed. Long practice effects are not restricted to naive subjects. They are also noted in many studies which use air pilots or naval servicemen, whose professional qualifications guarantee that they are practised trackers. Adaptive tracking was developed here in expectation that the task, although difficult, would be optimised for each subject and that training time would be minimised because of this. Inexperienced subjects would avoid wasting practice time in an undemanding (or, for that matter, overdemanding) and invariant training regime. Indeed, it seemed at first that adaptive tracking could offer a method for ensuring the maximum possible loading of a subject's limited processing capacity. However, the assessment of capacity is notoriously difficult: Moray (1979a) surveys the many problems. In tracking, as in many other tasks, there are readily observable "differences of personal style between those who like working near their limit and those who do not" (Moray, 1979b, p.17) and it is not certain how this relates to available capacity. Even accepting these limitations, it did seem feasible that a suitable adaptive technique could offer the prospect of an optimised, substantial mental workload for each subject. Experimental attempts to validate this claim will be described in this chapter.

Stages in the design of an adaptive tracking task

Procedures for the design of effective adaptive tasks are far from precise, in part because the criterion of skilled performance may be difficult to define. Many decisions have to be made on the basis of unique task-specific factors and other peculiarities unanticipated before the practical tests begin. This procedural uncertainty is, as yet, very characteristic of adaptive task methodologies and the difficulties it causes cannot be overstressed (Kelley, 1969b; McGrath & Harris, 1971; Wiener, 1973; Williges and Williges, 1978). While there is no definitive way to build an adaptive tracking task, it is usual for the preliminary considerations to include:

1. Selection of one or more performance variables for observation.
2. Selection of the adaptive variable: that is, the component of the task which will adapt.
3. Specification of the adaptation rule, or the criterion for instituting a change in the adaptive variable.

The procedure which relates these items to one another will be called the *adaptive algorithm*. The adaptive algorithm finally adopted here after various pilot projects is described by Figure 2-1. The task was designed to provide a difficult tracking load without exceeding the operator's capabilities. No specific target level of performance was demanded. The task was not to become uncontrollable nor to adhere cautiously to undemanding levels of difficulty. At the same time, any subjective impression of discontinuity or oscillation in the task demands was considered undesirable.

Assuming that the subject does not react to the rigours of task adaptation by abandoning the joystick in frustration, it is not necessarily easy to assess whether the subject is on the point of losing control over the tracking task. Individual differences in tracking strategy make it amply clear that tracking skill is multidimensional, with skilled execution reflecting an individual balance of performance criteria. Figure 2-2 suggests just one way of describing performance on the basis of the tracking error characteristics. This model proposes that precision of response and consistency of performance jointly contribute to competence but that these characteristics may not always coincide. For example, a subject may be capable of tracking very accurately but demonstrates this ability only intermittently. An adaptive task must cope with the expected variety of operator response by settling on the performance features it considers important.

In the adaptive algorithm, tracking absolute error was chosen as the performance variable on which the decision to adapt would be based. A

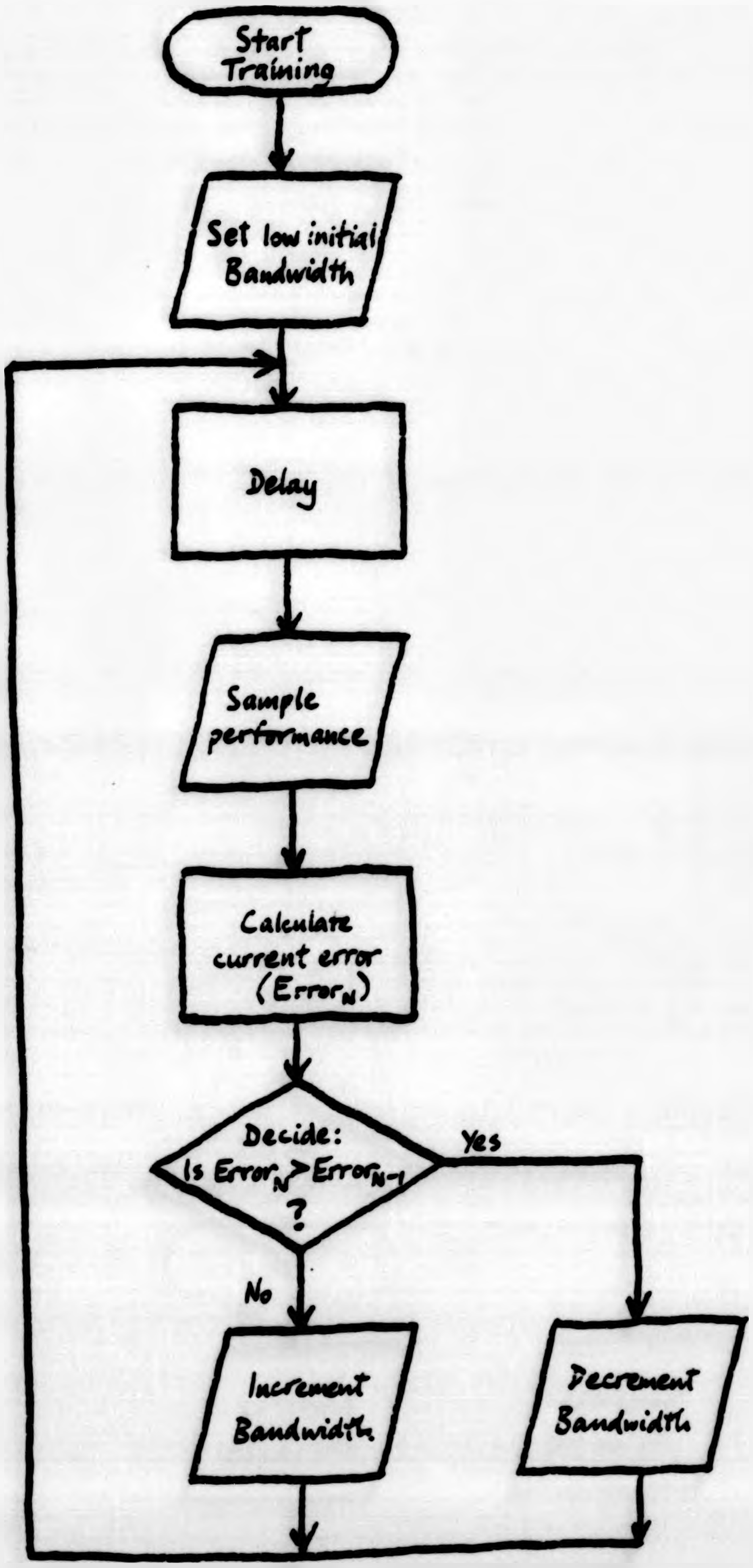

FIGURE 2-1 Adaptive algorithm for the tracking task.

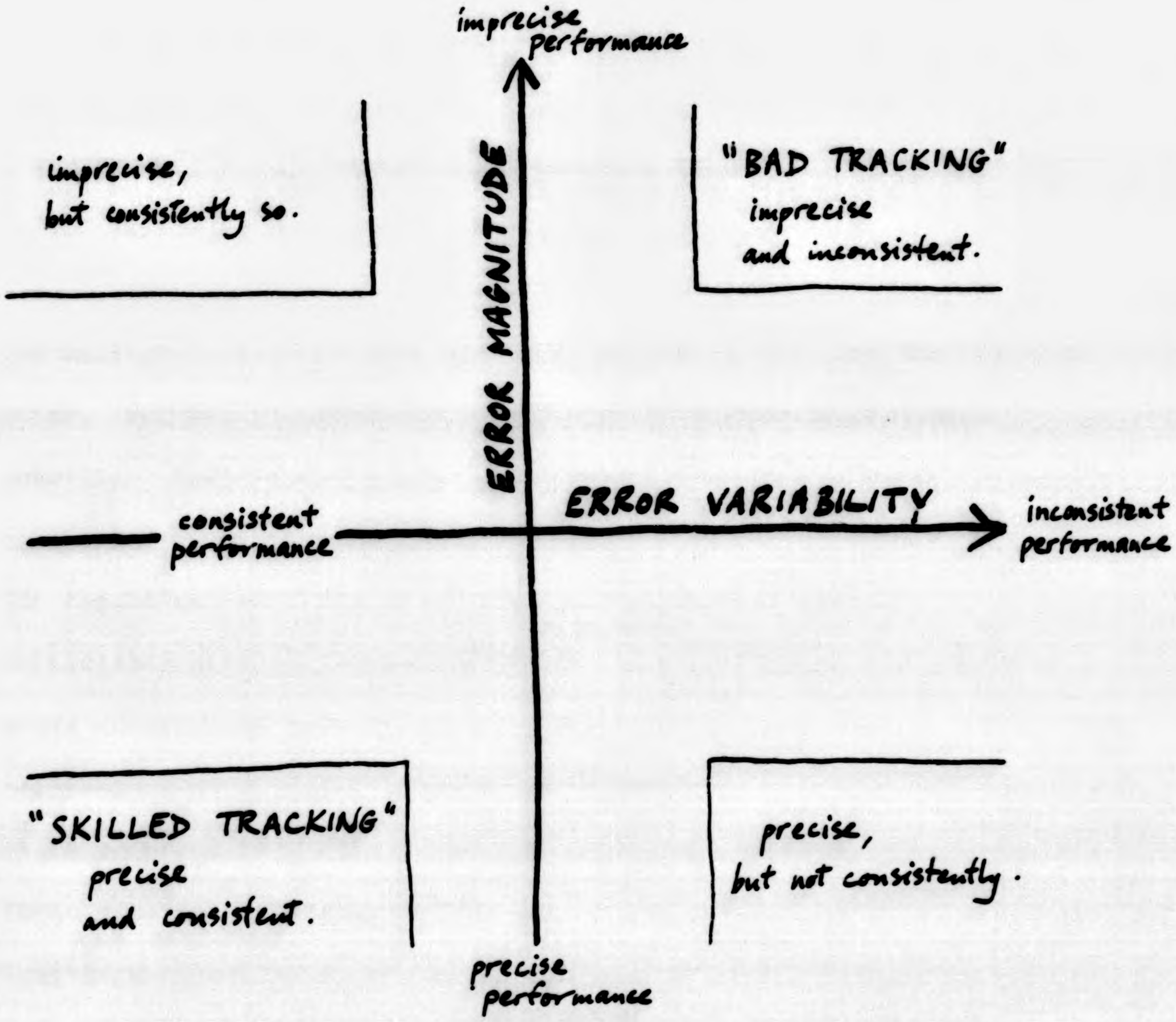

FIGURE 2-2 Characteristics of tracking performance.

single reading of performance quality was represented by the mean of ten samples of absolute error taken over a 1.5 second sampling window. Control of the task difficulty was implemented by adjusting the bandwidth of a Gaussian noise forcing function using an electronic filter. The bandwidth was the adaptive variable. The adaptation rule compared current performance to a reference based on past performance. The forcing function bandwidth was increased when current performance indicated improvement (that is, less error compared to the past) and it was decreased when performance showed some deterioration.

The adaptation rule in this task did not aim to minimise the tracking error. Rather, it attempted to provide the most difficult task possible. Adaptation toward higher bandwidths was driven by evidence of short-term maintenance of performance; that is, by recent constancy (or reduction) of tracking error. In the terms of the multidimensional performance model of Figure 2-2, such conditions would normally be promoted by a tendency toward consistency. The maintenance of a stable level of performance for a brief time could allow progress to a more difficult track, regardless of the absolute level of error. Adaptation did not rely on a relentless and permanent long-term reduction of the absolute level of error, because this would have granted subjects who were performing in a consistent, stable manner-- but were content to tolerate high error-- the luxury of remaining indefinitely at a low and probably insufficiently demanding bandwidth. Using an adaptation rule based on recent evidence that the error was no longer increasing helped to resolve problems caused by subjects' different tracking styles. The principle, then, was to make the task as difficult as possible without provoking the operator's error into wild oscillations, and to accept the operator's self-defined criterion of acceptable error.

How often to adapt?

The subject's own past performance provided the reference level of performance, so the reference was subject-specific, not an absolute level. Past performance was always that of the "recent" past. A recent score instead of a cumulative running average was used as the reference. With a cumulative historical measure, each additional score exerts increasingly less effect on the accumulated weight of past scores. The reference stabilizes and loses the ability to quickly incorporate shifts in the level of performance, so the reference can appear misleadingly discrepant when matched against a small local deviation. Pilot data suggested that this could cause adaptation at times when it was not really warranted by extremely short-term trends. When only the very recent past is used instead as the reference there is a risk, conversely, that current performance will appear so similar to the last reference measure that the course of adaptation may be excessively encouraged, causing the adaptive variable to run away in the more difficult direction. There is a real danger that this will occur if the change in the subject's measured performance lags somewhat behind a change in the task demands. Preliminary tests revealed that when the adaptive algorithm altered the bandwidth of the forcing function, performance scores were unlikely to react instantaneously to the change. If the adaptive algorithm is too impatient it may conclude erroneously that performance is unaffected by the change in the adaptive variable and consequently still more change is introduced before the performance measure has been able to catch up. In this tracking task a compromise evolved. It was concluded that using the most recent past error as the reference was appropriate, provided that there was a short delay (4.5 seconds was the duration finally adopted) between each adaptation and the next assessment of performance. This was adequate to improve the validity of the performance measure and so prevent adaptive runaway. This duration, added to the time of 1.5 seconds needed to sample performance, meant that an adaptation could occur every 6 seconds.

How much to adapt?

The effect of a change in the bandwidth of the forcing function on the subject's error depends a great deal on the mean bandwidth level. When tracking at a low average bandwidth, large alterations can be accommodated, but at a high bandwidth setting even small shifts can alter performance dramatically (Pew, 1974). For this reason, it proved necessary to ensure that adaptive adjustments to the forcing function bandwidth were of a magnitude appropriate to the current working bandwidth. To calculate the adaptive increment the algorithm used an arbitrary exponential function

$$i = 0.08\, e^{-w}$$

(where i is the adaptive step and w is the current operating bandwidth, both in Hz). An acceptable adaptation rate was determined empirically during the early tests of the adaptive technique by adjusting the constant term. This function is plotted in Figure 2-3 which shows how the amount of change imposed on the bandwidth by the adaptation process was a nonlinear function of the current bandwidth setting. This prevented abrupt shifts in the perceived task difficulty when the higher bandwidths were used. When the bandwidth adapted downward to present an easier task, it always retreated to the most recently used lower setting. The bandwidth was not allowed to reduce below the initial level imposed at the session startup.

Algorithm tuning and computer simulations

During the development of the adaptive tracking task, volunteers carried out short pilot sessions to test aspects of the adaptive algorithm. The results of these tests led to gradual improvements in the adaptive procedure. Although most of the tracking trials were carried out using a zero-order control law, some tests were also conducted with first-order controls to confirm that the adaptive task could be used with (the more difficult) first-order control dynamics.

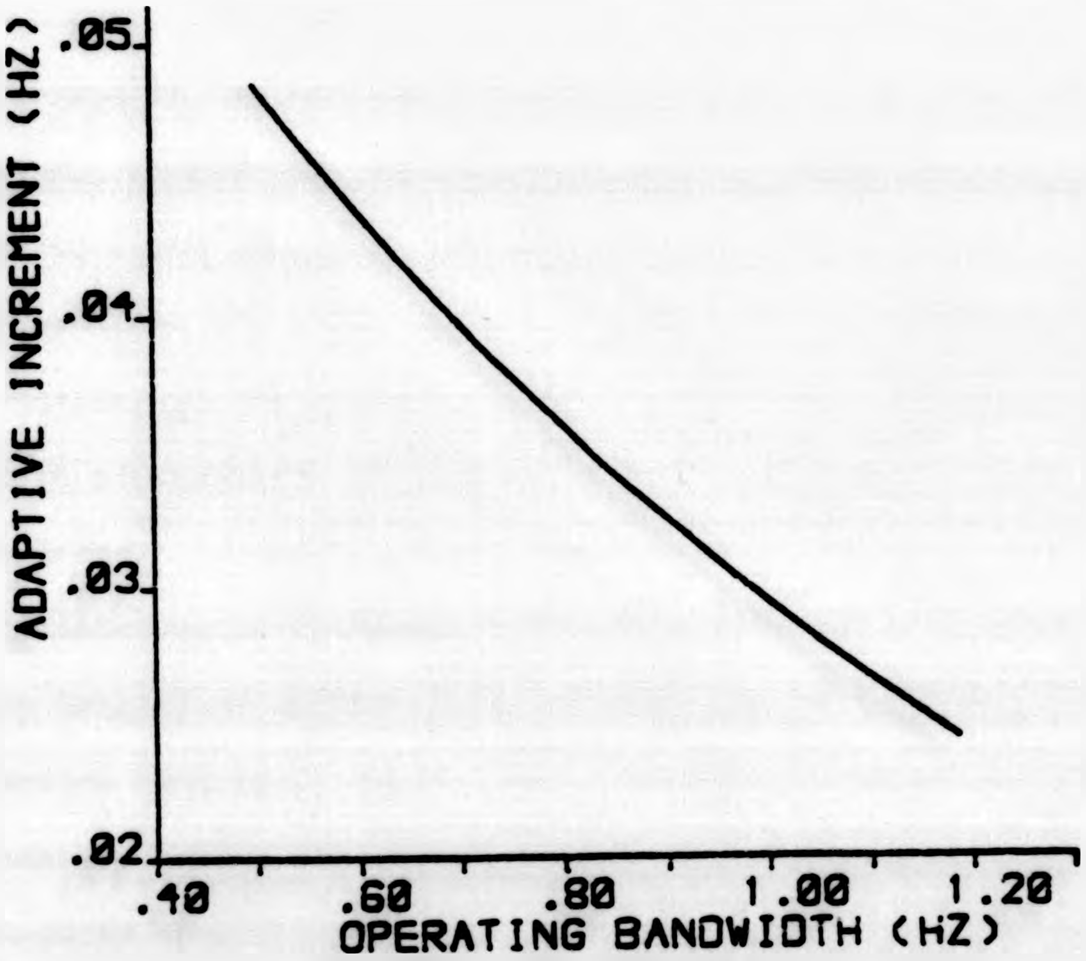

FIGURE 2-3 Calculation of the bandwidth increment in the adaptive task.

To obtain a thorough understanding of the response of the algorithm to different control techniques, a number of computer simulations of subject tracking behaviour were carried out in conjunction with the pilot studies. Artificial control responses were fed into the adaptive system in place of those normally produced by a human operator. These investigations included simulations of perfect tracking behaviour; perverse responding comprising control movements opposite to those required by the task; indifferent (random) control movements; and abandonment of control (that is, a motionless joystick). Of these, only perfect tracking resulted in a steady increase in the forcing function bandwidth. When the tracking was utterly contrary or the joystick was not moved, the bandwidth tended to oscillate around an average level near the starting value. In these cases the tracking error was a direct function of the forcing function, so the adaptive algorithm was actually responding only to local fluctuations in this signal; hence the oscillation without progress. If the simulated joystick response was a random signal representing an inattentive or indifferent subject, the tracking error was determined by the convolution of the random signal with the forcing function. Again, this provided only local variations and resulted merely in the oscillation of the adaptive variable. These computer simulations provided some reassurance that if subjects were motivated at least to attempt to keep the track on target, they should be able to benefit from the adaptive procedure.

Adaptive tracking task: General Procedure

The subject sat in a quiet comfortably-lit cubicle facing the tracking display across a small table. The display was about 60 cm away from the subject and was elevated to eye level. The joystick was mounted in a low metal box placed on the table before the subject. For many experiments, data collection was speeded up by equipping two cubicles with independent tracking systems so that two subjects could perform the task concurrently.

The tracking systems were maintained by an Electronic Associates Analogue Computer located in an adjoining room. This device processed the forcing and joystick signals and computed the error signal voltage which was sent to the tracking display. The tracking system design is shown in Figure 2-4. The forcing function and joystick data were also sent from the analogue computer to a Digital Equipment Corporation PDP-11/45 minicomputer which was programmed to support the adaptive algorithm and to store the voltage data in digital form. In addition, this computer summarised and printed the adaptive performance data. The PDP-11 was interfaced to a dual-channel Kemo Programmable Filter by which means it directly controlled the bandwidth of forcing signals sent to the analogue computer. The PDP-11 could also reset the analogue computer when necessary to halt the tracking task.

The forcing function was derived from a Hewlett-Packard Type 3722A Noise Generator. This instrument provided a time-varying random voltage output corresponding to a stationary Gaussian noise source

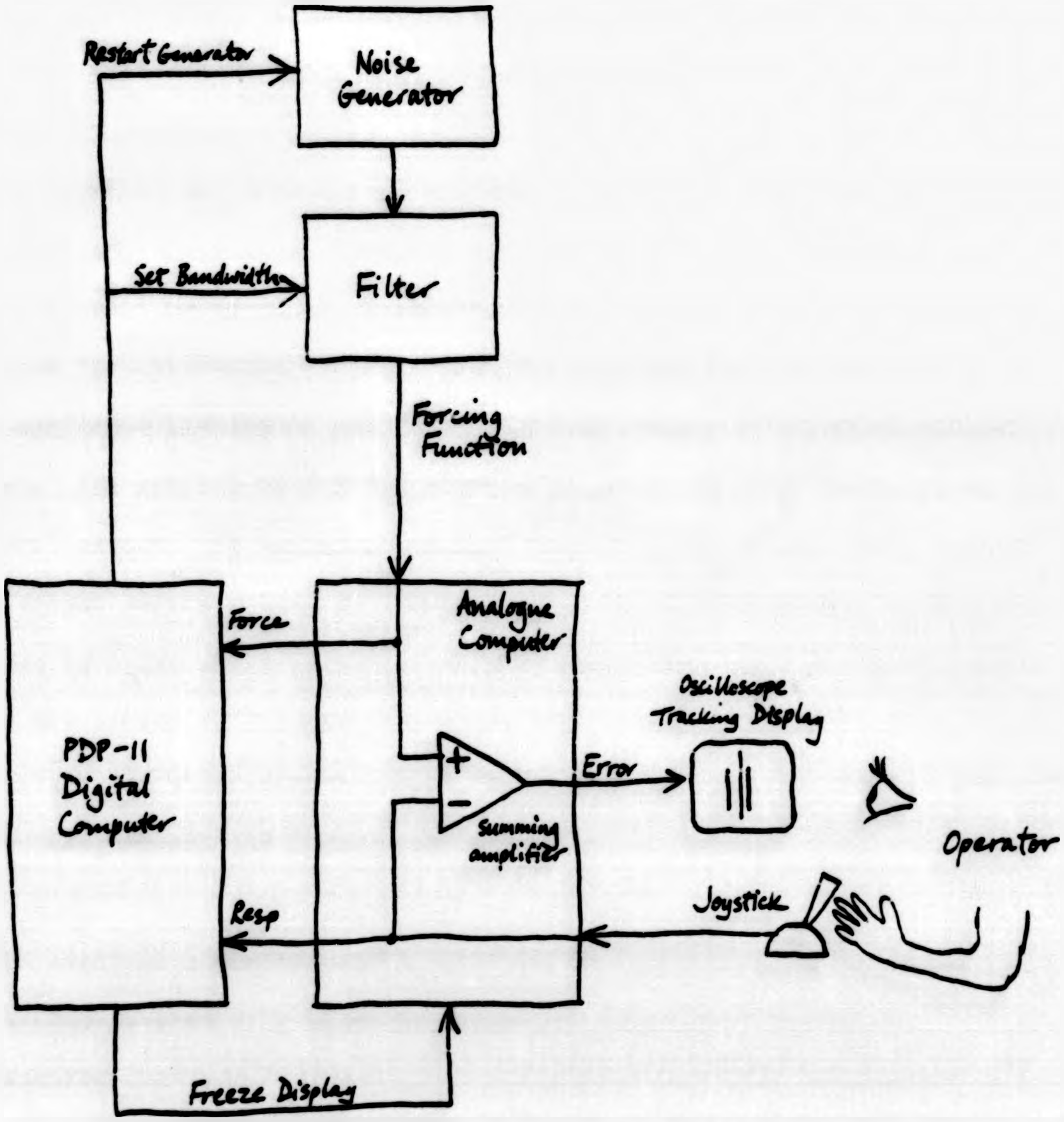

FIGURE 2-4 Zero-order compensatory tracking system.

with an average bandwidth of 5 Hz. This relatively high bandwidth signal was rolled off to a usable lower bandwidth by passing it through the Kemo filter enroute to the analogue computer. In the standard adaptive procedure, the program always set the initial filter cutoff to a starting bandwidth of 0.5 Hz, which supplied the subject with a sensible forcing function to begin training.

The joystick was a single-axis control. It was a precision servo potentiometer mounted in such a way that a short lever attached to its shaft could be moved in an arc from side to side. It was normally wired for zero-order control dynamics, although for certain experiments its signal was integrated to impose a first-order control law. The display device was a standard green-phosphor Telequipment oscilloscope. It was positioned so that an input signal would deflect a bright vertical line to the left and right of the screen's midpoint. Two vertical black cursor lines 2 cm apart were painted at the centre of the screen to bracket the target position of the track. The subject was instructed to move the joystick to keep the bright line centred between the cursor marks.

The adaptive algorithm was applied every 6 seconds. The events followed a precisely timed sequence. Tracking performance was measured over a 1.5 second sampling window. The mean absolute error was calculated and compared to the previous score and the forcing function bandwidth was adjusted as necessary by reprogramming the filter. Another 4.5 seconds of tracking was allowed to elapse before the sampling window was opened again to initiate the next adaptation.

Tracking sessions were administered in blocks, each block typically of three minutes' duration. During a three minute block, 300 samples of tracking performance would be recorded for later analysis and as many as 30 bandwidth adaptations could occur. In a full session, subjects might track for ten or perhaps fifteen blocks altogether. Subjects were not aware of block divisions except when rest breaks were granted between blocks. The breaks were signalled to the subjects by freezing the display, and usually lasted about one and a half minutes. Blocking simplified the postprocessing of data and allowed the computer necessary time to update the stored datafiles and print a performance summary.

The forcing function bandwidth always started at 0.5 Hz, and could not adapt below this setting. As subjects demonstrated more skill, the bandwidth increased by an increment appropriate to the ambient value. When the adaptation rule demanded it, the bandwidth retreated to an earlier lower level. Bandwidth settings carried over from block to block, so no discontinuities in the adaptation sequence were introduced by the blocking or the rest breaks.

Three experiments were conducted to investigate performance during the adaptive tracking task. The first experiment simply noted the progress of the adaptive task over three days of extended practice. The second experiment compared adaptive tracking to fixed-bandwidth tracking to determine whether the adaptive version provided a more effective training regime. The third experiment attempted to measure the impact of adaptive tracking on mental workload.

TRACKING EXPERIMENT 1: ADAPTIVE TRACKING PERFORMANCE

Procedure

Ten undergraduate students (five male, five female) answered a notice-board advertisement for subjects. None of them had volunteered for tracking work before.

The subjects were asked to practise the adaptive tracking task. They were advised that the task could become more difficult as the experiment progressed, but no mention was made of the adaptive process. The procedure followed the general description given earlier. All subjects held the joystick in the right hand. The control law was zero-order.

Training extended over three consecutive days. On each day, the subjects performed six three-minute blocks of tracking, providing a total of 18 minutes' data. A short rest break was granted after alternate blocks and the subjects were told of their progress.

On the first day the forcing function bandwidth began at 0.5 Hz. For the second and third sessions, the running mean bandwidth of the previous day was adopted as the startup bandwidth for the session.

Results

Adaptive task bandwidth

These subjects encountered a forcing function bandwidth which changed continually during the course of training. The bandwidth alternated between increasing and decreasing values as training accumulated, but the predominant adaptive trend for all subjects was a progressive increase in the mean bandwidth setting. The subjects did not all adapt to the same level, but all did manage to achieve some bandwidth excursions exceeding 1 Hz. Such high bandwidths define forcing functions which are difficult to track well. The overall pattern of adaptive task development is summarized by Figure 2-5 which shows mean bandwidth outcomes for successive three-minute blocks. The results show that the forcing function bandwidth rose during the experiment, with only minor discontinuities appearing between sessions. The adaptive process was unobtrusive: there were no abrupt changes in the task difficulty. At two points during the experiment (beginning with block 10 and again at block 16) the mean bandwidth tended toward stability for relatively long periods. Adaptation did not cease at these times; the adaptive adjustments were oscillating around the mean setting. The oscillations suggest that the subjects were finding it

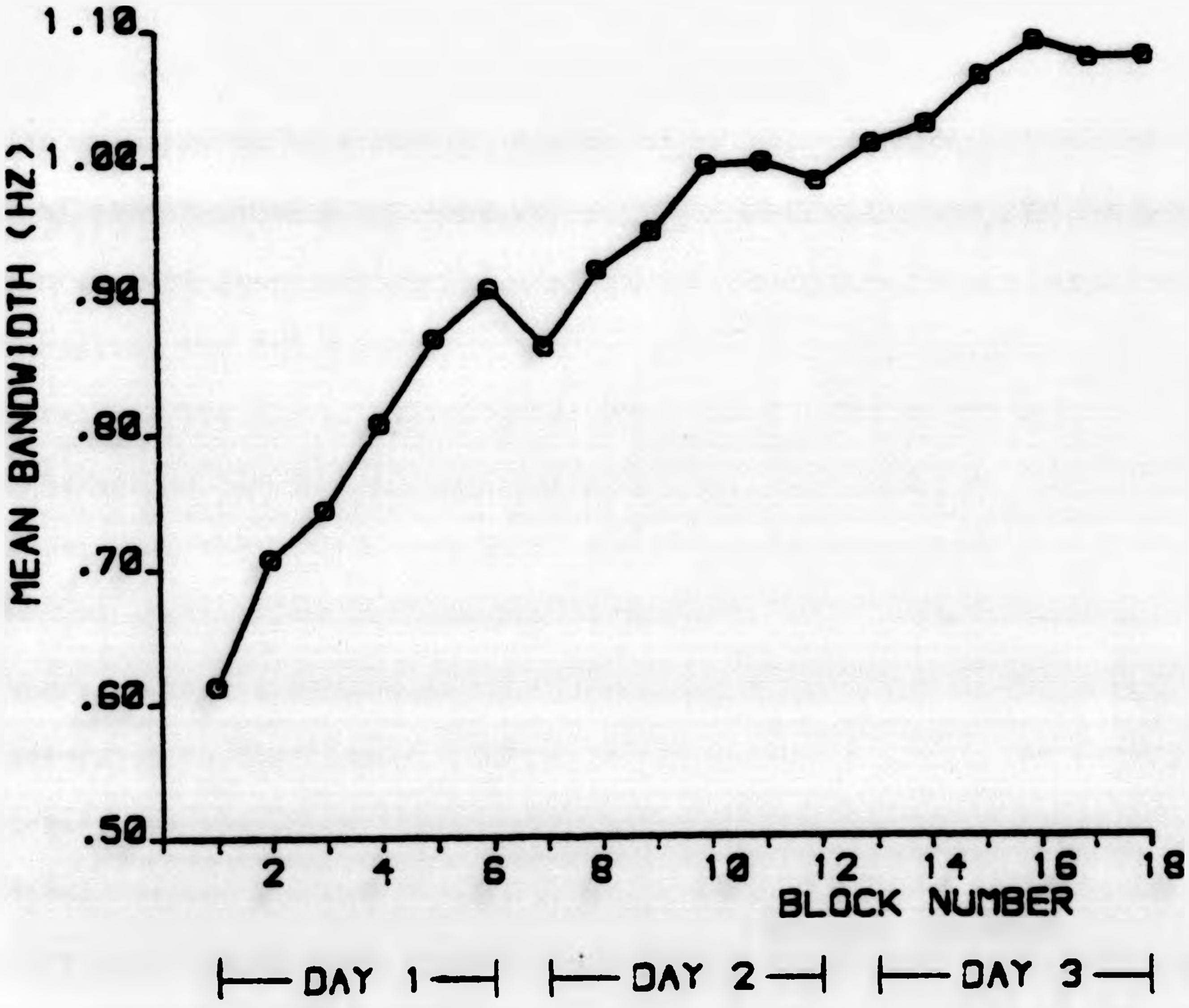

FIGURE 2-5 Tracking Experiment 1.
Effect of practice on the forcing function bandwidth.

truly difficult to improve their skill at these times. Certainly during the last three blocks (9 minutes) of the final practice session (when the task achieved its highest level of difficulty with a mean forcing function bandwidth of nearly 1.1 Hz) the task difficulty was beginning to approach conventional limits of human tracking skill (Elkind, 1956).

Tracking error

The most appropriate overall measure of tracking competence is the root mean squared error (Poulton, 1974). When comparing the RMS error scores from tasks having different track bandwidths, it is useful to normalise the RMS error with respect to the forcing function. To normalise the score, the observed error is divided by the error inherent in the forcing function (that is, the measure of the deviations from the target which the track would generate on its own if the subject did not move the joystick at all). The normalised score is thus a measure of the control error relative to the error generated by the track. Changes in the bandwidth affect the error created by the track. The normalisation is helpful when comparing these results because it cancels out a contribution of the forcing function-- which varies between and within subjects-- to the observed error. The RMS errors reported here have all been normalised with respect to the forcing function.

The RMS error tended to increase as the adaptive procedure increased the bandwidth of the forcing function. This corresponds to the observation in conventional tracking tasks that increasing the bandwidth promotes larger error (Elkind, 1956; Pew, 1974; Poulton, 1974). But Figure 2-6, which shows the error during each block, also suggests that during the final day at least, the mean level of error for the session as a whole did not change substantially although the block-by-block error was fluctuating a good deal. This agrees well with the earlier observation that the bandwidth was prevented from rising unlimitedly during the last session: at times performance was too inconsistent to allow unchecked adaptation to higher bandwidths. For some of the subjects the normalised RMS error at the end of training approached 1.0 or even exceeded this value; such scores signified that the subjects' energetic attempts to control the track were contributing as much to the observed error as were the disturbances imposed by the forcing function. This was a good indication that the task was imposing a difficult control load on these subjects.

Discussion

This experiment confirmed that the adaptive task was responsive to subjects' abilities but did not become uncontrollable even during extended practice. The task did not restrict the tracking RMS error but rather allowed it to assume the value dictated by the individual

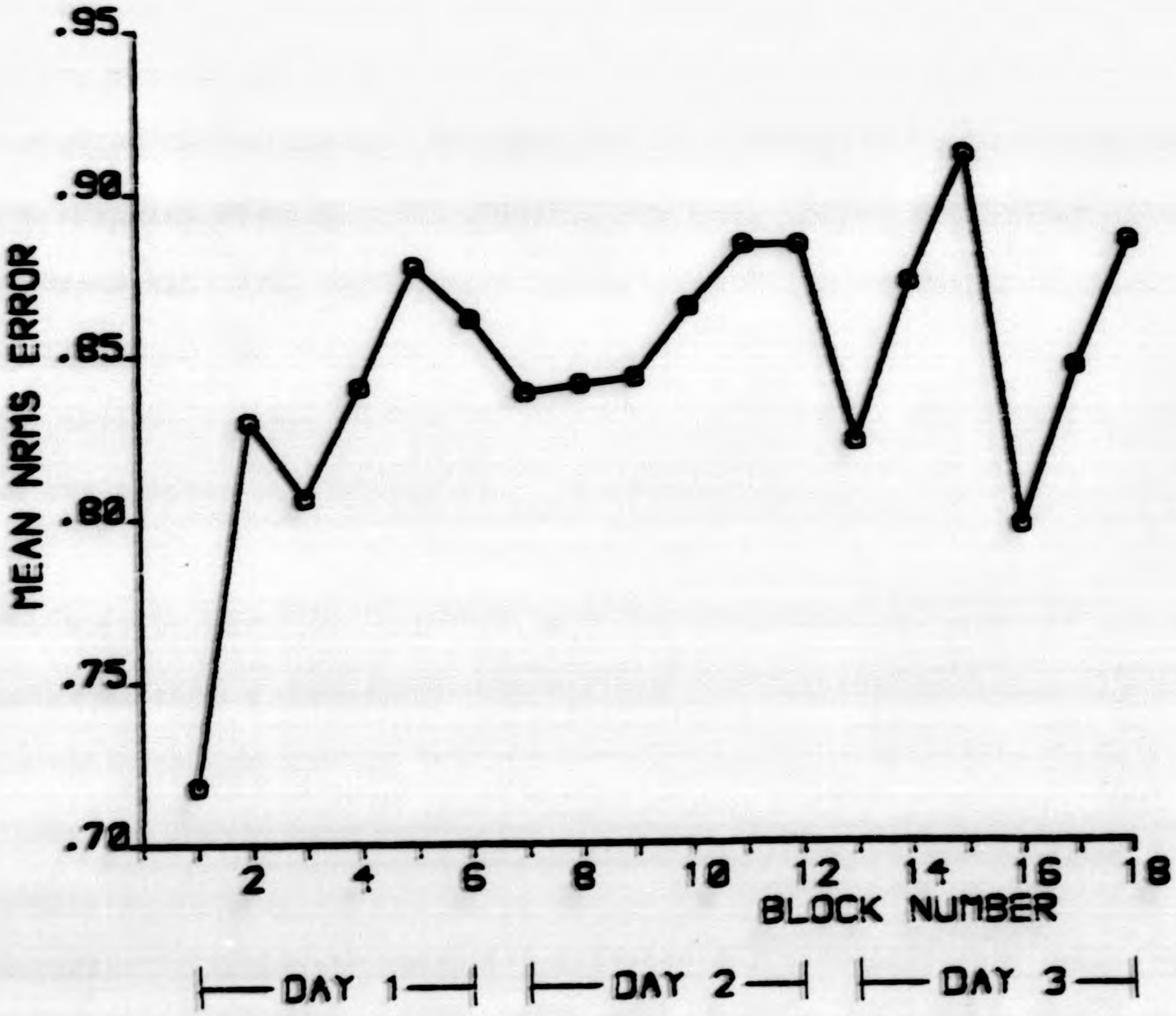

FIGURE 2-6 Tracking Experiment 1.
Effect of practice on the normalised RMS error.

subject's criterion of accuracy. This method of dealing with the error represents but one approach to adaptive training. For example, Kelley (1969a, 1969b) proposed an adaptive training technique to maintain the subject's error at a predefined level. However, this error maintenance approach introduces several genuine difficulties. The success of the procedure depends entirely on the validity of the error criterion, so a method must be devised to predetermine an appropriate level of performance. Such determinations are inclined to be arbitrary or pragmatic decisions (McGrath and Harris, 1971). Also, Kelley recommended that the running error level should be continuously indicated to the subject by somehow incorporating the information into the tracking display. The adaptive approach taken here instead stresses the local consistency of performance, treating relative stability of the error as an indication that the effort invested by the subject is appropriate to the task difficulty. The absolute level of the subject's tracking error is not controlled.

Although it was clear that the adaptive procedure provided the subjects with a demanding tracking task, it was not known how adaptive performance compared to that in nonadaptive fixed-bandwidth tracking tasks. A second experiment was therefore carried out to compare the adaptive tracking outcome directly to the results of conventional nonadaptive tracking methods.

subject's criterion of accuracy. This method of dealing with the error represents but one approach to adaptive training. For example, Kelley (1969a, 1969b) proposed an adaptive training technique to maintain the subject's error at a predefined level. However, this error maintenance approach introduces several genuine difficulties. The success of the procedure depends entirely on the validity of the error criterion, so a method must be devised to predetermine an appropriate level of performance. Such determinations are inclined to be arbitrary or pragmatic decisions (McGrath and Harris, 1971). Also, Kelley recommended that the running error level should be continuously indicated to the subject by somehow incorporating the information into the tracking display. The adaptive approach taken here instead stresses the local consistency of performance, treating relative stability of the error as an indication that the effort invested by the subject is appropriate to the task difficulty. The absolute level of the subject's tracking error is not controlled.

Although it was clear that the adaptive procedure provided the subjects with a demanding tracking task, it was not known how adaptive performance compared to that in nonadaptive fixed-bandwidth tracking tasks. A second experiment was therefore carried out to compare the adaptive tracking outcome directly to the results of conventional nonadaptive tracking methods.

TRACKING EXPERIMENT 2: COMPARISON OF ADAPTIVE AND NONADAPTIVE TRACKING TASKS

Procedure

Nineteen undergraduate students volunteered as subjects. Three had helped in previous pilot studies; the rest had no tracking experience.

Three tracking conditions were studied. Each subject served in only a single condition. The separate groups design was adopted to prevent the asymmetric transfer effects which can confound results if a subject tracks under several different experimental conditions (Poulton, 1974). Six subjects (four female, two male) were asked to carry out the adaptive tracking task. As before, this *Adaptive* group was given a forcing function with a starting bandwidth of 0.5 Hz and the task followed the general adaptive procedure. The remaining subjects (which included the few with some tracking experience) were divided into two groups, both assigned nonadaptive tracking tasks. The *Easy* group (six subjects: three female, one of whom had prior tracking experience; three male) received a fixed bandwidth forcing function of 0.5 Hz, identical to the starting bandwidth of the adaptive subjects. The bandwidth did not change during the session. The *Difficult* group (seven subjects: four female, including two with some experience; three male) worked with a 0.85 Hz bandwidth throughout. This bandwidth was arbitrarily chosen to present a challenging track. Except for the provision that the forcing function bandwidth remained fixed, the nonadaptive tasks followed the same procedure and used the same tracking system as the standard adaptive task. All subjects tracked using the right hand. Training was not as long as in the previous experiment. The entire experiment was completed in one session. Ten three-minute blocks of tracking were performed, for a total task duration of 30 minutes. Brief rests were allowed every two blocks, and during each break the experimenter spoke to the subject to give feedback and encouragement.

The participants were asked to rate the task difficulty at three points during the experiment. Following blocks 1, 5, and 10 each subject was handed a printed scale and asked to mark down his impression of the tracking demand. Simple nonadjectival scoring methods have proved useful for estimating subjective workload even in relatively complex tasks such as air traffic control (Philipp, Reiche, and Kirchner, 1971). Hess (1973) has developed a nonadjectival linear scale for assessing load during tracking tasks. The rating scale used in the present experiment, reproduced in Figure 2-7, was similar to the scale validated by Hess.

Results

Adaptive task bandwidth

The overall pattern of adaptive task development is summarised by Figure 2-8 which shows mean bandwidth outcomes for successive three-minute blocks. The results show that the mean forcing function bandwidth continued to rise during the whole experiment. The increase was approximately linear, with no abrupt task changes in evidence.

Think about the run of tracking you have just completed. Please indicate a number that describes the degree of difficulty you encountered in controlling the display during the run. Use this scale:

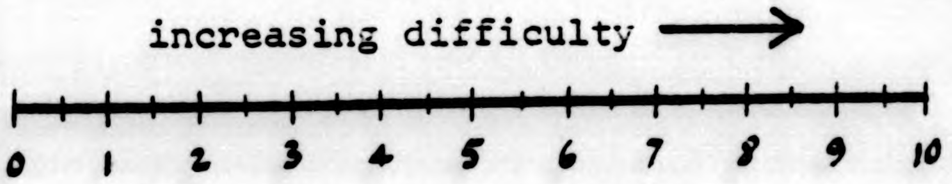

FIGURE 2-7 Tracking Experiment 2.
The subjective difficulty rating scale.

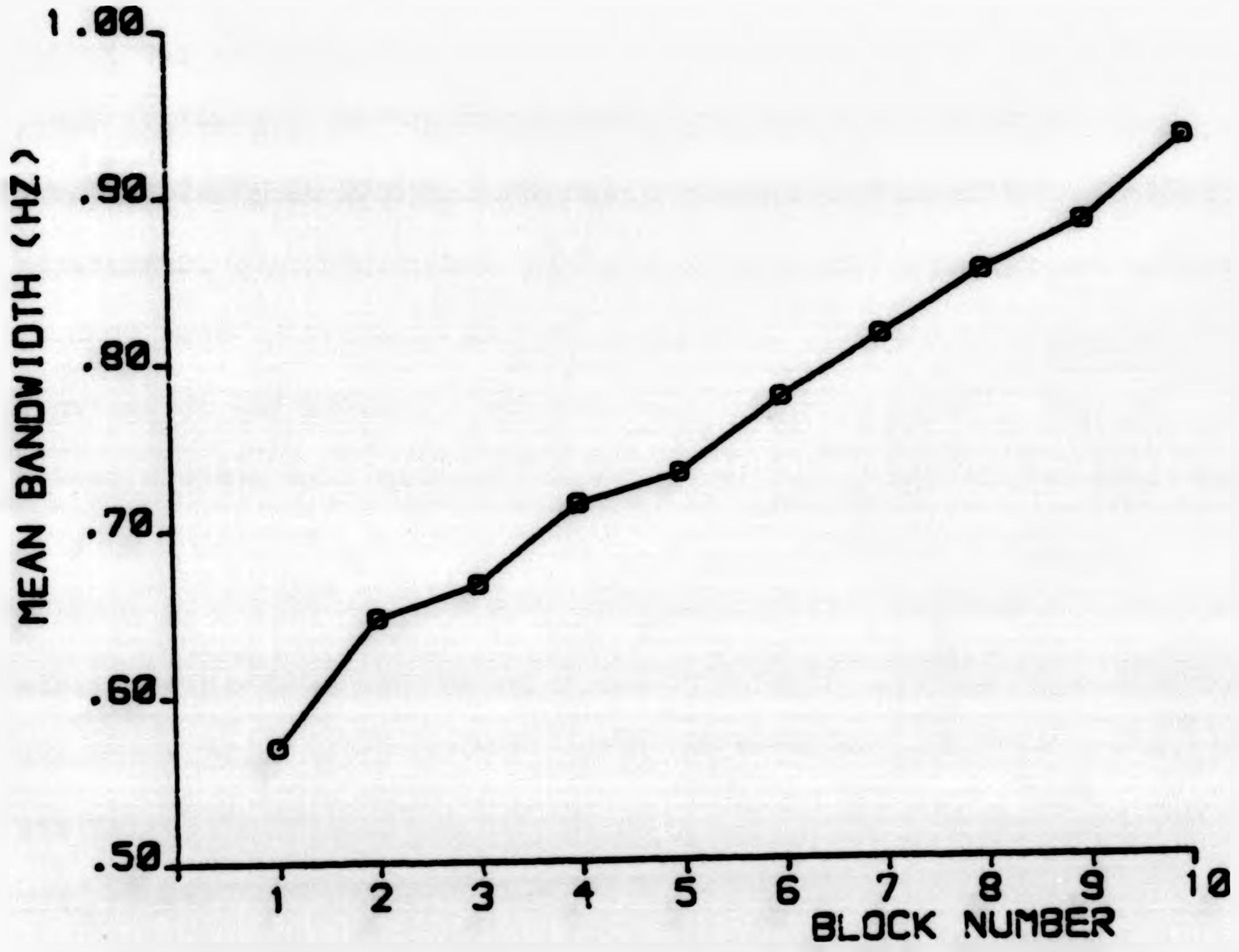

FIGURE 2-8 Tracking Experiment 2. Forcing function bandwidth.

The adaptive progress was slightly slower than in the previous experiment, but was otherwise comparable. The overall mean level of the bandwidth was 0.759 Hz.

Tracking error

For the Adaptive group, the RMS error tended to increase as the adaptive procedure advanced the bandwidth of the forcing function. During the extended training of Tracking Experiment 1, the error had shown little sign of levelling off until Days 2 and 3 of practice (corresponding to Blocks 7 through 18 in that experiment). In this experiment, tracking ceased after ten blocks and this was sufficient to show some settling of the error. During the last four blocks (12 minutes) of the session, the mean level of error clearly did not change a great deal despite the fact that the bandwidth was continuing to rise. The changes in the RMS error in the Adaptive condition are plotted in Figure 2-9, along with error outcomes from the two nonadaptive groups for purposes of comparison. The nonadaptive groups exhibited lower error scores overall than the Adaptive group, although the Difficult errors were nearly as large as the Adaptive measures and, indeed, did actually exceed the Adaptive errors during the earliest minutes of the session (when the Adaptive forcing function bandwidth was still relatively undemanding). Of the two nonadaptive conditions, the Easy trackers enjoyed considerably smaller errors than the Difficult subjects throughout the experiment. This was entirely as expected since the Easy group had the benefit of the lowest forcing

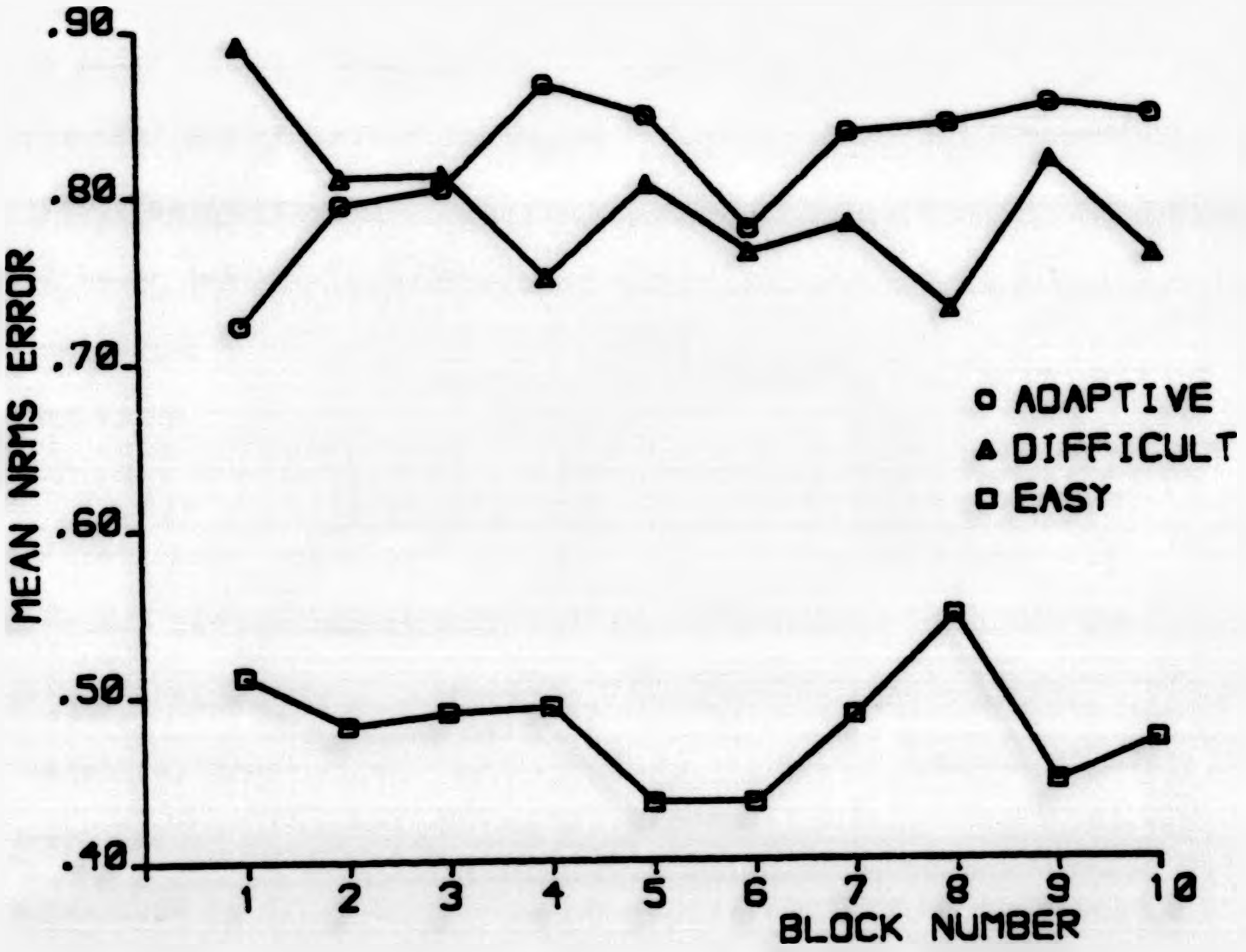

FIGURE 2-9 Tracking Experiment 2. Tracking normalised RMS error.

function bandwidth, and therefore received the least erratic track. But surprisingly, apart from some initial lowering of error mainly during the first few minutes of training, neither nonadaptive group seemed to benefit much from practice. Ignoring for the moment the warm-up effects evident in the first two blocks of the session, the overall level of error during the subsequent twelve minutes of the session (blocks 3 to 6) was very like the average level for the final twelve minutes (blocks 7 to 10) for both groups of nonadaptive trackers. This is seen in Table 2-1, which notes each group's mean RMS error during these two periods of tracking. The individual subjects' errors were examined more closely. Paired-difference t-tests confirmed that for this breakdown of the session (into earlier and later twelve-minute sections) there were no statistically significant changes in the error levels in any of the bandwidth conditions.

For an overall assessment of the error outcome, an analysis of variance (ANOVA) was carried out on the RMS errors from all the groups. Main effects were sought for the *Bandwidth* treatment (Adaptive, Easy, or Difficult) and for the amount of *Practice* (represented by the block number of the experiment). The ANOVA is summarised in Table 2-2. The analysis confirmed that there were certainly differences among the RMS errors attributable to the Bandwidth condition ($F_{2,44}$=13.8, p<.001). While Practice was not seen to have a general effect on the scores, the interaction of Bandwidth X Practice was significant ($F_{18,44}$=2.27, p<.01). Further one-way ANOVAs were applied to each group individually to identify any contribution

TABLE 2-1

TRACKING EXPERIMENT 2.
Effect of practice on the RMS error.

	mean normalised RMS error	
Bandwidth condition	*Blocks 3-6*	*Blocks 7-10*
Easy	.463	.488
Difficult	.784	.775
Adaptive	.825	.846

TABLE 2-2

TRACKING EXPERIMENT 2.
ANOVA Summary: Effects of bandwidth and practice on the average normalised RMS error.

Source	*df*	*SS*	*MS*	*F*	*Prob*
Between subjects					
Bandwidth (B)	2	4.394	2.197	13.8	0.0003 *
Subjects	16	2.548	0.159		
Within subjects					
Practice (P)	9	0.034	0.004	0.58	0.8102
B X P	18	0.267	0.015	2.27	0.004 *
Error Within	144	0.942	0.007		

* significant at $p<0.05$

of Practice. Only in the Difficult task could Practice account for changes in the RMS error ($F_{9,0}$=2.57, p<.02), an outcome reflecting the large decreases in the error made during the earliest minutes of this group's training. It was presumed that this warm-up effect was responsible for the B X P interaction in the Table 2-2 summary.

It is clear that practice did not produce dramatic benefits in nonadaptive tracking performance; in the Adaptive condition practice did not serve to reduce the error either. However, Figure 2-9 does suggest that during the final twelve minutes the Adaptive error remained essentially stable. It is important to recognise that during this period of near stability in the Adaptive condition error, the tracking demand (as defined by the bandwidth) was becoming steadily harder. Thus the Adaptive subjects were learning to cope during this time with higher and higher bandwidths with little effect on their error.

The average Adaptive bandwidth reached the same level as the fixed-bandwidth setting of the Difficult task in block 8 of the experiment. Yet apart from the very earliest minutes of the session, the Adaptive RMS error remained the same or slightly larger than the Difficult error even during the major portion of the session when the average Adaptive bandwidth was below the bandwidth of the Difficult task. This observation suggests that the frequent bandwidth transitions made the Adaptive task relatively complex even at low bandwidths. It appeared to be at least as challenging as a nonadaptive version of the task which relied only on a persistently elevated bandwidth to impose the tracking load.

Subjective estimates of workload

Subjects in each group were asked to make three ratings of task difficulty using a simple rating scale. The mean rating results are shown in Figure 2-10. The most striking characteristic of the ratings was that they indicated that all the subjects found their tracking task increasingly difficult as the session progressed. This effect was not confined to the Adaptive condition, and presumably was an effect of fatigue. When tested immediately following the first three minutes of tracking (block 1), the ranking of the subjective measures exactly coincided with the ranking of the RMS error results, showing that the Difficult subjects found the task most demanding, followed by the Adaptive trackers, with Easy subjects remaining least concerned. In later ratings, the Adaptive group declared their task to be the most difficult, but the Difficult group judgments stayed surprisingly low, not even exceeding the assessments in the Easy condition. Analysis of variance confirmed that subjects demonstrated a general tendency to judge later blocks of tracking as more difficult (Practice: F_{24}=11.17, p<.001). However, the ratings were too variable within each bandwidth group to establish by this means that the differences among the groups were attributable to the type of tracking task. The ANOVA summary is found in Table 2-3. Further checks were carried out using the Mann-Whitney U-test to compare just the Adaptive and the Difficult ratings at each of the three surveys.

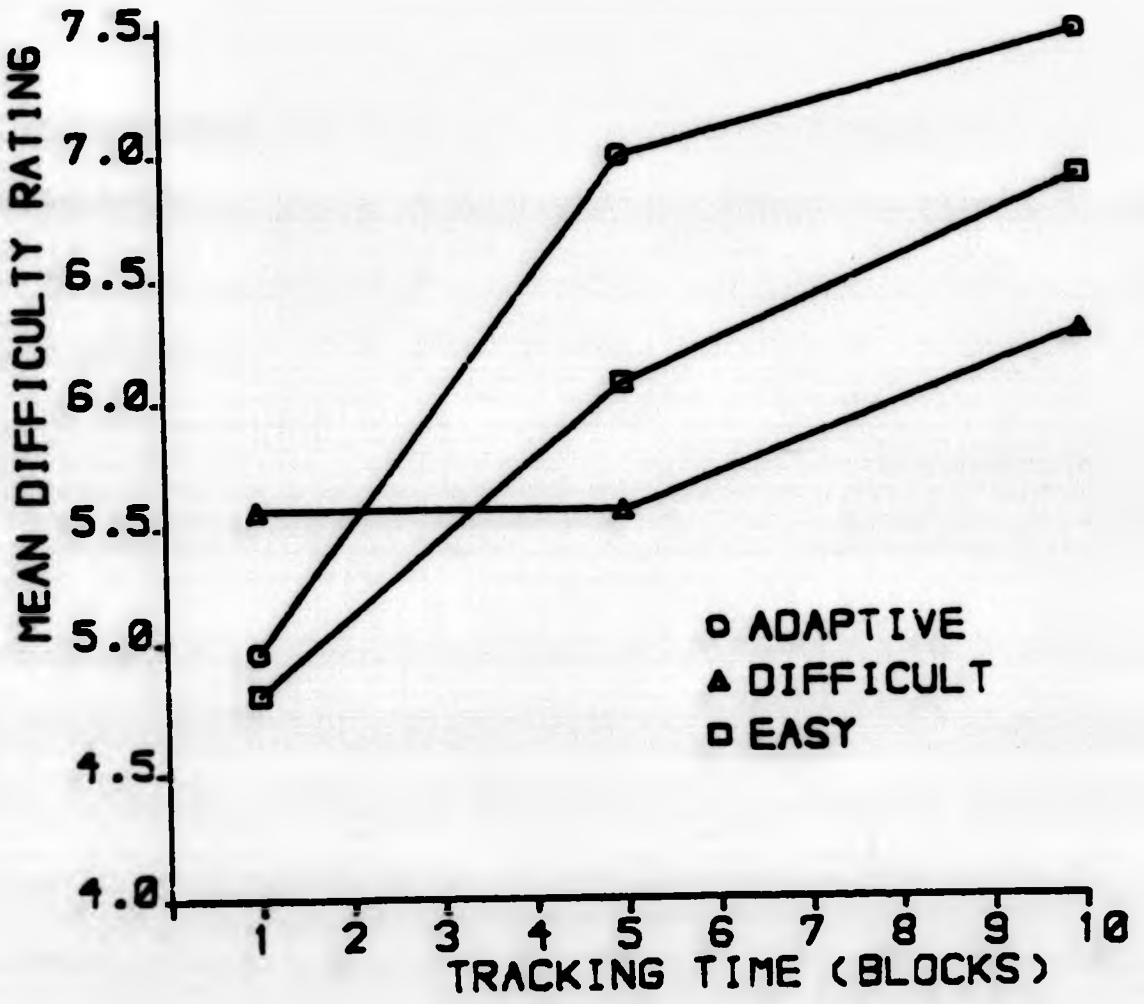

FIGURE 2-10 Tracking Experiment 2. Subjective difficulty ratings.

TABLE 2-3

TRACKING EXPERIMENT 2.
ANOVA Summary: Effects of bandwidth and practice on the subjective difficulty rating.

Source	*df*	*SS*	*MS*	*F*	*Prob*
Between subjects					
Bandwidth (B)	2	5.037	2.518	0.51	0.6087
Subjects	16	78.68	4.918		
Within subjects					
Practice (P)	2	29.974	14.99	11.17	0.0002 *
B X P	4	8.409	2.102	1.57	0.2069
Error Within	32	42.925	1.341		

* significant at $p<0.05$

None of the differences between these two conditions could be shown to be statistically reliable.

Discussion

The results confirmed that the tracking RMS error tends to become greater when the forcing function bandwidth is increased, whether the task is adaptive or fixed-bandwidth. Of the three bandwidth conditions examined here, the adaptive task resulted in the largest errors overall, indicating that this task was certainly difficult. None of the groups were able to reduce their error substantially through practice. Under nonadaptive conditions, it seemed that the choice of forcing function bandwidth determined the level of tracking error and that the subjects could not (or would not) improve much upon this. The adaptive task appeared to be unavoidably difficult regardless of the bandwidth value at any particular moment. The adaptive subjects did seem to be acquiring increasing skill as they worked because they were able to maintain a near-constant level of error even as they progressed to higher bandwidth tracks. The subjective ratings of task difficulty suggested that the adaptive task was the most challenging, but this result was not statistically significant. The usefulness of the subjective ratings was also compromised by their susceptibility to fatigue effects.

A third experiment was conducted in an attempt to measure the tracking workload more clearly. Adaptive and nonadaptive tracking performance were again compared, but this time a secondary measure of workload was taken.

Secondary tasks have often been used to gauge the availability of spare processing capacity during performance of a dedicated primary task (Pew, 1979; Rolfe, 1973). The basic premise of such techniques is that capacity is limited and that the subject's ability to service the secondary task reflects the extent of primary task demands on the same resources. Because tracking requires fairly continuous execution of finely measured movements, other skills which require precisely timed motor control are normally considered to be the most appropriate secondary tasks for the measurement of tracking workload.

Michon (1966) proposed that tapping regularity could provide an index of spare capacity during perceptual-motor performance. If subjects are asked to tap a switch as regularly as possible, they generally adopt a personal rate of two or three taps per second, which also approximates the duration most accurately reproduced by subjects in experiments on interval production (Michon, 1967). The individual tapping rate becomes nearly constant after practice. The task requires precisely timed movement, and has been shown to be sensitive to other ongoing sensorimotor processes. When another task disrupts the necessary timing processes the tap rate becomes irregular. Johannsen, Pfendler, and Stein (1976) demonstrated that a tapping index of this sort was a useful indicator of workload in a manual control task which incorporated tracking skills (a simulated aircraft landing task).

Michon (1966) pointed out that the variance of the tapped interval is not adequate as a measure of irregularity, because it may not be sensitive to very suggestive but extremely local instabilities. For example, the variance may not distinguish between a sequence comprising highly irregular taps and a sequence containing relatively stable intervals which encompasses a slow temporary shift in the mean. To better capture the transitional properties of successive taps, he recommended that the absolute values of the differences between successive intervals be scored. Michon suggested summing these absolute differences for a measure of average irregularity:

$$\sum|\Delta t|$$

where t is the interval between two successive taps. Because different subjects will choose different mean tapping rates, it is useful to correct the scores so that they will be comparable across subjects regardless of the average interval length t_{mean}. The corrected measure becomes

$$\frac{\sum|\Delta t|}{t_{mean}}$$

In practice, it is easiest to obtain t_{mean} by dividing the total tap sequence duration T by the number of taps N. Substituting, the computational formula for the *tap level* is thus

$$\frac{N}{T}\sum|\Delta t|$$

This standard tap level score increases when the tap rate becomes more irregular.

TRACKING EXPERIMENT 3: MENTAL WORKLOAD DURING TRACKING

Procedure

In this experiment, two indices of mental workload were obtained from subjects during a primary tracking task. The first measure was the same subjective difficulty rating scale which was used in Experiment 2. The second index was derived from performance in a secondary finger-tapping task. The loading measures were obtained under both adaptive and nonadaptive tracking conditions.

The primary task was the compensatory visual tracking task used in previous experiments. Three forcing function bandwidth conditions were studied. A separate groups design was used, so the subjects were divided among three groups. Sixteen undergraduate students volunteered as subjects. The *Adaptive* group (six subjects: four male, two female) was instructed to carry out the standard adaptive tracking task. The other two groups received an unchanging bandwidth.

Five subjects (three male, two female) were assigned to an *Easy* condition, like the one used in Tracking Experiment 2. This group was asked to track at a fixed bandwidth of 0.5 Hz throughout. Finally, the *Matched* group (five subjects: three male, two female) performed a fixed bandwidth task having the same average difficulty as the Adaptive group's task. This condition was intended to be similar to the Difficult task in Experiment 2; but to improve comparability with the adaptive data, the bandwidth was not just set to an arbitrarily high level. The choice of Matched bandwidth was determined only after all the Adaptive results had been obtained. The running mean bandwidth for each Adaptive subject was noted at the end of his session, and the overall mean for the whole Adaptive group was calculated. The result represented the average bandwidth setting during the Adaptive condition, and this value was used as the fixed level for the Matched condition.

The secondary task, performed by all subjects under all bandwidth conditions, was Michon's self-paced keytapping task. A response switch was provided alongside the tracking apparatus in the subject's cubicle. This was a sensitive Morse keyswitch (requiring 40 gm pressure to close a 0.2 mm gap) which the subject positioned comfortably beneath the left hand. To ensure that the subject would be aware of each keypress, the subject was asked to don headphones for this experiment and each depression of the key produced a distinct click in the headphones. The click was produced by a 1 KHz sinewave generator feeding a 70 dBA signal through an electronic switch which, when gated by the keypress, passed an 80 millisecond soundburst to the headphones. The time of each keypress was recorded by the PDP-11 computer. The subject was directed to tap the key throughout the session, choosing a tap rate satisfying personal preference but with the aim of spacing the taps as consistently as possible. It was explained that the regularity of the taps was being measured and that he was to try to maintain the regularity while tracking.

The primacy of the tracking task was established by stressing to the subject that the tracking was most important and could not be neglected, although it was necessary to tap throughout as well. The subject's wristwatch was removed as a precaution against overt timekeeping.

The experiment contained twelve three-minute blocks, for a total session duration of 36 minutes excluding rest breaks. The first three blocks were devoted to practising the tapping task alone. Starting with block 4, the tracking and tapping tasks were performed concurrently. Rest breaks were granted every two blocks. After eight of the dual-task blocks, the session was completed with a final tap-only block.

The subject was asked to fill in the task difficulty rating scale (Figure 2-7) at three times during the dual-task condition (following blocks 5, 8, and 11).

Results

Tracking task performance

As before, the Adaptive subjects progressed to high forcing function bandwidths. The bandwidth adaptations are depicted in Figure 2-11. The mean bandwidth during the final three minutes of tracking was 0.885 Hz, which was very close to the values achieved by the adaptive subjects in the previous experiments after comparable amounts of practice. The rate of bandwidth change (indicated by the slope of Figure 2-11) was similar to the earlier studies.

The overall mean bandwidth in the Adaptive condition (all subjects, all blocks) was 0.775 Hz. This estimate of the average forcing function became the fixed-bandwidth setting used for the Matched group.

Figure 2-12 shows the normalised RMS error history for all three tracking conditions. In general, the Matched subjects suffered from the largest errors and the Easy trackers had the smallest, with the Adaptive errors falling in between. During the Adaptive training the error increased along with the bandwidth, as it had done in the earlier experiments. In both of the previous experiments the Adaptive subjects had needed at least 20 minutes' practice before the steady rise in error began to abate. In this experiment the total tracking

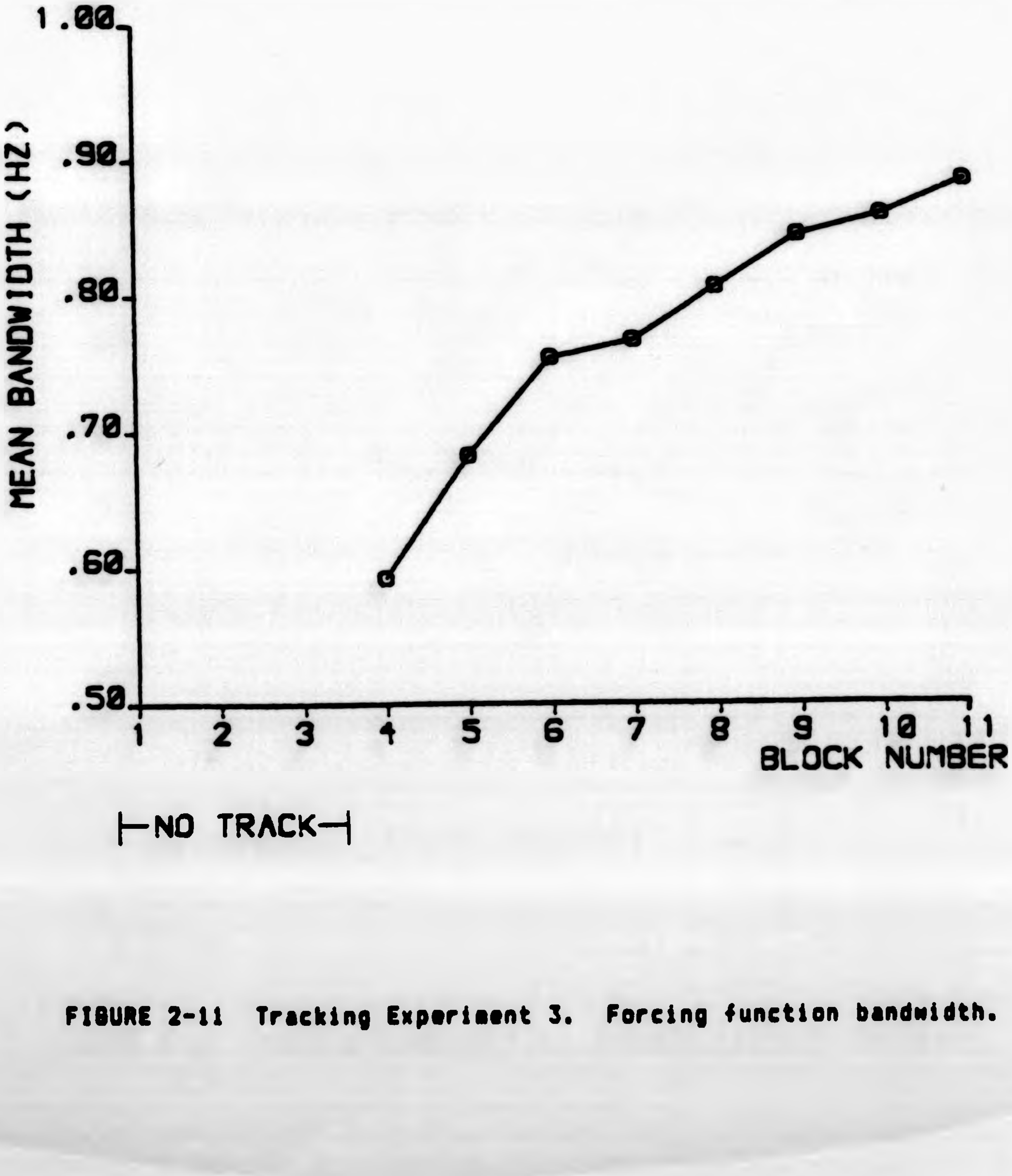

FIGURE 2-11 Tracking Experiment 3. Forcing function bandwidth.

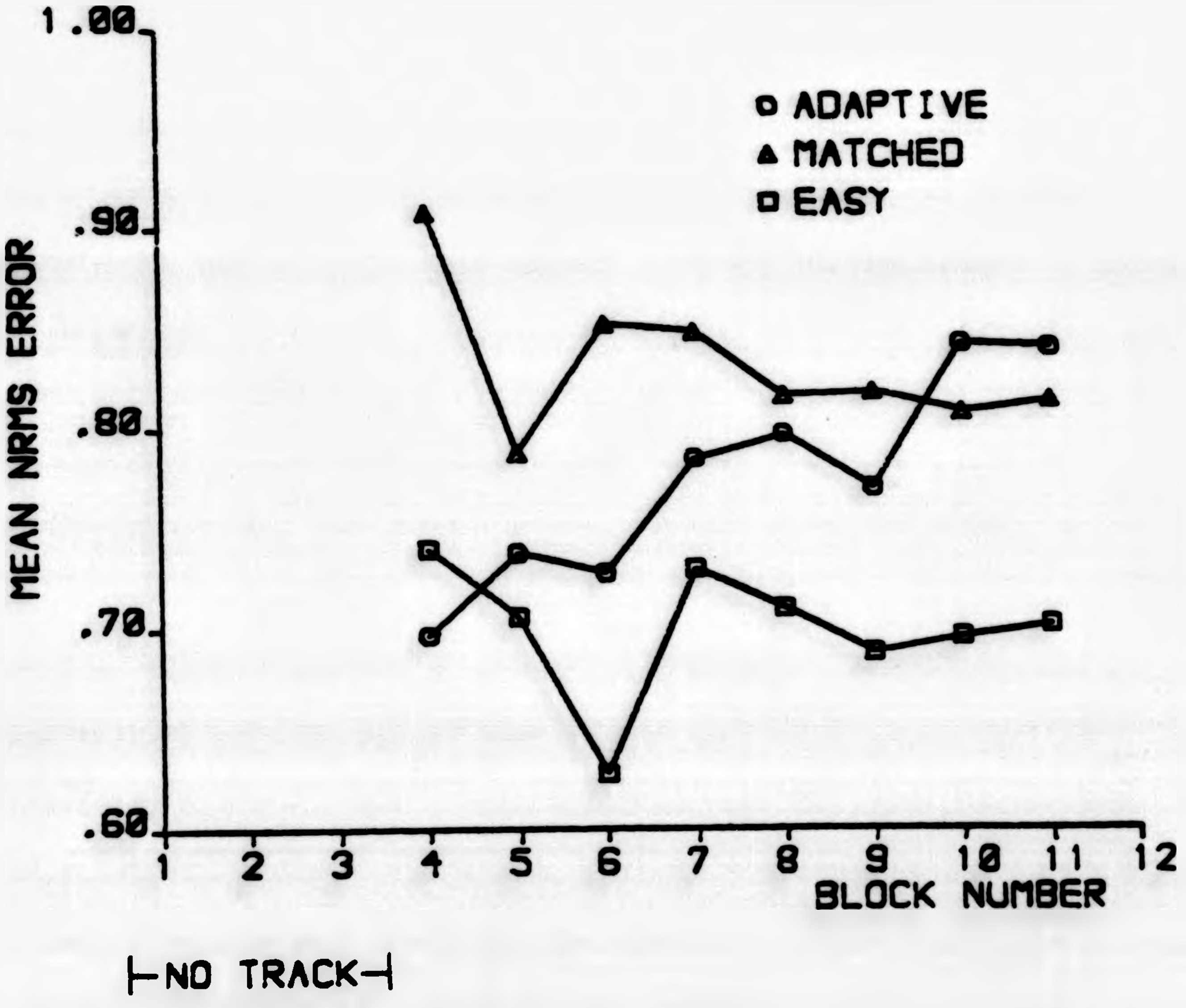

FIGURE 2-12 Tracking Experiment 3. Tracking normalised RMS error.

time was limited to 24 minutes. Figure 2-12 shows that the Adaptive error was increasing during all but the last few minutes of the session. In both the Matched and the Easy tasks the error exhibited only a slight inclination, if any, to diminish with practice. In these nonadaptive conditions the error was least consistent during the earliest blocks of tracking, as if the secondary tapping requirement had destabilised the tracking. Interference from the secondary task could also have been the reason for the slightly higher average level of error throughout the fixed-bandwidth tracking conditions in this experiment, when compared to the overall results for Tracking Experiment 2. The difference between the Matched error in this experiment and the error previously recorded for the Difficult group in Experiment 2 was not statistically significant, but the Easy task certainly did show a larger overall error here (that is, carried out in conjunction with the tapping task) than it did in the Easy condition of the earlier study ($t_{9df}=3.38$, $p<.02$, 2-tailed).

An analysis of variance of the error data from all groups could identify no general main effect of the Bandwidth group, nor of Practice; but there was a significant Bandwidth X Practice interaction ($F_{14df}=2.17$, $p<.02$). The ANOVA is summarised in Table 2-4. The B X P interaction probably reflected the tendency of the Adaptive error to change in response to the bandwidth movements. This interpretation of the interaction was supported by one-way ANOVAs of each group's errors, which showed that only in the Adaptive condition did error scores change reliably with Practice ($F_{7df}=2.98$, $p<.02$), with no significant Practice effects found in either nonadaptive condition.

TABLE 2-4

TRACKING EXPERIMENT 3.
ANOVA Summary: Effects of bandwidth and practice on the average normalised RMS error.

Source	*df*	*SS*	*MS*	*F*	*Prob*
Between subjects					
Bandwidth (B)	2	0.348	0.174	1.46	0.2682
Subjects	13	1.552	0.119		
Within subjects					
Practice (P)	7	0.044	0.006	1.24	0.2913
B X P	14	0.155	0.011	2.17	0.0146 *
Error Within	91	0.463	0.005		

* significant at $p<0.05$

The Matched error remained larger than the Adaptive error throughout most of the session. Examination of the bandwidth history in Figure 2-11 reveals that by block 7 of the experiment the mean Adaptive bandwidth had exceeded the value of the fixed setting used in the Matched task. Beyond this point, the RMS error in the Adaptive and Matched groups was, in average terms, very similar.

Mental workload: Subjective estimates

The subjects' ratings of the task load are shown in Figure 2-13. The data suggested that, as in Tracking Experiment 2, the perceived difficulty of the task increased as the session lengthened. The Adaptive subjects again judged their task to be the most difficult. The Matched trackers (apart from their first rating) judged their task to be more demanding than the Easy subjects. However, the ANOVA for these results, described in Table 2-5, failed to confirm consistent effects of Bandwidth or Practice on the ratings. The trends suggested by Figure 2-13 were not sufficiently robust to overcome the variance contributed by individual differences. The groups were analysed separately, but despite the apparent increases in the ratings over the course of the experiment, the ANOVAs failed to support any reliable effects of Practice on the difficulty ratings within any condition. So the confident interpretation of these subjective estimates of task difficulty remains difficult, as indeed it had been in the previous experiment.

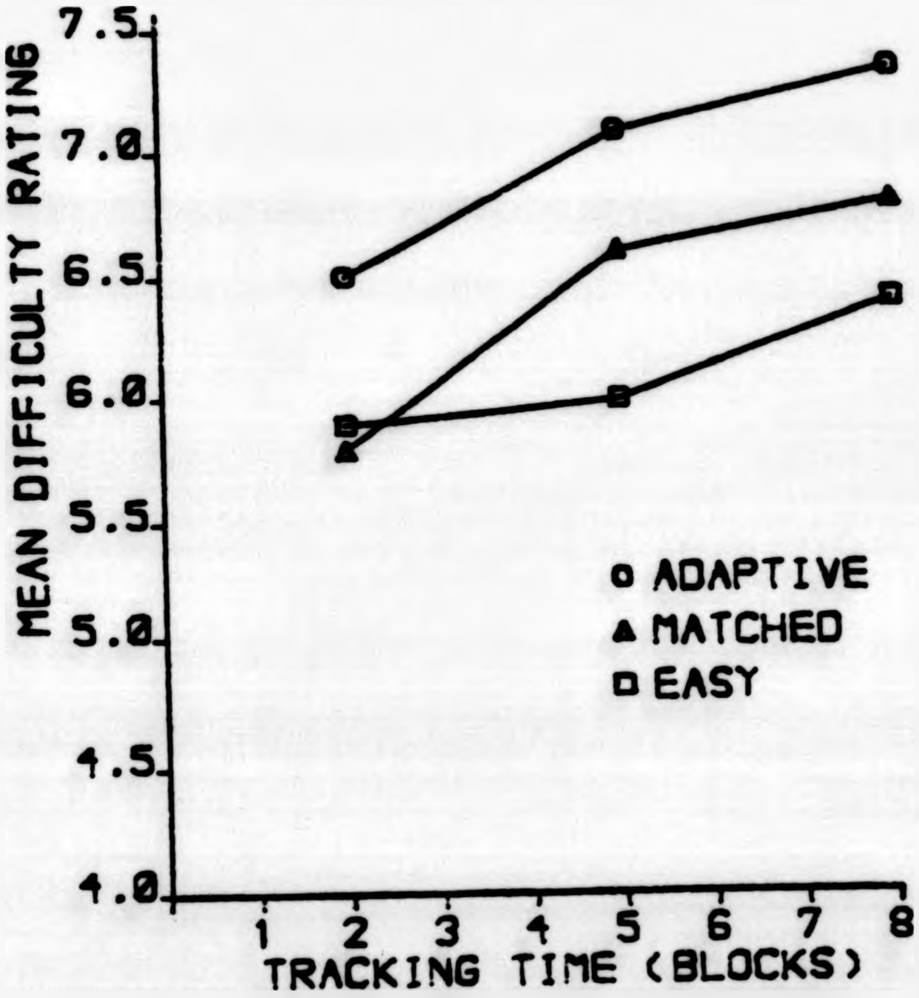

FIGURE 2-13 Tracking Experiment 3. Subjective difficulty ratings.

TABLE 2-5

TRACKING EXPERIMENT 3.
ANOVA Summary: Effects of bandwidth and practice on the subjective difficulty rating.

Source	*df*	*SS*	*MS*	*F*	*Prob*
Between subjects					
Bandwidth (B)	2	6.543	3.271	0.61	0.5562
Subjects	13	69.269	5.328		
Within subjects					
Practice (P)	2	4.922	2.461	2.91	0.0722
B X P	4	0.684	0.171	0.20	0.9348
Error Within	26	21.972	0.845		

Mental workload: Secondary task (tapping) measures

Before starting to track, subjects practised the tapping task for nine minutes (three blocks) to allow the personal tap rate to settle. Within a few minutes of practice the tap level had stabilised around a low value with good agreement across all three experimental groups. The tapping regularity was immediately affected when joined by the primary tracking task. The changes are apparent in Figure 2-14, which plots the tap level scores block-by-block. As soon as tracking began (at block 4), the tap levels rose to high values, indicating that the tap interval had become unstable. The increase in tap level between block 3 (last tap practice) and block 4 (first tap with tracking) was largest in the Adaptive condition, and the change was highly significant (t_{5df}=3.46, p<.01, one-tailed). The increase in the nonadaptive Matched condition scores was also reliable (t_{4df}=2.21, p<.05), as was the change observed in the Easy group (t_{4df}=3.86, p<.01).

The disturbance of the tap level persisted until the tracking finished in block 11. Figure 2-14 does suggest that the tap irregularity diminished a little over the course of testing, especially in the nonadaptive conditions. An ANOVA of the tap scores from the dual-task portion of the session (blocks 4 through 11) was carried out to determine whether the group differences in tap level could be attributed to the Bandwidth conditions, and to identify any general

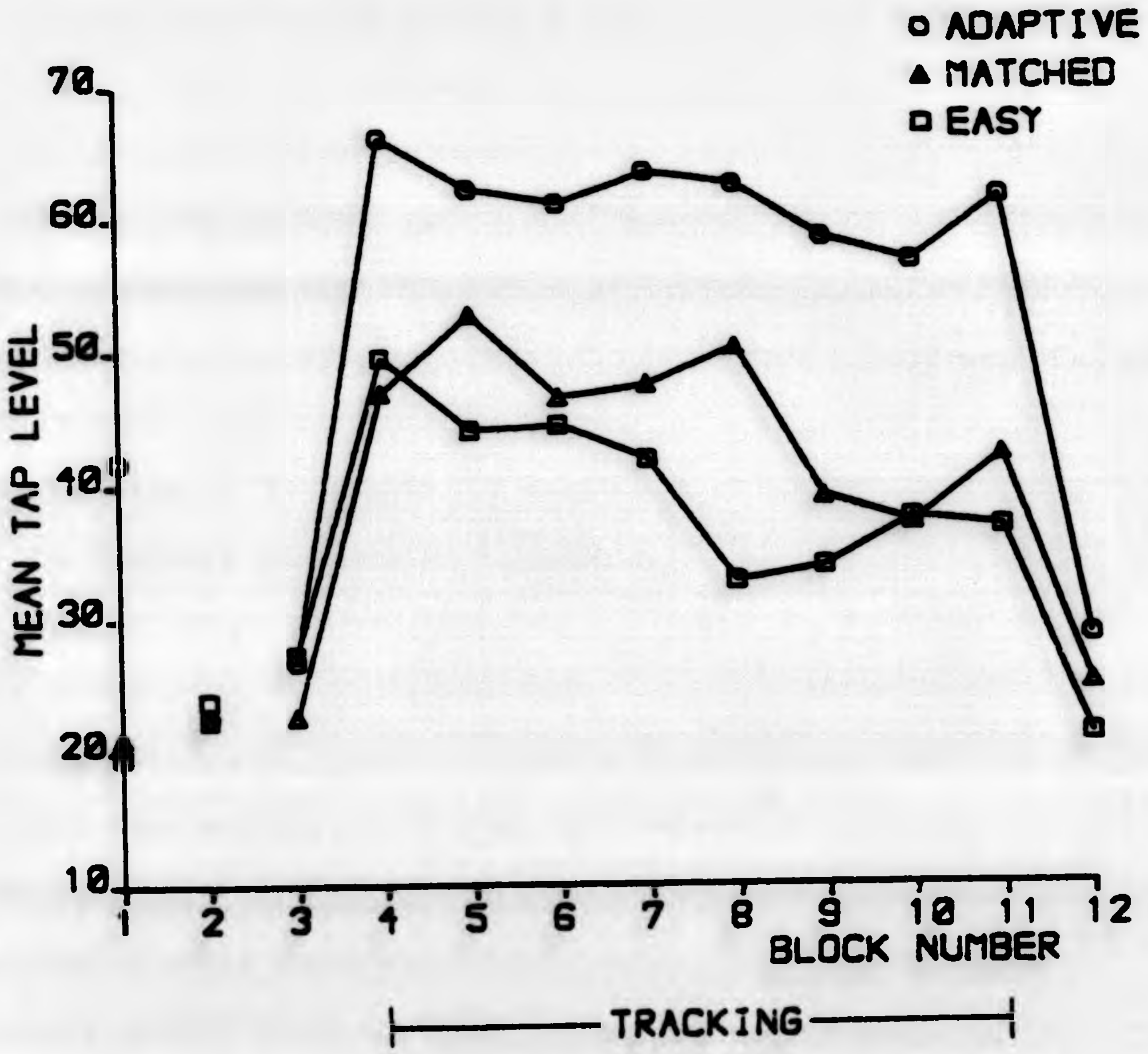

FIGURE 2-14 Tracking Experiment 3. Tap level in the secondary task.

tendency of Practice to reduce the tap level. The results of this analysis are shown in Table 2-6. No significant affect of Bandwidth could be confirmed. The analysis did find that the tap levels were altered by Practice ($F_{7,84}$=3.22, p<.004). Further investigation suggested that the Easy data were perhaps largely responsible for this Practice outcome. One-way ANOVAs carried out separately on each of the three groups showed that only the Easy tap levels were reliably altered by Practice; the effect was highly significant ($F_{7,84}$=4.98, p<.001). Figure 2-14 does imply that the tapping rate was more irregular during the Adaptive task than in the Matched tracking. To test this, another ANOVA was carried out in which the Adaptive and Matched conditions were compared to each other with the Easy data excluded. In this analysis, described in Table 2-7, neither Bandwidth nor Practice were found to have statistically significant effects on the tap levels. So despite the suggestion of greater tap irregularity in the Adaptive data, the differences were not sufficiently robust to allow statistical confirmation that the Adaptive tap levels exceeded the irregularity in the Matched data.

In the final minutes of the experiment (block 12), subjects again tapped without the tracking task and the tap levels returned to values similar to those noted before tracking had been imposed. The (post-tracking) final tap level in block 12 was compared to the (pre-tracking) baseline level of block 3. The tap irregularity did not differ significantly (according to t-tests) in any of the experimental groups. Thus the secondary task performance recovered completely with the removal of the primary tracking task.

TABLE 2-6

TRACKING EXPERIMENT 3.
ANOVA Summary: Effects of bandwidth and practice on the mean tap level.

Source	*df*	*SS*	*MS*	*F*	*Prob*
Between subjects					
Bandwidth (B)	2	11149	5574	1.13	0.3519
Subjects	13	63960	4920		
Within subjects					
Practice (P)	7	1710	244.4	3.22	0.0043 *
B X P	14	907.2	64.80	0.85	0.6106
Error Within	91	6911			

TABLE 2-7

TRACKING EXPERIMENT 3.
ANOVA Summary: Effects of bandwidth and practice on the mean tap level. Adaptive and Matched groups only.

Source	*df*	*SS*	*MS*	*F*	*Prob*
<u>Between subjects</u>					
Bandwidth (B)	1	5730	5730	0.95	0.3547
Subjects	9	54166	6018		
<u>Within subjects</u>					
Practice (P)	7	1138	162.7	1.72	0.1205
B X P	7	291.1	41.58	0.44	0.8735
Error Within	63	5958	94.58		

Discussion

The data confirmed that Adaptive tracking imposed a large perceptual-motor workload. The load was at least as great as that imparted by a high bandwidth nonadaptive task. Both the subjective ratings and the tap disturbance data hinted that the adaptive tracking was more demanding than either of the fixed-bandwidth conditions tested here, but the evidence was not strong enough to provide statistical confirmation. This was probably due mainly to the small numbers of subjects tested.

It is important to acknowledge that any comparison of adaptive and nonadaptive tasks is difficult to interpret. Because one of the tasks is adaptive, the differences between the tasks probably outweigh their similarities. For example, although in this experiment the Matched bandwidth was an attempt to simulate the "average difficulty" of the Adaptive forcing function, the Matched track would have been considerably harder to control during the earlier blocks of the session. The effects of this difference on the subject's set and the subsequent course of training are unknown. Kelley (1971) cautioned:

> The question of whether adaptive training is effective, compared with fixed training, has no answer. Whenever you test this question in the laboratory, you are always comparing a particular adaptive training system with a particular fixed-difficulty training system. One or the other may be better, but this says

> nothing about adaptive training as such...
> You simply cannot assess adaptive training per se on the basis of experiments which purport to compare adaptive versus fixed training. What you can do is ask, 'Is this particular adaptive system, in this or that situation, better than some particular fixed system for the training of a certain skill?' (pp.24-5)

Adaptive tracking: Conclusions

In conventional nonadaptive compensatory tracking tasks, skilled performance is usually acquired only slowly after much practice. The experiments described in this chapter confirmed that practice gains may certainly be very slight during training sessions of realistic length. In contrast, by using the adaptive tracking procedure, subjects readily learned to track high-bandwidth forcing functions with consistent error rates. The adaptive algorithm was designed to ensure that each subject individually set the standard of accuracy against which performance was judged. As a result, some subjects tracked with more precision than others. The procedure rewarded sustained effort with a more challenging track, without making adaptation dependent on the absolute accuracy of the subject's control strategy.

Undoubtedly the adaptive tracking task imposed a considerable perceptual-motor processing load, relative to compensatory tracking tasks in general. By tailoring the task difficulty to individual effort, the adaptive process provided an efficient form of training. This benefit, combined with the resultant high workload, makes the task suitable for experimental investigations of distractibility. Assuming that tasks which require much processing capacity are most likely to show the effects of additional processing demands (such as distractor analysis), the adaptive procedure should be appropriate for creating the sort of tracking task which is vulnerable to distraction effects. Just how susceptible tracking skill is to extraneous, involuntary distraction is the issue explored in the next chapter.

CHAPTER 3: EFFECTS OF AUDITORY DISTRACTION ON TRACKING

Overview

Three experiments were conducted to explore the effects of intermittent auditory distraction on visual tracking performance. It was demonstrated that loud noisebursts increased tracking RMS error during the earliest exposures and lengthened the tracker's response delay, but only when the noises were infrequent and irregularly separated in time. Even then, the disruption of tracking performance was a brief phenomenon coincident with each noiseburst rather than a lengthy disturbance. Moreover, the magnitude of the effect was influenced by the momentary demand of the tracking task. Repeated presentation of distractors led to very rapid diminution of their effect. This was especially noticeable if the distractors were frequent and arrived at fixed intervals: under such conditions main effects of distraction on error or delay were not significant. Distractors which took the form of verbal messages also produced an increase in the tracking RMS error. The effect of verbal distraction did not seem to be as localised as the effect of noise and it was more persistent in the face of repeated exposures. Again, any local effects were likely to be modulated by the prevailing tracking task demands. Subjective awareness of distractibility was generally minimal and shortlived, suggesting that subjects' own reports, while qualitatively appropriate, were probably not very precise indicators of the magnitude or duration of distraction effects.

It was concluded that tracking performance (mainly as gauged by the

RMS error) was susceptible to disruption by distraction, but that the effects were neither large nor consistent. Distractors were unlikely to produce noticeable effects unless they were unpredictable in a temporal or acoustic or informational sense. It was proposed that verbal distractors receive more persistent analysis than simple noise intrusions, leading to more sustained and diffuse effects on performance.

Introduction

The distractibility experiments assumed that distraction effects would be more detectable if the subject retained very little spare information processing capacity to divert to secondary information sources such as distractors. It was expected that a subject performing a task which demands much capacity, such as adaptive tracking, would be likely to show changes in task performance when a distractor intruded.

The evaluation of distraction effects during tracking proved to be problematic for several reasons. One difficulty, which had already been documented in the early literature on distractibility (such as Cassel and Dallenbach, 1918; Harmon, 1933; Kryter, 1950; Pollock and Bartlett, 1932; Woodworth and Schlosberg, 1954), was usually described as *adaptation*. In general, this term has referred to the tendency of any distraction effect to become less measurable with each additional presentation of the distractor. Because of adaptation, most experiments have suffered from the necessary constraint that distractors should be rare and intermittent.

Pilot data supported this conclusion. The irregularities in the tracking record which looked like plausible distractor-related effects (such as overshoots or response delays) tended to be inconsistent and unstable over multiple trials. Statistically this caused difficulty because small sets of distraction trials contained too much within-subject variability to be convincing. Attempts to overcome noisy results by testing the effects of many distractors failed because the effects would not persist through a sufficient number of trials. That is, many distractors seemed to invite adaptation. Consequently the experimental outcome was too often an unsatisfying collection of selective, anecdotal incidents.

In these preliminary studies, the distraction effects seemed to be brief transient deviations of control (of perhaps a second's duration or even less) embedded in much sampling noise. Sampling noise was always present because even if certain short-lived deviations appeared to be related to distractor scheduling, similar joystick events would also occur at other unrelated times. The response record would contain transient overshoots, for example, at times which were far removed from distraction as well as when a distracting event occurred. While it was often tempting to argue that a particularly compelling joystick event was a reaction to distraction, it was generally difficult because of the sampling noise to extend the observation across successive trials by that subject, and no less difficult to pool data across different subjects.

It seemed inevitable that unwanted transient effects would occur under these conditions because of uncontrolled factors within the human half of the tracking control loop. The difficulties stemmed from the dynamics of the track disturbance, or forcing function. To understand this, consider the subject's response to the forcing function at any given moment. The forcing function is, in these experiments, a time-varying random voltage taken from a Gaussian noise generator. Its moment-to-moment changes in amplitude determine the unstable target condition which the subject must track. Sometimes these amplitude changes are large and occur in rapid succession and the subject will tend to track them badly. Sometimes the changes will occur relatively slowly and by small amounts and the subject will tend to track these more easily. The *average* task difficulty is determined in large measure by the bandwidth of the forcing function signal, in this context a measure of the high-frequency components of the noise source. As the bandwidth increases, the track becomes more unstable and its movements more unpredictable. Having said this, it is important to recognise that the signal's randomness is measurable only as a function of time, so the *momentary* difficulty, and hence the subject's momentary error, will vary depending on what the forcing function happens to be doing at that moment in time. This is why tracking performance measures have traditionally been averaged over some reasonable time interval (say, several seconds at least): to avoid measuring a momentary error which is not a meaningful indicator of average task difficulty.

This has at least three implications for the measurement of distraction effects on tracking performance. First, if meaningful measures of tracking performance are necessarily ones that are averaged over time, they will tend to be insensitive to very brief control effects just as they are (intentionally) insensitive to the local forcing function effects. Measures which are not time-averaged will be noisy measures exhibiting much variance. This is a major problem: trying to single out one performance deviation as distraction-induced when it is surrounded by comparable variations induced solely by the forcing function dynamics. This suggests that a preferable scheme for detecting distraction effects would be one which uses an identical segment of the forcing function for both the distracted and nondistracted conditions. That is, the distracted tracking trials would be compared not with "just any" nondistracted performance, but rather with nondistracted trials using exactly the same sequence of forcing function events. Local forcing function effects would then be common to both conditions, but true distraction effects should only appear in the records of distracted trials. The second implication stemming from the time-varying nature of the forcing function is that the effect of an otherwise potent distractor might be negligible if it happens to occur at a time when the forcing function is comparatively well-behaved. If the track is not prompting the subject to make many control movements then a distraction effect measurable only in terms of altered control movement might pass undetected. This implies that we should try to schedule distractors during times when the tracking task is reasonably lively. Finally, it would seem sensible to give the subjects some time to practise the tracking task and learn about the system dynamics before applying distractors. The training serves two functions: it helps to reduce the within-subject variability which is typical of unpractised subjects and which threatens to swamp small distraction effects; also, appropriate pretraining can help to minimise the ongoing practice gains during testing which in many sensorimotor tasks may be substantial enough to counteract the temporary and only slightly deleterious effects of distraction (for example, Pollock and Bartlett, 1932).

The three experiments described in this chapter did use identical segments of forcing function for both the distracted and nondistracted conditions. Before testing for distraction effects began, subjects practised tracking using the adaptive task to ensure efficient progress to an appropriate level of difficulty for each individual. Indirect methods were used to ensure that a reasonable level of joystick movement would be required during the test sessions. Most importantly, during testing the tracking task was held to a difficult level by using a high bandwidth forcing function determined by the previous adaptive training. (Although training was adaptive, during the distractibility measurements the bandwidth could not be adaptively altered for reasons which will be explained in the experimental procedure.) Also, in the first two experiments the control dynamics of the joystick were altered as part of training procedure to promote additional joystick activity.

The purpose of the first experiment was to demonstrate any effect of pure noise auditory distraction on compensatory visual tracking. The experiment presented subjects with distractors after they had become well practised at the tracking task. The second experiment followed the same procedure but employed a markedly different schedule of distractor presentation to try to alter the timecourse of adaptation. The third experiment attempted to verify some of the findings of the previous studies using an altered and somewhat abbreviated procedure. Apart from a reduction in the amount of tracking practice the most notable change in the procedure was the use of verbal distractors in place of noisebursts.

DISTRACTION EXPERIMENT 1: RANDOM NOISE DISTRACTORS

Procedure

The subjects were six undergraduate students (four females, two males) who responded to a notice-board advertisement. None had previous tracking experience. All claimed to have normal hearing. They were paid £2 for their assistance.

All the subjects performed three practice sessions of adaptive tracking administered on three consecutive days. Each session contained 18 minutes of tracking with a short rest break at the halfway point. The first two days' practice used the standard adaptive tracking task with a zero-order control law. On the third day the control law was changed to first-order dynamics to encourage the subjects to use the joysticks very actively.

Distractors were introduced during the final session, on the fourth day. Subjects were warned beforehand that noises would occur and a sample of the noise was demonstrated. The purpose of the noises was not explained before the experiment; the subjects were told simply that there was no necessity to pay attention to them. During this distractibility test session the tracking task forcing function was held to a fixed bandwidth so that exact features of the track could be reliably repeated. An appropriate bandwidth was determined for every subject individually. The bandwidth was selected to approximate the subject's adaptive practice asymptote and so ensure that the task demanded a reasonable amount of processing capacity. The bandwidth was set 10% higher than the subject's mean adaptive bandwidth outcome from the final day of training. Since the bandwidth of the forcing function was tailored to each subject's requirements, the sequence of signals in the track differed from subject to subject, and no two subjects received identical tracks. The procedural change from adaptive to fixed bandwidth tracking was not announced to the subjects and the tracking task during the test session was in other respects comparable to the preceding practice sessions, lasting 18 minutes with a mid-task rest break. The control law was first-order, as it had been for the last practice session.

For the distraction measurements, each subject was exposed to loud noises according to a predetermined *Random* schedule. The subjects heard distractors infrequently, separated by intervals which were pseudorandomly drawn from a set of predefined intervals. This irregular schedule provided 12 distractors with a mean onset-to-onset interval of 82 seconds. The same presentation schedule, prepared beforehand, was used for all of the subjects. The distribution of the intersignal intervals used is shown in Figure 3-1.

Each distractor was a burst of white noise lasting 2.5 seconds, initiated by the PDP 11 computer. The noise source was a Grason-Stadler broadband (10 KHz) noise generator. The burst was timed and shaped by a custom-built electronic switch which ensured a transient-free 20 millisecond rise and fall time for each burst. After switching, the noisebursts passed through a Quad amplifier for presentation by a loudspeaker placed about 1.5 metres in front of the subject. The noises were quite loud, measuring 82 dBA against an ambient level of 32 dBA, calibrated in free air using a Bruel and Kjaer sound-level meter.

Tracking was performed at a fixed bandwidth so that the hardware generating the forcing function could be restarted regularly to provide an accurately repeatable track. By this means the computer could replicate particular segments of the forcing function signal whenever required during the test session. This meant that an overlaying or average transient analysis could be applied to the test trials. Pilot studies had shown that this technique helped to make the effects of distractors more measurable by improving the

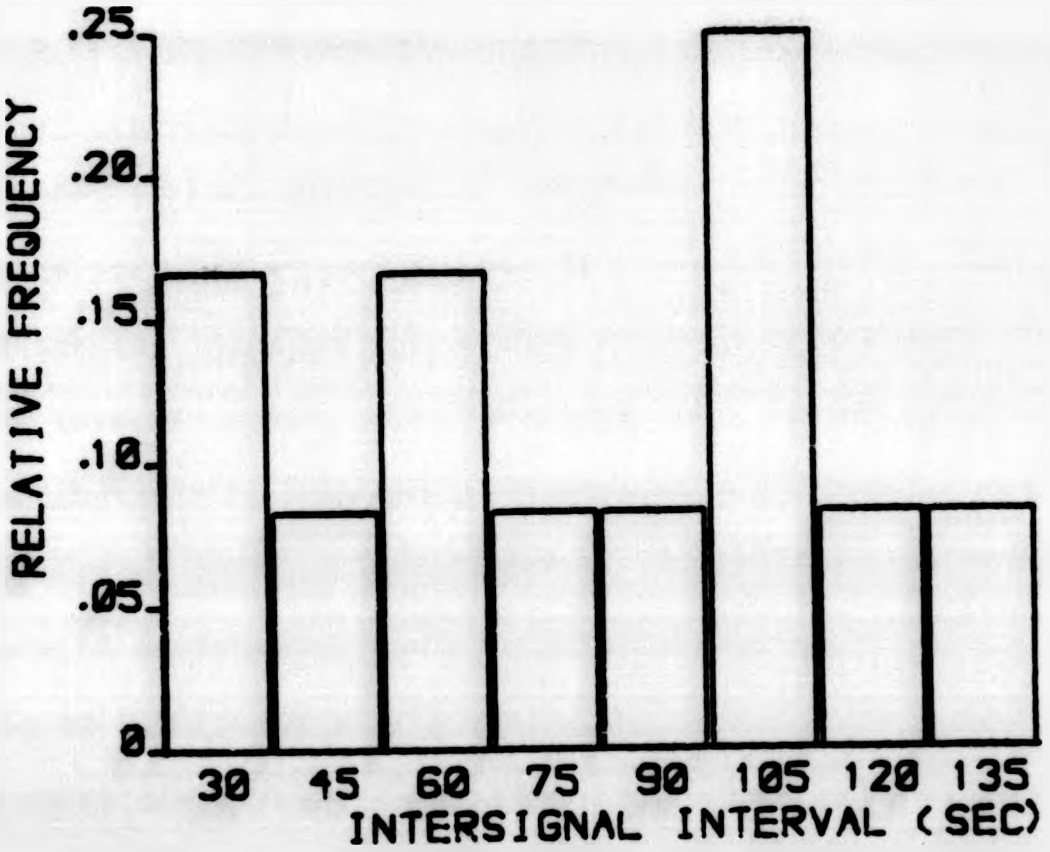

FIGURE 3-1 Distraction Experiment 1 (Random noise). Intersignal intervals.

signal-to-noise ratio of the effects. The reasoning behind the technique was established earlier in this chapter. When consecutive samples of tracking performance are taken, the subject's error exhibits variability because of fluctuations in the demands of the forcing function (sometimes the track is easy to follow, sometimes not) as well as because of irregularities in the skill of the subject. Afterwards, when studying small transient effects on performance, we seek only those differences arising from that crucial event, the distractor. It is useful to minimise the variability imparted by the continually changing demands of the track, because it may camouflage the variations caused by distractors. One effective way to achieve this is to offer the subject the same track each time.

By reusing certain portions of the track it was possible to make repeated measures of performance during a particular sequence of track disturbances. Because the forcing function events were the same across several trials, any differences seen in the joystick data would be attributable to variability within the subject, and not due to variations in the task requirements. A relatively noise-free average measure of performance could be obtained during the nondistracted trials by overlaying trials having the same forcing function. The same technique applied to trials during which a distractor event occurred provided a similar measure for the distracted condition. By comparing the subject's average tracking performance with and without a distractor during identical track events, it was expected that differences related to distractor occurrence would be more apparent than if data from unrelated track segments were compared.

In this experiment, the same forcing function events were recycled every 30 seconds. Subjects cannot normally recognize that a track is being repeated when it is of such long duration (Pew, 1974). Because all of the subjects had adapted to different signal bandwidths, the forcing functions were not the same across subjects, only within subjects. Precautions were taken to ensure that at the moment the signal restarted no transient conditions could appear on the displays to alert subjects to the technique; pilot subjects' reports confirmed that they were quite unaware that the track had been recycled. With each repetition of the forcing function several segments of the track were reliably regenerated: each segment appeared 24 times during the testing session. Data collection trials were always identified with one of these repeated sections of forcing function. In this experiment, two different track segments provided all the measurements. When a distractor occurred, it was always placed at a specific position within one of these forcing functions. Both forcing functions were tracked under distracted and nondistracted conditions. During the critical section of each trial the forcing function and joystick voltages were sampled every 100 milliseconds, these being subtracted to yield a measure of error. The full sampling window was six seconds wide. It extended from two seconds prior to the potential distractor onset until four seconds after onset.

In summary, tracking performance was measured many times but every trial used one of the two repeatable forcing functions, and examples of both distracted and nondistracted performance were accumulated for each forcing function. To achieve this, it was necessary to fix the forcing function bandwidth; so the adaptive task could be used prior to, but not during, the distractibility test trials. The method ensured that the closing adaptive practice trials and the subsequent fixed bandwidth test trials were experientially indistinguishable. The task presentation and performance requirements remained unchanged and there were no discontinuities in the signal sources or plant characteristics. The momentary track behaviour was similar in the two sets of trials, with the adaptive trials (operating to a bandwidth set-point comparable to the fixed level) imposing some additional variability on the upper bound of the track bandwidth as time passed. The adaptive and fixed trials were thus highly comparable.

Results

Training

During the practice sessions the subjects produced error scores consistent with previous observations of adaptive tracking. To assess performance, session data were tabulated into three minute blocks. The error is plotted block-by-block in Figure 3-2. The normalised root mean squared error (RMS error) settled between 0.7 and 0.9 volts during the period of practice using a zero-order control law. The error rose immediately when the control law changed to the more difficult first-order dynamics (third session: block 13) but the subjects' performance recovered gradually after this. The error levels and trends were comparable to other adaptive tracking results.

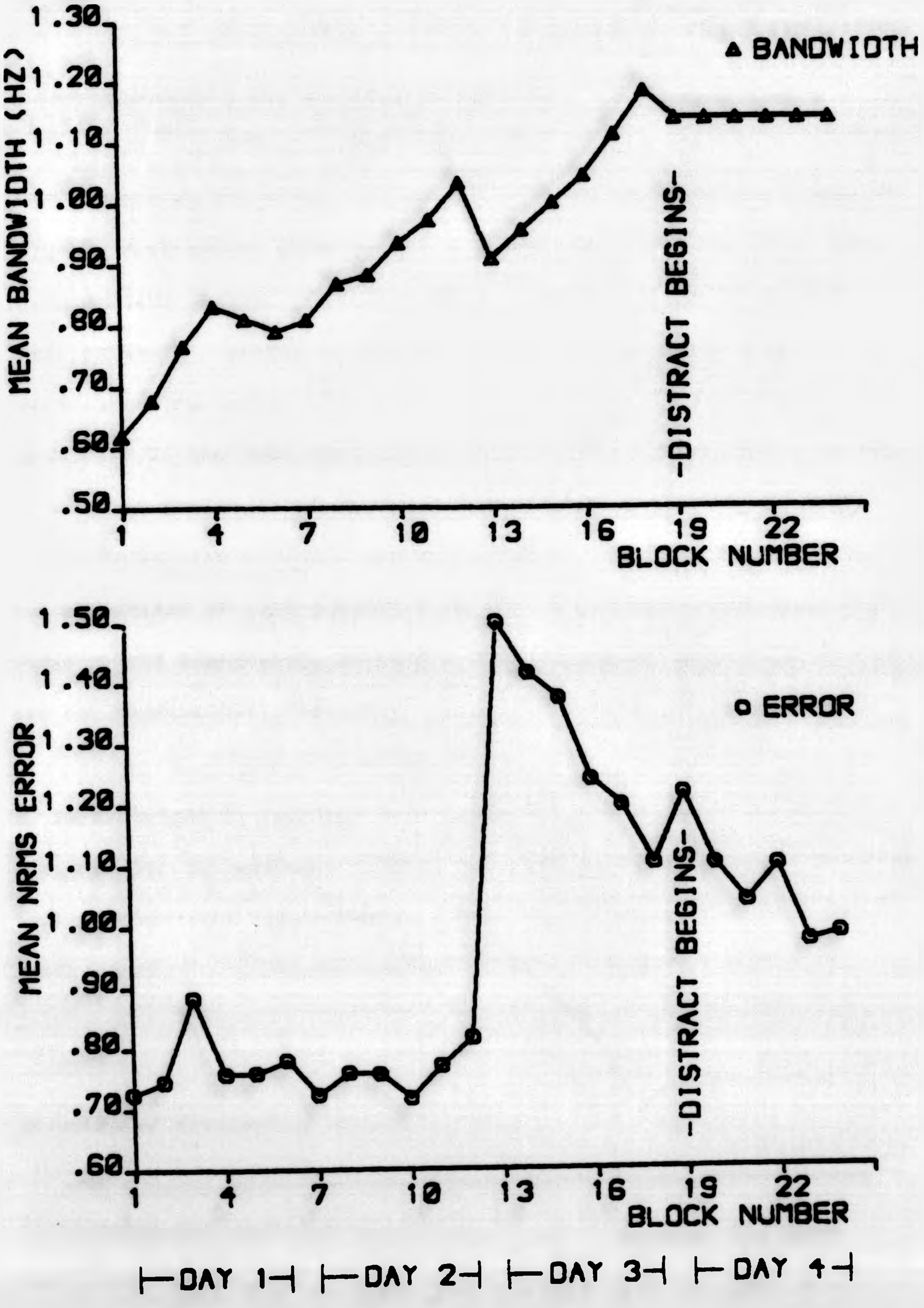

FIGURE 3-2 Distraction Experiment 1 (Random noise).
Tracking performance during the experiment.

Performance during the distraction session

The tendency of the error to diminish with further practice continued into the distraction session of the experiment. At this point the tracking task changed from an adaptive task to a fixed bandwidth tracking task. Looking at the RMS error for each three-minute block there is no indication that error was systematically affected by the imposition of the test conditions. Figure 3-2 shows a brief increase in error at the start of the distraction session (fourth session: block 19) but the increase was certainly not displayed by all subjects and comparison of each subject's error scores during this block with those of the immediately preceding block confirmed that the increase was not statistically reliable (t[illegible]=0.811, p>.4, not significant).

It is important to remember that these RMS error measures are global measures for three-minute blocks, and they do not separate distracted from nondistracted performance. They are gross measures of overall performance.

Analysis of distraction trials

Preliminary work had indicated that the size of any distraction effect was likely to be small. Various methods of analysis had been explored and it seemed likely that, to be sufficiently sensitive, the method of choice would have to incorporate at least two fundamental considerations. Firstly, the analysis would have to take into account the chronological order of the measurements, since it was probable that the interesting effects were very susceptible to the number of distractor exposures the subjects had received. Secondly, it seemed advisable to parcel out certain factors besides distraction which could lead to transient effects in the joystick data, such as local peculiarities of the track events within each trial.

The chronology of the experiment was retained by ordering the observations for each type of trial, Distracted and Nondistracted, into temporal sequence so that the earliest trials of one condition were paired with the earliest of the other. In this experiment simple pairing was not appropriate because Nondistracted trials had occurred three times as often as Distracted trials. Consequently, the Nondistracted trials were pooled by threes. In other words, the observations from the first Distracted trial were matched to the averaged data of the first three Nondistracted trials; the second Distracted trial was compared to the (averaged) fourth through sixth Nondistracted trials, and so on. This preserved the order of presentation and permitted a straightforward pairing of results for the two sorts of trials.

Having done this, it became feasible to study the differences between Distracted and Nondistracted trials using Analysis of Variance (ANOVA). This simplified the task of separating out the contributions of repeated exposure and track dynamics. Throughout a trial, forcing function and joystick signals were repeatedly logged every 100 milliseconds. The data for analysis came from the four-second window within each trial during which a distractor could appear. To measure performance changes across the time course of the trial, this window was subdivided. The basic measure of tracking performance was the RMS error measured over a one-second span. Thus each trial provided four such one-second time *zones* for analysis.

For every trial, then, each subject's raw data were grouped into four one-second zones containing the crucial events of the trial. These scores were tested by ANOVA. In the ANOVAs cited below, the main treatment effects of *Distraction* (its absence or presence) and *Exposure number* (in order of presentation) were examined. Also, to provide some understanding of the role of track dynamics the repeatable *Forcing function* was examined as a main effect. Differences were also anticipated due to the position of the one-second *Zone* selected from within the trial. These last two captured the contributions of global and local track characteristics, respectively, to differences in the scores. Recall that during training each subject had adapted to an appropriate task bandwidth. Consequently no two subjects received the same forcing function. While the forcing function was totally controlled within-subject during the test trials, the differences among forcing functions were all between-subject. Differences among forcing functions across subjects fell into the error term in the ANOVA.

None of the effects of distraction were sustained over an unlimited number of trials. For this reason, analysis of variance was always performed on two basic groupings of trials: the first analysis encompassing all twelve distractors (the whole session); the second ANOVA examining only the first six exposures (the first half of the session). The outcomes of these two fundamental analyses have always been reported. Further analyses have been presented taking more appropriate numbers of exposures where this enhances the distraction effect. Because of this ad hoc procedure, which helps to convey relative differences in the longevity of the effects, the significance levels attained should be treated with some caution.

For convenience in the descriptions which follow, Distracted trials-- those containing a noiseburst-- will be abbreviated to *D trials* and Nondistracted trials will be referred to as *ND trials.*

Effect of Random Noise distraction on tracking error

The indiscriminate block-by-block measures of performance described earlier were not capable of exposing a distraction effect. However, it became possible to detect effects when the distraction session was broken down into localized measures which directly compared D with ND trials while respecting chronological order. The effects can be seen despite the underlying global tendency (visible in Figure 3-2) for error to diminish during the distraction session as if overall performance was still being affected by practice.

When D trials were compared to ND trials having the same forcing function, it became clear that performance was selectively worsened by the presence of distractors. The D trials on average gave larger error scores than the ND trials. This is apparent in Figure 3-3 which shows the mean RMS error during the distraction session with D and ND trials plotted separately. Comparison of the scores from the (noisy) D trials with those of the (silent) ND trials indicates that the presentation of noise produced large RMS errors. The errors were largest in the earliest D trials. The difference in RMS error between the D and ND conditions diminished with subsequent exposures to noise.

The ANOVA for these results, Table 3-1, encompassing the entire distraction session (ie all twelve exposures to noise) did not show Distraction as a significant main effect on its own. Instead, Distraction interacted with other factors. The Forcing function which was used for a trial, the number of distractors already heard (Exposure number), and the variations in task demands across the trial's timespan (contributing to Zone differences) all had significant effects on the RMS error. The relevance of Forcing function and Zone simply confirms what is already known: that tracking error is never stable but rather it varies according to the particular track used and the local characteristics of the track. Taken over the whole extent of the distraction session, these powerful effects overshadowed any main effect of Distraction. However, Table 3-1 does show that the Distraction X Forcing function interaction was statistically significant ($F_{1,4}$=7.45, p<.04) which suggests that susceptibility to distraction depended on the particular forcing function used. Distraction also interacted with Exposure number ($F_{8,4}$=2.84, p<.04), confirming that the level of distraction effect

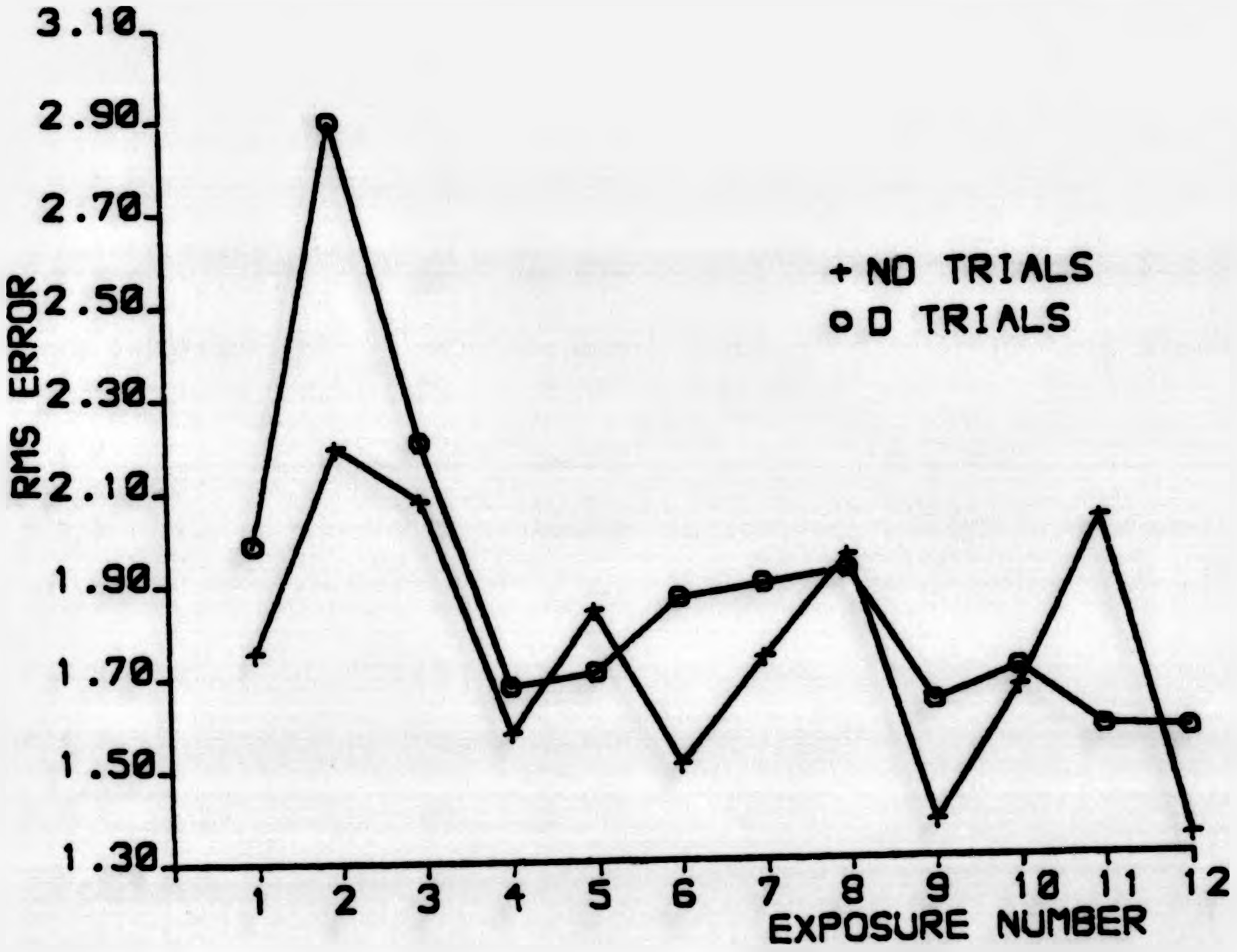

FIGURE 3-3 Distraction Experiment 1 (Random noise).
Tracking RMS Error: Distraction compared to no distraction.

TABLE 3-1

DISTRACTION EXPERIMENT 1 (Random noise).
ANOVA Summary: Effects of distraction, forcing function, number of exposures, and time zone in the trial on the average RMS Error. 12 exposures to distraction (entire session).

Source	*df*	*SS*	*MS*	*F*	*Prob*
Subjects (S)	5	116.14			
Distraction (D)	1	2.547	2.547	3.4131	0.1224
D X S	5	3.731	0.746		
Forcing function (F)	1	14.855	14.855	12.4143	0.0172 *
F X S	5	5.983	1.197		
Exposure number (E)	5	24.327	4.865	3.4019	0.0176 *
E X S	25	35.755	1.430		
Zone in trial (Z)	3	14.988	4.963	7.7562	0.0024 *
Z X S	15	9.597	0.640		
D X F	1	1.147	1.147	7.4465	0.0410 *
D X E	5	4.232	0.846	2.8357	0.0365 *
D X Z	3	0.146	0.049	0.1791	0.9086
F X E	5	6.657	1.331	2.7700	0.0398 *
F X Z	3	8.177	2.726	2.1857	0.1312
E X Z	15	2.975	0.198	0.9217	0.5446
D X F X E	5	4.923	0.985	2.4259	0.0631
D X F X Z	3	1.067	0.356	1.9757	0.1601
D X E X Z	15	3.153	0.210	1.2087	0.2843
F X E X Z	15	3.189	0.213	1.1393	0.3380
D X F X E X Z	15	1.962	0.131	1.2519	0.2541

* significant at $p<0.05$

was altered by the number of times the distractor had been presented. The significant Forcing function X Exposure number interaction ($F_{[illegible]}$=2.77, p<.04) was probably a manifestation of the slight practice effect alluded to earlier, indicating that the benefits of practice were not uniform for both tracks. This is not unusual and it is of no importance in the present context.

In order to see Distraction as a main effect on its own, it was necessary to confine the analysis to the earlier trials of the session, just as Figure 3-3 would suggest. The ANOVA of the first six exposures to noise (as opposed to the full twelve exposures), Table 3-2, did show Distraction as a significant main effect ($F_{1,4}$=8.04, p<.04). Table 3-2 also confirms that Forcing function and Zone (that is, both global and local track characteristics) were again major determinants of performance differences. Exposure number did not yet exert a conspicuous effect during these early trials, and the various interactions were not yet identifiable.

The impact of distracting noise did not appear to be evenly distributed across the whole duration of the D trial. The error effects are broken down into their one-second subsample zones in Figure 3-4. To simplify the presentation, only the difference between the D and ND RMS errors is plotted (a positive difference indicating that D error exceeded ND error). There is the suggestion here that the main influence of the noiseburst shows up a little over one second past the onset of the burst and lasts for about a second. The size of this distraction effect was undoubtedly sensitive to the number of

TABLE 3-2

DISTRACTION EXPERIMENT 1 (Random noise).
ANOVA Summary: Effects of distraction, forcing function, number of exposures, and time zone in the trial on the average RMS Error. First 6 trials only.

Source	*df*	*SS*	*MS*	*F*	*Prob*
Subjects (S)	5	59.160			
Distraction (D)	1	3.850	3.850	8.0427	0.0363 *
D X S	5	2.394	0.479		
Forcing function (F)	1	13.048	13.048	11.4684	0.0198 *
F X S	5	5.689	1.138		
Exposure number (E)	2	11.477	5.738	3.1790	0.0845
E X S	10	18.051	1.805		
Zone in trial (Z)	3	8.189	2.730	5.7470	0.0081 *
Z X S	15	7.125	0.475		
D X F	1	0.000	0.000	0.0003	0.9841
D X E	2	1.995	0.998	2.3326	0.1463
D X Z	3	0.387	0.129	0.5177	0.6796
F X E	2	4.426	2.213	2.7878	0.1080
F X Z	3	4.704	1.568	1.6397	0.2217
E X Z	6	1.971	0.328	1.0573	0.4098
D X F X E	2	2.809	1.404	2.4741	0.1329
D X F X Z	3	1.685	0.562	2.7241	0.0804
D X E X Z	6	2.205	0.367	1.5408	0.1985
F X E X Z	6	2.082	0.347	1.7111	0.1521
D X F X E X Z	6	0.375	0.062	0.5269	0.7847

* significant at $p<0.05$

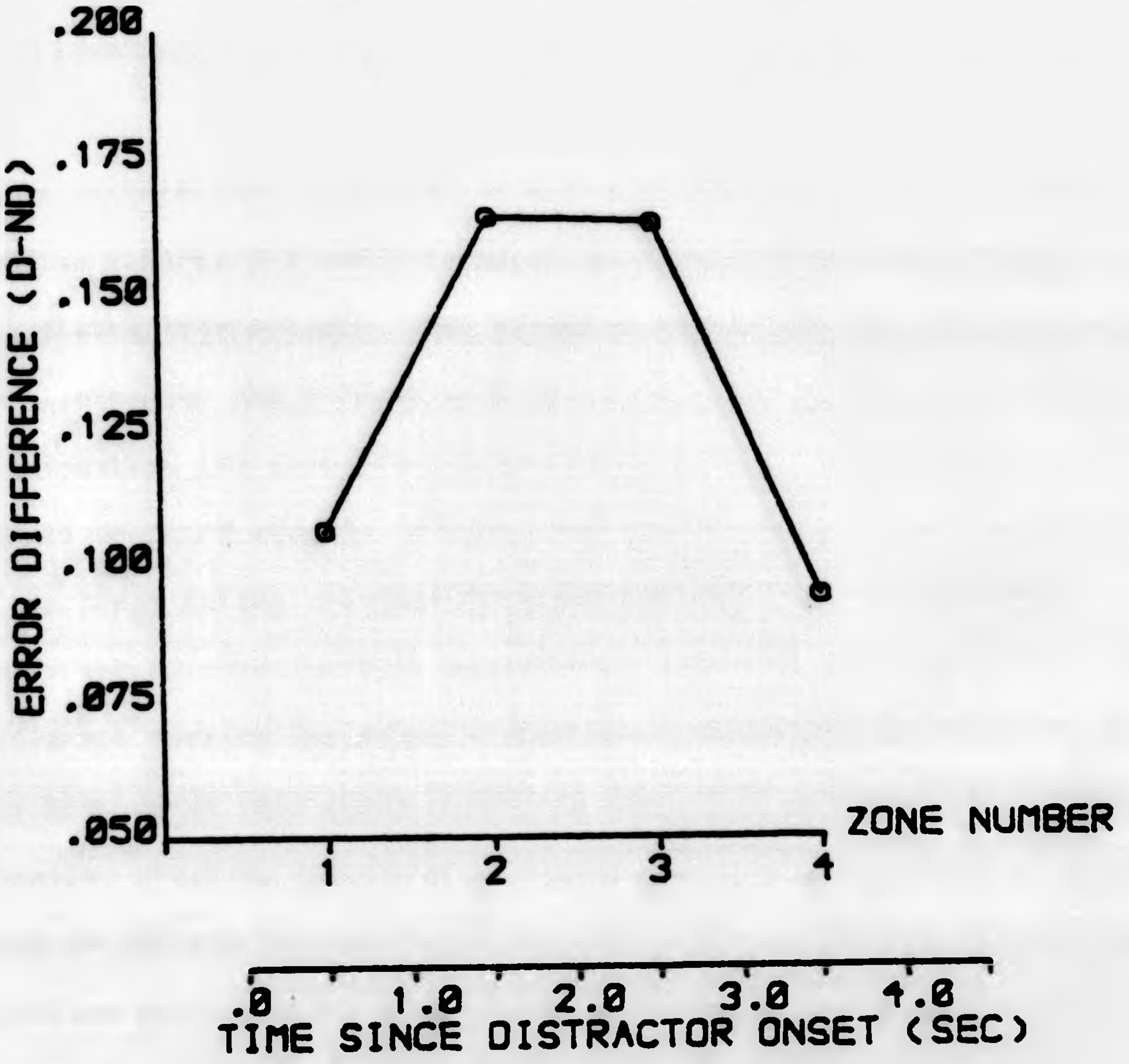

FIGURE 3-4 Distraction Experiment 1 (Random noise). Difference between Distracted and Nondistracted RMS Error: stability within the trial.

distractor exposures. The D minus ND difference decreased as the number of exposures increased. The very earliest exposures exerted the most noticeable effects, which were not so clearly localised and possibly extended even beyond the four-second duration of the sampling window. Later exposures tended to produce their peak effect about one or two seconds after distractor onset, while the last-presented noisebursts had no discernible impact. These observations are visible in Figure 3-5.

These descriptions of the error profiles are potentially misleading because Figures 3-4 and 3-5 depict only the average distribution of the distraction effect. They do not predict, for example, that for any distractor the effects will be minimal during the first second of presentation and then rise to maximum disruptive impact during the second or third second. A more cautious interpretation is called for since these curves are mean outcomes pooling different subjects, different forcing function bandwidths, and-- for each subject-- two different forcing functions. When the contributory data are separated out they often look quite different from this average result. For example, breaking the curve of Figure 3-4 into two parts by plotting each of the two forcing functions separately, gives two contrasting profiles which, before separation, had jointly provided a simple peak. This exercise is shown in Figure 3-6. The subdivision of the forcing function in Figure 3-6 is in turn arbitrary, since both forcing functions here are again collective results, obtained by pooling tracks from six subjects, tracks which we know are all slightly different (coming as they do from random noise of differing

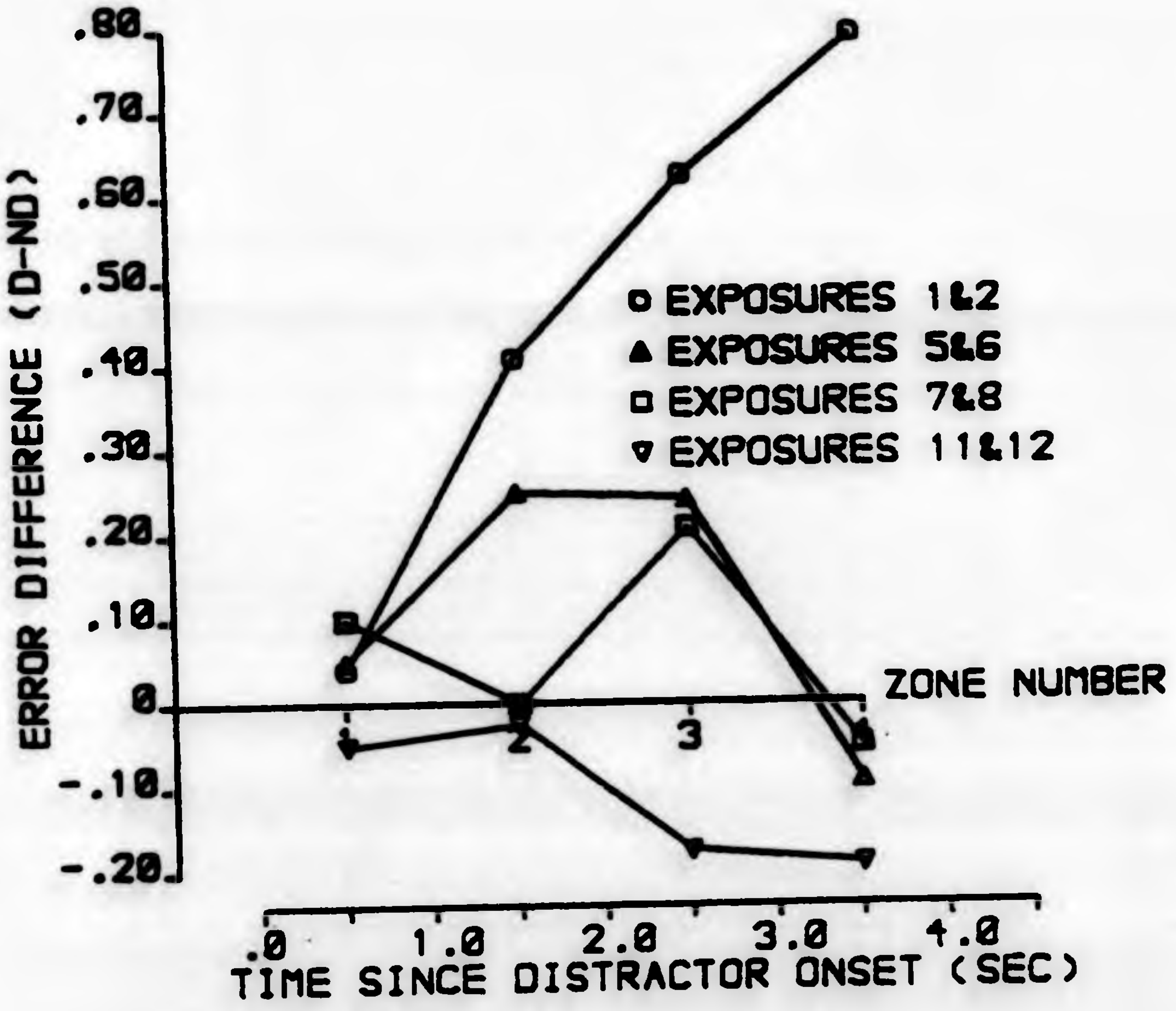

FIGURE 3-5 Distraction Experiment 1 (Random noise). RMS Error during the trial: effect of increasing exposure to distraction.

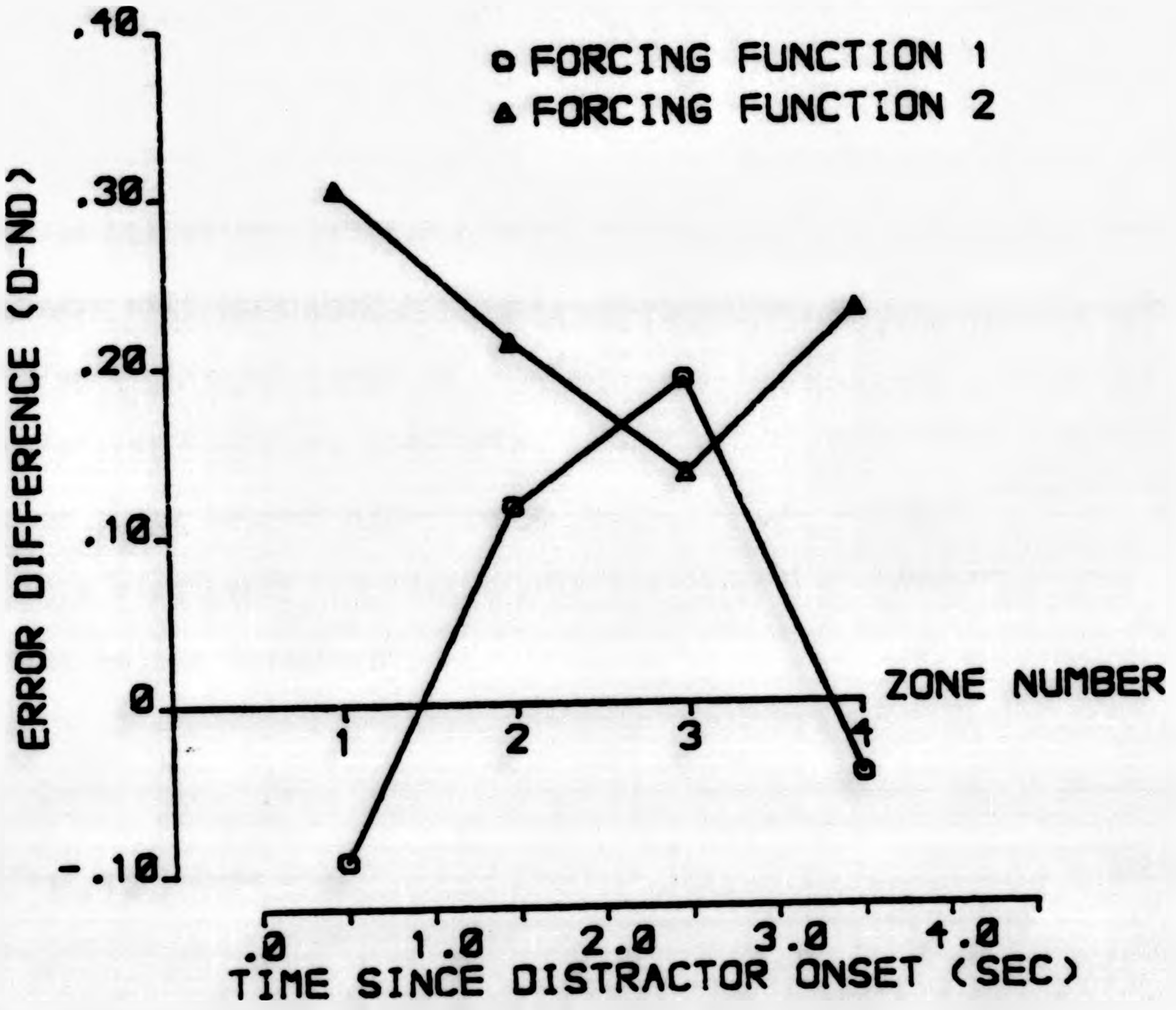

FIGURE 3-6 Distraction Experiment 1 (Random noise).
RMS Error during the trials dependency on track characteristics during distraction.

bandwidth). Following the exercise through to its conclusion warns us that the location of the maximum effect and its direction are likely to be influenced by the specific forcing function. This, after all, was what the Distractor X Forcing function and the Distractor X Zone interactions implied also. Quite apart from this, there are eccentric subjects for whom the average outcome can be an inaccurate description, as we shall see. What we do learn from these compilations based on averages is that (regardless of the chosen grouping of tracks and subjects) it seems likely that any localisation of the distraction effect within the trial is fairly stable for as long as the effect lasts but that the magnitude of the effect does decrease with continued exposure. On average, but not in every case, a noise distractor tends to increase the RMS error. On average, with the forcing functions used here, the effect is especially noticeable two or three seconds after noise onset. It may be worth noting that in this experiment the time of maximum impact corresponds to the offset of the noiseburst.

Effect of Random Noise distraction on amplitude of joystick movement

The interpretation is expanded slightly after studying the raw tracking data for individual subjects. Data from two of the subjects are displayed in Figures 3-7a and 3-7b. These show, for the first six exposures of the session, one forcing function and the simultaneous joystick voltages with D trials collected into a single plot and ND trials similarly averaged. The distractor onset corresponds to

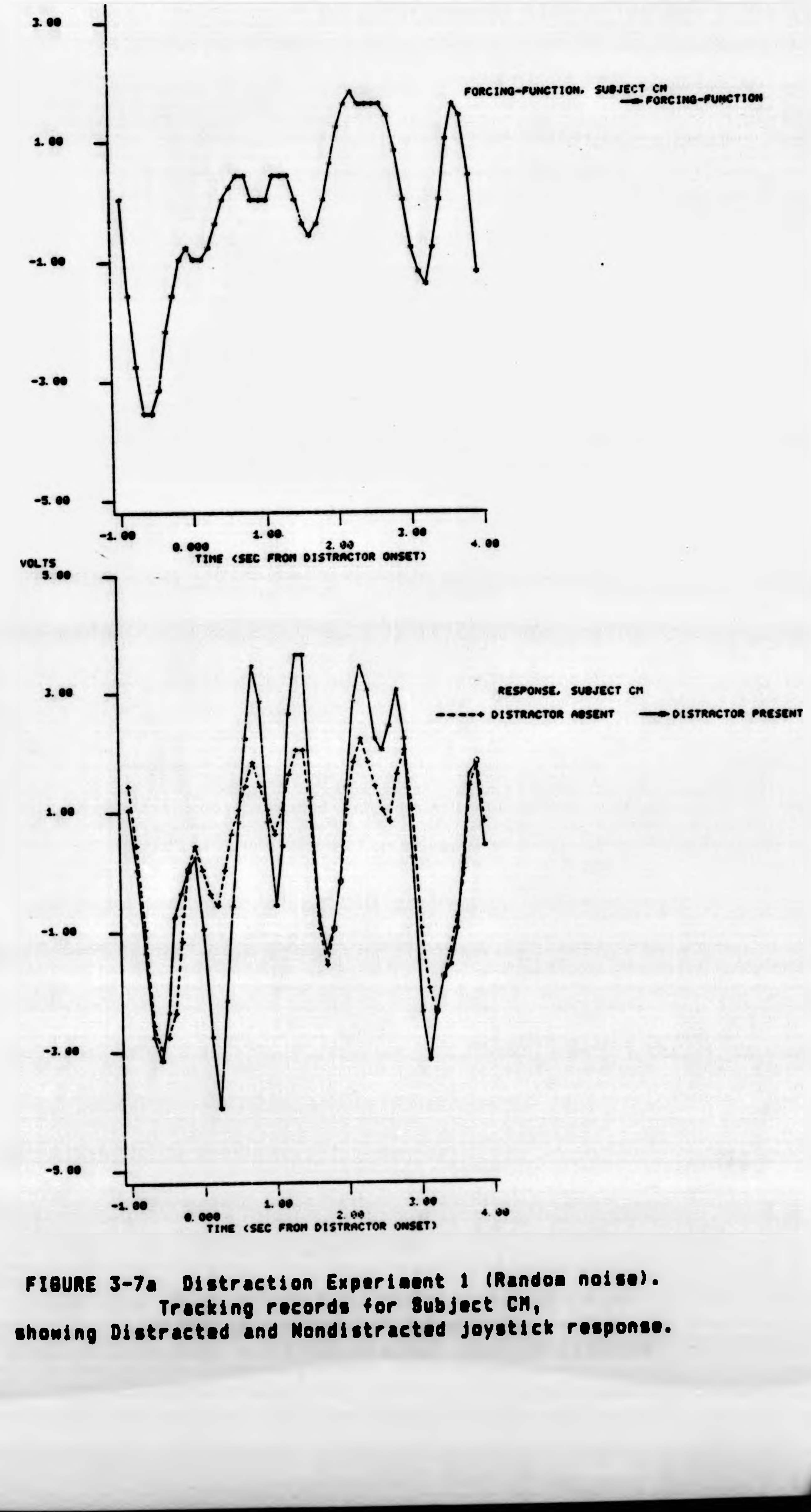

FIGURE 3-7a Distraction Experiment 1 (Random noise). Tracking records for Subject CM, showing Distracted and Nondistracted joystick response.

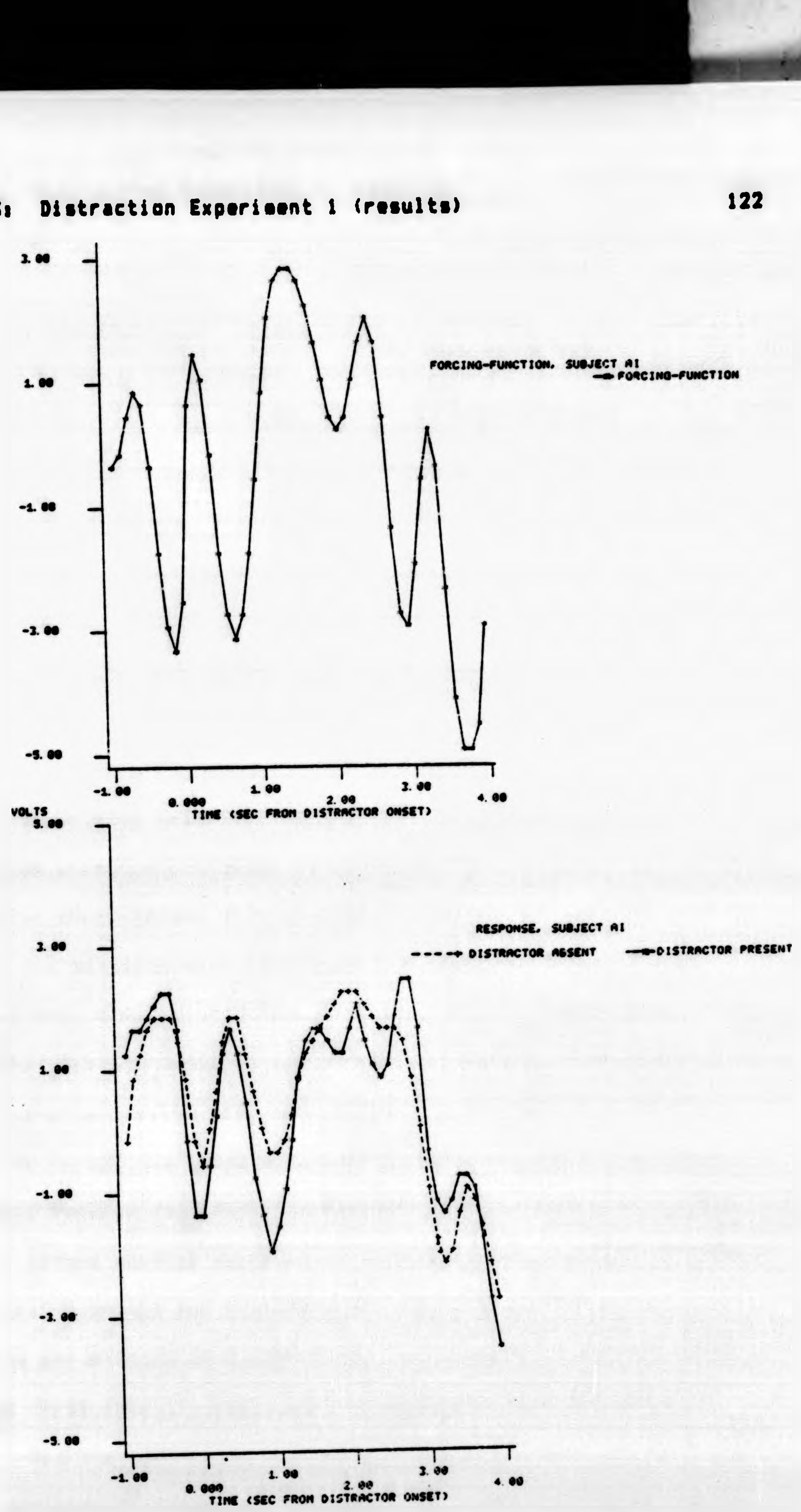

FIGURE 3-7b Distraction Experiment 1 (Random noise).
Tracking records for Subject AI,
showing Distracted and Nondistracted joystick response.

time=0. Before distractor onset, D and ND joystick records show good agreement, as would be expected. Following onset in the D trials Subject CM revealed a clear tendency to overshoot track reversals. For subject AI, on the other hand, the distractors seemed almost to have an alerting effect, manifested as a more accurate response to track reversals. That is, distraction tended to *reduce* this subject's error. (Indeed, for the first two exposures at least, this subject's RMS error over the four-second sampling window measured slightly lower during the noisy D trials than during the silent ND trials.)

Notice that in both examples, regardless of interpretation, distraction promoted increased movement of the joystick. Led by these and similar observations from several subjects, the joystick voltage data for all subjects were tabulated and analysed to see if they supported this conclusion. The joystick excursions were obtained by taking the absolute value of each joystick reading. Overshooting on track reversals implies that the joystick excursions are, on the whole, too large. (A large RMS error is not, in itself, evidence of overshoot, because consistent undershoot can also increase the RMS error.) In the initial analyses of movement size, neither the ANOVA of all twelve trials nor the ANOVA of the first six exposures supported any effects of distraction. An analysis was carried out on just the first four trials. The corresponding data are shown in Figure 3-8. The mean joystick movement amplitude was indeed increased by the earliest D trials but settled to ND levels with great speed. The extra movement was confined to the first couple of distractors. The ANOVA of these four trials, summarised in Table 3-3, confirmed

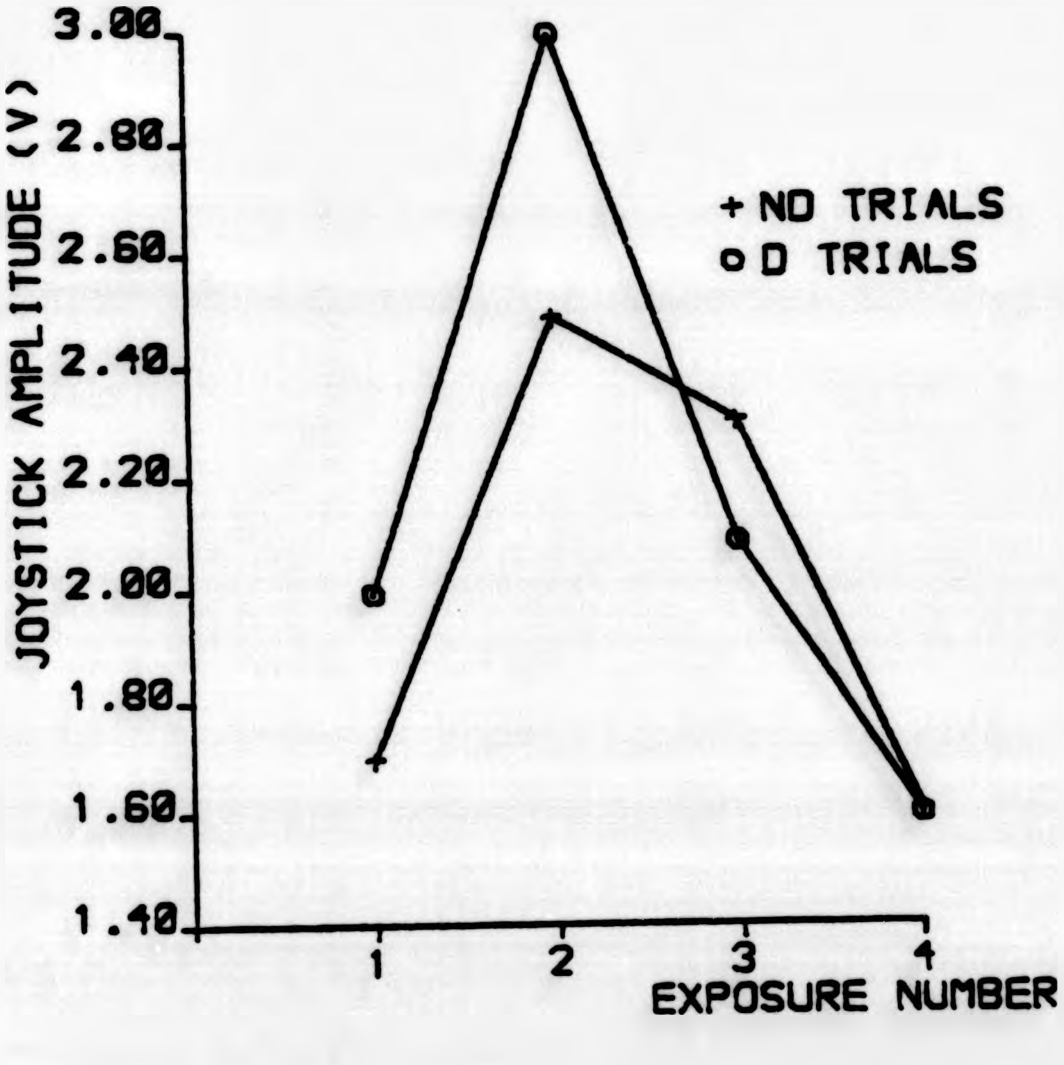

FIGURE 3-8 Distraction Experiment 1 (Random noise).
Joystick movement size: First four trials only.

TABLE 3-3

DISTRACTION EXPERIMENT 1 (Random noise).
ANOVA Summary: Effects of distraction, forcing function, number of exposures, and time zone in the trial on the average Joystick Amplitude. First 4 trials only.

Source	*df*	*SS*	*MS*	*F*	*Prob*
Subjects (S)	5	106.763			
Distraction (D)	1	0.808	0.808	0.4883	0.5202
D X S	5	8.276	1.655		
Forcing function (F)	1	25.486	25.486	19.0755	0.0078 *
F X S	5	6.680	1.336		
Exposure number (E)	1	6.938	6.938	4.2304	0.0935
E X S	5	8.200	1.640		
Zone in trial (Z)	3	5.396	1.799	4.5941	0.0179 *
Z X S	15	5.872	0.391		
D X F	1	0.010	0.010	0.010	0.9206
D X E	1	2.795	2.795	8.5178	0.0330 *
D X Z	3	0.972	0.324	1.8215	0.1857
F X E	1	0.878	0.878	1.7108	0.2473
F X Z	3	4.476	1.492	4.7321	0.0162 *
E X Z	3	0.470	0.157	0.4459	0.7266
D X F X E	1	0.384	0.384	0.2921	0.6151
D X F X Z	3	1.223	0.408	2.1440	0.1365
D X E X Z	3	0.466	0.155	0.4292	0.7378
F X E X Z	3	0.876	0.292	1.0938	0.3831
D X F X E X Z	3	1.606	0.535	1.7422	0.2005

* significant at $p<0.05$

that the effect was sensitive to experience: there was a significant Distraction X Exposure number interaction ($F_{1,64}=8.52$, $p<.03$). No main effect of Distraction could be verified by this analysis. Nor could any distraction effects be confirmed by ANOVAs of just exposure one and/or exposure two. The distraction effect was not sufficiently robust to overcome the handicap of such a small number of samples. The effect was undoubtedly a minor one of very brief duration. Inspection of the subjects' raw joystick data suggested that when large excursions occurred, they could mainly be described as overshoots at track reversals, but occasional examples of what appeared to be alerting (defined as improved accuracy of error correction) could be singled out as well. These descriptions are, of course, subjective interpretations of the data.

The variance of the joystick movements was also examined for differences arising due to distraction. Analyses of twelve, six, and three or fewer exposures were all carried out. No differences whatsoever could be discerned. The unchanged variance affirms that the distractors did not promote a tendency to inject extra spurious movements or cause the subject to "vibrate" the joystick through loss of control.

Effect of Random Noise distraction on response delay

The subject's response delay is a vital characteristic of his tracking skill and a contributor to his error. The corrective movements of the joystick will tend to lag behind the movements of the track. Exceptionally skilled trackers are able to reduce this lag when the track contains predictable elements by anticipating the next track disturbance and initiating movement before it begins. It is possible to estimate the subject's response delay by obtaining crosscorrelations between track and joystick voltages. The crosscorrelation must be calculated several times, on each occasion applying a different temporal offset between the track and joystick signals. That is, a crosscorrelation function can be computed. In this experiment the signals are sampled every 100 milliseconds, so it is feasible to crosscorrelate using this duration as the basic offset, or *lag*, interval. Thus for a lag 0 crosscorrelation there is no time offset interposed between the samples of track and response. In a lag 1 crosscorrelation, the joystick data are realigned before analysis so that they are compared to the track data which in reality occurred 100 milliseconds earlier. Lag 2 would define a 200 millisecond realignment, and so on. The crosscorrelation statistic attains its maximum value when the selected lag interval is close to the subject's true response delay. Normally the lag 0 crosscorrelation does not reflect the best possible correspondence between track and response, which can only be revealed by offsetting the joystick data by an appropriate amount. The optimum lag will depend on the skill of the subject, and task variables such as forcing function bandwidth normally influence the lag outcome. The measure of correlation at any arbitrary lag is not in itself very informative. It is more useful to calculate the crosscorrelation systematically for different lag intervals and so identify the lag associated with the highest correlation. If the best correlation is determined locally for each one-second subsample zone of the test trial, this yields an estimate of the subject's response delay during each second of the trial.

The mean outcome of such a lag analysis is given in Figure 3-9. This indicates that under ND conditions the joystick appeared to trail behind the track by about 150 milliseconds, but when distractors occurred within the trial the delay increased by a substantial amount, approximately 30 milliseconds more on average. The additional response delay appeared to persist for about eight exposures. The ANOVA summary in Table 3-4 confirms that this extra response delay attributable to Distractor presence was highly significant during the first eight exposures ($F_{1,44}$=13.85, p<.01). As might be expected, the lag was also influenced by specific characteristics of the track at the local level, as evidenced by the significant Distraction X Forcing function X Zone interaction ($F_{3,44}$=4.86, p<.01), where Zone as usual refers to the one-second subdivision of the trial.

These estimates are average response delays across all subjects. Even without distraction intervention, any single subject's response delay can be described only very approximately by an average lag duration, since response delay will normally show differences across very local track events, eg becoming greater for a brief time when the forcing

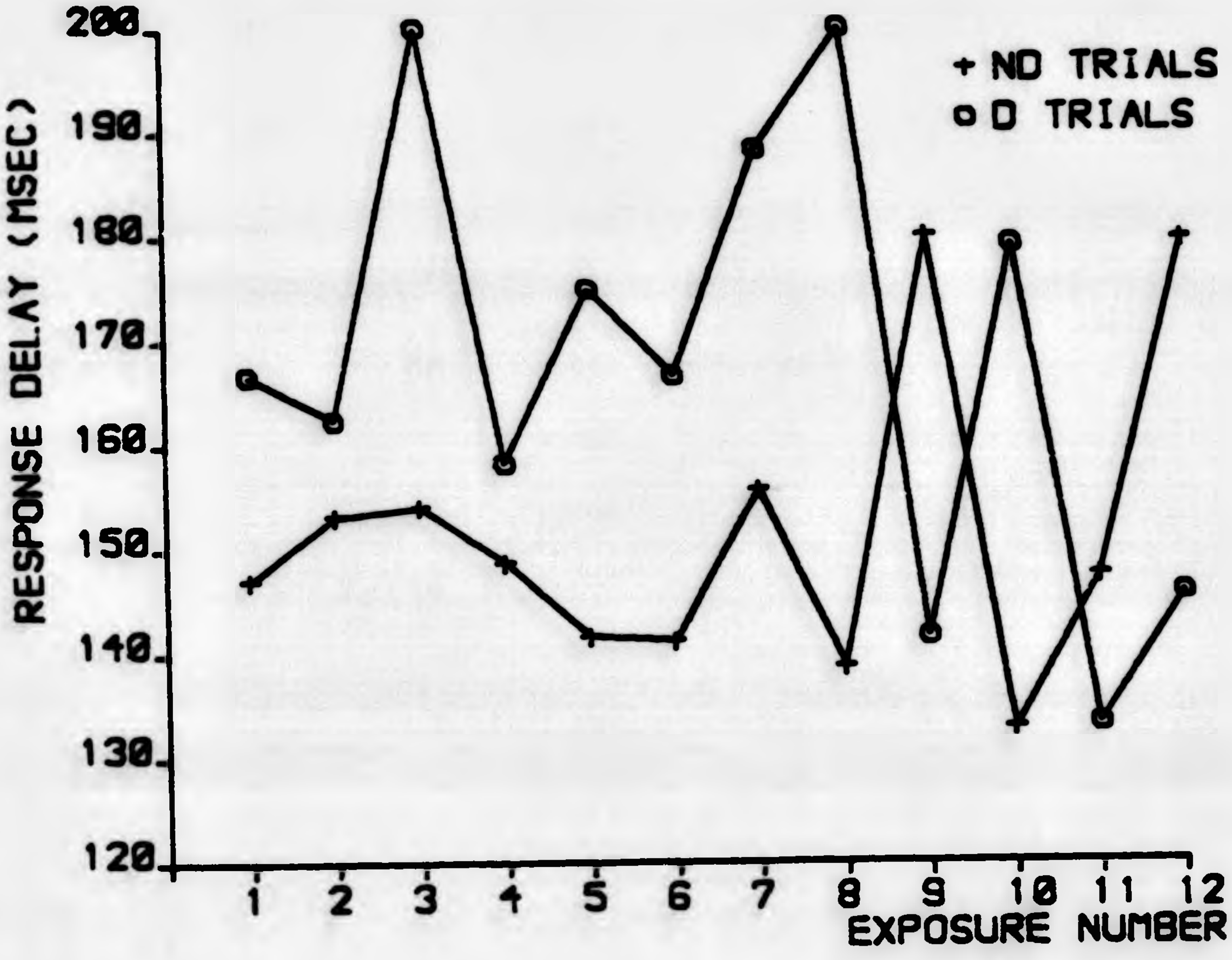

FIGURE 3-9 Distraction Experiment 1 (Random noise).
Tracking response delay: Distraction compared to no distraction.

TABLE 3-4

DISTRACTION EXPERIMENT 1 (Random noise).
ANOVA Summary: Effects of distraction, forcing function, number of exposures, and time zone in the trial on the tracking response delay. First 8 trials only.

Source	*df*	*SS*	*MS*	*F*	*Prob*
Subjects (S)	5	175.86			
Distraction (D)	1	8.050	8.050	13.8537	0.0141 *
D X S	5	2.906	0.581		
Forcing function (F)	1	0.496	0.496	0.7054	0.5571
F X S	5	3.516	0.703		
Exposure number (E)	3	1.175	0.392	0.7116	0.5629
E X S	15	8.254	0.550		
Zone in trial (Z)	3	14.236	4.745	3.4127	0.0447 *
Z X S	15	20.857	1.390		
D X F	1	0.683	0.683	0.9058	0.6125
D X E	3	1.085	0.362	0.2989	0.8271
D X Z	3	6.367	2.122	2.9252	0.0674
F X E	3	0.883	0.294	0.3520	0.7904
F X Z	3	1.255	0.418	0.1003	0.9577
E X Z	9	8.631	0.959	1.1683	0.3378
D X F X E	3	0.891	0.297	0.4152	0.7473
D X F X Z	3	11.692	3.897	4.8589	0.0148 *
D X E X Z	9	4.346	0.483	0.4779	0.8821
F X E X Z	9	3.198	0.355	0.3064	0.9682
D X F X E X Z	9	4.498	0.500	0.4727	0.8855

* significant at $p<0.05$

function causes a track reversal (Poulton, 1974). Mindful of this qualification, as well as the caution raised earlier about proceeding from the group mean to the individual when different forcing functions are involved, the average effects on lag across the trial's timecourse are displayed in Figure 3-10. The diagram shows the effect at its greatest, an increase in response delay of about 50 milliseconds, during the first two seconds following noise onset. The persistence of the extra delay during repeated exposure is indicated by Figure 3-11. For simplicity only the difference between the D and ND estimates of response delay is plotted here; the difference is positive when D lag exceeds ND. This diagram also shows how the lag effect became inconsistent during the final exposures.

Again, it is possible that any description of the response delay profiles is influenced by the groupings used to average the data from dissimilar tracks. It is more important that lag effects were still detectable well into the distraction session: these effects seemed not to be as sensitive to the number of exposures as had been the effects on joystick movement size.

Subjective reports

The subjects' own accounts of the distraction session tended to be compatible with the results already described. When asked if the noisebursts had bothered them, subjects invariably replied that the noises had startled them at first but that they very soon got used to

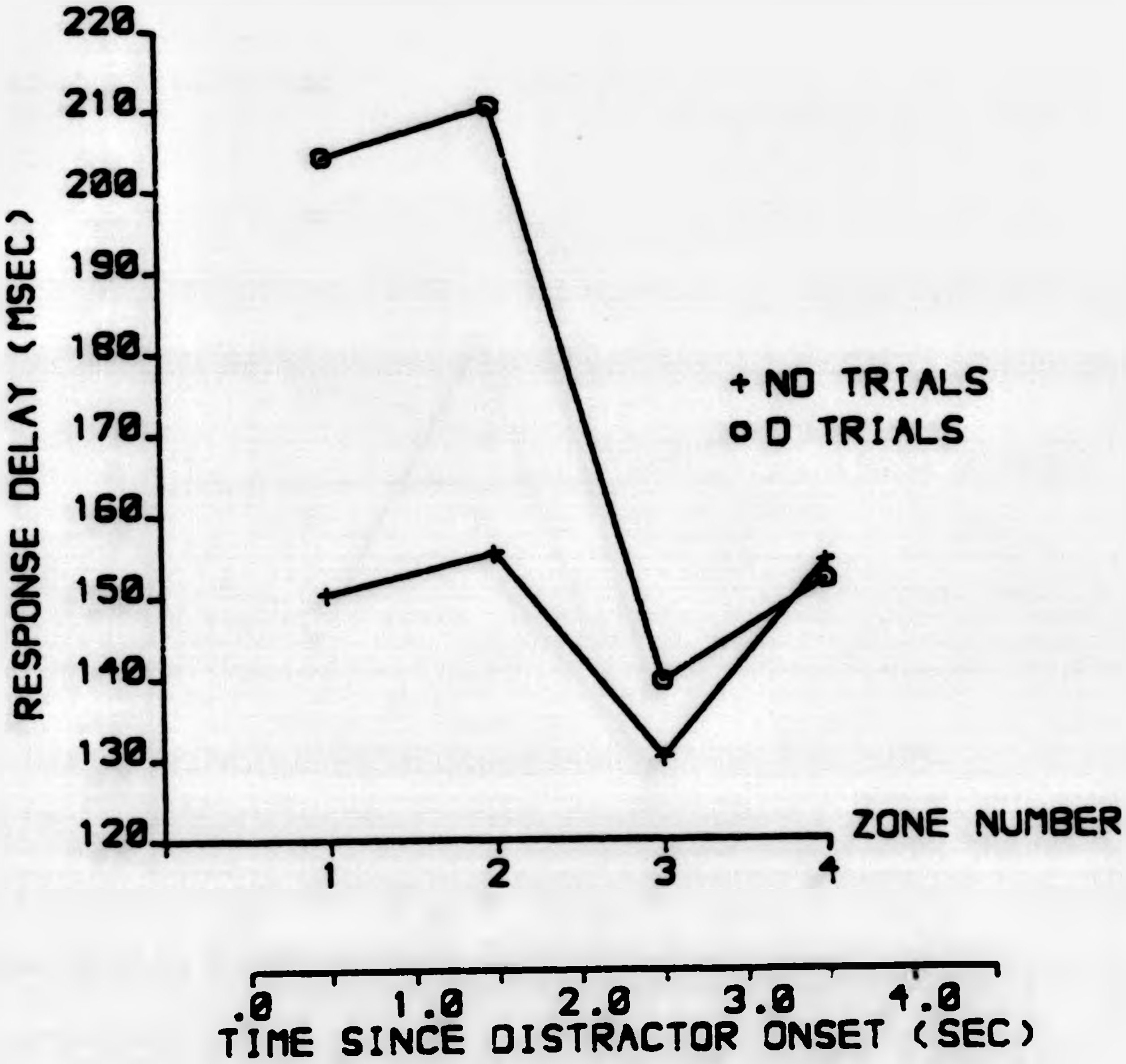

FIGURE 3-10 Distraction Experiment 1 (Random noise). Difference between Distracted and Nondistracted response delay: stability within the trial.

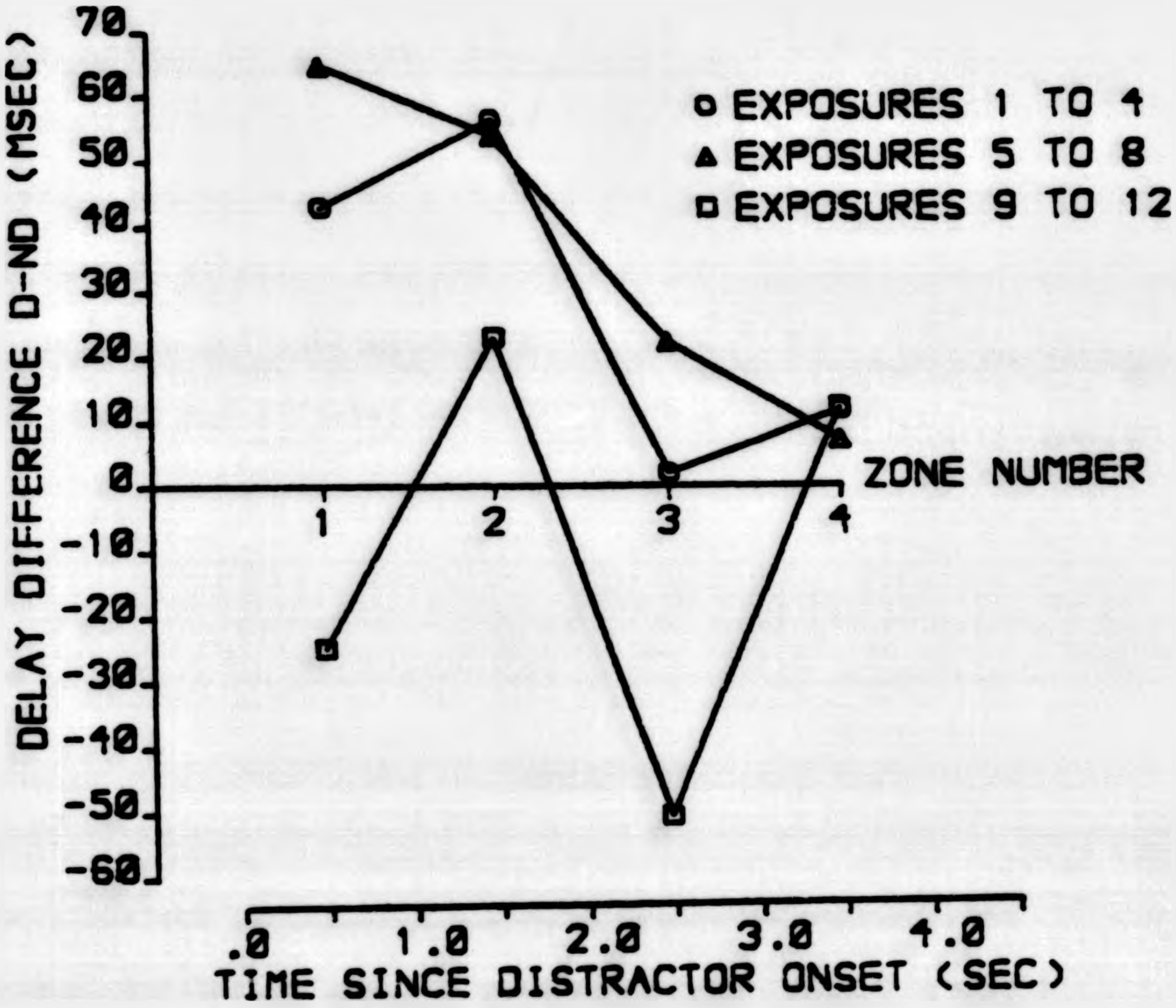

FIGURE 3-11 Distraction Experiment 1 (Random noise).
Response delay during the trial:
effect of increasing exposure to distraction.

them. Subject AI (whose distracted performance was earlier described as "alerted") claimed that the distractors did not affect her hand movements but had disrupted her train of thought. Subject MH pointed out that he had felt most bothered by a distractor just after it had gone silent; that is, at noise offset.

To most subjects, the disturbance was muddling or surprising in a vague way and for a brief time. For the most part, they did not believe that the distractors had interfered with their tracking. In general, the subjects were not good at describing the timescale of the distraction effect. They had retained an emotional impression of disturbance, but they seemed to have little sense of how long or in what way it had affected their tracking performance.

Discussion

For a subject confronting a substantial tracking workload, any of several strategies could be adopted to deal with a diversion of information-processing resources required by the arrival of a distractor. The simplest response would probably be to stop tracking for a time until the distracting event had been assessed. A motionless joystick would result in a constant voltage output in the joystick data for as long as the interruption lasted. This condition is actually quite difficult to detect with confidence unless it lasts for a long time, perhaps 20% of the measurement window (Moray, 1979b). Alternatively, a sudden loss of control over the joystick's movement could appear as a jerk or twitch of random size and/or direction. If the reallocation of resources did not require such a drastic loss of control, tracking would still suffer if the corrective movements of the joystick deteriorated because of undercorrection, overcorrection (overshoot), or increased response delay. All of these reactions to distraction would normally cause an increase in the RMS error, and would be detectable if they were of sufficient magnitude and duration. It is also possible that distractors could have an arousing or alerting effect on the subject resulting in enhanced control, either through improvement in his detection of error, a lowering of his criterion for taking corrective action, speeded motor control, or a temporary increase in his processing capacity. True alerting would be expected to reduce the RMS error.

In this experiment the distractors were shown to increase the RMS error. (A very few exceptional incidents of alerting-- with some attendant suppression of error-- might also have occurred early in the distraction session, but this is only a speculation based on visual inspection of individual tracking records.) At its maximum the increase in error was about 30% above the nondistracted level of error. The effect was a local one confined to the vicinity of a distractor and lasting no more than a few seconds. The introduction of distractors at random intervals did not seem to have a great impact on tracking at nondistracted times and no global increase in RMS error could be confirmed during the 10-minute distraction session taken as a whole. The repeatability of the effect was short-lived, becoming harder to measure with successive presentations of the distractors. The distraction effect interacted not only with the number of exposures, but also with the characteristics of the forcing function being tracked. This was taken as evidence that task demands had a profound influence on distractibility at any given moment.

The very earliest distractors were possibly associated with a temporary increase in the joystick movement amplitude. This was not accompanied by any measurable change in the variance of the movements. The additional movement was probably contained in a few exceptionally wide excursions, as if distraction encouraged perseveration of ongoing movements. This outcome was highly sensitive to the number of exposures and the effect very quickly disappeared. Since the effects on the RMS error were certainly more prolonged than this, the impact of distraction cannot simply be attributed to an inclination to swing the joystick farther in either direction. It is more likely that the initial very brief increments in joystick movement size corresponded to true startle reactions. This could be the best way to describe such short disruptions of motor control, associated only with the first one or two arrivals of unexpected loud noise.

Distraction had a more persistent selective effect on the subjects' tracking response delay. The response delay was lengthened by typically 20% when a distractor occurred. This effect was present for at least as long as the measurable increase in RMS error, and was presumably the major contributor to the increase. Again, the delay interacted with local features of the forcing function, underlining the importance of this task variable.

Subjects did not feel that they had exhibited much, if any, decrement in performance during the session, and all subjects were confident that the noisebursts had ceased to be distracting beyond the first one or two presentations. This would imply that subjects are not aware of much beyond their initial startle response. The subjective reports were comparable to others in the early distractibility literature (Cassel and Dallenbach, 1918; Pollock and Bartlett, 1932): beyond a sense of irritation or confusion, little knowledge of specific effects was demonstrated.

Certainly the most striking impression gained from both objective measures and subjective reports is of the extreme rapidity with which the distraction effect diminished. Despite the inclusion of only twelve distractors in the experiment, analysis could not detect reliable main effects beyond more than six or eight exposures. This exemplary demonstration of adaptation prompted a second experiment. The purpose of the experiment was to investigate the impact of nonrandom distractor scheduling on the distraction effect and the timecourse of its adaptation. It was assumed that a more predictable schedule of noise presentation would lead to still more rapid adaptation.

DISTRACTION EXPERIMENT 2: PERIODIC NOISE DISTRACTORS

Procedure

Six subjects (three females, three males) were recruited as in Experiment 1. Like the previous subjects, they had no tracking experience.

Training was administered as it had been in the earlier study, using the adaptive tracking task over three days, with the control law modification on the last day of practice.

For the final test session, on day four, the tracking bandwidth was fixed and the noisebursts were introduced. In this experiment the distractors could occur within any of three repeatable forcing functions rather than the two forcing functions previously used. This was merely a consequence of the distractor timing, which was more frequent than in the earlier study.

The noises were presented according to a *Periodic* schedule. The distractors were frequent and entirely regular in their arrival. This nonrandom schedule provided a noiseburst every 30 seconds. In all, 36 distractors were given.

The noise characteristics, the instructions to subjects, and the data-collection techniques were identical to those of Experiment 1.

Results

Training and test performance

Again training provided no surprises: the subjects achieved task bandwidths comparable to those in Experiment 1, with similar levels of performance. The normalised RMS Error for each three minute block is shown in Figure 3-12. This indicates that, as before, the error rose when the control dynamics changed during the training but settled during subsequent blocks. No overall performance decrement accompanied the introduction of distractors to the experiment.

Analysis of distraction trials

The D and ND trials were paired in order of presentation. Pairing was straightforward in this case because equal numbers of D and ND trials had been obtained. Analysis of variance examined the effects (across various groupings of trials) of Distraction, Forcing function, Exposure number, and Zone of the trial. ANOVA was always applied initially to the whole dataset of thirty-six exposures (the entire session) as well as to smaller subsets of trials comprising the first twelve and the first six exposures. The only difference between these ANOVAs and those of Experiment 1 was that three levels of Forcing function were involved here rather than two.

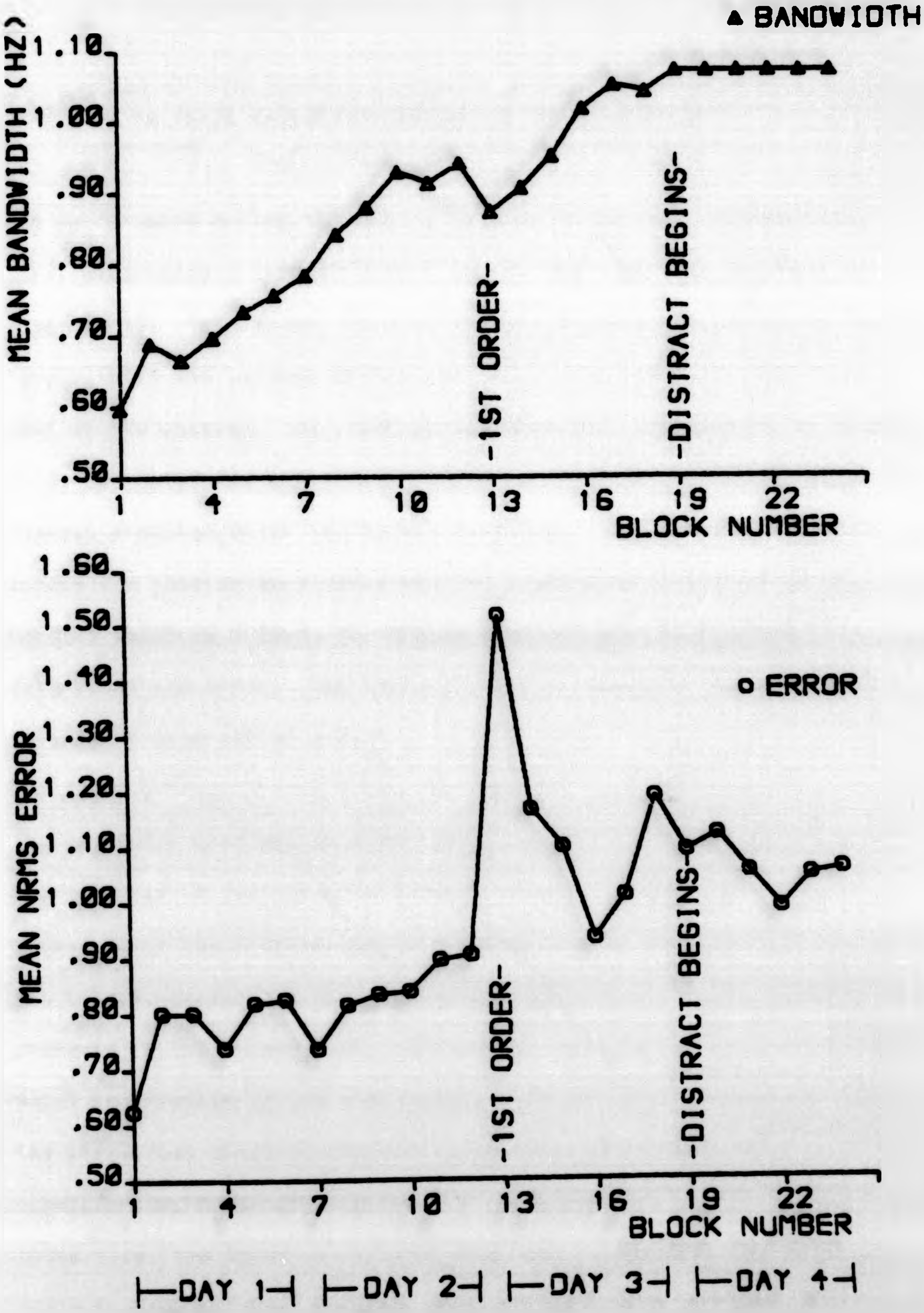

FIGURE 3-12 Distraction Experiment 2 (Periodic noise).
Tracking performance during the experiment.

Effect of Periodic Noise distraction on tracking error

The performance during the entire session of Periodic distraction, after separating D from ND trials in the usual way, is outlined in Figure 3-13. This shows the D error levels generally following the ND levels, with the largest differences occurring mainly in the first half of the session. As might be expected from examination of Figure 3-13, the ANOVA did not reveal effects of Distraction within this full session compilation of thirty-six exposures. The dominant effect across the session as a whole was the significant interaction of Forcing function X Zone, so forcing function characteristics determined the error. The ANOVA of twelve exposures was also unable to support distraction effects.

In the Random distraction experiment, the error effects had been concentrated in the first half-dozen trials. Similarly, here it looked as if the D error was larger than the ND error during the first six trials, except for the anomalous result at trial two (a mean decrease in the D condition error rather than an increase, an anomaly which contributed to the poor statistical outcome). Examination of the individual tracking records could not account for the trial two irregularity; no single subject was responsible. To aid inspection of these data, the first six trials have been replotted using a wider scale in Figure 3-14. Despite the differences between the mean levels for the first six exposures, the failure of the ANOVA to substantiate the distraction effect indicated that on a trial-by-trial basis the

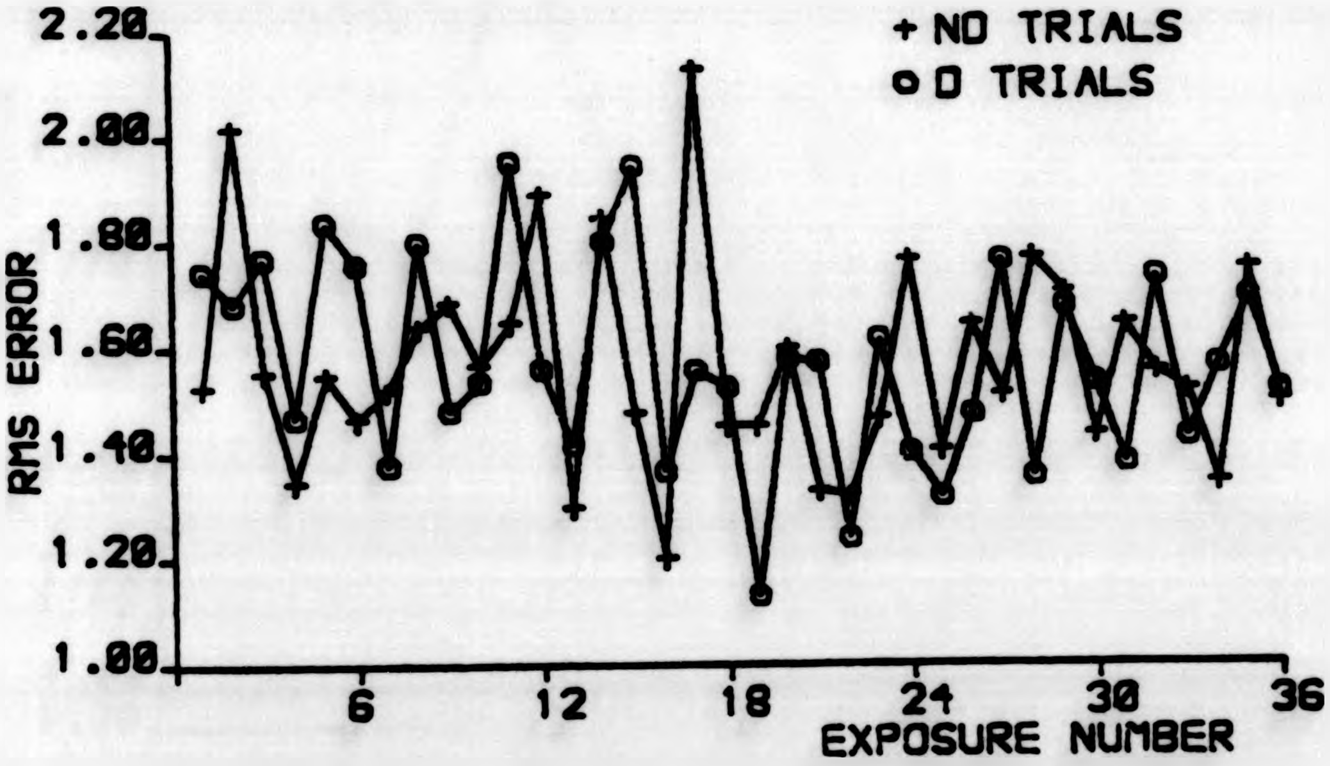

FIGURE 3-13 Distraction Experiment 2 (Periodic noise).
Tracking RMS Error: Distraction compared to no distraction.

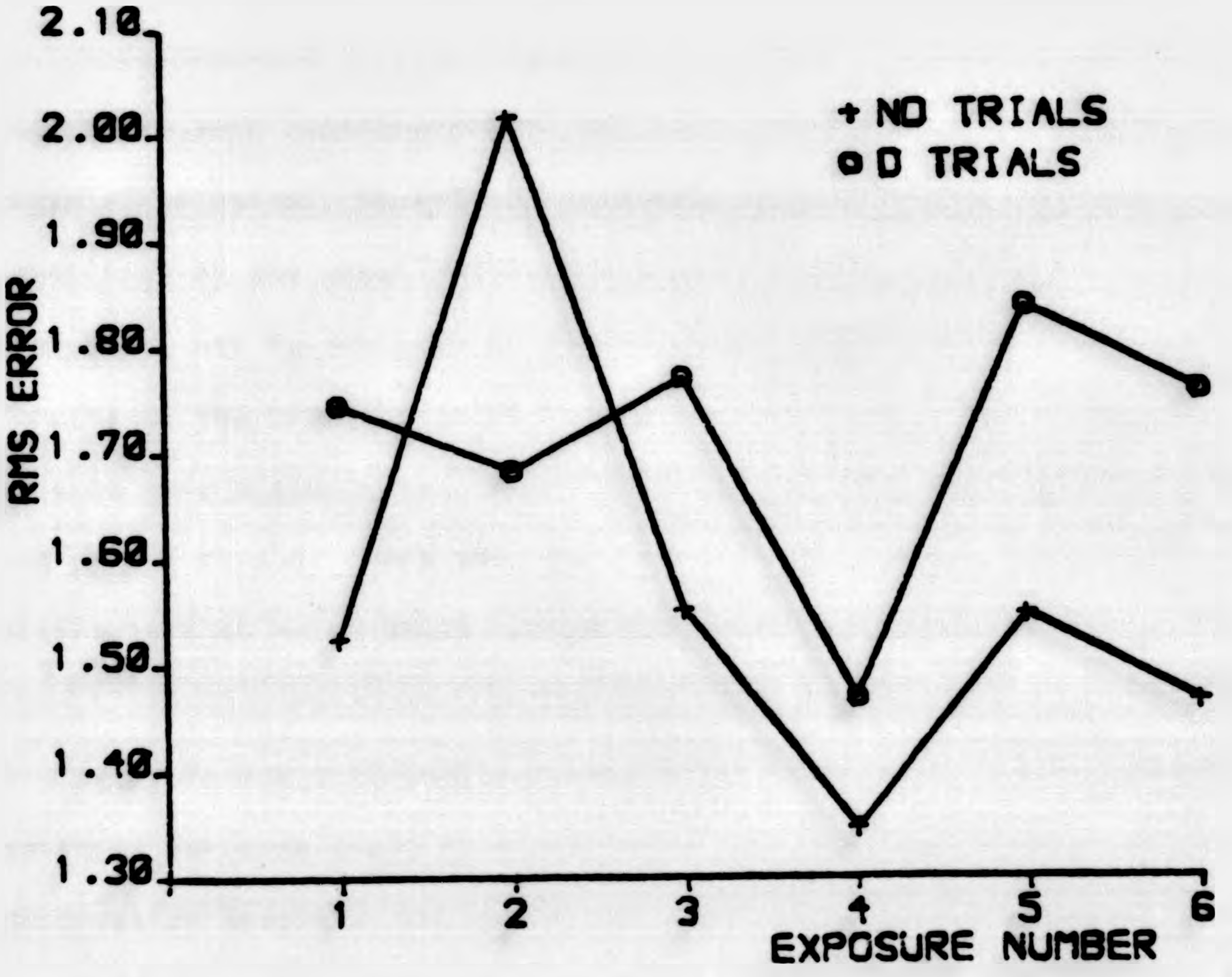

FIGURE 3-14 Distraction Experiment 2 (Periodic noise).
Tracking RMS Error: First six trials only.

effect was simply too erratic. The ANOVA of the first six exposures is summarised in Table 3-5. Similar analyses carried out on the first one, two, and three exposures likewise failed to yield any significant effects of distraction.

Effect of Periodic Noise distraction on amplitude of joystick movement

There was no evidence of any increase in the joystick movement amplitude during D trials. Joystick differences between D and ND conditions were inconsistent in the early trials, as Figure 3-15 clearly indicates. No effects dependent on Distraction were identified by the usual ANOVAs of thirty-six, twelve, and six exposures, nor by analyses of intermediate numbers of trials. The results of the previous study suggested that if movement amplitude effects were present, they would certainly be very small and last only one or two trials. Here analyses considering only the first three distractors or fewer still failed to detect distraction effects.

As in the Random distraction study, no differences in the joystick movement variance could be discerned at any time during the distraction session. No significant effects of any variable on the variance were detected by any of the ANOVAs which examined various numbers of trials in accordance with the standard procedure.

TABLE 3-5

DISTRACTION EXPERIMENT 2 (Periodic noise).
ANOVA Summary: Effects of distraction, forcing function, number of exposures, and time zone in the trial on the average RMS Error. First 6 trials only.

Source	*df*	*SS*	*MS*	*F*	*Prob*
Subjects (S)	5	18.622			
Distraction (D)	1	1.319	1.319	3.6805	0.1117
D X S	5	1.791	0.358		
Forcing function (F)	2	2.997	1.498	0.6766	0.5339
F X S	10	22.143	2.214		
Exposure number (E)	1	1.503	1.503	1.9912	0.2163
E X S	5	3.774	0.755		
Zone in trial (Z)	3	1.663	0.554	1.4326	0.2723
Z X S	15	5.804	0.387		
D X F	2	0.972	0.486	0.5060	0.6218
D X E	1	0.720	0.720	5.7876	0.0604
D X Z	3	0.327	0.109	0.2107	0.8875
F X E	2	0.386	0.193	0.3929	0.6889
F X Z	6	19.791	3.298	2.3315	0.0572
E X Z	3	1.051	0.350	1.0490	0.4008
D X F X E	2	1.699	0.850	0.6701	0.5370
D X F X Z	6	2.656	0.443	1.3628	0.2610
D X E X Z	3	0.990	0.330	1.1770	0.3521
F X E X Z	6	1.577	0.263	0.9129	0.5001
D X F X E X Z	6	1.237	0.206	0.9941	0.5522

* significant at $p<0.05$

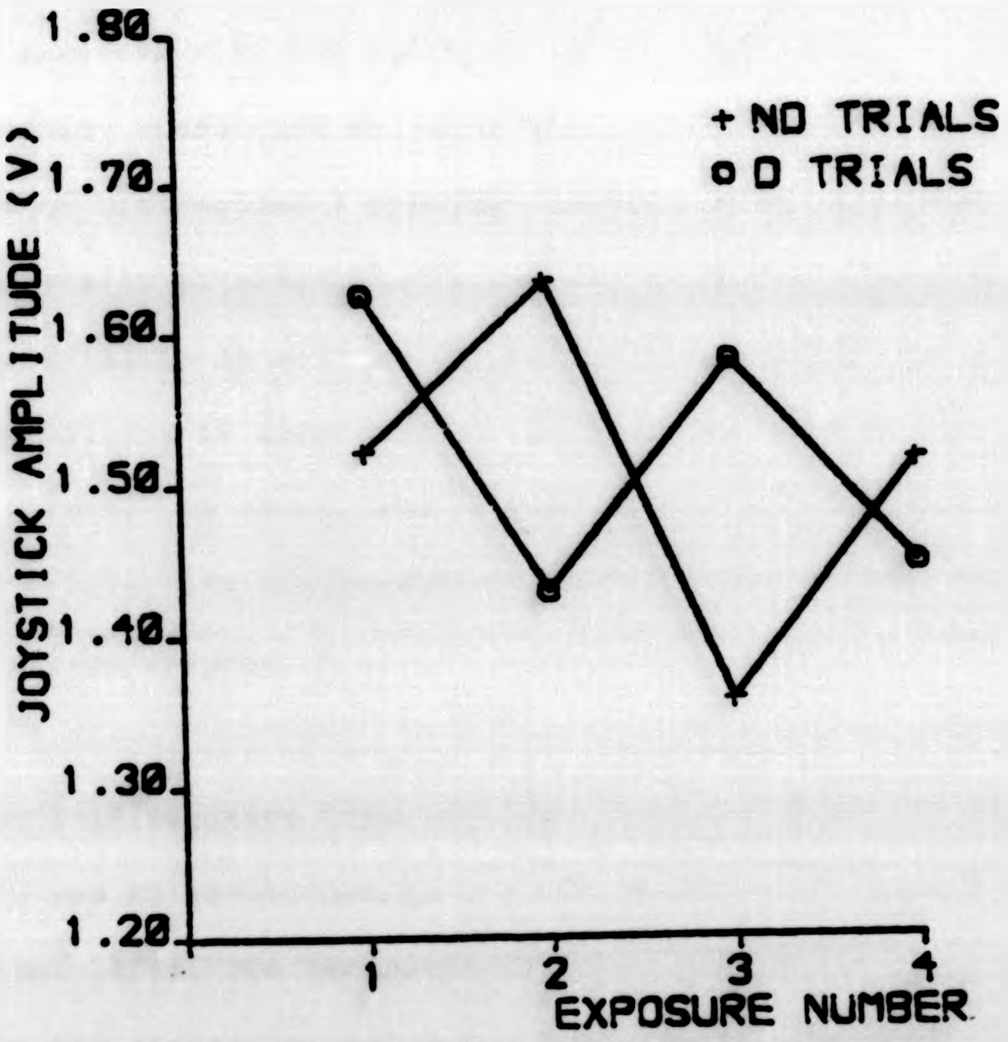

FIGURE 3-15 Distraction Experiment 2 (Periodic noise). Joystick movement size: First four trials only.

Effect of Periodic Noise distraction on response delay

There was some slight evidence, however, of a difference in response delay between the D and ND conditions. The lag data for the first twelve trials are found in Figure 3-16. The response delay effect was erratic, usually showing increased delay in the early D trials, but sometimes not. Where the effect was most consistent (trials six to eight), the average size of the increment was about 30 milliseconds, which compares with the observations of Experiment 1. The ANOVA for these data, summarized in Table 3-6, confirmed that the interactive effect of Distraction X Forcing function X Exposure number was statistically significant ($F_{6,47}$=2.49, p<.04), so the track characteristics as well as the number of exposures affected susceptibility to distraction. Distraction was not identifiable as a main effect. As usual, any influence of distraction was confined to early trials; the whole-session ANOVA did not contain significant distraction effects.

Figure 3-16 suggests that the mean level of delay tended to rise in both D and ND conditions during the earliest six trials of the session. Also, the estimated delays-- especially for the early ND data-- did appear to vary more from trial-to-trial in this experiment, compared to the results for Experiment 1 (cf Figure 3-9). Certainly the distraction effect was highly erratic during the earliest trials. The analyses of six and fewer exposures could not confirm any dependence on distraction.

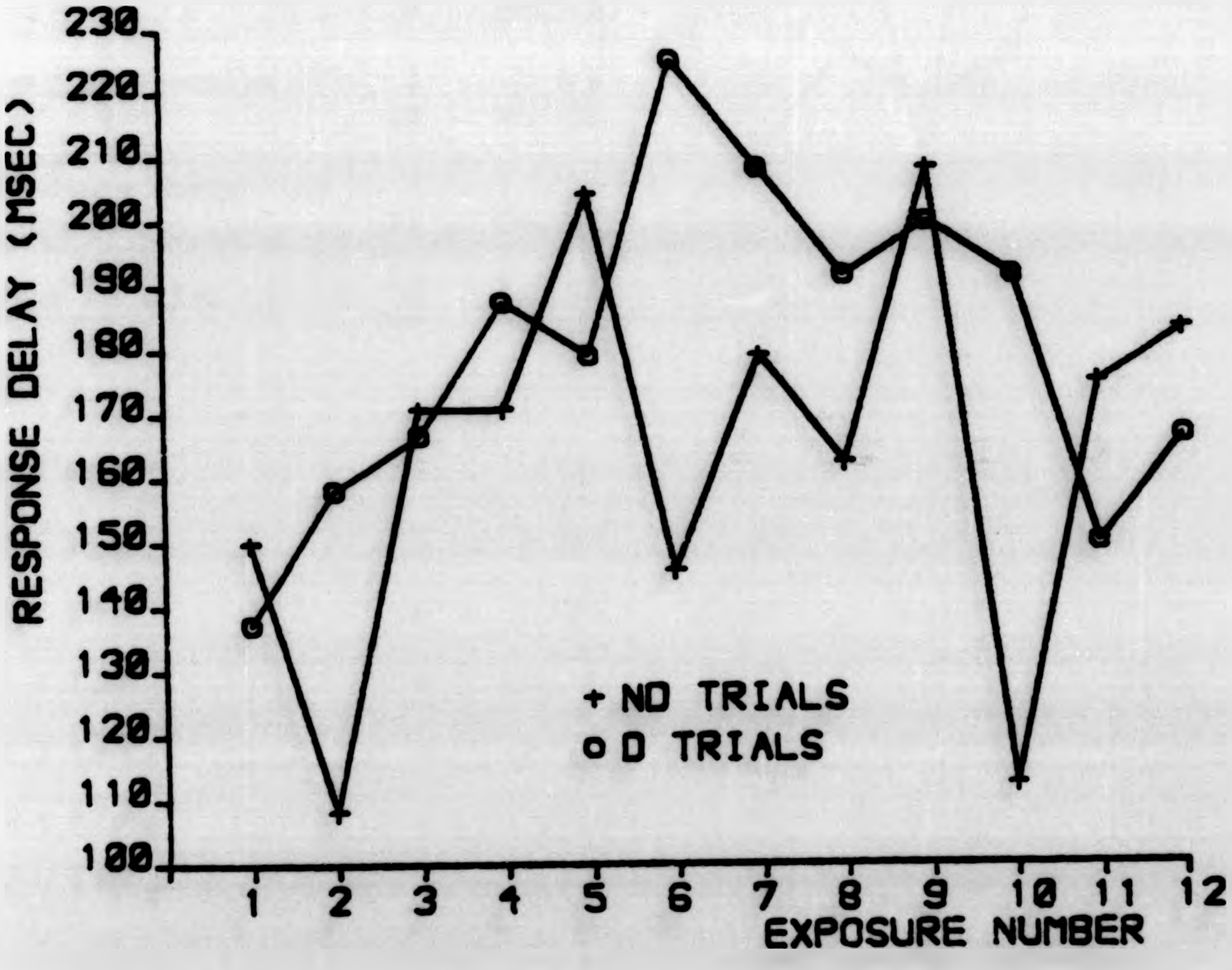

FIGURE 3-16 Distraction Experiment 2 (Periodic noise).
Tracking response delay: Distraction compared to no distraction.

TABLE 3-6

DISTRACTION EXPERIMENT 2 (Periodic noise).
ANOVA Summary: Effects of distraction, forcing function, number of exposures, and time zone in the trial on the tracking response delay. First 12 trials only.

Source	*df*	*SS*	*MS*	*F*	*Prob*
Subjects (S)	5	71.847			
Distraction (D)	1	3.674	3.674	1.4901	0.2767
D X S	5	12.326	2.465		
Forcing function (F)	2	3.566	1.783	0.3988	0.6852
F X S	10	44.705	4.470		
Exposure number (E)	3	17.153	5.718	2.5369	0.0950
E X S	15	33.806	2.254		
Zone in trial (Z)	3	13.125	4.375	1.2078	0.3413
Z X S	15	54.333	3.622		
D X F	2	1.129	0.564	0.3460	0.7191
D X E	3	0.340	0.113	0.1808	0.9075
D X Z	3	0.840	0.280	0.1664	0.9169
F X E	6	3.045	0.508	0.2651	0.9479
F X Z	6	62.281	10.380	2.3646	0.0543
E X Z	9	10.667	1.185	1.1358	0.3584
D X F X E	6	17.399	2.900	2.4918	0.0445 *
D X F X Z	6	8.191	1.365	0.7149	0.6420
D X E X Z	9	7.090	0.788	0.5259	0.8486
F X E X Z	18	34.052	1.892	1.2530	0.2384
D X F X E X Z	18	24.004	1.334	1.0220	0.4441

* significant at $p<0.05$

Subjective reports

Subjects again claimed that the noises had not been distressingly disruptive, and that after the first one or two they had grown accustomed to them. The reports were essentially the same as the accounts of the first experiment. Remarks made by several subjects did suggest that they had begun to anticipate the arrival of the distractors, basing their expectancy on an assumption that the intersignal interval was constant (for example, one subject reported that the arrival of distractors "seemed fairly regular at the end but not at the beginning"; another remarked that she "began to worry if the distractor felt like it was late"). Some subjects, then, had awareness that the distractor scheduling was periodic.

Discussion

In this experiment the evidence of a distraction effect was slight. The only reliable effect was confined to the tracking response delay. Noisebursts still retained some ability to lengthen this delay. Even so, distraction interacted with tracking demands and the effect disappeared early in the experiment. There were no consistent changes in the RMS error nor in the amplitude of the joystick movements, even during the earliest trials.

In contrast, the Random schedule of the previous experiment had brought about rather more robust effects directly attributable to distraction and to interactive processes involving track variables. Changes in RMS error, response delay, and-- very briefly-- joystick movement amplitude were all obtained under the Random conditions. When comparing the effects of the Random and Periodic schedules, it is striking that the Random distractors should have exerted so much more impact during an equivalent number of trials. The absence of measurable effects when Periodic scheduling was used, even for a comparable number of distractors, argues that the schedule itself, and not the absolute number of distractors, led to the rapid diminution of the distraction effects.

The failure to elicit effects during the very earliest exposures to Periodic distraction had not been anticipated. An hypothesis to account for this could be based on the early formation of expectancies. The distractors occurred at entirely regular intervals. The inherent predictability of the noisebursts (of which some subjects were consciously aware) may have reduced their surprise value, making startle responses and other distraction effects less likely. Another explanation could be that because the Periodic distractors arrived much closer together in time than the Random distractors, the earliest noisebursts may have affected arousal in a beneficial way. There have been suggestions that *continuous* noise of moderate loudness sometimes promotes better performance for this reason (Broadbent, 1971; Broadbent, 1979; Fisher, 1984a; Gawron, 1982). Such an effect would appear similar to the attentional alerting (illustrated by more precise responses to track reversals) which seemed to characterize individual tracking records occasionally. This could act to suppress the RMS error even while the distraction was tending to increase it, and could help to explain the erratic nature of the distraction effect. The wide variations in the estimates of response delay, observed for D and ND performance alike, could be evidence of a widespread instability of arousal encouraged by the regular arrival of noises in rapid succession. The results of the previous study had indicated that the adaptation to intermittent noise distractors had proceeded very swiftly. Even if the postulated arousal effect was very shortlived, it would have needed to interfere with only a few measurements to render the distraction effect undetectable.

Both of these experiments have established that the effects of distraction were unquestionably affected by the simultaneous demands of the tracking task. The interactions of the Zone and Forcing function effects with the Distraction treatment confirmed that susceptibility to distraction varied with local and not so local characteristics of the forcing function. Indeed, it is necessary to accept that in general terms the track conditions exerted greater control over the subjects' performance than did the distraction used in these experiments.

There has always been some uncertainty about how loud noise must be before it affects performance. Broadbent (1971) had concluded that the intensity of intermittent noise had to exceed 95 dB before performance would suffer. But Broadbent did not attribute such effects solely to distractibility, and developed a more general account of noise stress. He pointed out that at lower intensities the effects of noise were difficult to anticipate: any performance decrements were likely to be comparatively brief transient effects, and sometimes performance would actually improve. Later (Broadbent, 1979) it became clear that lower intensities could be stressful as well as distracting. The noisebursts used in the present experiments measured 82 dB, generally considered to be moderately loud. Tracking performance was certainly affected by this level of noise, and the effects have been shown to be very local. Fisher (1972) too had found that 80 dB noisebursts were sufficiently loud to evoke distraction effects during a serial reaction time task, and in this case also the effects were very closely bound to the distractors. In general, it does seem that in order to be distracting, white noise must be loud, even if research has not defined the necessary loudness terribly precisely. This, plus the evidence of rapid adaptation, suggests that noise in itself does not divert attention for long.

It follows that investigations of the effects of intermittent noise on performance have been tempted to attribute the influence of noise to an increase in stress and altered levels of arousal. Typically a change in attention or evidence of distractibility is viewed as just one of several consequences of the arousal state. Such descriptions often omit any reference to the information conveyed by the distracting event. Yet this can be the most striking property of a natural distractor. Everyday experience suggests that many distractors are effective specifically because of their rich information content or semantic properties. An example is hearing one's own name spoken, which is a demonstrably potent distractor (Moray, 1959). There is every chance, of course, that information-loaded distractors will have affective or arousing properties as well, even if they would not be considered stressful as such. White noise is not very interesting. Even if it is loud enough to alter arousal, its information content is low and its irrelevance easily verified, so its demands on attention and hence its distraction potential might differ from a realistic distractor.

The noisebursts of Experiments 1 and 2 typified "high-energy but low-information" distractors. In contrast to these loud (and perhaps stressful) noises, the third experiment used verbal distractors of only comfortable loudness. The distractors were designed to be interesting but to remain irrelevant to the task. The purpose of the experiment was to determine whether "low-energy but high-information" distractors such as these would produce effects on tracking performance in any way similar to the effects of loud white noise.

DISTRACTION EXPERIMENT 3: VERBAL DISTRACTORS

Procedure

The eight subjects (three females, five males) were unpaid volunteers from the Psychology Department undergraduate subject pool, all participating in a tracking experiment for the first time.

The entire experiment was completed in one session. Training used the adaptive tracking task, as before, but training time was shortened. This was necessitated by the difficulty of recruiting subjects for the lengthy four-day procedure used formerly. All subjects performed 15 minutes of adaptive practice, with rest breaks and encouragement after 3, 9, and 15 minutes. The task used a zero-order control law throughout. Following training, the distraction trials began, taking a further 7.5 minutes.

For the distraction measurements, the forcing function bandwidth was fixed at the mean level each subject had achieved during the final three minutes of adaptive practice. No indication was given to the subjects that the bandwidth was being held constant during the last part of the session. The controlled forcing function methodology adopted during the previous distraction experiments was again used. In this experiment the forcing function generator was restarted every 45 seconds. Three sections of forcing function were reliably replicated each time. Accordingly, during the whole of the distraction phase each forcing function appeared 10 times and these repetitions provided all of the data collection trials. When distractors occurred they were always synchronized to one of the three repeatable forcing functions. Nondistracted tracking data were also accumulated for each forcing function. The sampling window within each trial was wider in this experiment, extending from one second prior to potential distractor onset to eight seconds after onset. More samples were taken so that the distractors could be of longer duration than previously, and so that more data could be collected after offset. During the sampling, joystick and forcing function readings were taken every 100 milliseconds, as before.

Distractor timing was irregular and infrequent, but did not use the same random schedule as Experiment 1 because of the altered trial length. The mean onset-to-onset interval was 31.1 seconds. The intersignal intervals were distributed as indicated in Figure 3-17. In all, 15 distractors were presented. The same presentation schedule was used for each subject.

The verbal distractors were English sentences, prerecorded by the experimenter onto magnetic tape. Each message had been rehearsed until its duration was as close as possible to four seconds. All subjects received the same messages in the same order. The messages were played back one at a time on a Revox 77 tape deck. The computer controlled the tape movement during playback so the timing of each distractor onset was very tightly matched across subjects. Using the same equipment as in the previous experiments, the messages were amplified and presented by loudspeaker. Sound levels were comfortably

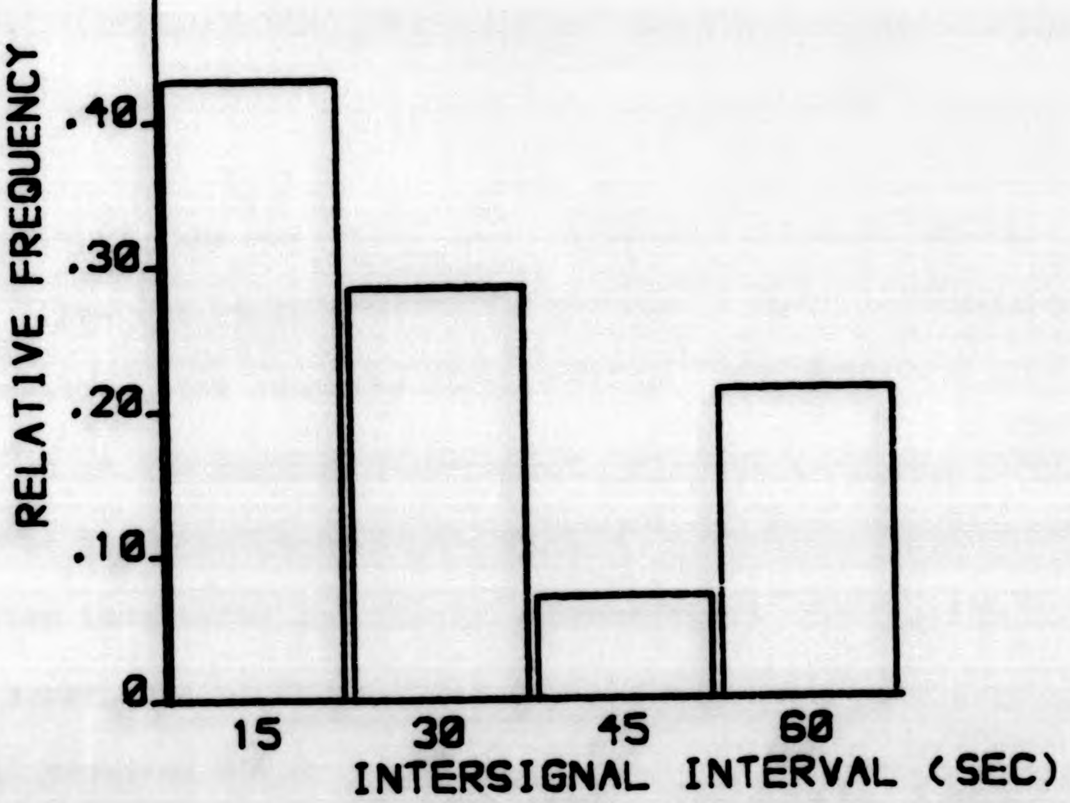

FIGURE 3-17 Distraction Experiment 3 (Verbal distractors). Intersignal intervals.

loud at around 65 dBA, against a quiet background of 32 dBA.

The distractors were intended to be interesting and likely to solicit attention but not to convey any information relevant to the experiment. The message texts are listed in Table 3-7. In this experiment the subjects were not forewarned of the distractors and the messages were not demonstrated.

Results

Training

The training time was relatively short in this experiment. Also, the control law was restricted to zero-order dynamics throughout. Consequently, the adaptive bandwidths did not reach the high levels obtained in the earlier experiments, and the absolute levels of RMS error at the conclusion of training were lower. The acquired forcing function bandwidths and the normalised RMS errors are shown for the full experiment in Figure 3-18. These RMS errors are indiscriminate global measures which do not separate D from ND performance. Matched against the training for the earlier experiments, it is apparent that the error measures are very similar for equivalent amounts of practice. So although these subjects were not allowed to become as skilled as the earlier subjects, they responded to adaptive training in much the same way.

TABLE 3-7

DISTRACTION EXPERIMENT 3 (Verbal distraction).
Texts of messages used as distractors, in order of presentation. The messages were prerecorded and each lasted four seconds.

(1) They are getting ready to test it now.

(2) We don't want it to break down this time.

(3) Performance is being measured by computer.

(4) They've all got certain psychological skills.

(5) No one has ever seen it like that.

(6) Some messages could be very important.

(7) They should finish before the lights go out.

(8) All of them were wearing shoes and socks.

(9) I think they will be able to hear it now.

(10) The purpose of the experiment is clear.

(11) Mental arithmetic can be difficult for some people.

(12) The chairs don't have any adjustability.

(13) Medical problems don't matter in this case.

(14) The arms and legs were still attached.

(15) These messages are trying to distract you.

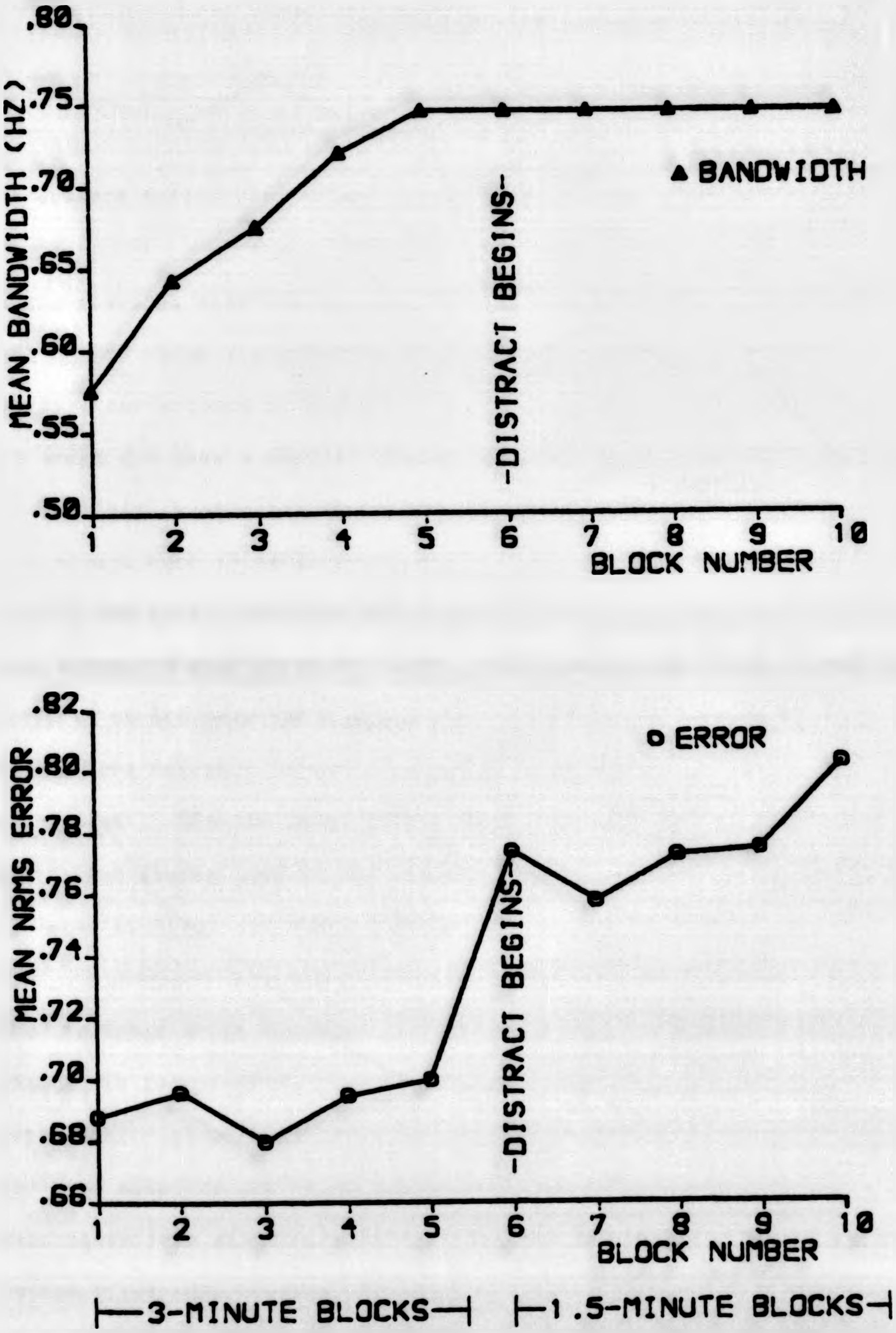

FIGURE 3-10 Distraction Experiment 3 (Verbal distractors). Tracking performance during the experiment.

Performance during distraction

In the previous experiments, the introduction of distractors to the task caused local disturbances but could not be shown to disrupt tracking performance in a general and diffuse way. Here, however, the RMS error did show a general increase as soon as the distracted trials began. This is apparent in Figure 3-18 at the point at which distractors were introduced (block 6). As distraction began, the overall RMS error increased and remained large. To verify the effect, each subject's mean error for block 6 (the first distracted block) was compared to his mean for block 5 (the last block of training). The errors were reliably larger in block 6 ($t_{7df}=2.375$, $p<.025$ one-tailed). Similar large errors were sustained during subsequent distracted blocks, and so the mean of blocks 6 through 10 inclusive likewise exceeded the block 5 level ($t_{7df}=3.196$, $p<.01$).

The increased error has been attributed to distraction and not to changes in the tracking task bandwidth during this part of the experiment. It will be recalled that during training the bandwidth had been adaptive and so had exhibited some variability around the mean bandwidths shown in Figure 3-18. The forcing function bandwidth became fixed when distraction began (at block 6), but the constant bandwidth adopted at that time was the mean value achieved during the final block of training (block 5). In other words, the tracking demands imposed by the fixed bandwidth were no more difficult than the average demands of the final three minutes of adaptive training. Yet the RMS error greatly increased during the fixed bandwidth task. In Chapter 2, Tracking Experiment 2 compared performance during fixed bandwidth and adaptive tasks. That experiment demonstrated that a fixed bandwidth forcing function would cause smaller tracking errors than an adaptive task which was using, on average, an even lower bandwidth disturbance. In this distraction experiment performance did not benefit from the switch from adaptive tracking to fixed bandwidth conditions as would have been expected. The error became larger when the bandwidth was held at a level matched to the previous adaptive mean. This is reasonable evidence that the introduction of distractors, not the constant bandwidth, was responsible for the change in performance. Under normal nondistracted conditions, we would have expected the fixing of the bandwidth in block 6 to have encouraged smaller or near-equal errors but not a significant increase in the error. Still further support for the view that the distractors, not the bandwidth manipulation, caused the change comes from the previous observations of tracking error during Distraction Experiments 1 and 2 (shown earlier in Figures 3-2 and 3-12, respectively). In neither of those experiments did the switch from the adaptive to the fixed bandwidth task (at block 19 in those studies) provoke a sustained increase in the RMS error as it did here. Even acknowledging the difference in training time here as compared to the earlier procedures, the introduction of distractors remains the simplest explanation for the substantial and prolonged increase in error in the present experiment.

Analysis of distraction trials

Again the D and ND trials were paired in presentation order. The effects of Distraction, Forcing function, Exposure number, and trial Zone were evaluated by analysis of variance. ANOVA was initially applied to the entire session's data, namely all fifteen distractors. Additional smaller analyses selected out the first three, six, nine, or twelve exposures for examination. As in the previous studies, this was done to help identify any effects which were not sustained across the full timescale of the experiment.

Effect of Verbal distraction on tracking error

The increase in tracking error specific to D trials was small and variable. Figure 3-19 shows the RMS error during the D trials with corresponding ND scores plotted separately. The mean RMS error was 0.346 in the D condition, compared to 0.339 in the ND trials, but the D error did not exceed the ND level consistently. The ANOVA for these data is summarised in Table 3-8. As usual, the global and local features of the forcing functions were major determinants of the error, as shown by the significant Forcing function and Zone main effects, respectively, and the reliable Forcing function X Zone interaction. The interaction of Forcing function X Zone was a larger effect than either the Forcing function or Zone alone, confirming that the error during any particular second within the sampling window was

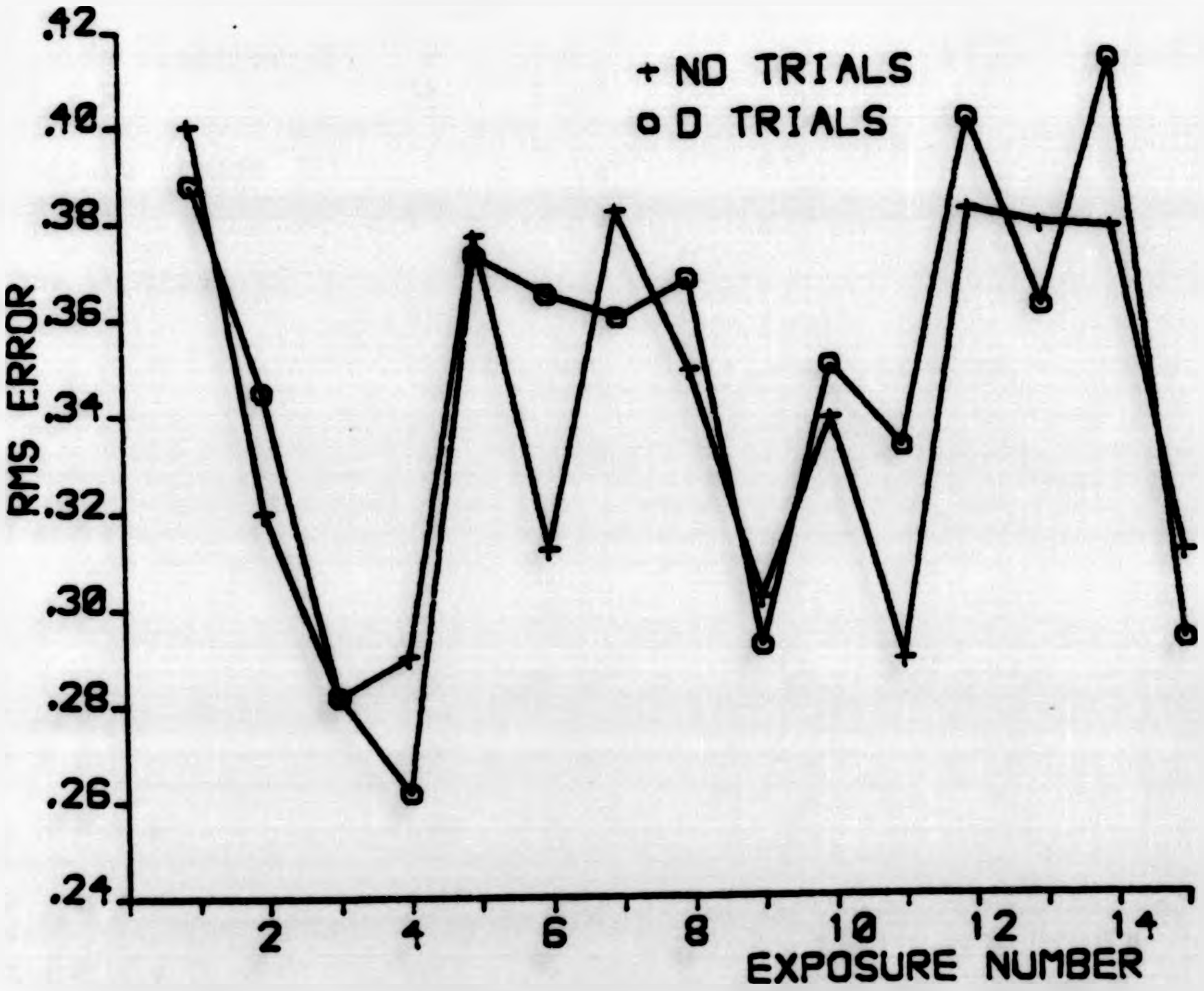

FIGURE 3-19 Distraction Experiment 3 (Verbal distractors). Tracking RMS Error: Distraction compared to no distraction.

TABLE 3-8

DISTRACTION EXPERIMENT 3 (Verbal distraction).
ANOVA Summary: Effects of distraction, forcing function, number of exposures, and time zone in the trial on the average RMS Error. 15 exposures to distraction (entire session).

Source	*df*	*SS*	*MS*	*F*	*Prob*
Subjects (S)	7	16.289			
Distraction (D)	1	0.019	0.019	0.3003	0.6049
D X S	7	0.446	0.064		
Forcing function (F)	2	2.736	1.368	7.4343	0.0064 *
F X S	14	2.576	0.184		
Exposure number (E)	4	0.165	0.041	1.4669	0.2381
E X S	28	0.788	0.028		
Zone in trial (Z)	7	5.017	0.717	4.5227	0.0006 *
Z X S	49	7.765	0.158		
D X F	2	0.036	0.018	0.7787	0.5186
D X E	4	0.048	0.012	1.0086	0.4207
D X Z	7	0.069	0.010	1.1717	0.3360
F X E	8	0.131	0.016	1.1893	0.3218
F X Z	14	25.291	1.807	17.6754	0.0000 *
E X Z	28	0.221	0.008	0.8844	0.6373
D X F X E	8	0.208	0.026	1.5095	0.1742
D X F X Z	14	0.216	0.015	2.3129	0.0086 *
D X E X Z	28	0.174	0.006	1.0584	0.3932
F X E X Z	56	0.571	0.010	0.9896	0.5005
D X F X E X Z	56	0.276	0.005	0.7763	0.8773

* significant at $p<0.05$

heavily influenced by the choice of forcing function. Distraction exerted its effects in a manner which was sensitive to ongoing track events. This was indicated by the significant complex interaction of Distraction X Forcing function X Zone ($F_{14,84}$=2.31, p<.01). The interaction of Distraction with these variables again supports the conclusion reached by the previous studies, that distraction effects on tracking are modulated to a very great extent by the momentary task requirements. All of the effects cited above were very robust and were significant in all of the ANOVAs of earlier trials as well (except in the ANOVA comprising exposures one through six, for which the D X F X Z interaction appeared only at the p<.07 level). Some of the differences due to track characteristics are suggested by Figure 3-20 which shows the error outcomes as pooled into three "average" forcing function categories. Forcing function 2 shows the distraction effect most consistently, while Forcing functions 1 and 3 share somewhat similar error histories and offer rather less consistent distraction effects. These results illustrate the smallness of the error effect and the basis of the interactions reported above.

The distraction effect on the error, however small, could still be seen eight seconds after the onset of the distractor. Error effects were apparent well beyond the immediate vicinity of distractor onset. The speech messages terminated at four seconds, yet a discrepancy between D and ND performance was in evidence toward the end of the sampling window, a considerable time after the messages had finished. This is suggestive of a longterm and diffuse influence on performance rather than transient disruption. Distraction did not interact

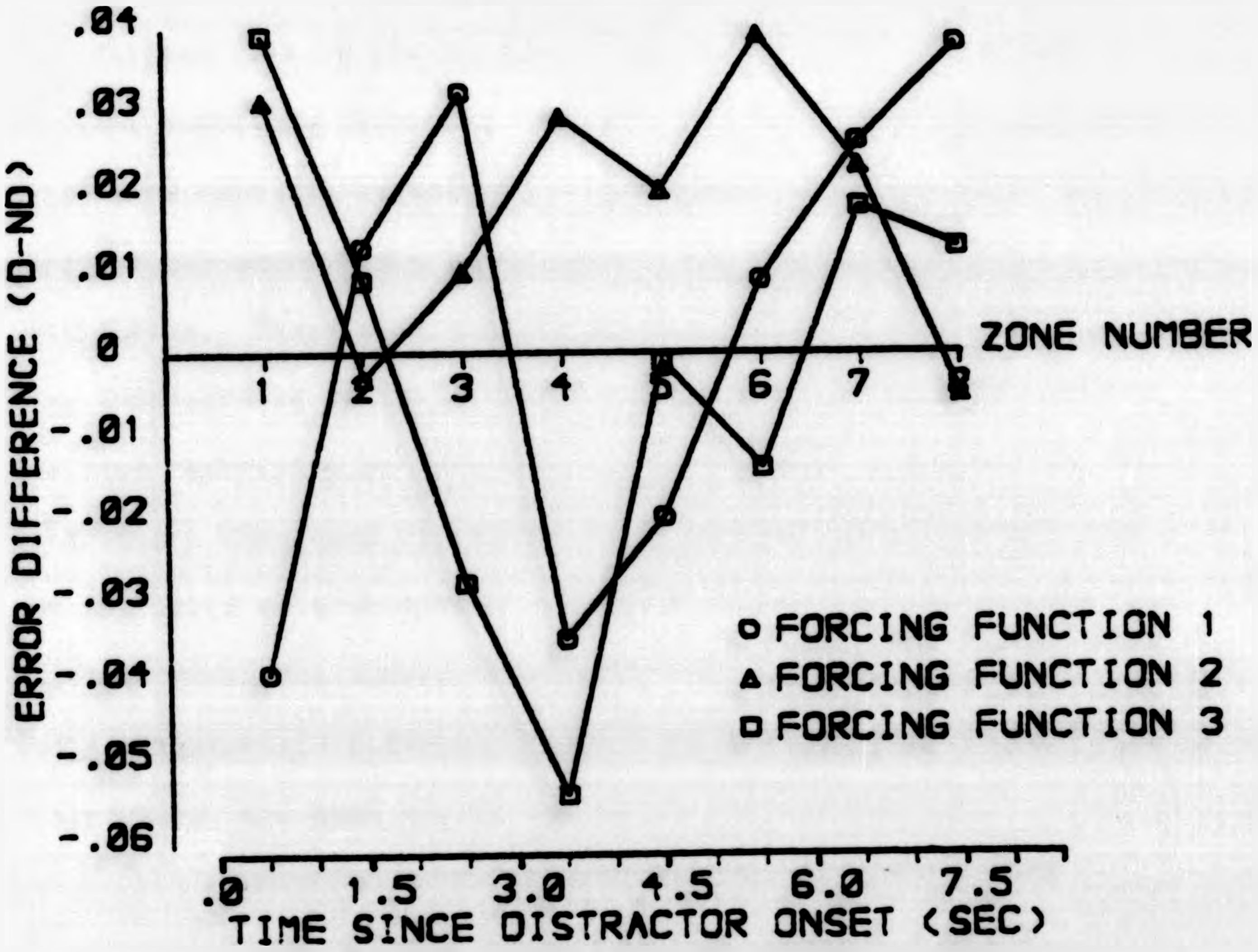

FIGURE 3-20 Distraction Experiment 3 (Verbal distractors).Distracted and Nondistracted RMS Error: dependency on track characteristics.

significantly with the trial Zone alone (ie without the interactive participation of Forcing function), so the purely temporal properties of the distractor such as onset or offset did not exert systematic effects on the error.

Effect of Verbal distraction on amplitude of joystick movement

The analyses were unable to detect any differences in joystick movement amplitude resulting from distraction. The only variables which were reliably associated with differences in movement amplitude were the one-second Zone partitioning and the Forcing function X Zone interaction. These were significant for all of the ANOVAs, whether they incorporated as few as three exposures or as many as fifteen. Analyses restricted to just the first and second presentations of distractors continued to confirm only these effects without revealing any influence of distraction. In other words, the size of the joystick movements simply depended on the behaviour of the track and was not measurably different in the immediate vicinity of the distractors, not even during the very earliest exposures.

Effect of Verbal distraction on response delay

The estimates of tracking response delay over the full course of the experiment are shown in Figure 3-21. Only the data for the first few trials carried any suggestion that there might be an increase in

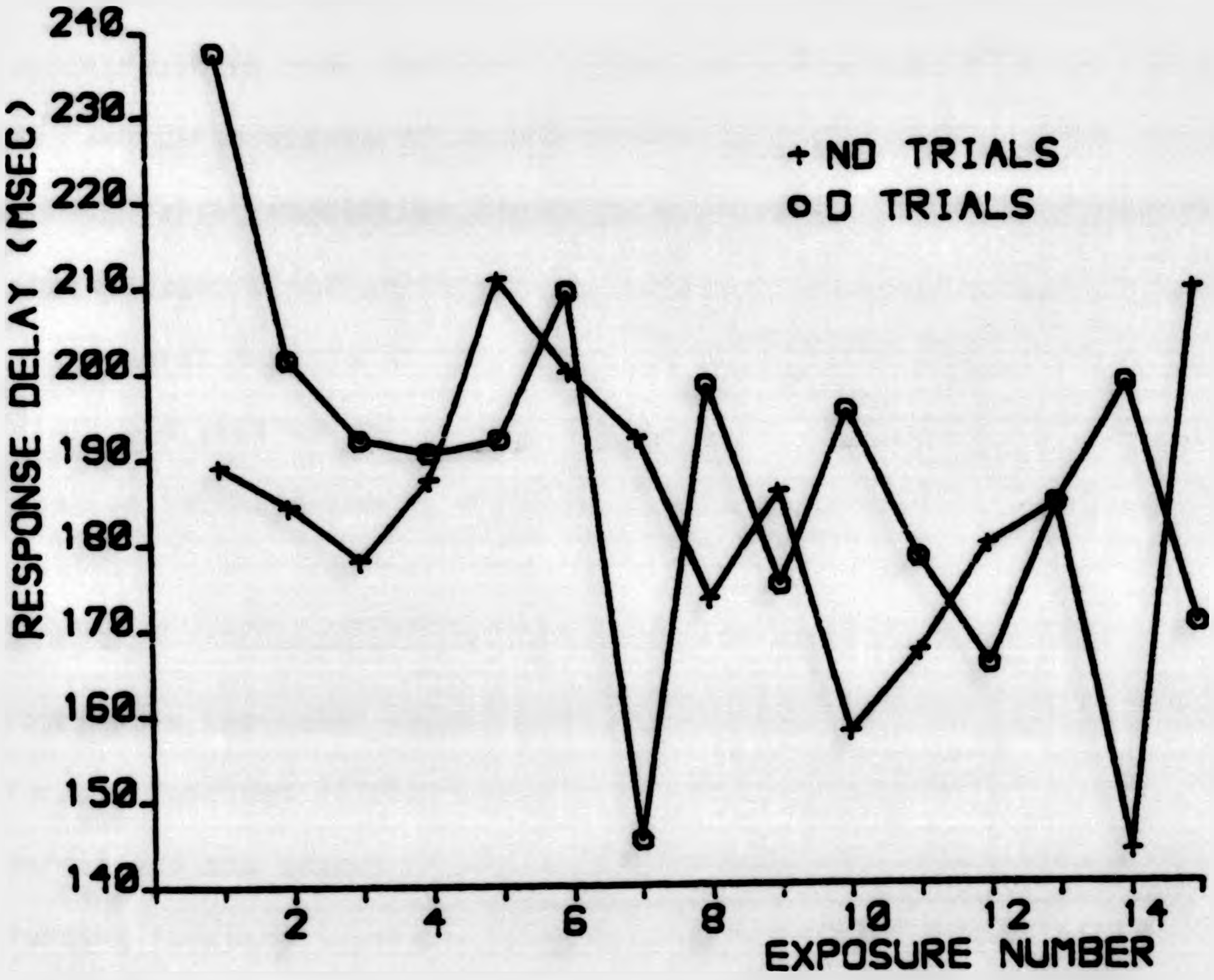

FIGURE 3-21 Distraction Experiment 3 (Verbal distractors).
Tracking response delay: Distraction compared to no distraction.

response delay specifically due to distraction; subsequent trials demonstrated little consistency. The ANOVA for these data, Table 3-9, identified a significant interaction of Distraction X Forcing function X Exposure number X Zone ($F_{\text{[illegible]}}=1.38$, $p<.05$), but no main effect of Distraction could be confirmed. In analyses of smaller groupings of trials even this complex interaction disappeared. Despite the impression of increased delay in the earliest results of Figure 3-21, the delay estimates conveyed a good deal of variability and no significant distraction effects emerged even when analysis concentrated on just the first three presentations (the earliest one, two, and three exposures were examined without success). The four-variable interaction indicated solely by the whole session ANOVA (which suggests that distraction effects on delay time were modified by the joint effects of track behaviour, temporal position within the trial, and experience) is such an isolated result that it cannot be taken as very convincing evidence of local distraction-evoked delay effects.

There were two other significant effects on response delay. The Forcing function X Zone interaction merely verified that the delay during any one second of the trial was affected by the prevailing forcing function (just as error had been). Exposure number also emerged as a significant factor. This outcome was more interesting, since it was not typical of the earlier distraction experiments. The Exposure number effect confirmed the impression conveyed by Figure 3-21 that the mean delay tended to decrease over the course of the experiment. This observation could pertain to the absence of

TABLE 3-9

DISTRACTION EXPERIMENT 3 (Verbal distraction).
ANOVA Summary: Effects of distraction, forcing function, number of exposures, and time zone in the trial on the estimated tracking response delay. 15 exposures to distraction (entire session).

Source	*df*	*SS*	*MS*	*F*	*Prob*
Subjects (S)	7	160.807			
Distraction (D)	1	1.752	1.752	0.8349	0.6056
D X S	7	14.689	2.098		
Forcing function (F)	2	0.978	0.489	0.2971	0.7508
F X S	14	23.047	1.646		
Exposure number (E)	4	19.182	4.796	3.3993	0.0218 *
E X S	28	39.500	1.411		
Zone in trial (Z)	7	26.814	3.831	0.8714	0.5367
Z X S	49	215.410	4.396		
D X F	2	3.982	1.991	0.9394	0.5835
D X E	4	7.732	1.933	1.3691	0.2694
D X Z	7	10.765	1.538	0.7481	0.6342
F X E	8	10.652	1.332	0.9638	0.5261
F X Z	14	107.964	7.712	2.2804	0.0097 *
E X Z	28	34.243	1.223	0.7314	0.8361
D X F X E	8	26.439	3.305	1.8110	0.0937
D X F X Z	14	28.526	2.038	1.2214	0.2721
D X E X Z	28	27.126	0.969	0.7456	0.8203
F X E X Z	56	79.447	1.419	0.8643	0.7446
D X F X E X Z	56	109.927	1.963	1.3771	0.0449 *

* significant at $p<0.05$

distraction effects on response delay. The point will be raised in the later Discussion.

Subjective reports

The subjects' own reports were of particular interest here because no warning had been given that distractors would occur, and the purpose of the messages was not explained until completion of the experiment. When questioned afterwards, subjects expressed some amusement about the messages and most did think that they were some form of distraction. Although it was generally maintained that only the first few exposures had commanded attention, most of the subjects were able to recall with reasonable accuracy messages from late in the session which they had found particularly intriguing or amusing, so these distractors had evidently been processed into memory. The final message, "These messages are trying to distract you," was recognised by every subject during questioning after the experiment.

Discussion

In this experiment the introduction of distracting material resulted in global as well as local worsening of performance. This was indicated by an overall rise in RMS error common to D and ND measurements alike. The performance decrement was sufficiently general that extremely local effects were not very conspicuous. Furthermore, because the distraction effect interacted with the local task demands, there was only a little evidence that the error became especially large in the immediate vicinity of distracting events. Evidence of local effects took the form of an interaction of Distraction X Forcing function X Zone. This observation supports the view that the relative efficacy of a distractor is affected by the momentary state of the track. The interaction was significant throughout most of the experiment. It appeared that differences between the D and ND error persisted well beyond the offset of the distracting messages (and up to the boundary of the sampling window in this experiment), emphasizing that the distraction effect was not exclusively tied to the simultaneous presence of a distractor.

The lack of strong local effects associated with distractor placement was also affirmed by the absence of changes in the joystick movement amplitude at the moment of distractor onset. The absence of such changes is compatible with the suggestion made earlier in this chapter that an initial transient increase in joystick movement size could be a startle reaction to loud noise rather than an essential feature of the distraction effect. An absence of startle reactions would be understandable here, where the distractors were considerably quieter.

Evidence of rapid adaptation to distraction was missing also. In contrast to the earlier experiments, there was no significant interaction of Distraction X Exposure number in any of the ANOVAs of tracking error. Actually, the failure to observe this interaction is not in itself adequate proof that the distraction effect resisted repeated exposure; after all, it was not easy to measure any systematic differences between the D and ND trials in these data. Even so, no main effect of Exposure number on the error was observed either, which confirms that the general increase in error common to both D and ND trials persisted throughout the experiment. It would be justified, therefore, to conclude that repeated exposure to these verbal distractors did not provoke dramatic adaptation effects.

In the previous experiments the distraction effect was very localised, and adaptation to distraction was unavoidable and swift. Yet here local effects and adaptation processes were difficult to identify. Some of the differences observed in the present study could have resulted from the use of verbal material with its different acoustic, temporal, and informational properties. Some differences may have been prompted by the abbreviated training period. These methodological alterations could have shaped the present results in several ways.

For example, one conjecture to account for the absence of adaptation is that the verbal messages had such varying potency as distractors that this made the effects of repeated exposure far from systematic. If one or two of the messages arriving late in the session were (in an acoustic or informational sense) particularly compelling distractors, they may have been capable of countering-- or reversing-- adaptation.

It would be reasonable to suspect that verbal distractors would lead to especially variable effects on performance. The verbal material brought at least two types of uncertainty to the task. As in the previous studies, there was temporal uncertainty associated with the distractor's onset time. Additionally in this experiment, the distractors had unpredictable content. Since the distractors here were not homogenous bursts of white noise but rather sources of continuous information, there is no reason to assume that the interruptions were all equally effective nor that each was uniformly disruptive across its four second duration. Variations in semantic or acoustic potency across the message would contribute variability to the observed distraction effects. This would be in addition to the variability arising from the interaction between the temporal boundaries of the distractor and the local demands of the tracking task. Semantic variation would tend to encourage analysis of every distractor. The subjects' ability to recognise the contents of distracting messages certainly argues that they were not completely ignoring or blocking out the verbal material even toward the end of the session. It would have been interesting to determine whether particular messages had been especially potent distractors, but the data from this experiment cannot describe pure message effects. With the tape apparatus used here, it was necessary to present the messages to all the subjects in the same order, so message content was bound up with the effects of exposure number in the experimental design. Future investigations of verbal distraction could better evaluate the role of message content in adaptation by varying the presentation order to balance for the contents effects.

Shortened practice time could also have promoted variability in the performance measures, leading to weak local measures of distraction effect. The estimates of tracking response delay suggested that tracking performance was not really as consistent during distraction trials as the error measurements had implied. The estimated response delays were long and they extended over a wide range. The delays had considerable variance, as evidenced by the large error-within terms in the ANOVAs. It is feasible that such instability was an outcome of the shortened training procedure. This impression of performance instability was supported by the observation that estimates of delay throughout this experiment appeared to be more sensitive to increasing experience than in either of the previous studies. The tendency here was for the delay to diminish as the testing proceeded, as was noted some pages earlier. This could be interpreted as a beneficial practice effect. The ANOVAs for the larger groupings of trials (incorporating nine exposures or more) verified that Exposure number exerted a significant influence on the response delay. The progressive shortening of the delay during the experiment is visible in Figure 3-21. The tendency suggests that tracking performance was not that stable. It is even possible that some distraction-induced increases in response delay could have been masked by practice benefits which were continuing to reduce delay throughout the experiment.

Finally, the verbal distractors could have exerted a general loading effect on the task as a whole, affecting D and ND trials alike, mainly because of stressful affective cues they conveyed. Fisher (1984a) has alluded to the "worry work" which can be imposed by some stressors over and above the information processing demands of the events themselves. For example, subjects could be devoting part of their processing capacity to assessing the relevance of certain verbal messages, or to seeking some kind of semantic continuity across successive messages. The role of the subject's "set" on arousal and performance in distracted tasks is far from clear, but the suggestion that it can complicate distractibility is a recurrent theme in the literature. In 1918, when introspective reports were still being avidly collected, Cassel and Dallenbach noted that the subject's "conscious attitude" seemed to be related to the susceptibility of reaction time to the disruptive effects of noise. Broadbent, in a study of the effects of continuous noise on sensorimotor reactions, commented that subjects were influenced by their initial experience with the noise: "subjects first meeting the task under the worst possible conditions [of noise] do badly throughout the experiment" (Broadbent, 1957, p23). Fisher (1983) offered evidence that subjects expect noise to degrade their performance and so come to experiments with this set. Such factors as worry work and set could have helped to spread the effects of the verbal distractors of Experiment 3 across D and ND trials indiscriminately and would have acted against adaptation.

In summary, these subjects displayed larger tracking errors when verbal distractors were introduced to the task, but the deterioration in performance was diffuse and not just confined to a brief period following distractor presentation. Within the brief period following presentation, the effects of distraction interacted with the simultaneous demands of the track, so local consequences were to a great extent task-determined. This general picture of impaired performance was not very sensitive to experience and the distraction effect survived repeated exposures. The subjects were not themselves conscious of long-term effects.

Auditory distraction: Conclusions

Clearly, increases in tracking RMS error are elicited by "low-energy but high-information" verbal distractors as well as by "high-energy but low-information" stimuli such as noisebursts. In either case the distraction effect is task-sensitive: the susceptibility to distraction is influenced by the tracking demand at any given moment.

In these experiments the effectiveness of white noise as a distracting agent depended very much on the schedule of distractor presentation. Intermittent noises produced only very local effects on tracking error, lasting but a few seconds beyond distractor arrival. These effects were much diminished by repeated exposure; moreso if the noises came at frequent fixed intervals. Indeed, unless the noises arrived infrequently and at random times it was not possible to measure any sustained effects of distraction on the RMS error. With suitably random noiseburstѕ, there was also an increase in the tracking response delay which was clearly associated with distractor presentation. In addition, there was a suggestion that the loudness of random noisebursts prompted wide deflections of the joystick during the first few exposures, as if the initial bursts had startled the subjects and caused a brief loss of motor control. The subjects' self-awareness of distraction effects was limited; indeed, it was as if conscious knowledge of distraction effects did not persist beyond the initial startle effects. In contrast to these observations, when the noisebursts arrived at frequent and entirely regular intervals the only reliable effect of the distraction was some lengthening of the tracking response delay, and the effect endured only briefly. Taken together, these results suggest that a loud noise disrupts tracking performance mainly because it increases the subjects' response delay, and the increase is a transient disturbance which follows the placement and duration of the distractor very closely. Repetition of the distracting noise, especially at regular intervals, leads to a rapid disappearance of the effect.

Verbal distractors also led to increases in tracking RMS error. These distractors were at much lower energy levels than the noise events, so distraction effects on tracking evidently are not restricted to situations which impose high levels of noise stress on the subject. The error effect was not so localised and was more persistent over repeated exposures. The increase in error was not accompanied by a reliable immediate increase in tracking response delay as it had been in the previous studies. Because it had been necessary to introduce other methodological changes in this study, it was not certain that the properties of the verbal distractors were entirely responsible for the differences in the results. The differences obtained do add weight to arguments that verbal distractors create special conditions and pose special problems for experimental analysis. It would be justified to treat verbal distraction as a somewhat separate issue demanding a comprehensive research strategy. The research programme would have to establish boundary conditions empirically for this difficult class of distracting material. For example, it would be helpful to clarify the (probably separate) roles of acoustic and semantic properties in verbal distraction. It would be necessary to study more carefully the relation between the distractor's duration and the duration of the distraction effect so that local versus global consequences might be anticipated. If such a research plan were based on the tracking methodology developed here, it would surely contribute to the understanding of distractibility in tasks other than tracking as well.

CHAPTER 4: TEMPORAL EXPECTANCY DURING TRACKING

Overview

An experiment investigated the ability of subjects to anticipate the temporal position of brief acoustic events. The subjects' expectancies were explored under several different conditions of intersignal timing. The effects of simultaneous tracking on temporal expectancy were also assessed. The results indicated that subjects adopted rational internal models of event timing which provided reasonable predictive success, regardless of whether they were asked to track at the same time. Indeed, the workload imposed by the adaptive tracking task did not appear to interfere with the formation of temporal expectancies or the generation of predictions. The tracking demands seemed mainly to encourage some quickening of expectancies, so that trackers tended to predict that events would occur slightly earlier than non-trackers. It is suggested that adaptation to the effects of auditory distraction could plausibly include the learning of temporal expectancies, and that the formation of such expectancies would probably not be inhibited by the processing demands of continuous visual tracking.

Introduction

The distractibility work reported in the previous chapter indicated that distraction effects on tracking were variable and sensitive to context. These observations were compatible with the traditional view that it is difficult to establish reliable effects of auditory distraction or to quantify the many factors which are presumed to influence adaptation. Adaptation, in particular, figured prominently in the distracting noise experiments. Historically, most investigations of distractibility have commented on adaptation processes, sometimes complaining about their interference with the collection of data on distraction effects (for example: Hack *et al*, 1965; Harmon, 1933; Kryter, 1970; Pollock and Bartlett, 1932; Woodworth and Schlosberg, 1954). Clearly, adaptation involves the alteration of expectancies of various sorts. In the context of a tracking task, the skill of anticipating what distractors are likely, or anticipating when they will occur, may be regarded as a benefit of extending the operator's internal model (Kelley, 1968) of the control task to incorporate additional contingencies. Even randomly occurring events can contribute to expectancies: it is sensible to ascribe predictability (at least locally) to events because events in the real world are rarely wholly random.

Distraction Experiments 1 and 2 jointly provided evidence that a number of periodic distractors was far less disruptive than a similar number of randomly scheduled distractors. In other studies also, the adaptation to distraction has progressed much more quickly with periodic distractors than with aperiodic ones (Eschenbrenner, 1971; Finkelman and Glass, 1970; Theologus *et al*, 1974). These results point to some sort of arrival-time expectancy as a likely factor in the adaptation to distraction. However, to participate in the process of adaptation during the distracted tracking tasks described in the previous chapter this expectancy formation would have to take place alongside the substantial tracking workload. It was not known to what extent the timekeeping (and any other necessary processes) for anticipation of distractor timing could be compatible with heavy loading. For example, Dimond (1965) demonstrated that a secondary loading task (tactile stimulus-detection) could impair useful temporal expectancies in a primary reaction time task; Schmidt (1968) has cited other similar examples. For this reason an experiment was carried out to investigate how *temporal expectancy* is accommodated during tracking performance.

The experiment outlined in this chapter had two basic aims. The first was to describe some characteristics of the expectancies people form about the timing of intermittent events: to determine whether the extent of temporal regularity in a signal time series would be learned in a rational way. Secondly, it was intended to discover whether people were capable of forming useful temporal expectancies while simultaneously attending to a tracking task. These questions were worth answering to decide if it were feasible that temporal expectancies could be aiding the tracker's adaptation to distraction. If, for example, tracking was found to impede expectancy formation, it would argue that adaptation to distraction is unlikely to depend heavily on temporal anticipation while engaged in such tasks. An exploration of this sort, while not addressing distraction phenomena directly, does contribute to a more complete understanding of the demands which tracking places on the operator. This in turn assists in formulating a theoretical account of distractibility during tracking tasks and the sources of adaptation.

The idea of expectancy is an old one in psychology and it has been applied in many different contexts. Sanders (1966) has provided a good summary of its various usages. But like involuntary attention, expectancy has suffered empirical neglect. Schmidt (1968), reviewing studies of anticipation learning in motor skills, complained that

> As important as they seem for skilled performance, anticipation and timing have received very little investigation. In fact, a great deal of the research on motor skills has sought to systematically prohibit the subjects from anticipating (p 631).

In the experimental literature that does pertain, temporal expectancy has mainly figured as a factor in reaction time. Reaction time experiments have confirmed that responses tend to be faster when the subjective expectancy of stimulus presentation is at a peak (Naatanen and Merisalo, 1977; Naatanen, Muranen, and Merisalo, 1974; Vroon and Vroon, 1973) but the main concern of these studies has generally been the time-course of response preparation rather than the information processing which underlies expectancy generation. In these investigations the examination of expectancy has usually been confined to short foreperiods of several seconds duration or less preceding the imperative reaction stimulus. Longer-term expectancies extending over many seconds or minutes have not figured greatly in the reaction time approach. Given the long timescale of most recurrent distracting events in the real world, expectancies of much longer duration would probably be quite important in the adaptation to distraction. However, even the few studies which have examined temporal expectancies in predictive movements (Naatanen *et al*, 1974) or synchronized movements (Bartlett and Bartlett, 1959) have limited themselves to expectancy intervals of, at most, four seconds.

For data on extended intervals, we must look elsewhere. A form of lengthy temporal expectancy has sometimes entered into theories attempting to account for the performance decrement which can appear during vigilance tasks. In its original formulation (Baker, 1962; Baker, 1963) the expectancy interpretation of the vigilance decrement described the observer's failure to detect signals as a consequence of erroneous temporal expectancies. In vigilance tasks the signals are rare and their discriminability is low. Expectancy theory suggested that after missing one or two signals the observer would have faulty knowledge of the signal schedule and that the resultant inaccurate expectancy of future event timing would lead directly to more missed signals and thus to further misleading expectancies. While it is now quite clear that expectancy failures are not sufficient to fully account for vigilance performance (Davies and Parasuraman, 1981; Mackie, 1977), these ideas pointed to the importance of the analysis and storage of sequential temporal information in expectancy formation. Curiously, the shaping of expectancies by the serial presentation of intervals has not found much discussion in the literature on timing behaviour and time perception (Michon, 1967). Both Baker and Michon have emphasized that an understanding of temporal expectancy requires analysis of the sequential dependencies present in the timing information received by the subject. In reaction time and anticipation studies, it does seem that temporal expectancy is heavily influenced by very recent experience (see, for example, Karlin, 1959); that subjectively the likelihood of experiencing extreme intersignal intervals is minimized; and that expectancy formation and learning effects appear very rapidly (Bartlett and Bartlett, 1959; Mowrer, 1940; Naatanen and Summala, 1976). These characteristics, especially the reluctance of subjects to incorporate extreme events into their expectancies, reflect some general properties of the way human beings perform statistical analysis and extrapolation (Peterson and Beach, 1967).

Expectancy is a term used to describe both an anticipated event and a psychological state corresponding to a degree of preparedness. In most situations some feature of the subject's performance has been singled out as representing the presence or level of an hypothesized state or, on some occasions, as indicating the actual content of the anticipated event. Sometimes both issues can be addressed by the same data, as when the subject is asked to wait until the expectant state reaches a certain level before he reveals what outcome he expects. Any usage of the term ought to define expectancy precisely within the observational context. In the experiment reported here the participants were asked to monitor intermittent signals. The subject was asked to try to anticipate each upcoming signal event and to attempt to predict its arrival with as much precision as possible. A single prediction was requested for each signal. The subject made a prediction by pressing a key only at the moment when he judged the next signal to be imminent. That is, the subject attempted to predict the intersignal interval by witholding the keypress until he felt that the likely intersignal duration had elapsed. With this procedure the temporal expectancy can be defined operationally as the time interval which has elapsed between the last presentation of a signal and a subject's keypress. This definition is appropriate because the interval terminated by the keypress should equal the representation of the intersignal interval supplied predictively by the subject's internal model. The operational definition assumes that the behaviourally-realized prediction expresses the content of the subject's expectancy. In the event of perfect prediction the subject's keypress would coincide with the target signal's onset, and so the *prediction interval* as measured from the previous signal to the keypress would just equal the intersignal interval. If his prediction were early, this prediction interval would be too short and some time would elapse before the signal arrived: a measure of the anticipation or early error. Alternatively, if the subject were to withold his prediction for too long a time, the signal would occur before any prediction had been entered. We might probe such late errors by asking the subject to immediately press the key as fast as possible whenever caught out in this way. The duration of the reaction latency relative to the signal onset would also be expected to yield some information about the level of preparedness at the moment of signal arrival (Naatanen and Merisalo, 1977; Vroon and Vroon, 1973).

Assuming that the number of late errors can be kept to a minimum, the response series thus accumulated is a sequence of subject-generated predictive estimates of the intersignal interval. The estimates are products of a decision made by the subject on the basis of his knowledge of the situation. Thus they may be regarded as manifestations of an internal model for temporal expectancies. The subject-generated intervals obtained in this experiment were analysed to determine how successfully such models could anticipate time-series of events. The subjects were asked to predict intersignal intervals which were drawn from specific parent distributions of time intervals. The investigation sought to determine how the actual distribution of intervals would influence the subjects' expectancies. The experiment also measured how the inclusion of a simultaneous tracking task affected the formation of expectancies under the same conditions.

EXPECTANCY EXPERIMENT: TEMPORAL PREDICTION DURING TRACKING

Procedure

Sixty undergraduate students were recruited as voluntary subjects. None of the subjects had been asked to serve in a similar experiment previously.

The subject sat in a small well-lit cubicle adjoining the computer room. The subject faced the tracking display across a small table, which supported the tracking joystick and a response key. The response key was a microswitch mounted in a small box. It was positioned beneath the subject's preferred hand with arms resting comfortably on the table. Wristwatches were removed beforehand.

Each subject listened to a time-series of signals presented through headphones. The PDP-11/45 computer was programmed to control the sequence of experimental events and to record observations. The signal series consisted of tonebursts each lasting 250 milliseconds, separated by silent intervals of adjustable duration. The signals came from an 800 Hz sinewave source, measuring 70 dBA at each ear if allowed to run continuously. The burst was gated by an electronic switch with a rise and fall time of 20 milliseconds. The electronic switch eliminated click transients and the resulting toneburst was heard as a short tick of comfortable loudness. The silent background between signals was very quiet, not exceeding 32 dBA.

All subjects read these instructions:

> In this experiment, you will be asked to put on the headphones and listen to a succession of signals. Each signal is a soft short beep. An interval of silence will separate each signal from the next. The duration of this interval might vary quite a bit, so sometimes it may seem a long time between signals and sometimes they may follow each other fairly quickly. You will not be given any warning when a signal is about to occur. Your job is to try to anticipate the signals as best you can. You will indicate your expectation of a signal by pressing the pushbutton at just that moment when, in your estimation, the signal ought to occur.
>
> This is what you must do. Listen for the first signal and then the one following it. But before yet another signal occurs, try to predict its arrival in the following way: wait until you think that the signal must be just about to come on, and then push the button firmly once. Listen for the signal and note whether it follows your prediction closely. If your prediction is good, the signal will come on just after you have pressed the button. To be as accurate as possible you must aim to respond as near to the approaching signal as you are able, but you must avoid being so hesitant that the signal arrives before your button-press has been carried out.
>
> Continue anticipating all the remaining signals, making a prediction for each signal which follows. Make your predictions without counting-out, or tapping, or resorting to other such methods. When you hear a very long tone-- you may stop and rest.

> I will come and tell you when the next signal sequence starts.
>
> You are trying to make predictions, so it will not do much good if you always wait so long that the signal arrives before you have pressed the button. Nevertheless, you may find that sometimes a signal does catch you unexpectedly. If the signal starts before you have indicated a prediction, then you must make your response then. But because the signal has already begun, you must press the button as fast as you can. Then go on and try to predict the next signal more successfully.
>
> Here is a brief summary of your job:
> For the first and second signals, just listen. Thereafter, try to make a prediction for every upcoming signal by preceding each signal with a button-press as closely as you can manage. If the signal happens before you've made your prediction, immediately press the button as fast as you can and proceed to your next prediction. The long tone marks the end of the signal series. Rest then and wait for me.
>
> If you have any questions, ask them now.

Two independent variables were studied: the subjects' mental workload, and the statistical properties of the set of intersignal intervals.

Mental workload was increased by imposing a visual tracking task. Half of the subjects were asked to carry out the standard adaptive compensatory tracking task simultaneously with the prediction task. These *loaded* subjects had to indicate predictions with one hand while continually manipulating the joystick with the other. The tracking task conformed to the adaptive procedure previously described, using zero-order control dynamics. The remaining *nonloaded* subjects did not track; the joystick was switched off and they saw only a static bright line display during their sessions.

In addition to workload, the experiment manipulated the probability distribution of the intersignal intervals which the subjects were trying to predict. Three populations of *intersignal intervals (ISIs)* were defined, each with a distinct probability density. The *gaussian* ISIs were drawn from a population of values weighted by a gaussian probability distribution. This distribution had a mean equal to 15.0 seconds and a standard deviation of 3.4 seconds. In the *uniform* conditions, the ISIs had a range of values and a mean duration comparable to the gaussian set, but the ISIs were uniformly distributed; that is, they were sampled from anywhere within the range with equal probability. The standard deviation of the uniform set was 4.6 seconds. Lastly, the *constant* conditions presented an unchanging ISI. The ISI was maintained at a constant duration equivalent to the mean value of the gaussian and uniform distributions (15 seconds). The actual distributions are shown in Figure 4-1. All three ISI distributions were tested under both nonloaded and loaded conditions. The subjects were not told about the ISI differences beforehand.

Each subject received a single combination of workload and interval distribution. A totally separate groups design was used, one group per experimental condition. The sixty subjects were randomly divided into six groups of ten subjects each. The groups contained equal or near-equal numbers of both sexes. Thus there were three nonloaded groups each receiving one of the three distributions of intervals without any tracking demands, and three loaded groups who received the same distributions but with the added tracking requirement.

The major dependent variable studied was the duration of the

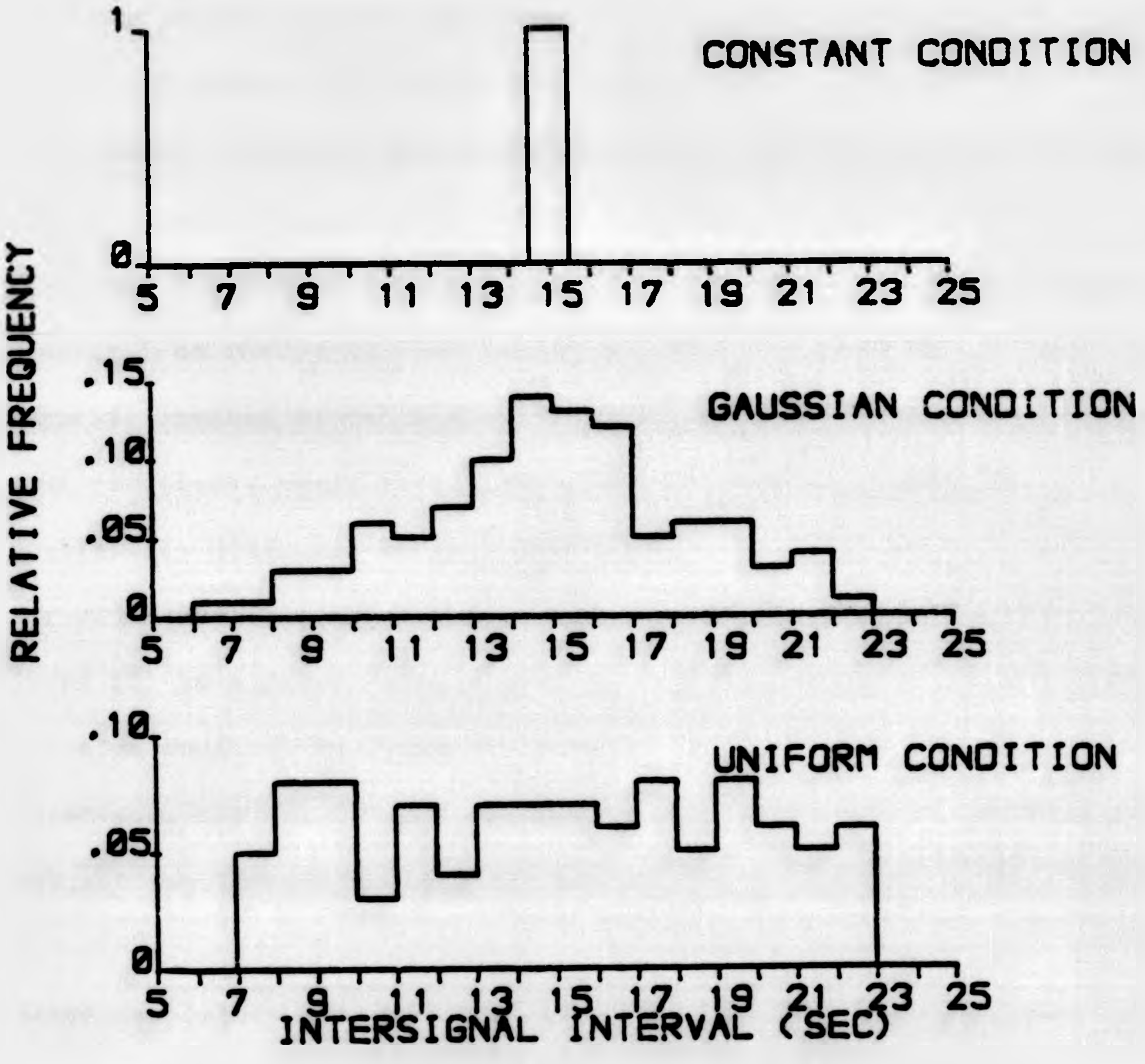

FIGURE 4-1 Expectancy Experiment. Intersignal intervals.

prediction interval, measured from the last-presented signal to the moment the subject pressed the response key. In addition, reaction time data were examined whenever the subject's keypress followed, rather than preceded, the target signal.

The experimental session comprised five blocks of trials. Each block took 5.5 minutes. The blocks were separated by two-minute rests. Every block consisted of a time-series of toneburst signals. Signal timing within a block was governed by a list of twenty intersignal intervals drawn up beforehand by choosing twenty independent random samples from the parent population of available intervals. Each interval bounded by its signals defined a *prediction trial.* Within each condition, then, five time-series drawn from a specific probability distribution were presented to each subject in randomly varied order. One constraint common to all conditions was that the first ISI in a block (always used as a demonstration interval) had a duration equal to the population mean. Allowing for the demonstration intervals, the total number of prediction trials offered to each subject was therefore ninety-five.

After completing the experiment, the subjects were asked to fill in a questionnaire designed to assess their awareness of the signal timing and their prediction strategy. The questionnaire is reproduced in Appendix A.

Results

All of the subjects, whether nonloaded or loaded, were able to follow the instructions in a satisfactory manner. The response histories confirmed that none of the subjects resorted to a superficial response strategy nor demonstrated any persistent reluctance to predict signals in advance of their arrival. The keypress was usually executed after the passage of an interval slightly shorter than the mean ISI of the signal series. There was no evidence to suggest that non-predictive strategies of convenience were used. That is, subjects did not persistently copy the shortest ISI they had encountered nor did they resolutely withold all responses until the awaited signal had actually arrived.

The responses were a mixture of early predictive keypresses and late reaction responses. The majority of responses-- on average, 66 out of the 95 trials offered to each subject-- were predictive; that is, they were made prior to the onset of the target signal. As will be seen, the number of predictions varied with the experimental conditions. The minority reaction responses were analysed separately and will be considered in detail later. There were also a very few trials for which no responses were made, presumably because subjects forgot that they had not yet made a prediction for that trial, or perhaps because they failed to push the key properly. There were, on average, two misses per subject.

Effects of the prediction task on tracking performance (loaded groups only)

Subjects in the loaded conditions carried out the secondary adaptive tracking task in addition to the primary prediction task. Examination of the tracking data revealed that the demands of the prediction task had not impaired tracking. Tracking performance was compared with the baseline tracking measurements obtained in the experiments of Chapter 1. Tracking Experiments 1 and 2 provided an appropriate reference for unimpaired adaptive tracking behaviour. It was clear that progress in the adaptive task in this expectancy experiment was very similar to the earlier results. The comparability is most easily illustrated by Figure 4-2, which plots the mean adaptive bandwidth and normalised RMS error alongside the previously obtained baseline results. (Time of sampling rather than the usual block number has been used for the x-axis in this figure because the experiments did not all use the same block duration.) For both the acquired bandwidth and the resultant error, little difference can be seen between the present adaptive outcome and the results of the previously studied (dedicated) adaptive tasks.

The tracking history of the loaded subjects was also checked to determine whether tracking performance had been similar across groups despite the use of different ISI distributions for the prediction task. The adaptive tracking outcomes are summarised in Table 4-1. The form of the ISI distribution used for the prediction task did not affect the progress of adaptation during the secondary tracking task.

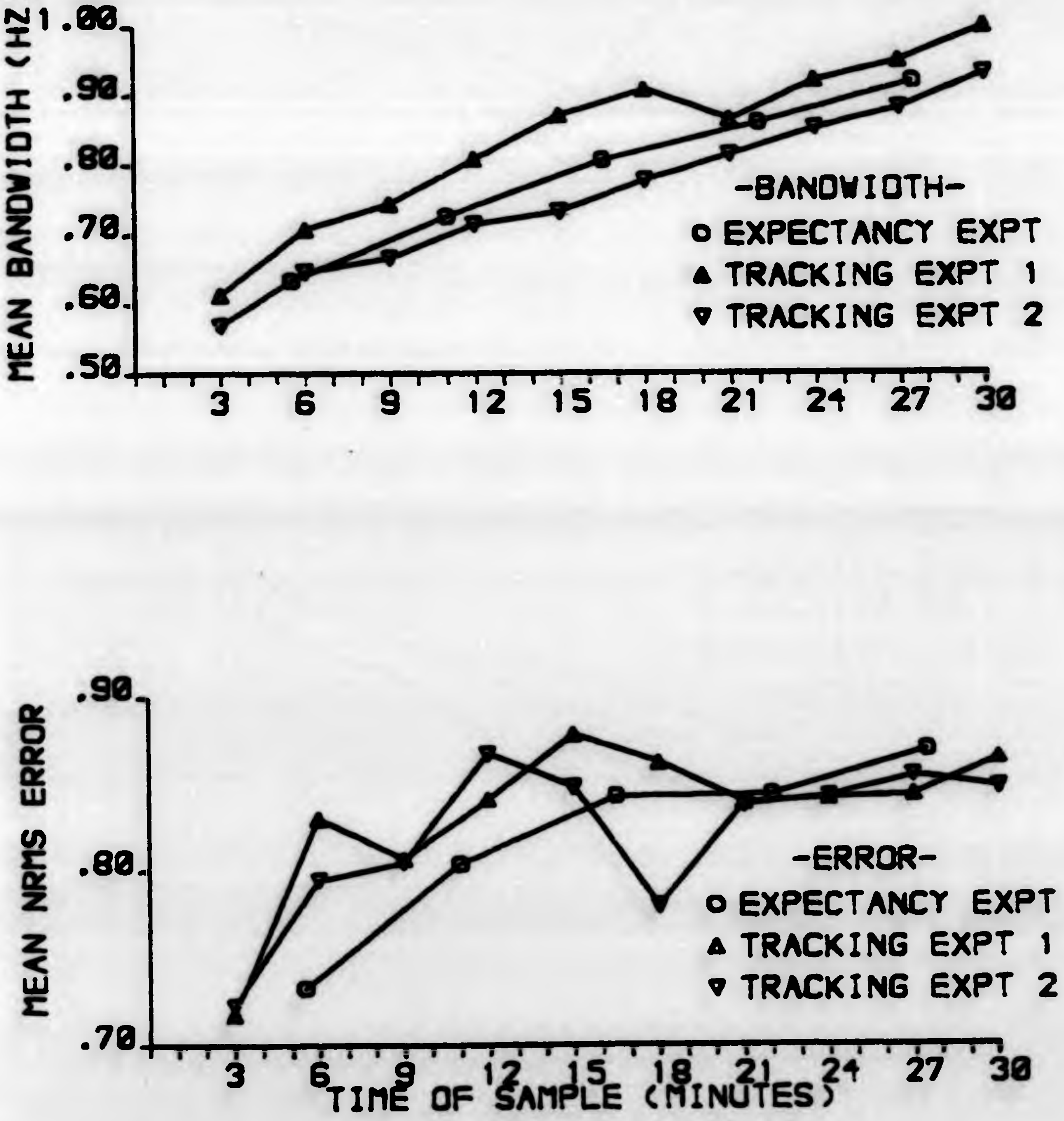

FIGURE 4-2 Expectancy Experiment.
Comparison of tracking performance (with the prediction task)
to performance in Tracking Experiments 1 and 2 (tracking alone).

TABLE 4-1

EXPECTANCY EXPERIMENT. Effect of ISI distribution on tracking

	ISI condition		
	LC	LB	LU
Adaptive bandwidth mean (hz)	0.779	0.800	0.798
Adaptive bandwidth std dev (hz)	0.111	0.161	0.148
Tracking RMS Error (normalised wrt force)	0.820	0.856	0.827

All the groups achieved similar average adaptive bandwidths (which established the tracking difficulty). The bandwidth history contained comparable variance in each condition, which suggests that the adaptation process exhibited approximately the same degree of stability in every group. The error results also indicated that the tracking competence of the subjects was similar across conditions. A one-way analysis of variance for each of these measures confirmed that there were no significant effects of ISI condition on bandwidth mean, standard deviation, or RMS error. The subjects all responded in a similar way to the tracking task, regardless of the primary task conditions, and the loading effects were probably much the same across ISI conditions.

Prediction interval

The prediction interval was the duration which the subject allowed to elapse between the onset of last signal heard and the keypress marking imminent expectancy of the next signal. The simplest description of the prediction intervals was obtained by comparing their distribution to that of the ISIs which the subject was attempting to predict. The distributions of predictions for the three ISI conditions are shown in Figure 4-3, with the nonloaded and loaded treatments drawn on the same axes. The ISI distributions themselves were presented previously in Figure 4-1. The prediction distributions shown here average all genuine predictions from all subjects within each condition. Keypresses were considered to be genuine prediction attempts if they

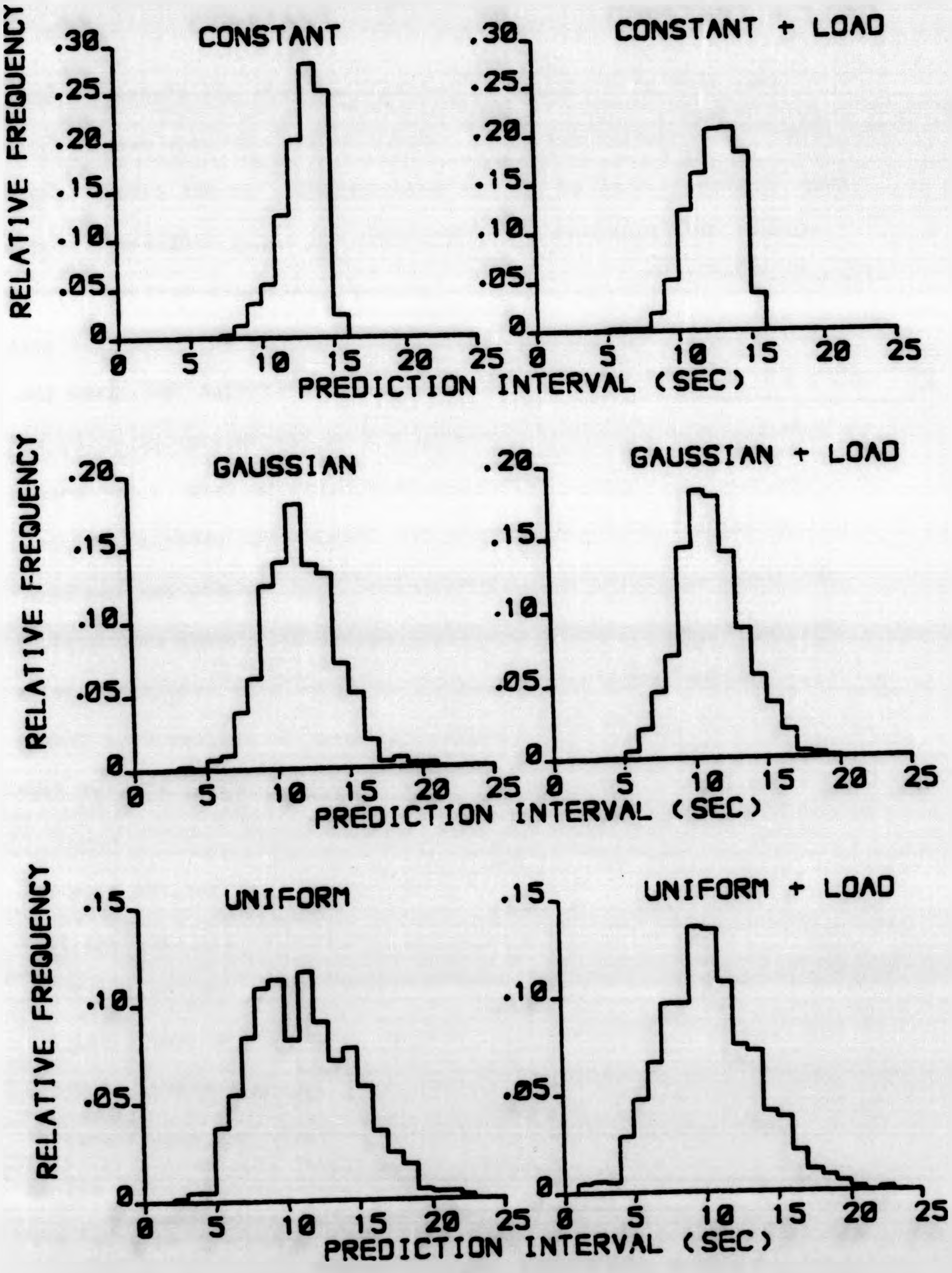

FIGURE 4-3 Expectancy Experiment.
Distributions of prediction intervals.

occurred before the onset of the target signal or if they were simultaneous with the signal onset. Also, any responses following the signal within 120 milliseconds were assumed to have originated just before the signal began and were treated like predictions, since normal reaction processes would be incapable of activating keypresses with such short latencies. (Virtually simultaneous responses of this sort were, not surprisingly, comparatively rare: a few are visible in the 15 to 16 seconds bin of the constant prediction distributions in Figure 4-3.) When no predictive keypress occurred during a trial, a reaction response was sought. An acceptance rule for reaction responses was adopted such that keypresses occurring at least 120 milliseconds beyond the target signal but no later than 2 seconds after the signal were taken as reactions. Responses following a signal by 2 seconds or more were regarded as attempts to predict the next trial's target signal.

The most obvious feature of Figure 4-3 is that the more peaked the ISI distribution, the more peaked the resultant distribution of prediction intervals became. The prediction distributions of Figure 4-3 include all individual differences, because they average the results from all subjects within a group. The individual prediction distributions had different mean durations. For this reason, the characteristic shape of the prediction distributions is more appropriately inferred from the "aligned" collective distributions of Figure 4-4, in which the differences between the individual subjects' means have first been subtracted out before pooling. This procedure preserves the relative spread of prediction values without the distortion due to baseline

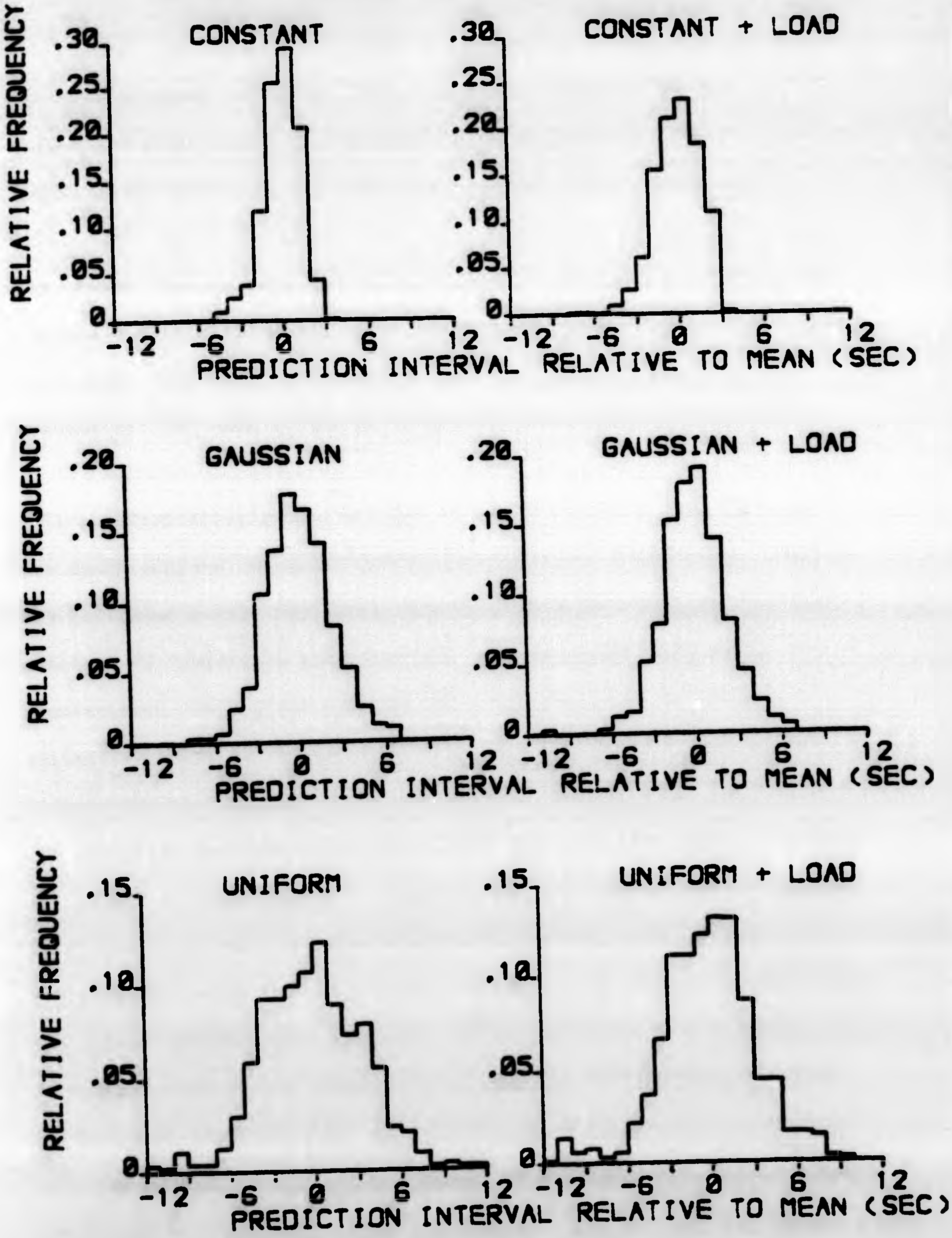

FIGURE 4-4 Expectancy Experiment.
"Aligned" distributions of prediction intervals.

differences among subjects. This heightens the peaks of the prediction distributions, representing more accurately the distribution shape for any individual subject within the group.

The loaded results were very similar to the nonloaded outcomes except for the distribution means. The loaded subjects exhibited lower mean durations. The mean, mode, standard deviation, and number of predictions for each of the distributions are listed in Table 4-2.

Before interpreting these results, it is important to consider that the distributions represent differing numbers of predictions. Taking the nonloaded conditions first, the table entries indicate that the constant (C) condition permitted the largest number of genuine predictions. This group managed 690 predictions out of their collective total of 950 trials. This performance was followed very closely by the gaussian (G) group with 673 predictions. The uniform (U) condition provided 549 predictions, the smallest result. In the the loaded groups there were somewhat more predictions under each ISI condition, but they followed the same rank order with the most predictions coming from the loaded constant (LC) group and least from the loaded uniform (LU) subjects. The lower means in the loaded conditions explain why these groups were able to make slightly more predictions: these subjects had adopted a more conservative (that is, earlier) prediction on average, and therefore were less likely to be caught out by the target signal before a prediction had been made. The loss of some prediction trials must be taken into account when interpreting the results shown in Figure 4-3 and Table 4-2. It is

TABLE 4-2

EXPECTANCY EXPERIMENT. Prediction interval characteristics.

	ISI condition					
	C	G	U	LC	LG	LU
mean	12.92	11.66	11.21	12.49	11.24	10.30
mode*	13.40	11.55	10.60	12.70	11.30	10.20
std dev	1.58	2.48	3.46	1.73	2.30	3.23
n	690	673	549	736	701	627

* average of individual subjects' results

important to evaluate how the varying losses affect the observed prediction distributions. For every trial there was a probability, defined by the corresponding ISI probability density function, that the target signal would arrive before a prediction could be made and so eliminate that prediction trial. In other words, whenever the subject wanted to mark out any prediction interval longer than the minimum ISI possible, there was always a chance that his prediction would be preempted by the arrival of a signal, and the chance of this happening depended on the particular distribution of ISIs. This artifactual trial elimination effect would occur in its simplest form under the constant ISI conditions. Here, because the ISI was invariant, all predictions exceeding fifteen seconds-- but only these predictions-- would be prevented. In this case a ceiling would be imposed on the range of results, truncating the right-hand tail of the prediction distribution. In the uniform ISI conditions, there was an equiprobable chance that any prediction interval, regardless of its duration, would have to be abandoned because of an early signal. This would eliminate trials uniformly across the range. Since the subjects had the freedom to predict earlier (that is, to place more predictions toward the left of the response distribution) but could not supply a prediction interval after the signal had occurred, the right-hand tail would again suffer. Under gaussian ISI conditions, the effect was similar to the uniform description but with the difference that trial elimination was not evenly apportioned across the range. For example, when the intended prediction interval was a relatively long one the loss of the trial would not be so likely to occur under gaussian conditions as in the uniform case, because the proportion of very short ISIs was reduced under gaussian conditions.

While the major impact of the trial elimination effect at first appears to be mainly the elimination of increasing numbers of predictions as the probability density of the ISIs becomes more uniform, especially from the right-hand tail of the prediction distribution, the effect must also act on the mean and variance of the observed prediction distributions. For example, if all predictions longer than the constant ISI are lost from the C and LC groups' results then both the mean and the variance of the observable prediction distribution must be reduced by the effect. There are at least three significant implications which follow from this discussion of trial losses. First, the mode of the prediction interval distribution, not the mean, provides a safer basis of comparison among the various conditions. With the constant and uniform distributions used in this experiment the trial elimination effect would be expected to suppress the right-hand tail of the prediction distribution (thus altering the mean) but not to alter the location of the distribution peak. In the gaussian conditions it is more difficult to anticipate whether the effect will promote some shift in the peak since prediction intervals of differing duration are not all eliminated in the same proportion. However, a means of checking the distribution mode will be described very shortly. The second implication is that the skewedness of the prediction distributions cautions against the use of parametric statistical tests for these data. Lastly, any relative differences between nonloaded and loaded treatments should retain their validity despite the trial elimination effect. As long as the ISI distribution was the same, trial loss would be expected to operate in identical fashion under either level of workload.

To estimate the extent of the trial elimination effect on the prediction data, the prediction distributions for the gaussian (G and LG) and the uniform (U and LU) conditions were replotted with a correction for lost trials. The loss of trials had been affected by ISI probability, so the restoration of the missing trials was based on the same information. Correction terms were obtained by multiplying the response counts in each subject's original prediction distribution by the corresponding probabilities from the appropriate ISI distribution. These weighted increments were added to the entries in the original prediction distribution, and the relative frequencies were recalculated. This procedure provided an adjustment to each relative frequency proportional to the expected magnitude of the trial elimination effect. The average "corrected" prediction distributions appear in Figure 4-5. It must be stressed that despite the correction procedure these distributions must still under-represent the longer prediction intervals, since it is not possible to estimate how many predictions would have extended beyond the limit of observability fixed by the right-hand extent of the ISI distribution. This enforced truncation of the prediction distribution will affect the mean and variance outcomes even in the "corrected" data so these statistics must always be treated with caution. However, the truncation will not affect the mode of the prediction distribution. The position of the modal prediction interval is therefore important. Figure 4-5 makes it clear that an allowance for the trial elimination effect resulted in

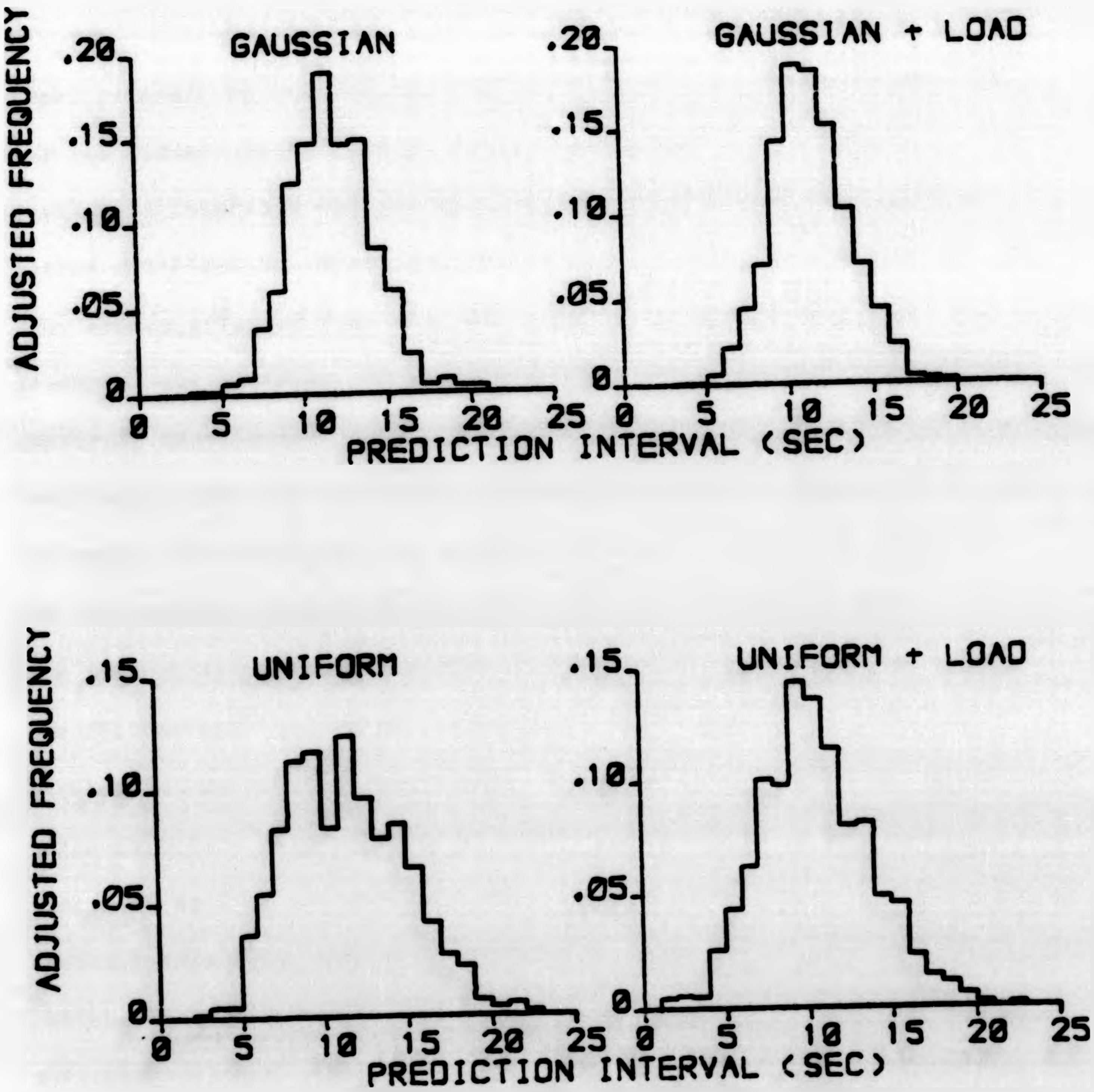

FIGURE 4-5 Expectancy Experiment.
Distributions of prediction intervals with correction
for trial elimination effect.

distributions differing only very slightly from the original uncorrected data. Some bin frequencies were slightly increased, but the modes were found to be unchanged by the correction procedure. This was confirmed individually for each subject in the experiment.

Finally referring back to the prediction distributions of Figure 4-4 and the summary statistics of Table 4-2, some general features of the prediction intervals may now be noted. The predictions followed a peaked distribution regardless of the form of the ISI distribution and they always extended over a wider range of durations than the presented set of ISIs. The fewest predictions were recorded for the uniformly distributed ISIs; the most occurred using constant intervals. The resulting prediction distributions suggested that the different ISI treatments had prompted changes in mean and variance. For the reasons already given, this description must be qualified. The prediction variance appeared to increase in the same direction as the ISI variance and so was smallest in the constant conditions and largest in the uniform conditions, irrespective of loading. It also appeared likely that predictive keypresses were, on the whole, committed earlier in the trial when anticipating gaussian rather than constant intervals, and that predictions came earlier still when contending with uniform rather than gaussian events. However, it is stressed that the changes in the prediction mean and variance under different ISI treatments could have been influenced by the artifactual trial elimination effect and these statistics cannot be taken as definitive. Even so, Figure 4-4 does illustrate that a greater proportion of predictions accumulated in the left-hand tail of the prediction distribution under uniform conditions than under gaussian conditions. That is, there was possibly a greater preference for short prediction intervals under uniform ISI conditions. This difference would not have been artifactual because the trial elimination effect would not have restricted the size of the left-hand tail. More reliable evidence that the duration of the prediction interval was affected by the ISI distribution came from examination of the modal prediction values. It has been established that the prediction distribution mode was not affected by the trial elimination effect and was a valid measure of central tendency in these data (even if the mean was not). The mode results for all of the subjects are listed in Table 4-3. These data argue that the prediction interval mode decreased as the ISI duration became more uncertain; that is, the prediction intervals did shift toward shorter durations. Within the three nonloaded groups (C,G,U) the mode was reliably affected by the ISI distribution (Kruskal-Wallis H_{2df}=14.74, p<.001). Within the loaded treatments (LC,LG,LU) the differences in the prediction mode were also statistically significant (H_{2df}=11.74, p<.01).

It was clear that despite the demands of the simultaneous tracking task, the loaded subjects were perfectly capable of executing sensible predictions. By inspection, the prediction distribution shapes were very similar under the two workload conditions. The loaded subjects actually executed more predictions overall than the nonloaded subjects (Mann-Whitney $U_{30,30}$=315, p<.023). It is understandable that the loaded subjects were able to complete more predictions because they tended to indicate their expectancy earlier than the nonloaded

TABLE 4-3

EXPECTANCY EXPERIMENT. Mode of the prediction interval distribution.

Subject	ISI condition					
	C	B	U	LC	LB	LU
1	13.5	10.5	11.5	12.5	12.5	9.5
2	14.5	11.5	8.5	12.5	10.5	10.5
3	14.5	13.5	11.5	13.5	12.5	7.5
4	13.0	13.5	9.5	12.5	11.5	12.5
5	13.5	12.5	8.5	13.5	11.5	11.5
6	12.5	10.5	11.5	13.5	11.5	13.5
7	12.5	11.5	10.5	13.5	11.5	9.5
8	13.5	12.0	11.5	11.5	10.5	6.5
9	13.5	10.5	13.5	12.5	10.5	11.5
10	13.5	9.5	9.5	11.5	10.5	9.5
average	13.40	11.55	10.60	12.70	11.30	10.20

subjects, and so were less likely to be caught out by the arrival of the target signal. However, the differences in prediction interval due to loading were suggestive rather than conclusive. Although the prediction mode was of shorter duration under loading in all of the ISI conditions, the decrease was statistically significant only for the constant ISIs (Mann-Whitney $U_{10,10}=26$, $p<.05$) and not for the gaussian or uniform conditions. Nor did loading appear to have any systematic effect on the standard deviation of the subjects' predictions. In any case, it would be difficult to draw conclusions from comparisons of standard deviation because the number of predictions contributing to this statistic differed with each group.

Sequential dependence measures: general comments

Comparisons of distribution shape, mean, and variance can provide a global description of the subject's expectancy. At a more detailed level, the prediction-to-prediction transitions and the correlation between the subject's attempted extrapolation and the history of ISIs provide insights of a different sort, because these data comment on the amount and source of information incorporated into the subject's internal model of the temporal events. Specifically, the *crosscorrelation function (CCF)* relating the response history to the signal events indicates the extent to which the temporal expectancy is a function of intersignal timing. The CCF displays the relationship for a range of ISI positions (lags) prior to the expectancy outcome. That is, the crosscorrelation can suggest how much each ISI in turn preceding a given expectancy might influence that expectancy.

In calculating the crosscorrelations for the experimental data, the five blocks of trials contained in each session were concatenated and treated as a single series on which the correlation analysis was performed. However, the actual block boundaries were respected in such a way that the prediction-ISI pairs being correlated could not be alignments of data from different blocks. (Recall that short rest-breaks intervened between blocks of trials). Therefore when it was necessary, trials at the block boundaries were omitted from the analysis so that prediction data from the opening trials of one block would not be paired with data from the closing trials of the previous block. This prevented inadmissible pairings and still permitted correlations to be based on as many observations as possible, ranging from 75 to 95 data-pairs per lag. Analysis normally extended to lag 5.

The general form of the notation adopted here to represent a crosscorrelation function will be $r_{AB}(t)$. This denotes the set of correlation coefficients obtained when series A is compared to series B at the superimposition defined by a lag or shift of t experimental trials.

In the prediction experiment, the subject generated predictions with varying success and sometimes failed to make a prediction. In such a case he would hopefully react following the signal, but the trial would necessarily be lost from the collection of prediction intervals.

Any sequential record of prediction intervals would be likely, therefore, to contain discontinuities. Analyses using crosscorrelation techniques are made more difficult when there are discontinuities or missing observations in the series of events under study. If a sequential analysis of prediction intervals were confined solely to data falling between trial gaps, there would be very many short series. This would greatly compromise the analysis both statistically and interpretively. A better alternative would be to eliminate discontinuities by choosing for each trial a measurement which at least approximates the subject's temporal expectancy. As measured in this experiment, temporal expectancy is an interval of time which is assumed to be nominated by the internal model as the anticipated extent of the intersignal duration. The expectancy interval represents the time required for a particular psychological state of readiness to reach criterion, being the state of expectancy which declares that a signal is imminent. This provides a rationale for approximating the subject's expectancy during "failed" predictions so that these trials may be included in the sequential analyses. We really want to measure and analyse the prediction interval. On those occasions when the signal arrives before prediction has been made this cannot be measured. If the subject follows instructions, he will react to the signal and provide a reaction latency. Although reaction latency could reasonably be expected to follow some function of expectancy (Karlin, 1959; Klemmer, 1956; Klemmer, 1957; Mowrer, 1940), the relationship is unknown a *priori* and reactions cannot be translated into covert prediction intervals. But in such cases we know that the subject was still waiting when the intersignal interval ended, so we can safely conclude that the prediction interval would have been at least as long as the intersignal interval for that trial. Therefore we may use the ISI as an approximation of the prediction interval. Estimates of expectancy derived from such trials will tend to be under-estimates because the approximation is always the minimum value we can identify under such circumstances.

Sequential dependence I: Dependence of expectancy on prior responding

The gaussian and uniform ISI series were generated by sampling random numbers with replacement and the ISIs were statistically independent. This was confirmed by examination of the *autocorrelation functions (ACFs)* for the ISI time-series, denoted $r_{II}(t)$, which varied nonsignificantly around zero. The ACF is comparable to a crosscorrelation function in which the event series is correlated with a time-shifted version of itself. The ACF for a random process tends toward zero for increasingly longer lags (Lynn, 1973) and this is analogous to saying that the events have no memory: beyond an adequate lapse of time there is no correlation between events. The rate at which ACF diminishes toward zero is controlled by the conditional probabilities which characterize the process.

A subject's predictions, on the other hand, were not independent of one another. The dependencies were identified by computing $r_{PP}(t)$, the ACF of the prediction series. Broadly speaking, interpretation of $r_{PP}(t)$ hinges on the sign of the correlation. Totally unvarying prediction intervals (extreme response perseveration) would produce $r_{pp}(t)=+1$ for all choices of t. Realistically, with a moderate amount of response perseveration, the magnitude and persistence of the effect would be gauged from the peak positive value attained by $r_{pp}(t)$ and its rate of decrease as earlier lags (larger t's) are examined. Negative autocorrelation values, however, could suggest tendencies to alternate in some regular fashion between long and short predictions.

An autocorrelation analysis was carried out on the nonloaded prediction intervals, up to lag 5. When the mean ACFs of the subjects' predictions are examined in Figure 4-6, it seems that the coefficient quickly reduces to a low value and remains near zero for the ACF beyond lag 1. The average results at $r_{pp}(1)$ were positive-- although small-- and these could indicate some response perseveration in the prediction sequence. The lag 1 coefficient was not always large enough to be significant for all subjects. The lag 1 data for individual subjects, Table 4-4, show this. From the ACF results it can be concluded that when subjects do persist in a particular mode of response (and not all subjects do), they are not likely to maintain any substantial degree of response consistency beyond one further trial. Unfortunately, the overlooking of some data files during analysis meant that of the loaded group data, only the loaded constant condition ACF could be examined. On inspection, Figure 4-6 suggests that the sequential dependencies in this condition were not noticeably different from the nonloaded constant outcome.

Of course, some indication of last-response perseveration is not to be

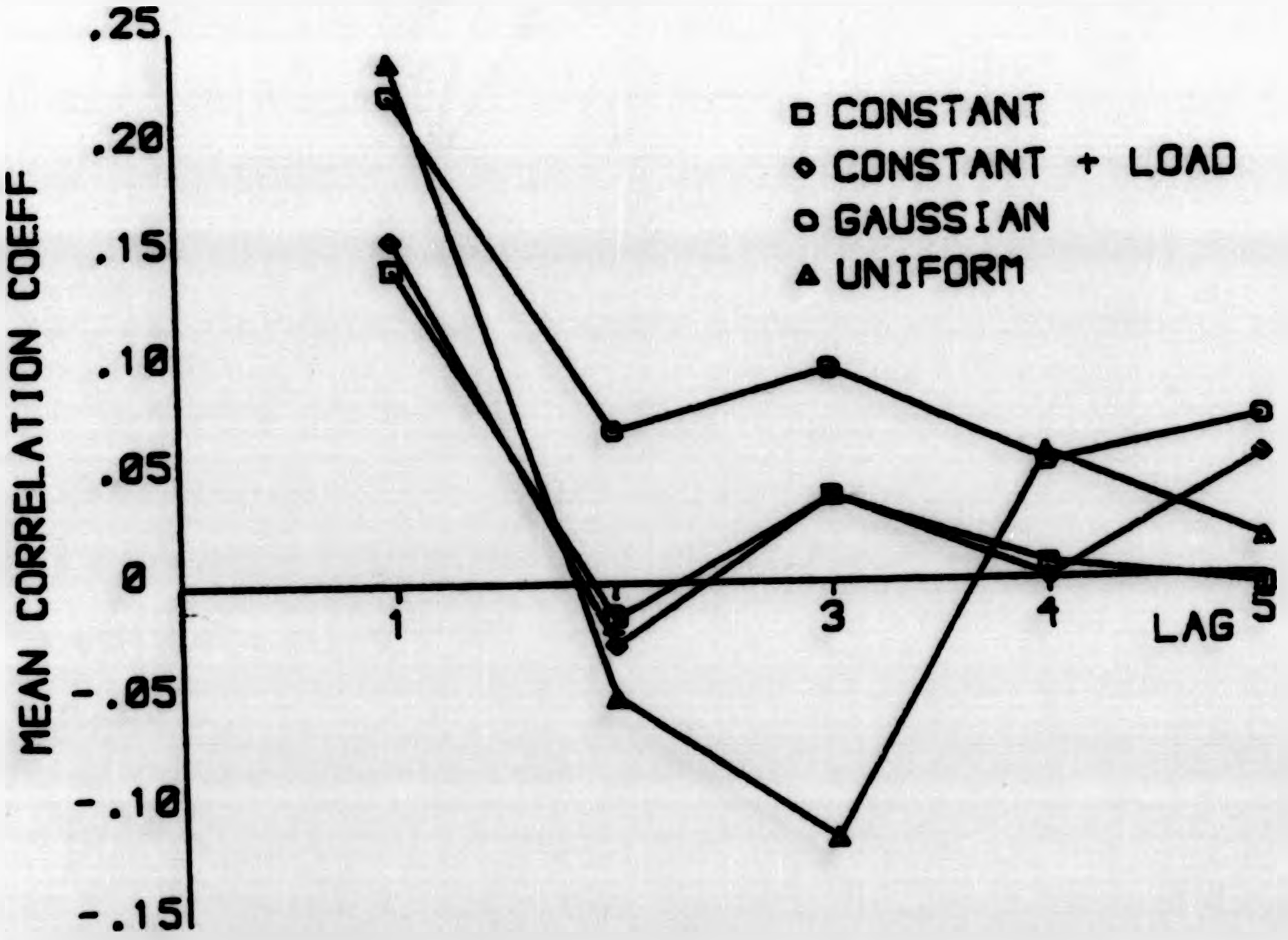

FIGURE 4-6 Expectancy Experiment.
Autocorrelation functions of prediction intervals.

TABLE 4-4

EXPECTANCY EXPERIMENT. Autocorrelation of prediction intervals, showing Lag 1 only. Results for individual subjects.

Subject	ISI condition			
	C	G	U	LC
1	+.204	+.284 *	+.155	+.103
2	+.418 *	+.140	+.235 *	+.050
3	-.044	+.157	+.218 *	+.168
4	+.310 *	+.239 *	+.194	+.165
5	-.068	+.074	+.317 *	+.004
6	+.241 *	+.235 *	+.264 *	+.335 *
7	+.035	+.163	+.273 *	+.203
8	+.143	+.203	+.230 *	+.078
9	+.029	+.202	+.135	+.301 *
10	+.148	+.503 *	+.332 *	+.142

Each correlation coefficient based on 95 data-pairs

* significant, p<.05 (2-tailed)

taken as evidence that the expectancies displayed little variability overall. Predictions from the uniform group were spread across a very wide range of durations, notwithstanding the correlation at $r_{pp}(1)$. The ACF result does indicate that some subjects display reluctance to wildly shorten or lengthen their predictions between two successive trials; but the observed wide range of the predictions demonstrates that the task is nonetheless one which compels subjects to adopt highly variable expectancies in the long term. Subjects did not simply tap out the same prediction trial after trial. But responses occurring close together in time were more likely to be similar than responses widely separated.

Sequential dependence II: Dependence of expectancy on prior signal timing

Subjects witnessed ISIs one by one. How did information from each ISI help to teach the internal model about population statistics? The subjects' predictions were not static representations of the underlying statistical structure of the time-series presented. The predictions were dynamic responses sensitive to signal spacing on a trial-to-trial basis. This was demonstrated by the crosscorrelations between the prediction interval and the ISIs preceding it. Figure 4-7 shows the mean crosscorrelation $r_{pI}(t)$ between prediction and preceding ISIs. Each curve plotted is an average CCF for the ten subjects in each experimental condition. (No CCF results are computed for the constant groups. Because the ISI never changed in the

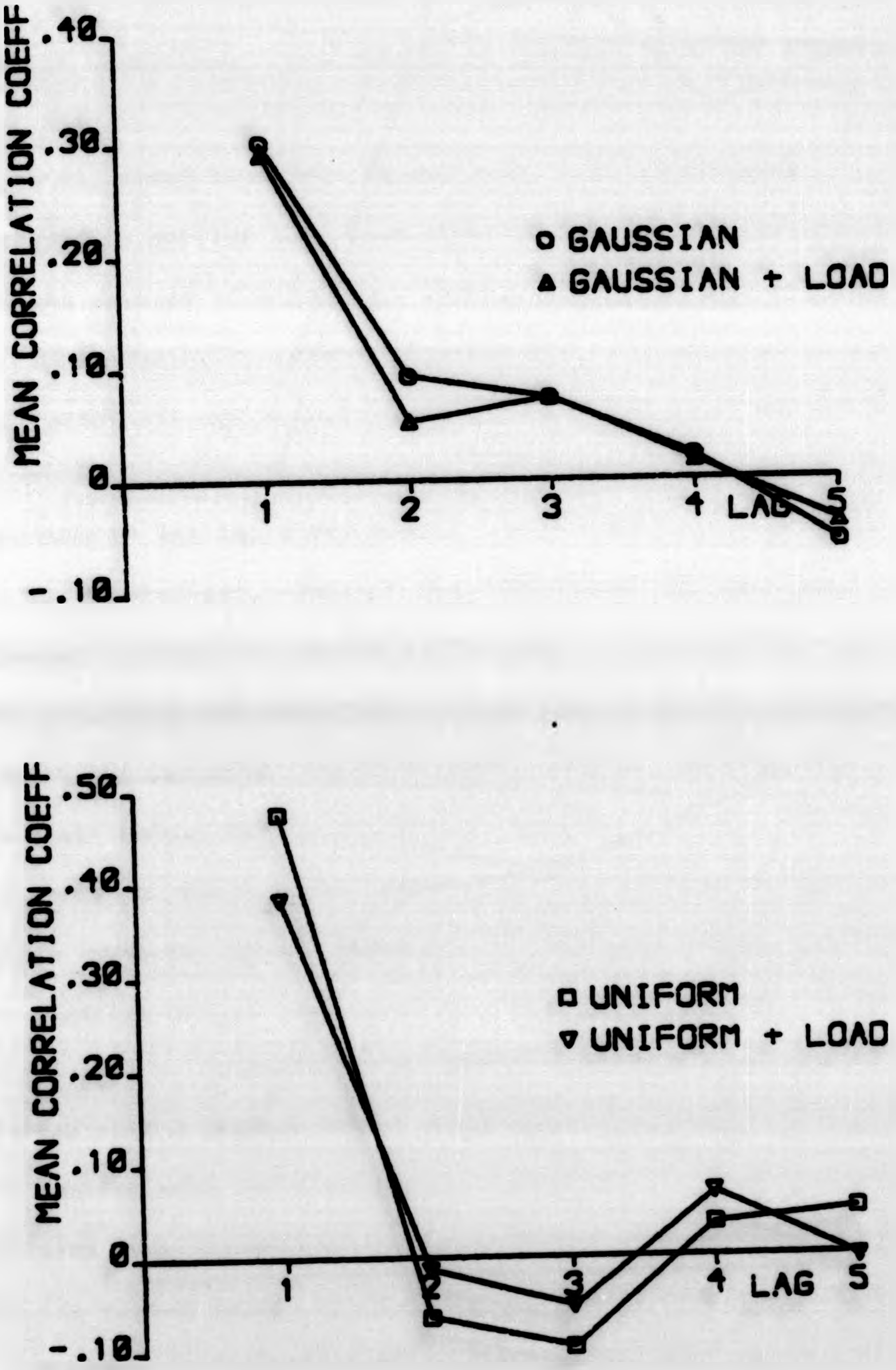

FIGURE 4-7 Expectancy Experiment.
Crosscorrelation of prediction interval
with previous intersignal intervals.

constant conditions, there can be no lag-dependent ISI effects.)

Figure 4-7 shows that for the most part $r_{\rho\iota}(t)$ decreased as successively earlier ISIs were considered. The results demonstrate that the setting of expectancy depends most on recent ISI sampling and less on historically remote observations. The correlation with the most recent ISI was always positive. This was verified for all subjects. Furthermore, for 31 of the 40 subjects involved, the magnitude of the lag 1 correlation coefficient $r_{\rho\iota}(1)$ was significant ($p<.05$, two-tailed). Thus if the most recent ISI was long, the subsequent expectancy tended to be long; if the ISI was brief, the next prediction was shortened. Inspection of Figure 4-7 suggests that loading did not alter the CCFs substantially. However, the choice of ISI distribution did influence the crosscorrelation. The CCFs looked different for the two ISI conditions. The maximum correlation was always found at lag 1. Beyond $r_{\rho\iota}(1)$ the coefficient diminished. For the gaussian groups the impact of previous ISIs seemed to fade away more gradually compared to the abrupt drop in the uniform CCF, which started with a greater lag 1 correlation but changed sign by lag 2. The coefficients for the gaussian condition (G and LG groups combined) differed significantly from the uniform (U and LU combined) results at all lag values except lag 4. See Table 4-5. The early appearance of a negative correlation in the uniform CCFs might indicate either that expectancy formulation under uniform ISI conditions simply does not look back any farther than one previous trial, or perhaps that the internal models used by these subjects find the older trials to be useful indicators of how long the next trial is *not likely* to be.

TABLE 4-5

EXPECTANCY EXPERIMENT. Crosscorrelation of prediction interval with previous intersignal intervals: comparison of gaussian and uniform results.

	mean value of $r_{PI}(t)$		
t	Gaussian	Uniform	Test of G vs U difference
1	+.297	+.430	sig ($U_{20,20}=111$, $p<.02$, 2-tailed)
2	+.073	-.039	sig ($U_{20,20}=69$, $p<.02$, 2-tailed)
3	+.074	-.075	sig ($U_{20,20}=42.5$, $p<.002$, 2-tailed)
4	+.019	+.052	not sig ($U_{20,20}=140$)
5	-.048	+.024	sig ($U_{20,20}=114$, $p<.02$, 2-tailed)

Under uniform conditions, in which the ISIs will be inclined to vary greatly from trial to trial, this could form part of a plausible internal model.

Reaction responses (Late errors)

Not surprisingly, the number of reaction responses that subjects were forced to commit (when the signal had arrived before a prediction had been entered) was related to the number of short-duration ISIs within the signal time-series. Short ISIs were most abundant in the uniform distribution, and when using this distribution reaction responses accounted for 37% of all trials. In the two gaussian conditions the average proportion of reactions fell to 25%. Under constant conditions reactions accounted for an average of 23% of all trials.

The reaction time distributions were peaked curves with long right-hand tails, as illustrated in Figure 4-8. While most of these forced responses had a latency of only a few hundred milliseconds, reactions were occasionally delayed by nearly a second. The reaction time means and standard deviations are summarized in Table 4-6. The large standard deviations testify to the skewed response distributions. The ISI conditions did not appear to exert any systematic effects on mean or variance, but the tracking workload did seem to alter the reaction times. Comparing the G and LG gaussian groups, there was a significant increase in the mean latency for the (loaded) LG condition (Mann-Whitney $U_{10,10}=18$, $p<.01$). The standard

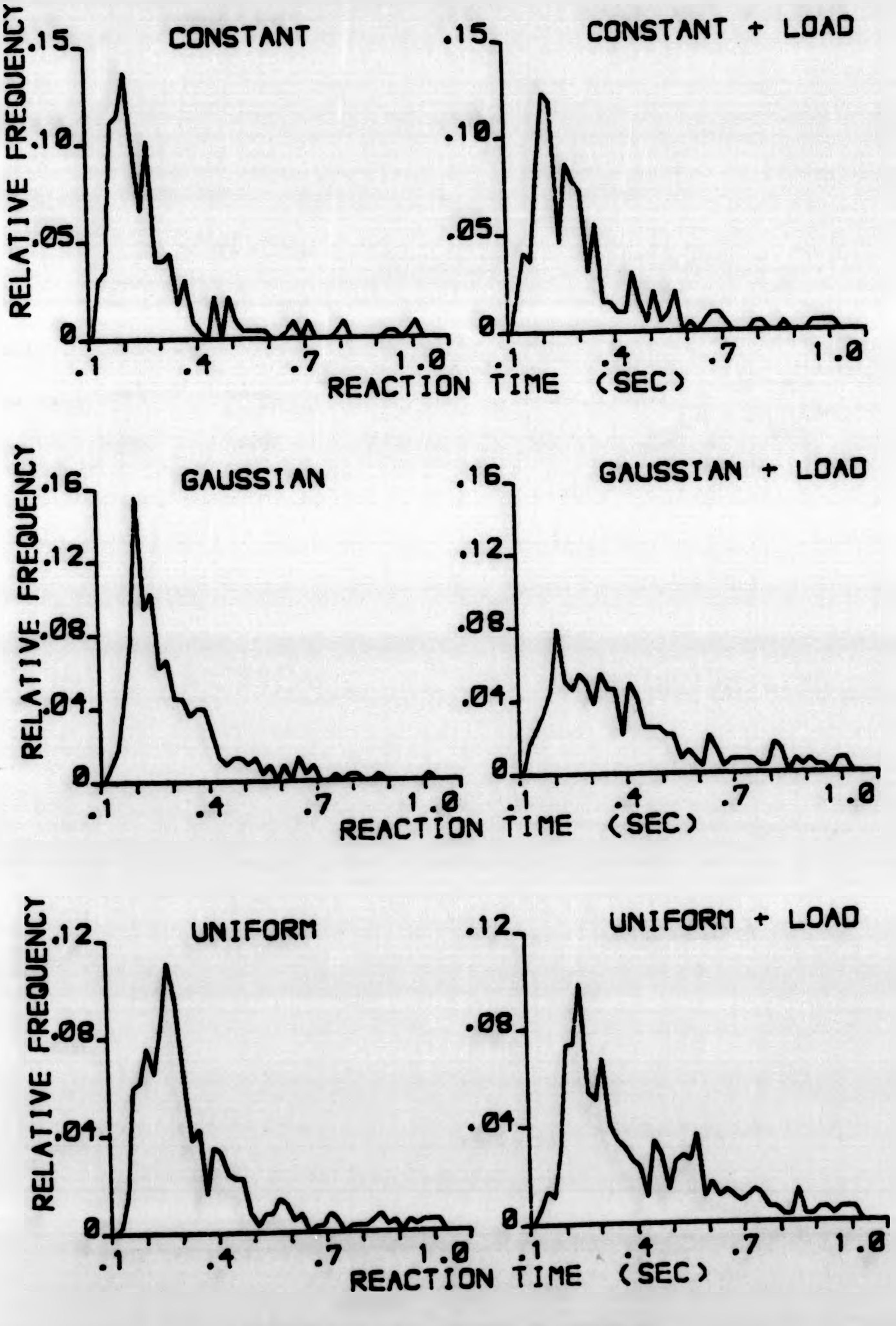

FIGURE 4-8 Expectancy Experiment.
Distributions of reaction responses.

TABLE 4-6

EXPECTANCY EXPERIMENT. Reaction time characteristics.

	ISI condition					
	C	B	U	LC	LB	LU
mean	305 msec	295	356	322	417	388
std dev	199	141	230	192	240	252
n	231	225	379	198	243	323

deviation of the reaction responses was also larger for the LG subjects ($U_{10,10}=19$, $p<.01$). Examining the nonloaded and loaded uniform ISI conditions, both the mean and the standard deviation of reactions again appeared to increase in the LU conditions but the differences were not statistically significant in this case. In the constant groups, the LC condition provided an increase in mean reaction latency over the C group which was nearly significant ($U_{10,10}=28.5$); but there was no indication that the standard deviation had been affected in any systematic way by the tracking workload.

It was hypothesised that whenever predictions were missed out the reaction times would measure longest when the subject's momentary expectancy differed a great deal from the concurrent signal timing. The subject would be more likely to be forced to react during the trials having the shortest ISIs. A test of the hypothesis reduces to: the shorter the ISI, the less appropriate the subject's expectancy would be, and the longer a reaction would require. To test whether this was so, the correlation between reaction time and the coincident ISI duration was examined. The analysis was applicable, of course, only to the conditions in which ISI was changing, namely the gaussian and uniform conditions. The results are summarised in Table 4-7. Under nonloaded conditions the correlation between reaction time and concurrent ISI was nearly always negative-signed. Negative correlation coefficients were obtained for all of the nonloaded G subjects, and for all but two of the nonloaded U cases. This argues that reaction latency did tend to become longer when the trial demanded a particularly short prediction interval. In the loaded

TABLE 4-7

EXPECTANCY EXPERIMENT. Correlation of reaction time with concurrent intersignal interval.

	ISI condition			
Subject	G	U	LG	LU
1	-.520	+.018	-.521 *	+.372 *
2	-.812	-.430 *	-.173	+.161
3	-.014	-.478 *	-.007	+.411 *
4	-.290	-.126	-.380 *	+.398 *
5	-.468 *	+.140	-.223	-.064
6	-.599 *	-.015	-.430 *	-.017
7	-.297 *	-.263	+.164	-.037
8	-.358	-.132	-.177	+.233
9	-.564 *	-.345 *	+.312	+.542 *
10	-.331	-.130	+.130	+.288
mean	-.425	-.176	-.131	+.229

* significant, $p<.05$ (2-tailed)

groups the correlation results were not as consistent. Although most of the LG correlations were still negative-signed, this was clearly not the case for the LU subjects. In general the inverse relation between ISI duration and reaction latency was not strongly demonstrated under loaded conditions. Indeed, for both gaussian and uniform ISI conditions the loaded groups' correlation coefficients reliably differed in sign and/or magnitude from the corresponding nonloaded outcomes ($U_{10,10}=20$, $p<.05$ for gaussian; $U_{10,10}=9$, $p<.002$ for uniform). One interpretation of this result might use the observation that the standard deviation of the reactions changed with loading: this would argue that the subjects' reaction latencies became more variable because of the secondary tracking task, and were therefore less reliable as systematic indicators of expectancy.

Questionnaire results

The subjects' own impressions of the prediction task reflected a respectably high awareness of the experimental conditions. Some of their responses to the questionnaire appear in Table 4-8. Among those subjects receiving variable intersignal intervals all but one recognised that the interval was not regular. However, more than half of the constant ISI subjects suspected that the interval was unvarying. Out of the total of sixty participants only seven acknowledged that they had attempted, at any point, to use a prediction strategy based on counting, tapping, or marking out time between signals. Despite the professed awareness of so many constant

TABLE 4-8

EXPECTANCY EXPERIMENT. Results of Questionnaire.

The text of the questionnaire is given in Appendix A.
In this Table (Qn) indicates the question number which supplied the data.
The results are based on ten subjects per condition.

	ISI condition					
	C	G	U	LC	LG	LU
Number of subjects who judged ISI to be of constant duration (Q5)	7	0	1	6	0	0
Number of subjects who admitted using a counting strategy at any time (Q1)	2	1	0	2	0	2
Subjects' mean estimate of average ISI duration in sec (Q6)	15.7	15.6	16.1	13.3	17.5	14.0
Mean estimate of number of signals in last block (Q7)	12.5	15.9	12.3	13.0	14.7	13.4
Block perceived as most accurate (Q3)	2.8	2.8	3.3	3.5	2.7	2.9
Satisfaction rating for predictions, using scale of 1=not at all satisfied; 2=satisfied; 3=very satisfied (Q4)	2.0	1.7	1.6	2.0	1.8	1.4

subjects, the strategists were not exclusively from the constant groups; indeed, only one of the constant subjects who suspected invariant ISIs confessed to testing a counting strategy. In this and in all other cases the subjects claimed that conscious measurement strategies were abandoned after only a few trials because they judged the methods to be ineffectual. There was no evidence to suggest that loading affected the subjects' perceptions of ISI regularity nor the temptation to try to cope by using counting techniques. Overall, the subjects' estimates of average ISI duration were quite close to the correct value of 15 seconds, although there were some quite extreme individual estimates. While there was no reliable universal effect of loading on the estimated duration, at least within the constant conditions the loaded LC subjects did perceive the ISI to be shorter than their nonloaded C counterparts (Mann-Whitney $U_{10,10}=25$, $p<.05$). This accords well with the earlier observation that LC subjects produced shorter prediction intervals than their C colleagues. Irrespective of the ISI conditions the participants in this experiment generally underestimated the number of signals in a block of trials by a wide margin (the actual number was 21) and they tended to believe that the "middle" of the experiment contained their most accurate predictions. Interestingly, most felt fairly satisfied with their predictions. Although the crude grading of the satisfaction measures adopted in Table 4-8 suggests that the constant subjects (both nonloaded and loaded) were the most satisfied, the small differences in satisfaction ratings did not turn out to be statistically significant.

The subjects' comments on the task were suggestive of some of the cognitive processes which could underlie expectancy. Initial difficulties with the task were somehow overcome after a bit of experience: "The first three or four tones in each run I found particularly difficult" (subject MK; LC group). Knowing when to push the button was "principally a matter of tension, not extreme, just a gradually [sic] growth" (subject PW; group LG). Subject WL (group LG) decided when to push "by considering time between other tones." Quite a number of subjects were prepared to be much more specific about the prediction process. Thus subject DZ (group LG) claimed that it was "just a case of remembering the previous prediction and compensating for error." Subject AR (group LG) said "I estimated when to push the button from the interval between the two preceding beeps" and subject EG (group LU) was even more precise about the same tactic when he reported that he had judged when to push "from the two previous times between the beeps-- take a rough average." It also emerged that the loading effect of the tracking task did sometimes interfere with the predictor's conscious efforts. The clearest expression of this came from subject PS (LU group) who described his frustrated attempts to identify some pattern in the signal timing: "Much of the time I was thinking 'not yet.' Mostly I relied on estimation of the time gap. I did also try to order the sequence according to rhythm but this never really came to anything and was interrupted by tracking."

A most striking characteristic of the subjects' remarks was that virtually all subjects had expressed some doubts about their ability to perform adequately in the experiment when the procedure was first described to them, yet at the conclusion of the session most of the participants expressed satisfaction with their efforts. Something had convinced them during the course of the experiment that the task was not impossible and that their predictions had been competent. It would not be unreasonable to suppose that this sense of competence derived from the formation and maintenance of an internal model (however flawed) for the temporal events in the experiment.

Discussion

When subjects are asked to predict the arrival time of intermittent signals, they generate predictions which are, on average, just short of the average intersignal interval. Their predictions are rational decisions which incorporate some knowledge of the statistical properties of the signal timing. The prediction intervals follow a peaked distribution, and usually contain a few extreme examples which actually fall outside the range of intersignal intervals; that is, the subjects typically underestimate the shortest possible intersignal interval and overestimate the longest duration. Temporal expectancy correlates with the subject's most recent experience of the intersignal timing; furthermore, one prediction may tend to be correlated with the next. This argues that expectancy, while inclined toward local consistency, is responsive to the sequential changes in signal timing. When the intersignal intervals become increasingly uncertain (when, for example, their distribution is uniform rather than gaussian) the modal prediction interval becomes shorter. This suggests that the subjects are sensitive to the distribution of event times and they recognise when brief intersignal gaps will be more frequent. When the intersignal interval is invariant, subjects can often detect the constancy (although their expectancies will be far from constant). But even when faced with much more difficult random timing intervals, subjects have considerable confidence in their predictions. Within their repertoire of information processing skills there is an accommodating provision for temporal expectancy; the subjects in this experiment do not feel that they have been unfairly called on to exercise clairvoyance.

When subjects are simultaneously loaded with the adaptive tracking task, they still predict competently and their expectancies cover a comparably wide range. But when subjects are tracking they probably tend to predict slightly earlier than they would if they were not loaded. Although this speeded responding might happen because the tracking puts some kind of diffuse time pressure on the prediction task, such an explanation is unconvincing because the reaction times collected from the loaded subjects were longer than those of the nonloaded participants. It is perhaps more plausible to suggest that the simultaneous tracking demands made the processing of temporal information more unreliable or noisy, producing an effect very similar to that of adding uncertainty to the intersignal timing. It is also possible that the intersignal intervals were simply perceived as being shorter in the loaded conditions. It has often been noted that durations spent in busy, demanding activity tend to be experienced as taking a briefer time than durations which are relatively inactive or unloaded. Block (1979) has reviewed various accounts of the phenomenon. This effect on conscious perception could have contributed to the shortened expectancies in the loaded groups.

Toward a model for temporal expectancy

A preliminary model for the temporal expectancy found in this experiment has been explored. The approach is fully described in Appendix B. The key idea behind the model is that expectancy during any intersignal interval is related to a moving-average of previously experienced intersignal intervals. Specifically, it is suggested that the prediction interval approximates a nonlinearly-weighted average of just a few preceding intersignal gaps. One interesting outcome of the model tested in Appendix B is that the averaging process describing the loaded subjects' behaviour does not differ significantly from that for the nonloaded subjects. Again, this argues that expectancy formation is not greatly affected by the extra processing demands imposed by the tracking task. The model also agrees with suggestions made elsewhere (Baker, 1963; Bartlett and Bartlett, 1959) that people are content to base their temporal expectancies on very few observations, perhaps after witnessing only two or three intervals, for example. If this is correct, it may help to explain why adaptation to distraction can occur after only a few exposures. Of course, if the distractors arrive at regular (and hence predictable) times, the expectancies are more likely to be accurate and adaptation would follow very quickly.

Temporal expectancy: Conclusions

It was suggested at the start of this chapter that adaptation to distraction involves the formation of expectancies about what distractors are possible and/or when they will occur. The work reported here describes how a kind of temporal expectancy is adopted even when subjects are asked to anticipate very irregularly-timed events. It is clear that subjects are very resourceful at generating temporal expectancies and that they extract useful information from experience. The most significant finding is that subjects are quite capable of formulating viable temporal expectancies while undergoing a heavy tracking load. Since the acquisition of temporal expectancy is not impaired by the processing demands of a simultaneous tracking task, it is certainly possible that temporal expectancy contributes to the adaptation process during repeated distraction. This possibility awaits future research.

Of course, a systematic investigation of adaptation would rightly devote much attention to the subject's expectations about the *content* of the distracting event; that is, its physical and informational properties. Such characteristics are far too wide-ranging to find a place in the present limited study which, after all, found that simply to demonstrate reliable distraction effects was surprisingly difficult. Expectancies about the contents of potentially distracting events naturally benefit greatly from learning, and they typically have a greater chance of predictive success than temporal expectancies, so they probably dominate in most adaptation processes.

CHAPTER 5: FINAL DISCUSSION

Overview

Distractibility is best viewed as a facility which influences the reallocation of processing resources rather than obstructing or destroying processing activity. This means that the effect of a distractor depends on the availability of spare capacity and the consequences of reassigning resources. If distraction interferes with tracking performance only when capacity is limited, and if, as seems likely, distractibility introduces only brief intermittencies into the operator's information processing, then much tracking performance is likely to survive distraction.

In at least some circumstances, distractors seem to compete for processing time with the processing demands of tracking, forcing the operator to increase his tracking response delay. Increased delay can have an additional knock-on effect on the tracking error. It is plausible also that some operators may try to compensate for extra delay by increasing transfer gain, thus making proportionally larger corrections for a given quantity of track disturbance in an attempt to enhance their sensitivity to error.

Problems of auditory distractibility measurement are attributable to fundamental difficulties related to the use of loud noise as distractors and, in the case of tracking, to the relative immunity of the measurement task and the insensitivity of its performance indices.

Recommendations for overcoming some of these problems are briefly noted, and suggestions for future research are outlined.

Many skills-- including the allocation of attention, monitoring and supervisory behaviour, and manual control tasks-- exemplify the strategic processing of information according to an internal model of world events. It is proposed that distractibility can contribute to the formation of internal models for attentional selection. Distraction is seen as a precursor to the purposive timesharing of selective attention, rather than an aberration in attentional functioning. In this view, distraction becomes an integral part of adaptive information processing skills.

The attentional demands of auditory distraction

There are incidents, such as aircraft accidents, in which a lapse in performance can have catastrophic effects. Distraction is often supposed to contribute to such lapses. Nevertheless, it should not be assumed that whenever distraction meaningfully degrades performance, it does so because it has plunged the nervous system into a sort of malfunctioning chaos. Distraction is more likely to act as a temporary factor in the allocation of a limited capacity resource. Broadbent (1958; 1971) has carefully considered the cumulative evidence of the effect of noise on behaviour which supports this intermittency interpretation. He has argued that noise does not impede performance as if it causes a sort of paralysis which prevents processing; instead, he has likened the effect of noise to a "blink" in the prevailing course of information processing, a moment in which the observer temporarily attends elsewhere. Broadbent's (1958) ultimate concern was with the effects of loud sustained noise, but his analysis began with the transient effects of novel stimuli (including noise onset and offset) which seemed to him to embody true distraction effects and to comply very convincingly with the "blink" analogy. Broadbent pointed out that even when we observe energetic "startle" responses to sudden events (the "jumping out of one's skin" characterised by stereotyped muscular contractions and so on) which may seem to be so severe that they disrupt task execution, these startle reactions should not be dismissed as aberrations which hinder the organism by swamping out or destroying the intended response to the task. Rather, the startle pattern represents purposeful competition for processing resources. Startle and orienting reflexes are very closely associated and are elicited by a wide range of transient stimulus events (Graham, 1979; Lynn, 1966); the orienting response reacts to the same signal variables (such as physical or contextual change) which, in voluntary selective attention tasks, will typically cause attention to switch to an unattended channel (see, for example, Hulstijn, 1979). The components of such reflexes serve to enhance the acquisition of information about a novel source. It is the momentary devotion of resources to this new goal which leaves the primary task impaired in the case of distractibility.* That is, the effect of distraction is most appropriately viewed as a limited capacity phenomenon.

---

*Startle and orienting responses provide vivid indicators of resource reallocation, but they do not go far enough to convey all of the cognitive implications of distractibility. Indeed, it is reasonable to separate the effects of these biological priming mechanisms from the more covert effects of cognitive distraction. It is precisely because the priming mechanisms manifest themselves in specific highly characteristic physical reactions and are physiologically recognisable, that they need not become the crucial subject matter of distractibility research in an information processing (that is, a non-physiological) context such as this. Thus startle and orienting have their place in these discussions as essential but relatively low-level biological tuning processes which subserve-- but are not to be seen as the same as-- full attentional distractibility with its rich cognitive connotation.

In Broadbent's interpretation distraction is a consequence of limited capacity, not a kind of perceptual crippling or trauma. His blink analogy carries with it some interesting implications:

> A blink usually lasts a definite brief period, though the frequency of blinks may be high or low. Each blink cuts off the incoming information to the nervous system instantaneously, but this does little harm for three reasons. First, if a man looks at some novel and important object he will temporarily suspend blinking. Secondly, although he cannot keep up this suspension of blinking indefinitely, he has some control over the time of occurrence of his blinks. As he goes on looking at an object he is more and more likely to blink, but if he knows that no crucial information will arrive at a certain time he can usually blink at that time, and so avoid interference with the task. Thirdly, he is able to continue acting even though he is not taking in information: he may follow a regularly wandering line with a pointer even while his eyes are shut, although he cannot of course do this with an irregularly wandering line (Broadbent, 1958, p.96).

These opportunities to compensate for the effects of blinking are comparable to the human operator's facilities for organising capacity and mobilizing effort during attending and performance (Kahneman, 1973). That is, we may expect that the attentional demands of distractors will not be constant and that human beings will have some freedom to mitigate their effects. As a result, when distractors are found to have more serious effects on performance, the task is likely to be one in which the rate of information transmission is high and capacity is somewhat restricted. Such tasks encompass those incorporating complex responses, difficult decisions, unanticipated events, or time pressure, for example.

Noise and verbal distractors compared

Laboratory measurements of distraction effects have long been inclined to use acoustically simple stimuli as distractors, particularly broadband white noise and occasionally pure tones. These stimuli are favoured not least because they are easy to generate, possess precisely defined onset and offset characteristics, and have an energy content which is simple to measure and easy to control. The experiments described in this study, as well as other published accounts, have indicated that when used as irrelevant distractors their effect on performance is to create a very brief impairment but that they lose this disruptive effect after relatively few exposures. These characteristics suggest that such distractors are 1) analysed and rejected quickly and that 2) the recognition/decision time is rapidly shortened by repetition. A number of experiments have demonstrated that repetition brings greater benefits if the distractors arrive at predictable nonrandom times (Eschenbrenner, 1971; Finkelman and Glass, 1970; Theologus *et al*, 1974). In the present study, distraction effects were largely unverifiable in (periodic) Distraction Experiment 2, although they had been clearly identified in (aperiodic) Experiment 1. The absence of effects during the second experiment cannot be considered strong proof of the role of periodicity (comprising as it does an essentially negative result) but the difference between the experiments was at least in the appropriate direction to support the contention that nonrandom repetition leads to a more rapid accommodation of distraction affects. It has also been proposed that the immunity derived from experience may be attained even more quickly if the operator's temporal expectancy can accurately anticipate the distractor's arrival.

Because simple noise distractors tend to have such fleeting effects, experimenters must make noises loud to obtain repeatable consequences. The use of loud noise can, in carefully controlled circumstances, make effects on performance detectable-- but the effects are probably not purely distraction-based. The relations between noise stress and noise distraction, along with the confusion accompanying the use of both continuous and intermittent noise in research, are important and have already received some discussion previously in Chapters 1 and 3. It seems likely that when the intensity of noise exceeds some level, as yet uncertain but probably between 80 and 90 dB (note the remarks on noise stress in Chapter 1), it induces a complex change in the subject which combines the arousal concomitants of stress, the muscular reflexes characteristic of startle, and the transient effects of distraction into a compound effect on attention and performance (see Graham, 1979, for the amplitude and rise-time properties of sounds which elicit startle; see also Kryter, 1970, for his summary of the amplitude levels of noise likely to alter performance). These internal events may give rise to increased error generally, and in tracking may even have specific effects on the operator's gain (visible as excessive joystick movement) and on the response delay (evidenced by increased control lag).

In contrast to noise, verbal distractors are awkward to generate under automated conditions, more uncertain in their onset/offset features and energy content, and difficult to categorize. However, as distractors their greater potency is assured by their semantic ambiguity. They cannot be dismissed as irrelevant without some sort of analysis, and if analysis requires a deviation of processing resources they may give rise to distraction effects. While the distraction effects of noisebursts seem to be very localised, it seems likely that verbal distractors invariably create longer disruptions, simply because of the necessity to attend to the entirety of each message in order to evaluate its relevance. This additional processing time is probably just one variable affected by the use of verbal distractors. The processing resources which are recruited probably differ also; access to semantic memory is an obvious example. Consequently the effects of verbal distraction on tracking may be due not so much to temporary changes in specific transfer function parameters such as the operator's response delay, but to a more general effect of capacity restriction-- leading to a more chronic neglect of control over ongoing processes, including ongoing movements.

This view hypothesizes that at more moderate levels of loudness the relevance of a pure noise distractor is not sufficiently ambiguous to elicit repeatedly a large deviation of processing resources. The event comes to be recognised as unimportant so early in analysis that very quickly there is little or no interruption of the ongoing task. The recognition process is probably enhanced by temporal expectancy-- the formation of which, conveniently, seems not to be hampered by a heavy tracking load (Chapter 4). Loud unpredictable noise works mainly because it is loud: intensity seems often to be associated with stress, and stress is often linked to altered performance. A corollary to this hypothesis is that any invariant sounds are likely to be limited in their distractibility potential in the same way and for the same reasons as white noise. Their disruptive properties will be quickly weakened by their inherent recognizability and, when it applies, their temporal predictability.

In this context, it is interesting to note that when electronically synthesized speech is used in hazard warning systems, it evokes behavioural responses associated both with simple tone annunciators and with authentic spoken warning messages (Hakkinen and Williges, 1984; Simpson and Williams, 1980). This mixed effect of synthesized speech has been attributed to its idiosyncratic acoustic features which readily distinguish it from genuine speech. It is as if the synthetic utterance-- with its slightly odd acoustic properties-- lies somewhere between a spoken and a noise-based distractor.

Acting as a stressor, a loud noise distractor may create conditions which are difficult to analyse: its direct effects on attending and capacity may interact with other psychological processes. As Fisher (1986) concluded about noise effects:

> Taken collectively, we are left with an account of a stimulus which at sufficient intensity may distract, activate, mask, or increase mental demands. One temptation is to build models that account for one influence in terms of the effect of another: For example, increased arousal causes increased tendency to distraction; increased mental demand causes increased arousal; increased mental demand is the result of distraction; increased arousal creates connotative signals which absorb capacity and increase mental load.

> Noise research provides a useful illustration of the problems of competing explanations, perhaps bolstered by the varying evidence that has arisen because noise has been a favoured stress for experimentation. There seems no good a priori reason for neglecting evidence identifying a particular kind of influence (p.131).

As an aid to interpretation, Fisher (1984a; 1986) proposed a composite model of stress which assumed that

> any *one* stress has a *number* of influences, all of which can operate simultaneously. Thus, to use the example of noise again, it may variously distract (mode 1), mask data (mode 2), arouse (mode 3), increase mental demand (mode 4). The modes of influence identified are assumed to be only *potential*. Situational factors such as the task and the instructions may determine which modes actually operate to influence performance (Fisher, 1996, pp.131-132).

Demand awareness and subjectively attributed relevance

The "situational factors" mentioned above include determinants such as the subject's *private* perception of the distractor's potential relevance and the subject's impression of the *experimenter's* expectations. One of the differences between noise and verbal distractors is likely to be that verbal material is not so convincingly irrelevant or innocuous to the subject. Probably all laboratory experiments to measure distractibility are affected, to some extent, by the subjects' "demand awareness"; namely, the subjects' tendency to try to assess and fulfil what the experimenter is expecting of them. One implication is that subjects may privately attribute significance to distracting events despite the experimenter's belief that the distractors have no legitimate contextual relevance. In complying with the assumed demand characteristics of the experiment, the subjects may actively attend to distractors as if they formed part of the task.

Fisher (1986) has surveyed some of the evidence that subjects commonly strive to build up "demand awareness" of the experimenter's intentions. Fisher considers that demand awareness plays an important part in a person's efforts to overcome the effects of stress. The significance of demand awareness emerges from her proposal that in stressful situations, the individual will typically seek to reduce the psychological effects of stress by attempting to control relevant aspects of the environment. Furthermore, Fisher argues that to alleviate stress it may be sufficient that the individual believes that control is possible, even if it is not exercised. Because Fisher is concerned with the individual's perception of control effectiveness, she has looked for confirmation in the learning literature that awareness and evidence of control may be acquired by subjects even in the simplest learning tasks. Fisher (1986) pointed to a number of studies, involving both classical and instrumental conditioning, which have demonstrated that awareness mediates responding. She noted that an awareness hypothesis need not necessarily claim that the subject's awareness of contingencies matches the experimenter's, as long as the outcome convinces the experimenter that the subject has been conditioned. Fisher concluded that it was important to distinguish between awareness of what we might call the "declared" contingencies in an experiment, and the subject's own formulation of what the experimenter is seeking. The subject uses these (possibly different) assessments to decide, in some strategic fashion, how to cope with the experiment. Fisher has suggested that the subject's perception of control grows out of a two-stage processing of "suspicions" formulated before and during the experiment. So in a conditioning experiment, for example, she suggests that first suspicions about the probable stimulus-response contingencies will draw on previous experience and perhaps socially-learned examples; second-stage suspicions will arise from the subject's appraisal of demand. The stages may interact: demand awareness will probably be influenced by the events and contingencies which are observed early in the situation. The outcome of this suspicion assessment is taken into account when the subject defines the discrepancy between present conditions and the final goal and plans what action to take to reduce the discrepancy. In the context of distractibility research, the major implication of Fisher's hypothesis is that because subjects assess demand (particularly when stressed) they are less likely to ignore "irrelevant" distractors which convey a rich set of interpretive possibilities (such as verbal nonsense messages), than to ignore "irrelevant" distractors which offer less material for speculation (such as noisebursts). If the demand awareness interpretation is accurate, then the irrelevance of a distractor is impossibly difficult to guarantee. An "uncontaminated" distraction effect is elicited only by the very first exposure of a repeated distractor; after the initial exposure some element of intention may be driving the subject's attention-switching to the intrusion.

This means that our definition of an "irrelevant" distractor must be tempered by the possibility that subjects will misconstrue its genuinely innocuous nature, so that the motivation to attend to it may not be completely involuntary (except perhaps in the first instance). This possibility is best illustrated by example. In Distraction Experiment 3, subjects were not forewarned of the distracting messages and when they were presented the subjects may have thought that they were expected to devote considerable attention to them for the purpose of, for example, remembering them. This sort of supposition on the part of the subjects would not devalue the results of the experiment. Our position must be that a laboratory demonstration of distractibility can control demand awareness only imperfectly: the inability to determine whether these subjects were disinclined to regard the messages as "just distractors" simply reflects the fact that the subjects' motivation to attend to distractors in the real world is continually liable to change. To continue the illustration, we might imagine that subjects in the two earlier noise distraction experiments thought that they were supposed to remember how many noisebursts were presented, or to judge how often they occurred; perhaps they wondered if the experimenter hoped the noises would make them anxious. The point of these examples is that subjects were expected to attend occasionally to these laboratory events for some of the same reasons which motivate attention to distractors in real life. Distractibility exists as a psychological phenomenon because people are unable to judge the relevance of perceptual events until some kind of analysis is carried out and sometimes this analysis is so demanding that performance is affected. It is not apparent in any distraction experiment why subjects attend to an objectively irrelevant event (an event which does not, after all, have to be processed to carry out the prescribed task). We must presume that the event retains some subjective significance, if only for part of the experiment. Verbal material probably interacts importantly with subjects' demand awareness, because they are likely to try to build up their awareness basing it not just on the initial instructional bias, but also by surmising what they can from the contents of the messages themselves. This may sustain the distractor value of such material for a considerable time, although strictly speaking the process may no longer be entirely involuntary. This issue, the transition from involuntary to voluntary attending, will be discussed more fully in the final section of this chapter.

The attentional demands of tracking

Tracking integrity during competing demands

Tracking performance is undoubtedly sensitive to capacity limitations. This has been most clearly demonstrated by dual-task experiments which have exercised graded control over capacity with attendant effects on tracking competence. Trumbo, Noble, and Swink (1967) reported a number of dual-task experiments in which the primary task was pursuit tracking of a visual target that moved in unpredictable steps. The secondary task normally required subjects to analyse spoken sequences of digits and to call out expected digit-names before their next presentation. The experimenters recorded the tracking integrated error. The error was significantly greater under dual-task conditions compared to performance without the secondary digit task. In one investigation (condition "S" in Trumbo *et al*'s second experiment) subjects merely heard the secondary task digits while they tracked: there was no digit task response requirement at all. The subjects were told that the spoken digits provided "a noise condition" and that this test was to "simulate a noisy working situation" (Trumbo *et al*, 1967, p.237) during tracking. This condition really constituted a kind of distraction experiment, although Trumbo *et al* did not refer to it as such. This listening-only condition produced no significant effects on tracking error (in other words, there were no distraction effects) when compared to the tracking-only control condition, but a significant effect on the error was found if the test situation was altered only slightly. This alteration (treatment "R") asked the subjects to call out any digit-name they wished after each hearing of the experimenter's digits; that is, to issue some response (no matter that it had no apparent relevance to what the subjects were doing). Because this made the tracking performance worse, a reasonable interpretation of the result is that by adding a response requirement, the capacity devoted to the tracking task was compromised and performance could not be sustained. The full dual-task condition made tracking performance worse still. Later experiments by Trumbo's group (Trumbo and Milone, 1971) succeeded in showing that tracking error could be increased even by the relatively undemanding requirement to encode secondary task stimuli (somewhat comparable to the responseless listening condition of the 1967 experiment) or by the secondary task of retaining in memory the order in which signal lamps were lit. Trumbo and Milone suggested that the effect of a secondary loading task on primary tracking performance was dependent on central processing demands imposed by the secondary task (and not just by more obvious sensory and motor competition), with information retention having least effect on tracking, encoding of new information having relatively greater consequences, and recall-- that is, response selection and/or execution-- having the greatest impact. Under the appropriate conditions all of these secondary processing requirements were shown to have significant effects on tracking error. In another series of dual-task experiments Johnston, Greenberg, Fisher, and Martin (1970) examined compensatory visual tracking of a complex mixed-sine track while their subjects carried out a simultaneous verbal retention task. In this case, the subjects were urged to treat the verbal task as the main task and the tracking requirement was a subsidiary task. As in Trumbo's laboratory, the tracking integrated error was measured, and as Trumbo and Milone had reported, the error increased as a function of the difficulty of the alternative task's demands whether this involved encoding, retention, or recall. It is interesting to note one of their results which reiterates the now-familiar finding that effects of moderately loud noise on tracking may go undetected. Like Trumbo *et al* (1967), Johnston *et al* studied one control condition which had at least a superficial resemblance to a distractibility measurement. In treatment "TN" of their first experiment, Johnston *et al* asked subjects to track without any verbal task requirement but they exposed the subjects to continuous 86 dBA noise over headphones. Comparing these trackers to similar subjects receiving no noise (condition "T") they concluded that the loud noise had no effect on tracking error. It was assumed that mere noise imposed little demand on processing capacity.

Tracking characteristics during competing demands

When secondary task demands compete with tracking for processing capacity, one effect is that the operator's response delay is increased. This was demonstrated by Trumbo *et al* (1967). Because their tracking task used a discrete step function to move the target, the target motions were not continuous but only occurred at identifiable time-points. This allowed Trumbo *et al* to hand score the local spatial and temporal characteristics of their subjects' tracking movements at the moment of each stepwise change in the target's position. After hand scoring they found that while there were no systematic effects on the spatial properties of the tracking responses (that is, overshoots and undershoots of the desired target position), they could observe that the temporal properties of the operator's tracking movements had been systematically affected. The frequency of making responses which lagged behind the step movements increased at the expense of leading or anticipatory movements, which decreased in frequency. Also, the mean lag time was longer (by perhaps as much as about 80 milliseconds) when the trackers were loaded by the secondary verbal task. This result may be compared to the very similar effect of noise distractors on tracking response delay reported in Chapter 3, where the increased lag was estimated to be at least 30 milliseconds.

When response delay increases, the error is likely to increase as well. This is because even if the operator is still faithfully following the magnitude and direction of the track disturbance, the delay has in effect introduced a phase shift between the forcing function signal and his response. In a closed-loop negative feedback system such as compensatory tracking "any delay [in the feedback path] will mean that error cannot be compensated, since the response will be appropriate to a state of affairs that actually obtained earlier, rather than the one obtaining presently (Moray, 1981, p.27)." It is probably true that some of the increased error observed during distracted tracking is directly attributable to additional response delay.

In addition to response delay, a human transfer function representing tracking behaviour normally incorporates an element of operator gain. This is a description of how much the operator responds to a given amount of error. Demands on capacity could lead to changes in gain which would affect tracking error. One conjecture is that an operator who becomes aware of increased error (and it should be remembered that in compensatory tracking the display is indicating error) may increase his transfer gain to try to compensate more adequately for the build-up of error. The operator is, in effect, making himself more sensitive to error. One of the individual tracking records included in the results of Distraction Experiment 1 (Figure 3-7a) may be an illustration of this gain effect. Ironically, the tactic of increasing gain can lead to more rather than less error. An operator's corrective movements invariably contain some overshoot and undershoot, as well as tremor-induced movements. These sources normally augment the disturbances brought by the forcing function, and the operator does not distinguish well between the display movements caused by these factors and those due to the forcing function. If the gain is too great, the operator may end up trying to compensate for his own movements, and so increase the final error outcome.

The problem of distractibility measurement

The effects of distraction are obscured by their great dependence on local variations in the task demands. This was very noticeable in the second-by-second breakdown of the tracking task used in this study. There are probably a number of important limitations on tracking as a distraction-sensitive task.

The timescale of measurement

First, there is a fundamental incompatibility between the long-term time averages (such as RMS error) used to measure tracking performance and the short time scale of the transient changes provoked by distraction. If the duration of the distraction effect is short relative to the measurement period the effect may simply not be detectable. This dilution of very brief effects by the averaging process could be made even worse if, within the measurement span, the subject tries to compensate for the harmful effects of distraction. The distractibility literature has observed that subjects often seem to invest extra effort to regain good performance after distraction.

Task immunity

Secondly, in spite of adaptive task techniques, increasingly skilled sensorimotor performance probably has a generally liberating effect on spare processing capacity. It has been noted that as performance becomes highly skilled, that skill leads to a form of automaticity:

> One generalisation which can safely be made is that with practice performance improves and effort declines. This seems to be linked to the fact that prolonged practice in highly motivated human subjects makes the single channel limited capacity model of the human operator less and less appropriate. High levels of practice lead to "automatic" behaviour in which no conscious control is exercised. Simultaneously with the development of automaticity it seems as though the decision stages of information processing which are so evident early in practice disappear, and little or no effort is felt to be expended; and no effort is experienced as a feeling of stress. Behaviour and performance flow smoothly to their goal, and only if there is an emergency, a sudden change in system properties requiring a change in strategy by the operator, does a significant work load reappear (Johanssen *et al.*, 1979, pp.108-109).

In tracking skill particularly this automaticity may signify that the operator depends less on the closed-loop feedback model of manual control and rather more on a partially open-loop predictive strategy. That is, instead of continuing to function like an error-correcting servomechanism, the skilled tracker may begin to rely on predictive control of track behaviour sustained by subsidiary control loops (Pew, 1974). In this mode of control, error correction still remains vital, of course, but it would not necessarily require the same large allocation of processing capacity. Such a freeing of capacity could greatly reduce the effects of distraction on tracking performance. It has been observed that when track prediction is feasible, control can remain accurate even during short absences of the tracking display (Flowers, 1978). Even with a random forcing function, a track exhibits local predictability. "Some of the learnable characteristics of any irregular track are probably the average position on the display, the average and approximate maximum amplitude, the average frequency of reversals, and the average and approximate maximum rate of movement" (Poulton, 1974, p.121). Knowledge of this sort and predictive strategies which exploit it may greatly compromise the suitability of tracking tasks for the assessment of distractibility.

Lastly, it may be that a skilled operator can afford to monitor the tracking display using a quite discontinuous visual sampling strategy without losing control. Gaps between samples could offer the operator enough time to analyse distractors without harming performance. This proposal derives from models of visual sampling processes which describe how the physical properties of an information source and the internal model of the observer influence sampling behaviour (for a survey, see Moray, 1984 and also Sheridan and Ferrell, 1974). Such models build on the work of Senders (1964), who used the predictions of information theory to construct a testable hypothesis about when the observer should attend to an instrument display. Senders assumed that attention would not remain fixed on a display indefinitely, and that in a visual task, eye-movements (controlling foveal focus) would approximately indicate the redirection of attention. Senders noted that the Sampling Theorem showed that, for a time-varying signal which is limited to a bandwidth of *W* Hz, it is only necessary to sample the signal at 2*W* Hz to fully reconstruct the information in the signal. His proposal, basically, was that for a visual display fluctuating at the maximum rate of *W* changes per second, an optimal observer should, using a sampling rate of *2W*, look at it every *1/2W* seconds. In this case the optimal observer would be one who has full knowledge of the statistical properties of the signal, remembers observations perfectly, and whose goal is to reconstruct the signal. Even so, the model has proved to be very useful in the analysis of real tasks, and can provide a reasonable fit to eye-movement data. The exceptions occur at extreme bandwidths. At very low bandwidths, observers sample more frequently than the model predicts. At high bandwidths, observers take fewer samples than would be expected. It is important to appreciate how limited, in absolute terms, the range of signal bandwidths actually is in the context of human control tasks: in visual scanning studies, as in tracking work, even the most demanding tasks cannot use signal rates much faster than 1 Hz. Probably at low bandwidths the observer needs information more often because if he adheres to longer intersample intervals which better agree with Senders' model he will have problems with forgetting. At high bandwidths, observers may be economising on their sampling by extracting rate-of-change information from the signal and using the rate information to reconstruct the signal events. Fogel (1955) demonstrated that if the observer has rate information as well as amplitude data about a signal, it is sufficient to take samples exactly half as frequently as the Sampling Theorem otherwise dictates (that is, at *W* rather than *2W* Hz).

How may these points about visual sampling strategies relate to the distractibility of tracking tasks? People cannot track forcing functions with bandwidths much exceeding 1 Hz. If the operator extracts only position information under such circumstances, we might expect the sampling rate to be two observations per second (that is, 2W), and actually less if the forcing function bandwidth is below 1 Hz. A really efficient operator might be able to exploit rate information as well. (Personal experience with the tracking task described in this thesis certainly suggests that the operator is very conscious of local changes in the velocity of the track.) Taking Fogel's extension to the Sampling Theorem, the skilled operator might need to sample only at a rate of W Hz for a W Hz forcing function. These proposals imply that even for a difficult 1 Hz bandwidth tracking task, the operator could require no more than a brief sample every half second to produce adequate performance, especially if the accuracy criterion is undefined and normalised RMS errors approaching 1.0 are tolerated (as in the adaptive tracking task). At lower forcing function bandwidths, the observation rate could be even lower. In the extreme cases where the bandwidth exceeds 1 Hz, it seems likely that velocity information would become most apparent and so tend to drive the sampling of the optimal observer down to the W Hz rate. Senders has shown that the human observer can behave optimally in sampling tasks. Also, after collating data from many eye-movement studies, Allen, Clement, and Jex (cited in Moray, 1984) concluded that eye-movements rarely occur faster than twice per second in tracking tasks.

It has already been suggested that time-averaged measures of tracking performance are inherently insensitive to shortlived fluctuations; and that skilled performance implies a degree of automaticity which, in a tracking task, could help the operator to maintain the trajectory of a control movement even when attention is briefly diverted. If during tracking the operator can also adopt the low observation rates discussed here, it would account for the insensitivity of tracking to the brief transient effects of distracting noisebursts. The attentional demands of the distractors are readily accommodated because they rarely inconvenience the relatively leisurely visual sampling requirements and they do not interfere with semiautomatic ballistic movements. The account of sampling strategies could also explain the lack of effects of tasks like temporal prediction on tracking.

It is important to acknowledge that this model of sampling behaviour during compensatory tracking assumes an ideal operator under optimum conditions. In practical situations the displayed error sampled by the operator includes elements of the operator's own control movements, including random motor noise. Because of this, the actual bandwidth of the display events may exceed the forcing function bandwidth and the sampling rate may be higher than the hypothetical limit. Even with this qualification, this analysis of sampling behaviour may suggest a conclusion of great methodological importance: namely, that when it is used to measure transient effects, tracking may have a built-in deficiency because of the human operator's ability to sample optimally.

Recommendations

Future work on distractibility would profit by extending this research in two ways. Firstly, in seeking effects beyond the initial first exposure, it now seems advisable to relinquish loud noise distractors and concentrate on information-loaded distractors such as verbal messages. This approach should make possible a more thorough experimental investigation of adaptation processes. Initially, it would be appropriate to assess the effects of periodic versus aperiodic presentation on adaptation, using verbal messages. We would expect, from the accumulated evidence, that periodic exposure would lead to more rapid adaptation. When a confident understanding of the temporal factors has been obtained, it should be easier to construct repeatable distraction environments using these distractors. This would allow further work to expose the cognitive determinants of adaptation. A crucial issue, clearly, is to ask whether there is a necessary connection between the ongoing cognitive analysis of distractors and the gradual process of adaptation. One experimental approach might be to present verbal distractors which are correlated with respect to a single semantic category and compare the rate of adaptation to that when uncorrelated material is presented. If we assume that adaptation requires continual cognitive analysis, and that similarity should encourage particularly efficient analysis, categorization, and rejection, then we would predict that related distractors will lead to more rapid adaptation. Other experimental approaches to this question could exploit the ease with which verbal distractors lend themselves to "probe" techniques. For example, relatively slow or complex distractor analysis (with an accompanying performance decrement) might suddenly result when a series of comparatively consistent distractors terminates with a very distinctive probe distractor, such as a message in a completely different language. Experiments like these would help to explain how deep the analysis extends when distractors are processed. But such experiments must be preceded by work to develop an appropriate paradigm which uses information-rich distractors that enjoy some resistance to adaptation. The selection of appropriate verbal distractors remains difficult; the potential problem of subjects' demand awareness suggests that messages and instructions to subjects should be devised with care to prevent artifacts. However, even using normal prose as distractors allows for interesting and useful manipulations: tapes of material may be played backwards to eliminate semantic content; the information content can be altered by transforming normally coherent text into statistical approximations to English (such as the samples described by Taylor and Moray, 1960).

Secondly, while it does seem advisable to seek "complex, behaviourally rich" (Johannsen *et al.*, 1979, p.110) alternatives to laboratory tasks such as tracking, there are several avenues still unexplored in the tracking paradigm. It may be possible to prevent automaticity in tracking performance by making the plant inherently unstable, as in the "critical task" techniques advanced by Jex and his colleagues (Jex and Allen, 1970; Jex and Clement, 1979). In such tasks instability is built into the controlled element and it is increased until the operator eventually loses control. Also, the sensitivity of tracking performance measures might be improved by tightening the relation between distractor presentation and suspected high-susceptibility portions of the task. For example, synthetic forcing functions could be generated to control local features (such as track reversals) more precisely while still appearing random to the subject. It would also be worthwhile extending some of the distractibility work from compensatory tracking to pursuit tracking, because of the importance of pursuit skills in certain real-life tasks: most notably, driving an automobile. Driving is widely regarded as a "submaximal task" (Naatanen and Summala, 1976, p.152) and so may not be very distractible; but the social cost of car accidents is such that a reliable methodology of distractibility assessment applied to this problem could make a valuable contribution to the public good.

The last point may be taken further. The identification of distraction effects has clear practical significance for human factors engineering, both because of the need to assess the reliability and safety of man-machine systems and also because it could lead to methods for reducing operator susceptibility to distraction. This is especially pertinent in light of the evidence that there can be a very low level of subjective awareness of distraction effects. At the other extreme, there may be occasions when it is actually desirable to enhance distractibility in certain ways rather than to eliminate it. Warning systems tend to rely on loud and strident signals which interfere with problem-solving and performance, and the design of less stressful warning annunciators is becoming increasingly important as man-machine systems become more complex (Patterson, 1982). A better understanding of the basis of distractibility and its effects would provide guidance for this work.

Distractibility, internal models, and attention

In the introduction to Chapter 1 it was suggested that much of the contemporary neglect for involuntary aspects of attention such as distractibility could be traced to the popular cognitive approach and its preference for active information-processing interpretations of performance. Ironically, strong incentive to reexamine distractibility now comes from the original beneficiaries of the information-processing approach: selective attention, monitoring behaviour, and manual control.

Accounts of manual control skills such as tracking increasingly make reference to *internal models* which guide performance (Kelley, 1968). Internal models are cognitive frameworks which embody knowledge of those features of the real world which are relevant to a particular task. The models are used to generate operating strategies. In the case of a tracking task, the operator's internal model must incorporate knowledge about the properties of the forcing function. The internal model evolves with experience to generate various expectancies about the task. Thus, the skilled tracker will try to anticipate the movements of the track, often successfully. Current approaches to attention, monitoring, and supervisory control are also beginning to stress the importance of strategies and internal models (Moray, 1978; Sheridan and Johannsen, 1976). This perspective makes it possible to fit distractibility more meaningfully into the hierarchy of attention, placing it somewhere between the crude alerting which serves biological necessity and the more goal-oriented information-gathering strategies.

In this account, attention is a skill to control processing resources, a skill in which distractibility and learning help to build allocation strategies. For some time, evidence has been accumulating to suggest that internal models guide attentional selection when a task is well-practised. The data originate in a number of selective listening experiments in which the listener had to respond to targets on one or two auditory channels (Moray, Fitter, Ostry, Favreau, and Nagy, 1976; Ostry, Moray, and Marks, 1976). The results argue strongly that the skilled voluntary selection of messages, regardless of whether the information consists of pure tones or semantic material, is guided by a rational sampling strategy. The sampling strategy is rational in that it is precisely appropriate to the statistical structure of the information sources. In other words, the directing of attention among competing sources occurs in a sensible way, much as any skilled performance relies on the efficient deployment of resources. These studies exploited innovative, short-term measures of performance which demonstrated that whole-session estimates of signal detection statistics during divided attention experiments could not always reveal the truly labile quality of the ongoing information processing.

This insight only emerged when observations of performance relative to any one information channel were categorised on the basis of events occurring in the contralateral channel. The contralateral contingency analysis indicated that when a useful statistical relationship existed between messages, the signal detection measures at any one instant for a given message were modulated by what was happening on the other channel. When attention was divided between two messages, both the detectability of the signals being accepted at that moment and the response bias were affected by the events occurring in the other message. The observation that the instantaneous signal detection characteristics were regulated by the conditional probabilities within the whole stimulus set was an important one. It indicated that attention switching was a rational, skilled procedure in these experiments. Given sufficient practice, the timesharing observer can predict from one message content the other's ongoing target probability and adjust his detection strategy appropriately. The evidence supports the original proposal of Moray and Fitter (1973) that observers extract information concerning the statistical structure of the stimulus sources and use this knowledge to build predictive internal models of the events which demand attention. It does not seem to matter whether the messages are trains of pure tone bursts (Moray *et al.*, 1976) or contain semantic material (Ostry *et al.*, 1976). But achieving a near-optimal sampling strategy requires practice (see Moray *et al.*, 1976) in dealing with the local and general statistical properties of the stimulus world.

One observation from these experiments leads indirectly to distractibility. Curiously, in tests of dedicated selective attention-- in which the listener concentrated on one channel and did not want to divide attention between sources-- the effects of the "unattended" message on the signal detection measures were present as well. Ostry *et al.* (1976) stressed that this could not be accounted for by prior divided listening practice nor by carry-over of statistical models; nor did it seem plausible that in the dedicated task the listener confused his messages because he could not really tell one ear from the other. The significant point is that some form of involuntary analysis of unwanted information kept occurring. Moreover, the effect was more pronounced with semantic and alphanumeric targets than with pure tones. Therefore, transient events in the unwanted message stream (particularly when signaled by the temporal asynchronies and rise-time peculiarities of verbal material) were apparently triggering attention-switching involuntarily.

An obvious function of distractibility in this account of strategic selection is to provide a form of transient detection which operates involuntarily. This function is greatly extended by the inclusion of the idea of an internal model guiding skilled attention. Consider that forming an appropriate internal model to guide attention-- and revising it to keep up with a changing world-- likely relies on susceptibility to a range of stimulus cues which prompt involuntary switching. This proposal states that when learning how to attend, involuntary intrusions can advise about the appropriateness of the prevailing sampling strategy. When we genuinely want to divide our attention among multiple sources, distractibility can help to distinguish those sources and draw notice to their contingencies. When well-practised, this mix of voluntary and involuntary selection may evolve into genuinely skilled timesharing guided by an internal model. By implication, distractors cease to distract when they have been incorporated into an internal model of ongoing events.

This overview of attention incorporating distractibility provides a conceptual magnet to draw together the voluntary and involuntary phenomena of attention. The main themes are illustrated by the root-diagram in Figure 5-1. In this drawing, the higher attentional functions are dependent on the selection of information by lower processes, with voluntary and involuntary attending shown as joint contributors to purposive skilled selection.

The traditional research emphasis on voluntary selection is incomplete without the inclusion of involuntary and spontaneous aspects of attentional activity, including distractibility. In the traditional approach intrusions from an unwanted source seem to be significant mainly because they demonstrate the failure of dedicated attention. The attitude proposed here is that the occurrence of these breakdowns is more than just an incident of attentional misbehaviour; it is an important part of the total description of how attention is used. A complete understanding of attention must take account of the involuntary action inherent in normal attentional functioning, because it must form the substrate on which acquired attentional skills are built.

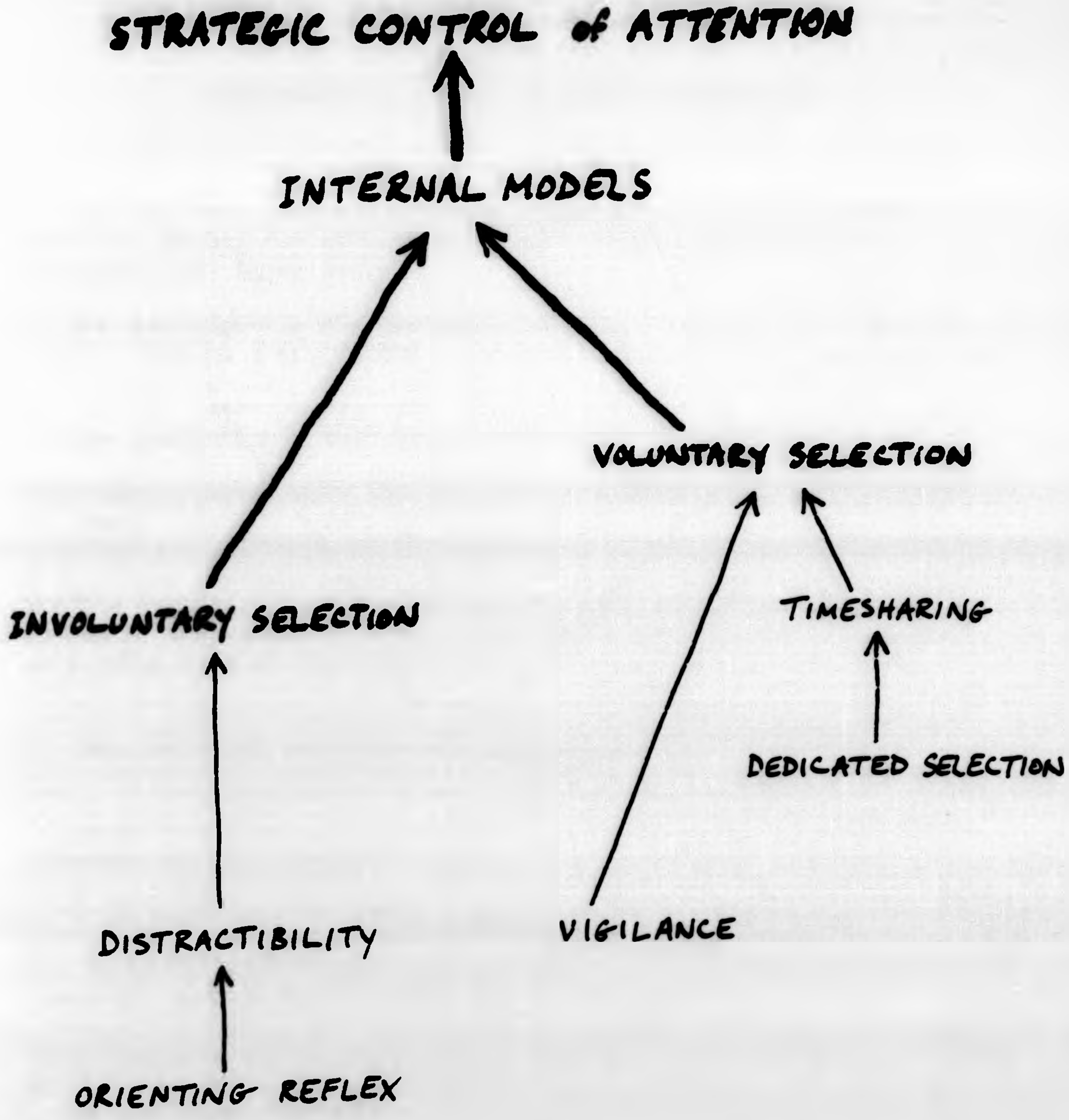

FIGURE 5-1 Functions serving attentional selection.

APPENDIX A

QUESTIONNAIRE USED IN EXPECTANCY EXPERIMENT

1. Did you find yourself counting, tapping, or using any other conscious method for marking out time? Never? Occasionally? Throughout the experiment?

If you did, describe the method(s):

2. Can you describe what you were thinking during the time leading up to your prediction? How did you judge when to push the button?

Were there moments when you felt you were under particular strain? If so, when?

3. The experiment was divided up into five sections, with rests in between. Did your predictons seem to be more accurate during any particular part of the experiment?

4. How satisfied were you with your predictions in general?

To answer the remaining questions, just think about the LAST signal sequence you were given:

5. Consider the quiet interval between each signal and the next. Did the quiet interval between signals always seem to have the same duration, or did its duration seem to fluctuate?

If it did seem to you that the quiet interval was changing, was there any pattern you could describe?

6. Estimate the average number of seconds that this quiet interval lasted:

7. How many signals do you think you heard during the last sequence?

APPENDIX B

TOWARD A MODEL FOR TEMPORAL EXPECTANCY

Given a series of intersignal intervals which are selected at random from some distribution of ISIs, how can an observer predict the onset of a signal? There are a number of internal models which might be adopted and which point to observably different outcomes.

The simplest internal models can lead to hopeless or inflexible strategies. For example, the subject might despair of predicting altogether, and react to each signal instead of anticipating its arrival. The subjects in the Expectancy Experiment clearly did not resort to such a simplistic strategy. Alternatively, the subject might adopt a rigid approach, always making the prediction after the same fixed interval on every trial and ignoring any discrepancy between the prediction and the actual signal arrival time. In fact, such a strategy would have served well in the constant ISI conditions. To use this strategy effectively in the other (varying ISI) conditions, the subject could avoid forfeiting an excessive number of trials only by selecting a fixed prediction interval less than or equal to the shortest duration in the ISI distribution. None of the subjects in this experiment demonstrated any inclination to settle with such brief prediction intervals and (even among those constant subjects who professed awareness that the ISI had a fixed duration) their predictions always retained considerable variance. Of course, there are limitations inherent in the timing skills demanded by the prediction task, namely time interval estimation and interval production, which must contribute to this variance. The psychophysical limitations would have prevented very stable prediction intervals even if the subjects had intended to supply them. Indeed, the constant ISI condition in this experiment provided a baseline indication of the psychophysical variability. While such perceptual-motor factors would have contributed something to the prediction variance in all ISI conditions, in the constant condition it was the only truly unavoidable source of variance. It is generally accepted that for durations ranging from about 4 to 30 seconds, such as those used in this experiment, an unpractised subject's attempts to reproduce the intervals would be expected to exhibit a standard deviation which is about 20% of the interval length (Woodrow, 1951). As Table 4-2 had indicated, the results for the nonloaded constant (C) condition were certainly compatible with this estimate, with the average standard deviation of the predictions (1.58 seconds) measuring close to 12% of the mean interval length (12.92). The proportion was very similar, approximately 14%, for the loaded constant (LC) group. Predictions in the other ISI conditions were more variable. In all of the other conditions, the standard deviation of the predictions exceeded 20% of the mean prediction interval. So, even allowing for some uncertainty in the variance estimates due to the trial elimination effect discussed in Chapter 4, it is fair to conclude that subjects in the constant ISI condition were operating respectably close to accepted psychophysical limits of timing accuracy, whatever the internal model underlying their expectancies.

there are limitations inherent in the timing skills demanded by the prediction task, namely time interval estimation and interval production, which must contribute to this variance. The psychophysical limitations would have prevented very stable prediction intervals even if the subjects had intended to supply them. Indeed, the constant ISI condition in this experiment provided a baseline indication of the psychophysical variability. While such perceptual-motor factors would have contributed something to the prediction variance in all ISI conditions, in the constant condition it was the only truly unavoidable source of variance. It is generally accepted that for durations ranging from about 4 to 30 seconds, such as those used in this experiment, an unpractised subject's attempts to reproduce the intervals would be expected to exhibit a standard deviation which is about 20% of the interval length (Woodrow, 1951). As Table 4-2 had indicated, the results for the nonloaded constant (C) condition were certainly compatible with this estimate, with the average standard deviation of the predictions (1.58 seconds) measuring close to 12% of the mean interval length (12.92). The proportion was very similar, approximately 14%, for the loaded constant (LC) group. Predictions in the other ISI conditions were more variable. In all of the other conditions, the standard deviation of the predictions exceeded 20% of the mean prediction interval. So, even allowing for some uncertainty in the variance estimates due to the trial elimination effect discussed in Chapter 4, it is fair to conclude that subjects in the constant ISI condition were operating respectably close to accepted psychophysical limits of timing accuracy, whatever the internal model underlying their expectancies.

Certainly with ISIs that continually change, the statistical properties of the ISI distribution could influence predictions in more meaningful ways. A number of dynamic or adaptive internal models are imaginable. Probably the simplest model would have the subject adopt an expectancy which simply copies the just-previous ISI. The possibility that subjects used a "simple copy" model was explored briefly by carrying out a computer simulation of the strategy. For each block of trials for each subject, an artificial series of predictions was generated by copying the ISI series one trial late. The constraint was added that the prediction thus generated could never exceed the ongoing trial duration. Any attempt to copy which resulted in a too-long prediction interval would be treated as a prediction failure or reaction response. This corresponds precisely to the way the real data were categorized. The number of successful simulated trials was then noted. The simulated data produced a great many more missed predictions than were actually observed in the original data. While nonloaded gaussian subjects had failed to predict an average of 24% of the trials, the simulation resulted in forced reaction outcomes for 56% of the trials. Similarly, while uniform subjects had been compelled to react 40% of the time, simulation indicated that if they had merely copied the previous ISI these subjects would have missed at least 62% of their predictions. It was concluded that subjects were unlikely to have been using such a simple copying strategy.

More realistically, the subject would depend on more sophisticated cues arising from the last trial's events. The most obvious source of information would be the error feedback obtained by the subject during the last prediction attempt. The error, as such, is difficult to analyse. When the subject makes genuine predictions the early errors are readily available (as the temporal difference between the prediction interval and the actual intersignal interval); but when the subject fails to predict before the target event, the late error in the expectancy is unknown (although it may be presumed to be a factor in the reaction latency). The error feedback experienced by the subject is a function of his momentary expectancy, the ongoing intersignal interval (that is, the temporal position of the target event), and psychophysical factors. In broad terms, the intersignal interval history is the source of the subject's information. Practical analysis of expectancy formation is therefore based on study of the effects of the ISI series. To propose an internal model which could be responsible for temporal expectancy it is necessary to describe how the subject's internal model abstracts useful assumptions about the statistical structure of the signal schedule from sequential observations as events unfold. Once the simplest, most rigid models have been discounted, any attempt at understanding expectancy must examine sequential dependencies within the data. Michon (1967) has pointed out that in prediction tasks of this sort, the model is most likely to be identified by studying the crosscorrelation function which relates the prediction time-series to the series defined by the signal schedule.

The autocorrelation of the prediction intervals (Figure 4-6) revealed cues arising from the last trial's events. The most obvious source of information would be the error feedback obtained by the subject during the last prediction attempt. The error, as such, is difficult to analyse. When the subject makes genuine predictions the early errors are readily available (as the temporal difference between the prediction interval and the actual intersignal interval); but when the subject fails to predict before the target event, the late error in the expectancy is unknown (although it may be presumed to be a factor in the reaction latency). The error feedback experienced by the subject is a function of his momentary expectancy, the ongoing intersignal interval (that is, the temporal position of the target event), and psychophysical factors. In broad terms, the intersignal interval history is the source of the subject's information. Practical analysis of expectancy formation is therefore based on study of the effects of the ISI series. To propose an internal model which could be responsible for temporal expectancy it is necessary to describe how the subject's internal model abstracts useful assumptions about the statistical structure of the signal schedule from sequential observations as events unfold. Once the simplest, most rigid models have been discounted, any attempt at understanding expectancy must examine sequential dependencies within the data. Michon (1967) has pointed out that in prediction tasks of this sort, the model is most likely to be identified by studying the crosscorrelation function which relates the prediction time-series to the series defined by the signal schedule.

The autocorrelation of the prediction intervals (Figure 4-6) revealed that in many cases the subjects' predictions were positively correlated over pairs of trials: subjects were trying to be consistent, tending not to produce a long prediction interval after a short one, for example. At the same time, the crosscorrelation of the predictions with prior ISIs confirmed that the subjects did continually adjust their predictions to suit the evidence of the intersignal intervals they had just witnessed. The description of the internal model which emerges, then, is of subjects trying to incorporate their perceptions of ISI behaviour into their expectancies, but tempering this long-term ISI-driven performance with a mistrust of very extreme amendments to their expectancy in the short-term.

The crosscorrelation functions (Figure 4-7) suggested that preceding ISIs could have provided information to build an internal model. Abandoning the simplest copying models, more than one ISI might be used as a computational base for expectancy: some weighted or unweighted average of several previous trials could provide the current prediction. Baker (1963) studied a temporal expectancy task and proposed that prediction could entail a running calculation based on an averaging of recent events to provide an estimator for upcoming events. In his formulation, an expectancy would be assigned a duration equivalent to the mean duration of the last few inter-event intervals which the subject had observed.

To test whether such a moving-average procedure could account for the predictions measured here, a correlation function may be plotted which compares predictions to I_N, the mean of the N preceding intersignal intervals. This is not a crosscorrelation analysis, because no time-shift has been imposed upon the data. The results comprise a set of correlation coefficients, produced by incrementing the size of the moving-average before calculating each coefficient.

Since the crosscorrelation analysis suggested that an ISI-dependent model would exhibit a preference for more recent experience, the basic moving-average formulation may be refined further to improve its chances of success. Assume that the CCFs of Figure 4-7 testify that subjects forget older experimental events so that intervals in the distant past are remembered rather less well than more recent durations. By *weighting* the ISI durations before calculating the moving-average it becomes possible to incorporate forgetting for older trials. That is, the correlation of prediction with a weighted moving-average of prior ISIs may be plotted as a function of N (the moving-average size). Taking the general form of a weighted moving-average, the ISI statistic would be

$$I_{N\text{-}weighted} = \frac{1}{\sum_{t=1}^{N} w(t)} \cdot \sum_{t=1}^{N} (I_t \cdot w(t))$$

where $w(t)$ is a weighting coefficient dependent on recency. This differs from a straightforward moving-mean calculation in that each ISI is weighted according to its position prior to the prediction. (In keeping with the notation used in Chapter 4, prior position has been denoted by t so that $t=1$ indicates the first ISI prior to prediction, $t=2$ points to the second previous ISI, and so on. The degree of forgetting is controlled by the weighting coefficient $w(t)$. Much forgetting corresponds to relatively low weight. The rate at which past trials are forgotten depends on the kind of decay function we choose for $w(t)$. A useful decay function is one of the form $w(t)=t^{-c}$ where c is a constant. This states that the weight assigned to an ISI is an exponential function of that ISI's position preceding the trial. For $c=0$ we obtain $w(t)=1$ for all values of t and we have an unweighted mean. For $c>0$ we obtain $w(t)<1$ when t extends back more than one trial. Because the rate of weight decrease is nonlinear, the severity of forgetting accelerates as t gets larger.

Following the analysis of the prediction experiment, several unweighted and weighted averaging models of these sorts were applied to the data. The most encouraging results were obtained when the weighted moving-average process just described was used to generate correlation functions comparing predictions to $I_{N\,weighted}$ for N ranging from 1 to 5. The ISIs preceding each prediction were weighted according to the exponential decay law. In order to test a suitable range of decay rates for each subject, a computer program was written to increment the constant of exponentiation c in orderly fashion. Following each adjustment to c the weighting coefficients for each I_t were computed and the statistic $I_{N\,weighted}$ was recalculated to provide a new set of correlation coefficients.

The correlation functions thus generated for a particular subject were examined to identify the best outcome and so specify the optimum decay constant for that subject. The correlation function was required to display a peak for some moving-average size beyond *N=1* if the averaging process was to be considered useful. The *best-peak* correlation function was identified as that containing the largest positive correlation coefficient out of all the coefficients computed. This optimum correlation function and the value of *c* which produced it were considered to comprise the best possible test of the moving-average procedure for a particular subject.

Why the "best-peak" correlation?

The optimum size of *N* for forming the moving-average I_N is not apparent *a priori*, although in his study of temporal extrapolation Baker (1963) concluded that probably no more than five of the previous intervals were implicated. Since the Expectancy Experiment of Chapter 4 reported that the crosscorrelation of predictions with prior ISIs was not identical for the gaussian and uniform conditions used there, it also seems likely that the choice of ISI distribution would affect the size of *N*. One logical method for determining the most suitable moving-average size would be to calculate the correlation function over a range of values of *N* and to note where the correlation ceases to benefit from increasing the size of the moving-average. The fall-off would mark the upper limit on the number of prior ISIs that might be contributing to the model's estimator.

This reasoning does not go far enough. Other factors could restrict the value of the correlation coefficient as N increases in size. Consider first a simple unweighted average of several ISIs. The ISIs were selected by random sampling; according to the Central Limit Theorem the mean of a subset of ISIs will more closely approximate the mean of the parent distribution of ISIs as the subset takes in more samples. It is obvious that as N increases, the value of the moving-average will converge on the overall mean of the ISI series, so the correlation will eventually be comparing the expectancies with a near-constant I_N equal to the ISI distribution mean. Since the correlation of a fluctuating quantity with a constant is necessarily zero, we may expect that the correlation coefficient considered here must tend toward zero for larger values of N simply because of the statistical properties of the ISI series feeding the moving-average procedure. But if the moving-average formulation is indeed appropriate we would expect to observe a positive-signed *peaking* of the correlation coefficient for some I_N when $N>1$. This peaking would confirm that an average of several prior ISIs is actually a better predictor of the expectancy outcome than any single ISI preceding the expectancy, and it would suggest a specific way in which the internal model pools and remembers data from past experience. Considering next a weighted moving-average, there is a similar tendency for $I_{N\mathrm{weighted}}$ toconverge on a stable value as N increases, and this would likewise introduce an artifact into the correlation function. Here $I_{N\mathrm{weighted}}$ will not actually approach the population mean but instead, if c is very large and thus weighting is very slight even for relatively recent ISIs, all I_e will be be very similar to I_1, so all the moving-averages will come to resemble the ISI immediately preceding prediction. This must clearly be the case, since the weighting of I_1 will be $t^{-c} = 1^{-c} = 1$ for any value of c; while the earlier trials will be weighted negligibly because as c gets large the limit of t^{-c} is zero for all $t>1$. In other words, if forgetting is very severe (c is big), a weighted moving-average of the last five ISIs will not differ substantially from the value of the last one ISI because the contribution of historical events will be weighted so slightly. In examining the correlation function under these circumstances, we would expect that the correlation attained when $N=1$ would be maintained at much the same level even when $N>1$. Again, the confident interpretation of the analysis must rely on the detection of a peak positive value. If a clear peak is found for $N>1$ then there is evidence that a moving-average incorporating forgetting-- even if the forgetting of older trials is relatively severe-- provides a better predictor of the subject's expectancy than any single prior ISI.

To summarize: increasing the size of the moving-average unlimitedly should either force the correlation coefficient toward zero, in the case of an unweighted average; or hold the correlation to a constant value, in the case of a weighted average. The only explanation for a peak in the correlation function in either circumstance is that a moving-average of particular size is an unusually appropriate predictor of the expectancy outcome.

Fitting the model to the data

The best-peak correlation analysis was applied to the data from each subject in the Expectancy Experiment. Nearly all of the correlation functions displayed a maximum for a moving-average containing two or more previous ISIs, confirming that the weighted pooling of historical data offered some predictive benefit. The advantage gained by using $I_{Nweighted}$ in comparison to using I_1 alone was small, but some increase in correlation was demonstrated by practically all subjects. Moreover, the location of the correlation maximum was fairly consistent across subjects, at least within the gaussian ISI conditions. The correlation functions shown in Figure B-1 are the average best-peak outcomes for the subjects in each condition. In the gaussian results, the position of the peak suggested that, on average, three ISIs could be usefully contributing to gaussian group expectancies regardless of loading. In the uniform conditions the peaking was not as consistent from case to case and the group outcomes were not alike: when nonloaded the peak occurred with two ISIs, but when loaded the largest correlation corresponded to a moving-average of five ISIs. However, statistical analysis suggested that the loading treatment did not really make a difference in either ISI category. For each of the ISI conditions, the nonloaded and loaded correlation outcomes were compared using analysis of variance to determine the effects of *Loading* and the *Size of the moving-average* on the correlation coefficients. The ANOVAs confirmed that loading had no overall effect on the magnitude of correlation in either ISI condition. But the number of trials incorporated into $I_{Nweighted}$

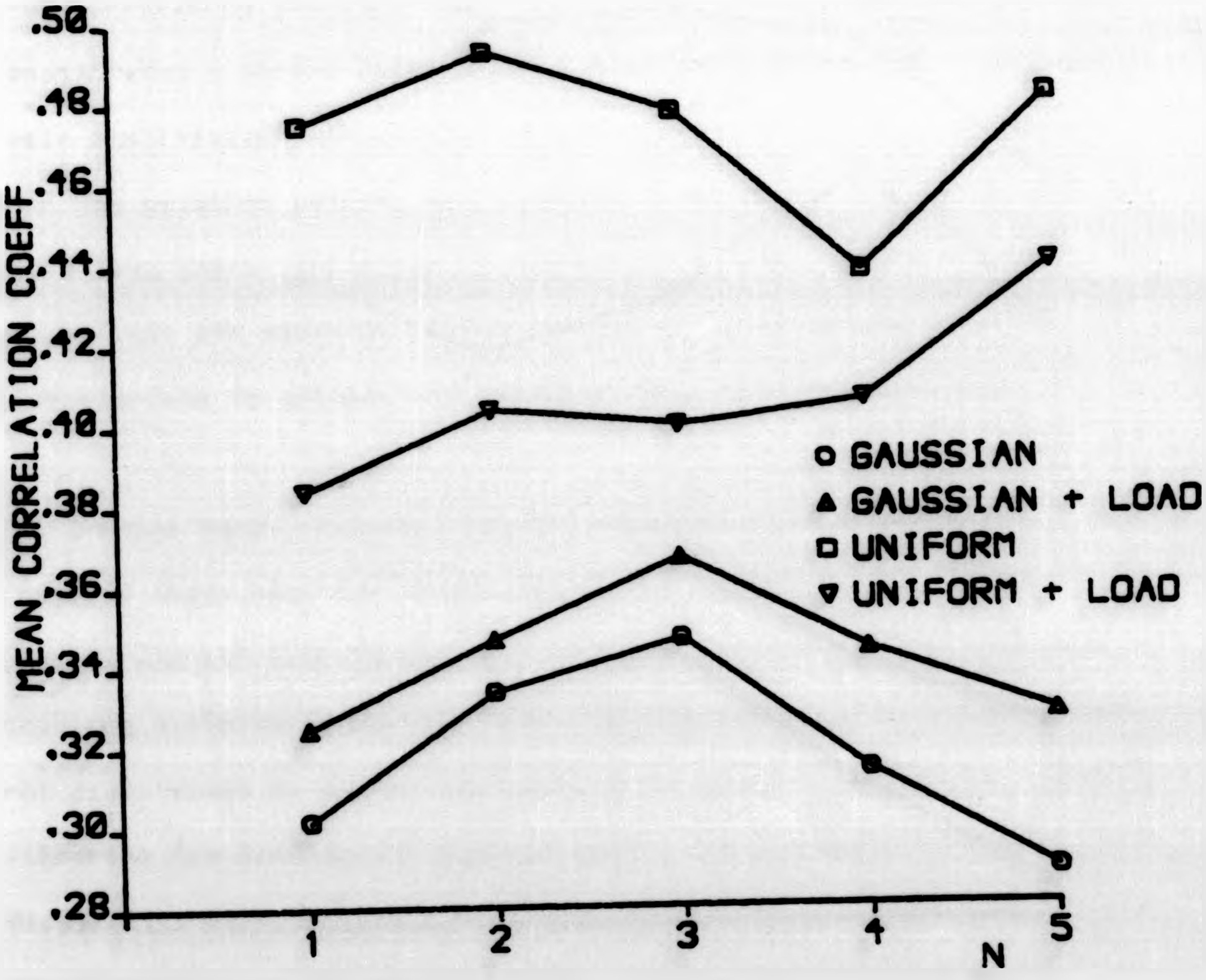

FIGURE B-1 Expectancy Experiment. Correlation functions showing optimum exponential decay model for expectancy.

definitely had significant effects in both the gaussian (F_{4df}=5.87, p<.001) and uniform (F_{4df}=6.64, p<.001) results. The ANOVA summaries are found in Table B-1. The uniform analysis also found a significant Loading X Size interaction (F_{4df}=5.1, p<.01), reflecting the disparity between the nonloaded and loaded correlation functions which had been noted above. Certainly when compared to the gaussian outcome, the uniform moving-average result is not so satisfactory. The earlier crosscorrelation analysis $r_{rI}(t)$ had shown that the uniform subjects demonstrated a strong dependence on only one previous ISI. Because of their distribution the uniform ISIs exhibited more local variability than the gaussian series, so in objective terms the recent ISI history was a less useful indicator of expectancy to the uniform subjects. It may be that the weighted moving-average model described here is inappropriate to account for the expectancy of these subjects.

The average decay constants for the forgetting functions which provided these best-fit models are listed in Table B-2. Again, loading was not seen to have any reliable effect: t-tests confirmed that the differences between nonloaded and loaded decay constants were not significant in either the gaussian or uniform conditions. Overall (ignoring the distinction between loaded and nonloaded subjects), the decay rates were clearly more severe in the uniform model than in the gaussian (t_{30df}=4.19, p<.01), supporting the conclusion that the uniform subjects ignored (or forgot) past intersignal timing more readily than the gaussian subjects.

TABLE B-1

EXPECTANCY EXPERIMENT.
ANOVA Summary. Effects of loading and weighted-average size on the optimum correlation of prediction with $I_{unweighted}$.

Gaussian ISI conditions.

Source	*df*	*SS*	*MS*	*F*	*Prob*
Between subjects					
Loading (L)	1	0.003	0.003	0.0192	0.8865
Subjects	19	2.355			
Within subjects					
Size of average (A)	4	0.029	0.007	5.8712	0.0004 *
L X A	4	0.002	0.001	0.4910	0.7450
Error Within	72	0.089	0.001		

Uniform ISI conditions.

Source	*df*	*SS*	*MS*	*F*	*Prob*
Between subjects					
Loading (L)	1	0.110	0.110	0.7908	0.6107
Subjects	19	2.622			
Within subjects					
Size of average (A)	4	0.018	0.005	6.6438	0.0002 *
L X A	4	0.014	0.004	5.0978	0.0012 *
Error Within	72	0.050	0.001		

* significant at $p<0.05$

TABLE B-2

EXPECTANCY EXPERIMENT.
Values of c, the decay constant, to yield optimum correlation of prediction with $I_{unweighted}$.

	ISI condition			
	G	U	LG	LU
mean value of c	1.50	3.62	0.86	2.96

REFERENCES

Baker CH (1962). On temporal extrapolation. *Canadian Journal of Psychology, 16*, 37-41.

Baker CH (1963). Further toward a theory of vigilance. In DN Buckner and JJ McGrath (eds), *Vigilance: a symposium*. (New York: McGraw-Hill).

Bartlett N and Bartlett S (1959). Synchronization of a motor response with an anticipated sensory event. *Psychological Review, 66*(4), 203-218.

Berlyne DE (1960). *Conflict, arousal, and curiosity*. (New York: McGraw-Hill).

Block RA (1979). Time and consciousness. In G Underwood and R Stevens (eds), *Aspects of consciousness, vol 1* (London: Academic Press).

Boggs DH and Simon JR (1968). Differential effect of noise on tasks of varying complexity. *Journal of Applied Psychology, 52*(2), 148-153.

Broadbent DE (1957). Effects of noises of high and low frequency on behaviour. *Ergonomics, 1*, 21-29.

Broadbent DE (1958). *Perception and communication* (London: Pergamon).

Broadbent DE (1971). *Decision and stress* (London: Academic Press).

Broadbent DE (1979). Human performance and noise. In CM Harris (ed), *Handbook of noise control, second edition* (New York: McGraw-Hill).

Broadbent DE (1983). Recent advances in understanding performance in noise. In G Rossi (ed), *Proceedings of the Fourth International Congress on Noise as a Public Health Problem* (Milan: Centro Richerche e Studi Amplifon).

Burns W (1968). *Noise and man* (London: John Murray).

Cassel EE and Dallenbach KM (1918). The effect of auditory distraction upon the sensory reaction. *American Journal of Psychology, 29*, 129-143.

Cohen A (1969). Effects of noise on psychological state. In WD Ward and JE Fricke (eds), *Noise as a public health hazard, ASHA Report 4* (Washington, DC: American Speech and Hearing Association).

Cohen S and Weinstein N (1983). Nonauditory effects of noise on behavior and health. In GW Evans (ed), *Environmental stress* (London: Cambridge University Press).

Conrad DW (1973). The effects of intermittent noise on human serial decoding performance and physiological response. *Ergonomics, 16*(6), 739-747.

Davies DR and Parasuraman R (1981). *The psychology of vigilance* (London: Academic Press).

Dimond SJ (1965). Storage of information about time. *Perceptual and Motor Skills, 21*, 261-262.

Elkind JI (1956). Characteristics of simple manual control systems. *Lincoln Laboratory Technical Report 111*, Massachusetts Institute of Technology.

Eschenbrenner AJ (1971). Effects of intermittent noise on the performance of a complex psychomotor task. *Human Factors, 13*(1), 59-63.

Finkelman J and Glass D (1970). Re-appraisal of the relationship between noise and human performance by means of a subsidiary task measure. *Journal of Applied Psychology, 54*(3), 211-213.

Fisher S (1972). A 'distraction' effect of noise bursts. *Perception, 1*, 223-236.

Fisher S (1973). The 'distraction effect' and information processing complexity. *Perception, 2*, 79-89.

Fisher S (1983). 'Pessimistic noise effects': the perception of reaction times in noise. *Canadian Journal of Psychology, 37*(2), 258-271.

Fisher S (1984a). *Stress and the perception of control.* (London: Lawrence Erlbaum Associates).

Fisher S (1984b). The microstructure of attentional deployment on a dual task in loud noise. *Canadian Journal of Psychology, 38*(4), 561-578.

Fisher S (1996). *Stress and strategy.* (London: Lawrence Erlbaum Associates).

Flowers KA (1978). The predictive control of behaviour: appropriate and inappropriate actions beyond the input in a tracking task. *Ergonomics, 21*(2), 109-122.

Fogel LJ (1955). A note on the sampling theorem. *IRE Transactions on Information Theory*, IT-12, 47-48.

Gawron V (1982). Performance effects of noise intensity, psychological set, and task type and complexity. *Human Factors, 24*(2), 225-243.

Graham FK (1979). Distinguishing among orienting, defense, and startle reflexes. In H Kimmel, E Van Holst, J Orlebeke (eds), *The orienting reflex in humans* (Hillsdale, New Jersey: Lawrence Erlbaum Associates).

Hack J, Robinson H, and Lathrop R (1965). Auditory distraction and compensatory tracking. *Perceptual and Motor Skills, 20*, 228-230.

Hakkinen MT and Williges BH (1984). Synthesized warning messages: effects of an alerting cue in single- and multiple-function voice synthesis systems. *Human factors, 26*(2), 185-195.

Harmon FL (1933). The effects of noise upon certain psychological and physiological processes. *Archives of Psychology, 23*, no. 147, 5-81.

Hartley L (1981). Noise, attentional selectivity, serial reactions and the need for experimental power. *British Journal of Psychology, 72*, 101-107.

Hess RA (1973). Nonadjectival rating scales in human response experiments. *Human Factors, 15*(3), 275-280.

Hockey GR (1970a). Effect of loud noise on attentional selectivity. *Quarterly Journal of Experimental Psychology, 22*, 28-36.

Hockey GR (1970b). Signal probability and spatial location as possible bases for increased selectivity in noise. *Quarterly Journal of Experimental Psychology, 22*, 37-42.

Hockey GR (1984). Varieties of attentional state: the effects of environment. In R Parasuraman and D Davies (eds), *Varieties of attention* (Orlando, Florida: Academic Press).

Holahan CJ, Culler RE, and Wilcox BL (1978). Effects of visual distraction on reaction time in a simulated traffic environment. *Human Factors, 20*(4), 409-413.

Hovey HB (1928). Effects of general distraction on the higher thought processes. *American Journal of Psychology, 40*, 585-591.

Howell WC and Briggs GE (1959). Effects of visual noise and locus of perturbation on tracking performance. *Journal of Experimental Psychology, 58*, 166-173.

Hulstijn W (1979). Selective attention and the orienting reflex. In H Kimmel, E Van Holst, J Orlebeke (eds), *The orienting reflex in humans* (Hillsdale, New Jersey: Lawrence Erlbaum Associates).

James W (1890). *Principles of psychology* (New York: Henry Holt).

Jex HR and Allen RW (1970). Research on a new human dynamic response test battery. *Proceedings of the Sixth Annual Conference on Manual Control*, 743-777.

Jex HR and Clement WF (1979). Defining and measuring perceptual-motor workload in manual control tasks. In N Moray (ed), *Mental workload: its theory and measurement.* (New York: Plenum Press).

Johannsen G, Moray N, Pow R, Rasmussen J, Sanders A, and Wickens C (1979). Final report of experimental psychology group. In N Moray (ed), *Mental workload: its theory and measurement.* (New York: Plenum Press).

Johannsen G, Pfendler C, and Stein W (1976). Human performance and workload in simulated landing-approaches with autopilot failures. In T Sheridan and G Johannsen (eds), *Monitoring behavior and supervisory control.* (New York: Plenum Press).

Johnson AW and Cole BL (1976). Investigations of distraction by irrelevant information. *Australian Road Research, 6*(3), 3-23.

Johnston WA, Greenburg S, Fisher R, and Martin D (1970). Divided attention: a vehicle for monitoring memory processes. *Journal Of Experimental Psychology,* 83, 164-171.

Kahneman D (1973). *Attention and effort* (Englewood Cliffs, New Jersey: Prentice-Hall).

Kalsbeek JWH (1964). On the measurement of deterioration in performance caused by distraction stress. *Ergonomics, 7*(2), 187-195.

Karlin L (1959). Reaction time as a function of foreperiod duration and variability. *Journal of Experimental Psychology, 58*(2), 185-191.

Kelley CR (1968). *Manual and automatic control.* (New York: Wiley).

Kelley CR (1969a). The measurement of tracking proficiency. *Human Factors, 11,* 43-64.

Kelley CR (1969b). What is adaptive training? *Human Factors, 11,* 547-556.

Kelley CR (1971). Fundamental problems. In JJ McGrath and DH Harris (eds),Adaptive training. *Aviation Research Monographs, 1*(2). University of Illinois at Urbana-Champaign.

Klemmer ET (1956). Time uncertainty in simple reaction time. *Journal of Experimental Psychology, 51*(3), 179-184.

Klemmer ET (1957). Simple reaction time as a function of time uncertainty. *Journal of Experimental Psychology,* 54(3), 195-200.

Kryter KD (1950). The effects of noise on man. *Journal of Speech and Hearing Disorders*, Monograph Supplement 1.

Kryter KD (1970). *The effects of noise on man.* (New York: Academic Press).

Lynn R (1966). *Attention, arousal, and the orientation reaction.* (Oxford: Pergamon).

Lynn PA (1973). *An introduction to the analysis and processing of signals.* (London: Macmillan).

McGrath JJ and Harris DH (1971). Adaptive training. *Aviation Research Monographs 1*(2), University of Illinois at Urbana-Champaign.

Mackie RR (ed). (1977). *Vigilance: theory, operational performance, and physiological correlates.* (New York: Plenum Press).

Michon JA (1966). Tapping regularity as a measure of perceptual motor load. *Ergonomics, 9*, 401-412.

Michon JA (1967). *Timing in temporal tracking.* (Soesterberg, the Netherlands: Institute for Perception RVO-TNO).

Miller G (1948). The perception of short bursts of noise. *Journal of the Acoustical Society of America, 20*, 160-170.

Moray N (1959). Attention in dichotic listening: affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology, 9*, 56-60.

Moray N (1969). *Attention: selective processes in vision and hearing.* (London: Hutchinson Educational).

Moray N (1976). Attention, control, and sampling behaviour. In T Sheridan and G Johannsen (eds), *Monitoring behavior and supervisory control.* (New York: Plenum Press).

Moray N (1978). The strategic control of information processing. In G Underwood (ed), *Strategies of information processing* (London: Academic Press)

Moray N (ed). (1979a). *Mental workload: its theory and measurement.* (New York: Plenum Press).

Moray N (1979b). Models and measures of mental workload. In N Moray (ed), *Mental workload: its theory and measurement.* (New York: Plenum Press).

Moray N (1981). Feedback and the control of skilled behaviour. In D Holden (ed), *Human skills.* (Chichester: Wiley).

Moray N (1984). Attention to dynamic visual displays in man-machine systems. In R Parasuraman and D Davies (eds), *Varieties of attention* (Orlando, Florida: Academic Press).

Moray N and Fitter M (1973). A theory and the measurement of attention. *Attention and Performance IV*, 3-19.

Moray N, Fitter M, Ostry D, Favreau D, and Nagy V (1976). Attention to pure tones. *Quarterly Journal of Experimental Psychology, 28*, 271-283.

Morgan JJB (1916). The overcoming of distraction and other resistances. *Archives of Psychology*, no. 35.

Mowrer OH (1940). Preparatory set (expectancy): some methods of measurement. *Psychological Monographs, 52*(2), whole no 233.

Myers CS (1909). *A text-book of experemental psychology.* (London: Edward Arnold).

Naatanen R and Merisalo A (1977). Expectancy and preparation in simple reaction time. In S Dornic (ed), *Attention and Performance VI.* (Hillsdale, New Jersey: Lawrence Erlbaum Associates).

Naatanen R and Summala H (1976). *Road-user behavior and traffic accidents.* (Amsterdam: North-Holland).

Naatanen R, Muranen V, and Merisalo A (1974). The timing of expectancy peak in simple reaction time situation. *Acta Psychologica, 38*, 461-470.

Ostry D, Moray N, and Marks G (1976). Attention, practice, and semantic targets. *Journal of Experimental Psychology: Human Perception and Performance*, 2(3), 326-336.

Patterson RD (1982). *Guidelines for auditory warnings on civil aircraft* (London: Civil Aviation Authority).

Peterson CR and Beach LR (1967). Man as an intuitive statistician. *Psychological Bulletin, 68*(1), 29-46.

Pew R (1974). Human perceptual-motor performance. In B Kantowitz (ed), *Human information processing.* (Hillsdale, New Jersey: Lawrence Erlbaum Associates).

Pew R (1979). Secondary tasks and workload measurement. In N Moray (ed), *Mental workload: its theory and measurement.* (New York: Plenum Press).

Philipp U, Reiche D, and Kirchner J-H (1971). The use of subjective rating. *Ergonomics, 14*, 611-616.

Plutchik R (1959). The effects of high intensity intermittent sound on performance, feeling, and physiology. *Psychological Bulletin, 56*, 133-151.

Plutchik R (1961). Effect of high-intensity intermittent sound on compensatory tracking and mirror tracing. *Perceptual and Motor Skills, 12*, 187-194.

Pollock KG and Bartlett FC (1932). Two studies in the psychological effects of noise. Part I: Psychological experiments on the effects of noise. *The Industrial Health Research Board, London (HMSO)*, Report no. 65.

Poulton EC (1974). *Tracking skill and manual control.* (New York: Academic Press).

Poulton EC (1978). A new look at the effects of noise: a rejoinder. *Psychological Bulletin*, 85(5), 1068-1079.

Poulton EC (1979). Composite model for human performance in continuous noise. *Psychological Review*, 86(4), 361-375.

Rasmussen J and Rouse W (eds). (1981). *Human detection and diagnosis of system failures.* (New York: Plenum Press).

Rolfe JM (1973). The secondary task as a measure of mental load. In WT Singleton, J Fox, and D Whitfield (eds), *Measurement of man at work* (London: Taylor and Francis).

Rossi G (ed). (1983). *Proceedings of the Fourth International Congress on Noise as a Public Health Problem* (Milan: Centro Richerche e Studi Amplifon).

Sanders AF (1966). Expectancy: application and measurement. *Acta Psychologica*, 25, 293-313.

Sanders AF (1979). Some remarks on mental load. In N Moray (ed), *Mental workload: its theory and measurement.* (New York: Plenum Press).

Schmidt RA (1968). Anticipation and timing in human motor performance. *Psychological Bulletin*, 70 (6), 631-646.

Senders JW (1964). The human operator as a monitor and controller of multidegree of freedom systems. *IEEE Transactions on Human Factors in Electronics*, HFE-5, 2-5.

Sheridan TB and Ferrell WR (1974). *Man-machine systems: information, control, and decision models of human performance.* (Cambridge, Mass: MIT Press).

Sheridan TB and Johannsen G (eds) (1976). *Monitoring behavior and supervisory control* (New York: Plenum Press).

Simpson CA and Williams DH (1980). Response time effects of alerting tone and semantic context for synthesized voice cockpit warnings. *Human Factors*, 22(3), 319-330.

Smith AP (1985). Noise, biased probability and serial reaction. *British Journal of Psychology*, 76, 89-95.

Smith KR (1951). Intermittent loud noise and mental performance. *Science, 114*, 132-133.

Taylor A and Moray N (1960). Statistical approximations to English and French. *Language and Speech,* 3, 7-10.

Theologus GC, Wheaton GR, Fleishman EA (1974). Effects of intermittent, moderate intensity noise stress on human performance. *Journal of Applied Psychology, 59*, 539-547.

Trumbo D and Milone F (1971). Primary task performance as a function of encoding, retention, and recall in a secondary task. *Journal of Experimental Psychology, 91*, 273-279.

Trumbo D, Noble M, and Swink J (1967). Secondary task interference in the performance of tracking tasks. *Journal of Experimental Psychology, 73*, 232-240.

Vroon PA and Vroon AG (1973). Tapping rate and expectancy in simple reaction time tasks. *Journal of Experimental Psychology, 98*, 85-90.

Wickens CD (1979). Measures of workload, stress, and secondary tasks. In N Moray (ed) *Mental workload: its theory and measurement.* (New York: Plenum Press).

Wiener EL (1973). Adaptive measurement of vigilance decrement. *Ergonomics, 16*, 353-63.

Wiener EL (1985). Beyond the sterile cockpit. *Human Factors, 27*(1), 75-90.

Williges RC and Williges BH (1978). Critical variables in adaptive motor skills training. *Human Factors, 20*(2), 201-214.

Woodhead M (1966). An effect of noise on the distribution of attention. *Journal of Applied Psychology, 50*(4), 296-299.

Woodrow H (1951). Time perception. In SS Stevens (ed), *Handbook of experimental psychology* (New York: Wiley) pp 1224-1236.

Woodworth RS and Schlosberg H (1954). *Experimental psychology* (London: Methuen).