THE UNIVERSITY OF HULL

THEORETICAL AND PRACTICAL CONSIDERATIONS IN THE DESIGN OF SPECIAL-PURPOSE KEYBOARDS AND OPERATOR TRAINING PROCEDURES

being a Thesis submitted for the degree of

DOCTOR OF PHILOSOPHY

in the University of Hull

by

Olga Feggou (BA, The American College of Greece, MSc)

February, 1988

THEORETICAL AND PRACTICAL CONSIDERATIONS IN THE DESIGN OF SPECIAL-PURPOSE KEYBOARDS AND OPERATOR TRAINING PROCEDURES

by

Olga Feggou (B.A., M.Sc.)

SUMMARY

This thesis deals with theoretical and practical issues of special-purpose keyboard design and operator training precedures. It is in two parts.

The research discussed in Part A was carried out in support of two contracts from the Post Office and had two main objectives: first, to provide information needed for the selection of a new keyboard design for use on foreign destination coding desks and second, to evaluate one component of a computer-based training system for operators of the new keyboard.

In relation to the first objective, four keyboard designs were evaluated, namely the 2x6, 4x4, 10-key chord and 10-key sequential. On the basis of the results of three experiments, the 10-key chord keyboard was chosen and recommended to the Post Office as the most efficient of the four layouts.

The computer-based training system for operators of foreign destination letter coding desks consists of three stages. In this thesis only the first such stage is discusssed, namely the Keyboard Familiarization, which involves developing the skill of keying chord patterns. Four different computer-based training methods are presented, discussed and evaluated in an experiment

specifically designed for this purpose.

The main issue of Part B is speed-accuracy tradeoff relationships in a keyboard task. Following a discussion of the limitations of typical Choice Reaction Time models and the benefits of studying speed-accuracy tradeoff relationships, another analysis of the data of Experiment 4 is presented. In this analysis, the speed and accuracy of any given response are examined in terms of those of the preceding response.

These findings are discussed in the light of established theories and it is concluded that no known theory can fully account for the obtained results. A new theoretical model is presented to explain them. Finally, this new model is assessed and its implications for keyboard operation and operator training are considered.

This thesis is gratefully dedicated to my parents Constantine and Irini Feggou for their financial and moral support throughout my education and for much much more...

CONTENTS

Page

ACKNOWLEDGEMENTS i

LIST OF TABLES ii

LIST OF FIGURES vi

PREFACE xii

PART A

Chapter 1: Special-Purpose Keyboard Designs 1

1.1 Definition of the Problem 1

1.2 Theoretical Background 2

1.3 Experimental Research on Keyboards 6

1.4 History of Chord Keyboards 12

1.5 Comparison between Sequential and Chord Keyboards 17

1.6 Recent Research on Chord Keyboards 19

1.7 Ergonomic Considerations in Keyboard Design 22

Chapter 2: Pilot Examination of Experimental Keyboards 31

2.1 Description of the Four Keyboards 31

2.2 The Need for a Pilot Examination 34

2.3 Displays 34

2.4 Apparatus 35

2.5 Experiment 1: A Pilot Investigation of the 10-Key Chord Keyboard 39

2.6 Experiment 2: A Pilot Investigation of the 10-Key Sequential Keyboard .. 56

Chapter 3: Evaluation and Comparison of Four Keyboard Designs 68

Experiment 3

3.1 Introduction 68

3.2 Method 69

3.3 Results 75

3.4 Conclusions and Recommendations 98

Chapter 4: Design of an Efficient Training Method for Operators of the 10-Key Chord Keyboard 100

Experiment 4

4.1 Introduction 100

4.2 Method 123

4.3 Results 130

4.4 Discussions and Conclusions 142

PART B

Chapter 5: Examination of Speed-Accuracy Relationships 145

5.1 Limitations of CRT Models 145

5.2 The Study of Speed-Accuracy Tradeoff 153

Chapter 6: Speed-Accuracy Relationships: A New Analysis of the Data of Experiment 4 159

6.1 Method of Analysis 159

6.2 Results of the Fixed-Random Mode of Training 163

6.3 Results of the Fixed-Adaptive Mode of Training 179

6.4 Results of the Cumulative-Random Mode of Training 196

6.5 Results of the Cumulative-Adaptive Mode of Training 207

6.6 Summary of Speed-Accuracy Variations Across Training Modes 219

Chapter 7: Discussion of Main Findings 226

7.1 Speed of Correct and Error Responses 226

7.2 The Effect of Accuracy of the Preceding Response on the Speed and Accuracy of the Current Response 230

7.3 Prediction of Speed and Accuracy 238

Chapter 8: Speed-Accuracy Tradeoff Theories 240

8.1 Required Characteristics of a Speed-Accuracy Model 240

8.2 Signal Detection Theory 241

8.3 The Fast Guess Model 247

8.4 The Random Walk Model 250

8.5 The Accumulator Model 256

8.6 The Tracking Model 260

8.7 Conclusions 264

Chapter 9: Further Examination of Subject Responses 266

9.1 RT Distribution Across Subjects 266

9.2 Classification of Error Responses 272

Chapter 10: Presentation of New Model 285

10.1 Probability Functions 285

10.2 Probability of a Response Being Correct if it is Made 293

10.3 Probability of a Response Being Made 303

10.4 Choice of the Cutoff Point 307

10.5 The Occurrence of an Error 312

Chapter 11: Assessment of the New Model 319

11.1 Basic Characteristics of the New Model 319

11.2 How It Accounts for the Main Speed-Accuracy Relationships 321

11.3 Limitations of the New Model 326

Chapter 12: General Conclusions and Recommendations 328

12.1 Introduction of the 10-Key Chord Keyboard .. 328

12.2 Operator Training Procedures 331

12.3 The Study of Speed-Accuracy Relationships .. 332

12.4 The Development of a New Theoretical Model . 336

References 339

Additional Bibliography 358

Appendix 1 363

Appendix 2 365

Appendix 3 379

Appendix 4 394

ACKNOWLEDGEMENTS

I would like to thank Dr. Dave Bartram for his consistent supervision, his challenging comments on successive drafts, his encouragement and patience. I also want to thank Dr. Bartram for writing the computer programs without which this reasearch would not have been possible.

Very special thanks go to Dr. Martin Crawshaw for his invaluable help and support throughout my research as well as for his incisive comments on the penultimate draft of this thesis.

Special mention is due to Ms. Margaret Lee who collected the data of Experiment 1 and to Mr. Johann Hasenauer with whom I carried out Experiment 4.

I would like to thank Mr. Richard Phillips for his computer advice and his assistance every time I was stuck with a computer or with ... finding the right word.

Thanks are also due to Mr. Ray Wallis for all his assistance every time the keyboards broke down.

I am indebted to Mrs. Linda Randerson for all her kindness and for typing a few of my references. I am also grateful to Mrs. Beverley Leak for her friendship and for allowing me to use her typewriter. Special thanks are due to Mr. Mark Youle for his help with the printing and all the paper shuffling.

A very big "thank you" goes to the whole of the Psychology Department in Hull for everything they gave to me, from equipment to hospitality.

I want to thank all my friends and colleagues in England, Greece and the U.S. for their encouragement all these years. Finally I want to thank my brother Spyros for his understanding, kindness and support.

LIST OF TABLES

1.1 Maximum Tapping Rate (Over a 15 secs Period) 28

1.2 Maximum Finger Pushing Force (lbs) 28

2.1 Mean Reaction Time (RT), Standard Deviation (SD) and Percentage of Errors (ERR) for each hand for each chord pattern. Also shown are the rank orders obtained by Ratz and Ritchie, 1961 (R&R) and Seibel, 1962, together with Seibel's RTs and error rates (all RTs in secs) 47

2.2 Correlation matrix among the variables
1: Mean RT for the left hand
2: Mean percentage of errors made by the left hand
3: Mean RT for the right hand
4: Mean percentage of errors made by the right hand
5: Rank order of stimuli by Ratz and Ritchie, 1961
6: Rank order of stimuli by Seibel, 1962
7: Mean RT from Seibel's (1962) data
8: Average error rate from Seibel's (1962) data
9: Number of fingers required for the appropriate response
10: Finger separation 48

2.3 Comparisons between left-hand and right-hand chords as well as between one-handed and two-handed chords (N=number of trials, RT=reaction time, SD=standard deviation, ERR=percentage of errors; RT in secs) 49

2.4 Relationship between the number of fingers required for the depression of a stimulus and mean RT and percentage of errors from Experiment 1 and Seibel's (1962) data (RT in secs) 50

2.5 Relationship between finger separation (the numbers denote the degree of finger separation) and mean RT and percentage of errors from the data of Experiment 1 and Seibel's (1962) data (RT in secs) 51

2.6 Information content and number of keys per response for constrained and unconstrained conditions for the 10-key sequential keyboard 57

2.7 Mean RTs (secs) and error rates for all fingers as a function of practice and ordinal position of the keystroke on the digit sequence 60

3.1 Subjects' mean reaction time (RT), number of errors (N of Ers) and adjusted scores (ASC) on the pretest (RT in csecs) 72

3.2 Assignment of subjects to matched groups on the basis of their ASC on the pretest 72

3.3 Experimental Design 73

3.4 Mean reaction time in csecs (RT), percentage of errors (ERR) and adjusted scores (ASC) for the four groups on the pretest 76

3.5 Correlation coefficients between residual adjusted scores on the pretest (across subjects) and residual adjusted scores during the ten training sessions 76

3.6 Mean adjusted scores (ASC), reaction time in secs (RT) and percentage of errors (ERR) for each group, averaged across subjects (AF=Auditory Feedback, VF=Visual Feedback) 77

3.7 Intercorrelations of mean RTs during the ten blocks of sessions 81

3.8 Intercorrelations of mean percentage of errors during the ten blocks of sessions 81

3.9 Intercorrelations of adjusted scores (ASC) during the ten blocks of sessions 82

3.10 Intercorrelations of the residuals of the adjusted scores during the ten blocks of sessions 82

3.11 Differences in mean RTs and percentage of errors between key sequences as a function of key repetition, adjacency and reversals (point-biserial correlations, r, have 254 degrees of freedom) 90

3.12 (a) Right-hand chord pattern (left-hand chord patterns are exact mirror images of the right-hand ones), mean RT (secs), standard deviation (SD) and percentage of errors (ERR), for each hand and each chord pattern, plus RT and ERR from Seibel's data, the rank order obtained by Ratz and Ritchie, and the number of fingers required for the depression of each stimulus 94

3.12 (b) Intercorrelations of all of the above variables 94

4.1 Eight one-hand chord patterns for the left and right hand 123

4.2 Average Adaptive Scores of the four training groups on the pretest 127

4.3 Correlations between pretest ASCs and RT, ERR and ASC for the six sessions (N=40) 133

4.4 Total training time over five training sessions and standard deviations for each group (n=10), averaged across subjects 134

4.5 Mean ASCs for each training group (averaged across subjects) for each one of the first three training sessions (n=10) 139

6.1 Mean residual RT of the current response as a function of accuracy, level of difficulty and level of practice of the current response 167

6.2 Residual RTs for current response as a function of stimulus difficulty (DIFF1), accuracy of the response and accuracy of the previous response 171

6.3 Errors on current trials as a function of the accuracy of the preceding response 175

6.4 Residual RTs for responses as a function of current stimulus difficulty and accuracy 183

6.5 Residual RTs for current response as a function of stimulus difficulty, accuracy of the response and accuracy of the preceding response 187

6.6 Errors on current stimuli as a function of accuracy on the preceding response 193

6.7 Errors on current stimuli as a function of accuracy on the preceding response 203

6.8 Errors on current stimuli as a function of accuracy on the preceding response 216

9.1 Number and proportion of errors as a function of handedness 276

9.2 Number and "type of errors" overall and for each training group 277

9.3 Observed and expected number and proportion of "type of errors" overall and as a function of the number of keys involved in one-handed stimuli 278

9.4 Number and proportion of errors as a function of "stimulus complexity", overall and for each training group 279

9.5 Observed and expected number and proportion of errors as a function of "type of errors" and "stimulus complexity" 280

9.6 Summary table of a factorial chi-square analysis of the factors "type of errors" (T), "Fixed versus Cumulative" (F vs C) and "Random vs Adaptive" (R vs A) 282

10.1 Speed-accuracy tradeoff functions for two different levels of difficulty and a "compromise" between speed of performance and error variance 311

LIST OF FIGURES

1.1 QWERTY Keyboard 7

1.2 Dvorak Simplified Keyboard 7

1.3 Alphabetic Keyboard 7

1.4 Maltron Keyboard 9

1.5 Telephone Layout 11

1.6 Adding Machine Layout 11

1.7 AEG-Telefunken Numerical Keyboard 11

1.8 Levy Keyboard 13

1.9 Conrad's 2x6 Chord Keyboard (1960) 13

1.10 "Double-binary" Chord Keyboard (Cornog, Hockman & Craig, 1963) 13

1.11 Keyset Units Used by Hillix and Coburn (1961) 16

1.12 Chord Alphanumeric Keyboard (Conrad and Longman, 1965) 18

1.13 Keyboard Designs Used by Bowen and Guinness (1965) 18

1.14 ANTEL Chord Keyboard 20

1.15 IBM Chord Keyboard 20

1.16 Microwriter (Endfield, 1978) 23

1.17 Keyboard Design Recommended by Kroemer (1972) 25

2.1 2x6 Chord Keyboard and Finger Position on Individual Keys (L=Little, R=Ring, M=Middle, I=Index) 32

2.2 4x4 Chord Keyboard and Finger Position on Individual Keys (L=Little, R=Ring, M=Middle, I=Index) 32

2.3 10-Key Chord or 10-Key Sequential Keyboard and Finger Position on Individual Keys (L=Little, R=Ring, M=Middle, I=Index, T=Thumb) 32

2.4 Examples of Stimuli for the Chord Keyboards 36

2.5 Example of Stimulus for the 10-Key Sequential Keyboard 36

2.6 Type of Stimulus and Visual Feedback for the 10-key Chord Keyboard 44

2.7 Obtained and Predicted Mean RTs for SEP=0 and SEP=1 as a Function of the Number of Fingers Involved in Each Chord Pattern (SEP= Finger Separation) 52

2.8 Type of Visual Feedback for the 10-Key Sequential Keyboard. (The keys 2,7 and 5 should have been depressed but the keys 1,7,4 were depressed instead) 58

2.9 Mean RT for Each Keystroke as a Function of Practice Blocks 63

2.10 Percentage of Errors for Each Keystroke as a Function Practice blocks 63

2.11 Mean RT for Each Finger and Each Keystroke Averaged across Practice (L=Little, R=Ring, M=Middle, I=Index, T=Thumb) 64

2.12 Mean Percentage of Errors for Each Finger and Each Keystroke Averaged across Practice (L=Little, R=Ring, M=Middle, I=Iindex, T=Thumb) 64

3.1 Examples of Stimuli and Type of Visual Feedback for Each Keyboard Group 74

3.2 Mean Reaction Time (csecs) as a Function of Practice for Each of the Four Groups 78

3.3 Mean Percentage of Errors as a Function of Practice for Each of the Four Groups 79

3.4 (a) Mean Reaction Time for Each Key of the 2x6 Chord Keyboard during the Last Twenty Training Sessions 86

3.4 (b) Mean Percentage of Errors for Each Key of the 2x6 Chord Keyboard during the Last Twenty Training Sessions 86

3.5 (a) Mean Reaction Time for Each Key of the 4x4 Chord Keyboard during the Last Twenty Training Sessions 87

3.5 (b) Mean Percentage of Errors for Each Key of the 4x4 Chord Keyboard during the Last Twenty Training Sessions 87

3.6 (a) Mean Reaction Time for Each Keystroke, for Each Key of the 10-Key Sequential Keyboard during the Last Twenty Training Sessions 89

3.6 (b) Mean Percentage of Errors for Each Keystroke, for Each Key of the 10-Key Sequential Keyboard during the Last Twenty Training Sessions 89
T=Thumb)

3.7 (a) Mean Reaction Time for Each Key (Averaged across All Chord Patterns) of the 10-Key Chord Keyboard during the Last Twenty Training Sessions 93

3.7 (b) Mean Percentage of Errors for Each Key (Averaged across All Chord Patterns of the 10-Key Chord Keyboard during the Last Twenty Training Sessions 93

3.8 Mean Reaction Time as a Function of the Number of Fingers for the 10-Key Chord Keyboard (Predicted and Obtained Values from the Last Twenty Training Sessions) 96

3.9 Mean Percentage of Errors as a Function of the Number of Fingers for the 10-Key Chord Keyboard (Obtained Values from the Last Twenty Training Sessions) 97

3.10 The 10-Key Chord Keyboard Design Adopted by the Post Office 99

4.1 Eckstrand's (1964) Training System 102

4.2 Pask's (1964) Adaptive Teaching System 116

4.3 Conventional Training System (System A) and Adaptive Training System Developed by Kelley (1969) (System B) 118

4.4 Example of Stimulus and Visual Feedback 129

4.5 Mean Reaction Time of Each Training Group as a Function of Practice 131

4.6 Mean Percentage of Errors of Each Training Group as a Function of Practice 131

4.7 Mean Adjusted Score (ASC) of Each Training Group as a Function of Practice 131

4.8 Mean RT of the Adaptive and Random Groups during the First Three Training Sessions 136

4.9 Mean RT of the Cumulative and Fixed Groups during the First Three Training Sessions 136

4.10 Mean Percentage of Errors of the Fixed and Cumulative Groups during the First Three Training Sessions 138

4.11 Mean Adjusted Score (ASC) of the Fixed and Cumulative Groups during the First Three Training Sessions 138

4.12 Mean RT of Each Training Group as a Function of the Level of Stimulus Difficulty 141

4.13 Mean Percentage of Errors of Each Training Group as

a Function of the Level of Stimulus Difficulty 141

4.14 Mean Adjusted Score (ASC) of Each Training Group as a Function of the Level of Stimulus Difficulty 141

6.1 Mean RT of the Current Response as a Function of the Level of Stimulus Difficulty and the Level of Practice (FR Group) 164

6.2 Mean Residual RT of the Current Response as a Function of the Level of Stimulus Difficulty and the Accuracy of the Current Response 166

6.3 Mean Residual RT of the Current Response as a Function of the Current Stimulus Difficulty and the Accuracy of the Preceding Response 166

6.4 Mean Residual RT of the Current Response as a Function of the Current Stimulus Difficulty, the Accuracy of the Preceding Response and the Accuracy of the Current Response 170

6.5 Mean Percentage of Errors as a Function of the Level of Stimulus Difficulty and the level of Practice 174

6.6 Mean RT of the Current Response as a Function of the Level of Stimulus Difficulty and the Level of Practice 180

6.7 Mean Residual RT of the Current Response as a Function of the Level of Stimulus Difficulty and the Accuracy of the Current Response 182

6.8 Mean Residual RT of the Current Response as a Function of the Level of Practice and the Accuracy of the Current Response 182

6.9 Mean Residual RT of the Current Response as a Function of the Level of Stimulus Difficulty and the Accuracy of the Preceding Response 185

6.10 Mean Residual RT of the Current Response as a Function of the Accuracy of the Current and the Preceding Response 185

6.11 Mean Residual RT of the Current Response as a Function of the Level of Stimulus Difficulty and the Accuracy of the Current and the Preceding Response 186

6.12 Mean Residual RT of the Current Response as a Function of Practice and the Accuracy of the Preceding Response 188

6.13 Mean RT of the Current Response as a Function of the Level of Stimulus Difficulty and Level of Practice 197

6.14 Mean Residual RT of the Current Response as a Function of the Level of Stimulus Difficulty and Accuracy of the Current Response 198

6.15 Mean Residual RT of the Current Response as a Function of Practice and Accuracy of the Current Response 198

6.16 Mean Percentage of Errors as a Function of Practice and Level of Stimulus Difficulty 202

6.17 Mean RT of the Current Response as a Function of Practice and Level of Stimulus Difficulty 208

6.18 Mean Residual RT of the Current Response as a Function of the Level of Stimulus Difficulty and Accuracy of the Current Response 210

6.19 Mean Residual RT of the Current Response as a Function of Stimulus Difficulty and Accuracy of the Current and the Preceding Response 212

6.20 Mean Residual RT of the Current Response as a Function of Practice and Accuracy of the Preceding Response 213

6.21 Mean RT of the Current Response of Each Training Group as a Function of Practice Session 220

6.22 Mean RT of the Current Response of Each Training Group as a Function of Stimulus Difficulty 221

8.1 Graphical Representation of Signal Detection Theory 243

8.2 Receiver Operating Characteristic (ROC) Curves 245

8.3 The Form of Speed-Accuracy Operating Characteristic Assumed by the Fast Guess Model 251

8.4 Graphical Representation of the Random Walk Model 251

9.1 An Example of the Distribution of RTs of Correct (White Columns) and Error (Black Columns) Responses (Data from Subject 37) 268

10.1 P(C|R) and P(R) Functions for a Simple RT Task 287

10.2 P(C|R) and P(R) Functions for a "Power" Test 289

10.3 P(C|R) and P(R) Functions for a Typical Choice Response Task 291

10.4 Distributions of the Amount of Information Accumulation the Means of Which Define the P(C|R)

Function 294

10.5 (a) P(C|R) Functions for "Easy" Stimuli 295

10.5 (b) P(C|R) Functions for "Difficult" Stimuli 295

10.6 P(C|R) of Fast-guess Responses as a Function of the Number of Equiprobable Response Alternatives 297

10.7 P(C|R) Functions for Subjects A and B, Where Subject A Has Received More Practice than Subject B 297

10.8 P(C|R) Functions for Stimuli S1, S2, S3, ..., Sn Which Are Selected from Sets with 1, 2, 3, ..., n Equiprobable Stimulus Alternatives 300

10.9 Optimum and Actual Functions of P(R) when the Ideal Cutoff Point Is x 305

10.10 P(C|R) Functions as Affected by Subjects' "Confidence". (When "confidence" is constant from trial to trial the P(C|R) is a single curve) 315

10.11 P(C|R) Functions and Cutoff Points for "Easy" and "Difficult" Stimuli and Various Speed-Accuracy Tradeoffs 316

PREFACE

This thesis deals with theoretical and practical issues of special-purpose keyboard design and operator training procedures. It is in two parts.

In Part A applied research carried out for and funded by the Post Office is described and discussed while in Part B theoretical issues which emerged from the applied research are investigated and a new theory is presented.

The research discussed in Part A was carried out in support of two contracts from the Post Office, one in 1984 and the other in 1985. (A summary of the work is presented in Bartram, 1987). The research had two main objectives: first, to provide information needed for the selection of a new keyboard design for use on foreign destination coding desks; and second, to evaluate one component of a computer-based training system for operators of the new keyboard. Given that the research was related to specific contracts, certain constraints were, inevitably, imposed on the work.

A literature review of keyboard designs is presented in Chapter 1, where evidence from experimental studies on keyboards is cited and ergonomic considerations of keyboard design are pointed out. In particular, some theoretical background on keyboard operation is provided, followed by experimental work on sequential and chord keyboards. The use of chord keyboards in the past and in recent years is examined in more detail. Human factors issues which affect keyboard operation are pointed out and it is suggested that they should

be considered in the design of any keyboard.

The four keyboard layouts given by the Post Office to be evaluated, namely the 2x6, 4x4, 10-key chord and 10-key sequential, are described in Chapter 2. Furthermore, two pilot experiments investigating questions regarding the operation of the 10-key chord (see also Lee, 1984; Bartram and Lee, 1985) and the 10-key sequential keyboards are reported (see also Bartram and Feggou, 1985a). An evaluation and comparison of these four keyboard layouts was made by means of a six-week experiment and is reported in Chapter 3. On the basis of the results of these three experiments the 10-key chord keyboard was chosen and recommended to the Post Office as the most efficient of the four layouts (see also Bartram and Feggou, 1985b; 1986).

The computer-based training system for operators of the foreign destination letter coding desks consists of three stages: (1) Keyboard Familiarisation, which involves developing the skill of keying various chord patterns; (2) Area-Name Training, during which trainees learn the chord pattern associated with each area-name; and (3) Address Encoding, which is a simulation of the actual letter sorting task operators are expected to carry out in the end of their training (Bartram, 1987). In the present thesis, only the first stage of keyboard familiarisation is examined. Four different computer-based training methods are presented, discussed and evaluated in an experiment specifically designed for this purpose. The investigation of the training procedures for keyboard familiarisation stage is reported in Chapter 4.

The main issue of Part B is speed-accuracy tradeoff relationships in a keyboard task. The limitations of typical Choice Reaction Time models and the benefits of studying speed-accuracy tradeoff relationships in choice response tasks are discussed in Chapter 5. A new analysis of the data from experiment in Chapter 4 is presented in Chapter 6, where the speed and accuracy of any given response is examined in terms of speed and accuracy of the preceding response. The main findings of these analyses are discussed in Chapter 7.

These findings are further discussed in the light of established theories in Chapter 8 and it is concluded that no known theory can fully account for the obtained results. Two further aspects of the data are explored in Chapter 9.

A new theoretical model is presented and assessed in Chapters 10 and 11. Finally, a brief summary of findings and conclusions is presented and implications for keyboard operation and operator training are made in Chapter 12.

PART A

Chapter 1

SPECIAL-PURPOSE KEYBOARD DESIGNS

The automation of data handling has been accompanied by the development of a number of data entry devices. These include keyboards, switches, levers, knobs, light pens, "mice", etc. Keyboards have been popular data entry devices with their use ranging from ordinary typewriters, numeric keypads, more sophisticated computer panels and keyboards for Letter Sorting Machines, to ones designed for specialized needs. The first three chapters focus on the first part of the Post Office project and investigate keyboards, with special emphasis on those designed for Letter Sorting Machines.

1.1. Definition of the Problem

All international mail is handled by a separate department at the Post Office. Letters are sorted and grouped according to their destination. This task is carried out mechanically by the Letter Sorting Machines. The Letter Sorting Machines are presently operated by operators using a special keyboard. All letters are fed into the machine which forwards them, one by one, to a "window" in front of the operator. The operator reads the envelope and presses a specific combination of keys on the keyboard which codes the specific destination of the letter according to one of a number of geographical areas. This operation enables each letter to be forwarded automatically to the appropriate

"pigeon hole" from where they are taken and sent to the appropriate area.

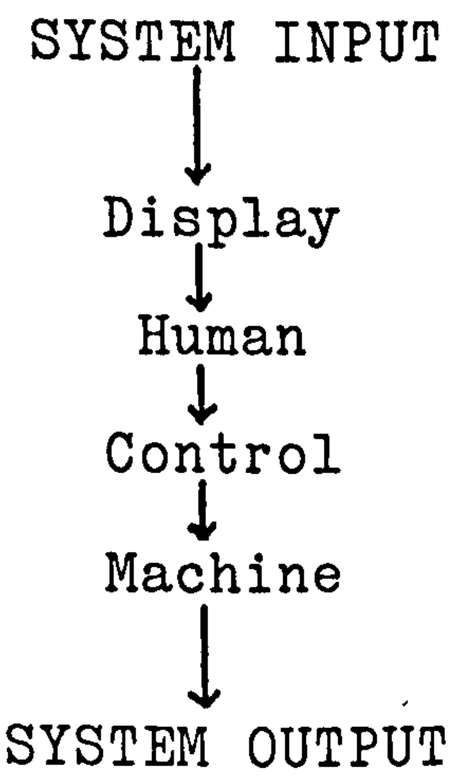
Currently, these machines are operated through a keyboard with two banks of twelve keys, one operated by the left hand and the other by the right. Operators have to press a specific key combination for each destination area. Key combinations involve the simultaneous depression of one key by each hand. This keyboard has a total capacity of 144 combinations (12x12=144). The need of the Post Office to increase the number of destination areas has necessitated the replacement of this keyboard by one with a bigger capacity. For this purpose, three new keyboards have been designed and certain aspects of them are examined in the light of previous experimental findings on keyboards. Furthermore, these keyboards are evaluated and compared with the existing keyboard.

1.2. Theoretical Background

The ultimate criteria for the selection of any keyboard for a given operating environment are efficiency of performance in terms of speed and accuracy, the need for comfortable operation that enables users to work on the keyboard for prolonged periods of time as well as that the keyboard is acceptable to workers. Phrased differently, the optimal keyboard should allow comfortable, high keying frequency and error free operation. Seibel (1970) refers to data entry as "a man to machine communication". By implication, it is very important for the whole "keying

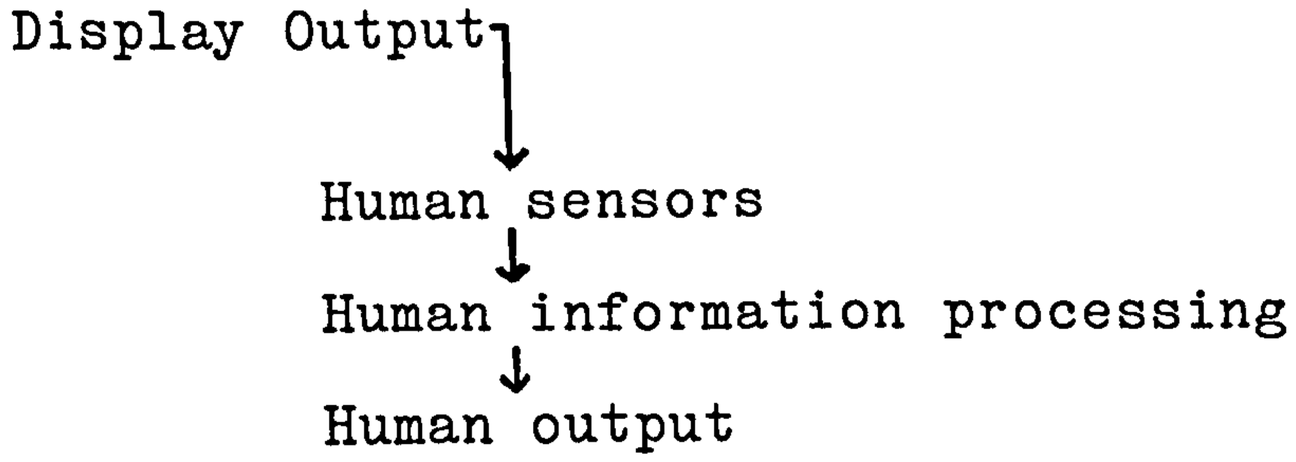
operation" to be considered when a new keyboard is being chosen. Levine and Teichner (1972, 1973; Teichner 1969) have developed the Information-Theoretic Approach, which deals with "man-machine communication". (A review of the theory can be found in Fleishman and Quaintance, 1984, chapter 10).

Levine and Teichner (1973) emphasise the difference between a task and a process. They define a task as the information transfer between two components: i.e. an information source and a receiver. A process is defined as the operation carried out on the information or data within a component. Teichner (1969) claims that a process can be described in the same terms as tasks at more detailed levels. Their Information-Theoretic Approach involves two levels of description: (1) the "system" and (2) its components, the "human" and the "machine". On the first level, the system transfers the information from the system input to the system output.

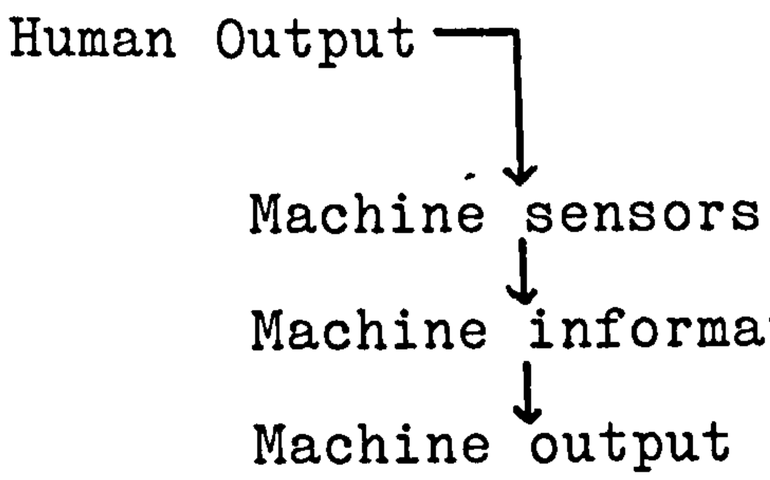


The "human" and the "machine" components of the system involve processes and consequently can be described at the second, more detailed level. The function of the "human"

component is to receive the display output through the human sensors and transfer it through human information processing as human output.



The function of the "machine" component is to receive the human output through the machine sensors, process it and produce a machine output.



By this description Levine and Teichner designated four major tasks: machine-machine, machine-human, human-human and human-machine. The advantage of this theoretical model is that it is not concerned with processes which are difficult to describe, but with tasks which can be operationally defined. Levine and Teichner say that all tasks can be classified according to four dimensions: (1) the nature of constraints that are imposed, that is the limitations on which events are sampled and how they are sampled in the stimulus-response situation; (2) the location of the constraints, that is whether the restrictions are placed on the input or on the output; (3) the amount and form of redundancy introduced by

the constraints; and (4) the relationship between the amounts of input and output information. Levine and Teichner claim that by analysing the constraints imposed on any task this model is useful in predicting human performance.

It is clear that the efficiency of any system of human-machine communication depends on all the system components. In the present series of studies, however, the machine components have to be assumed as constants since they have already been designed to meet the engineering system requirements. The emphasis of this project is on the human component that affects the human information transfer function in the operation of keyboards at the International Mail Coding-Desks. The human information processing component can be broken down to three basic parts:

Display output

Human information processing:

(1) DISPLAY ENCODING

(e.g. Read "SWEDEN" and encode as indicating DESTINATION=SWEDEN.)

(2) DISPLAY-RESPONSE MAPPING

(e.g. Relate SWEDEN to internal representation of letter-code for SWEDEN.)

(3) RESPONSE PREPARATION

(i.e. Retrieve the appropriate motor-response programme and prepare the response.)

Human output:

(4) RESPONSE EXECUTION

(i.e. Execute the response programme by pressing the appropriate keys.)

In the present series of experiments all emphasis is placed on the response preparation and response execution. The reason for concentrating on only these two aspects of the human component is to minimize the differential effects of learning related to display encoding and display-response mapping, as effects associated with these stages are likely to be independent of factors relative to keyboard differences.

1.3. Experimental Research on Keyboards Typewriters

The first keyboards that were constructed and used for industrial purposes were sequential, that is they required the depression of one key for each character. The conventional typewriter, QWERTY (figure 1.1), designed by Desmore and Scholes in 1873, is the most familiar example of such keyboards. Since then other keyboards have been examined aiming at an advantageous spatial re-arrangement of the keys. A strong alternative to the QWERTY is the Dvorak Simplified Keyboard (figure 1.2). The key arrangement of the Dvorak keyboard increases the workload of the right hand from 43% on the QWERTY to 56% (Snyder, 1984). Seibel (1970) claims that the Dvorak is a more efficient keyboard but admits that a change might not be cost-effective. An experimental study sponsored by the U.S.A. Government (Strong, 1956) showed that there is no advantage of changing to the Dvorak keyboard. Seibel (1972, in Snyder, 1984) questions Strong's methodology

Figure 1.1 QWERTY Keyboard

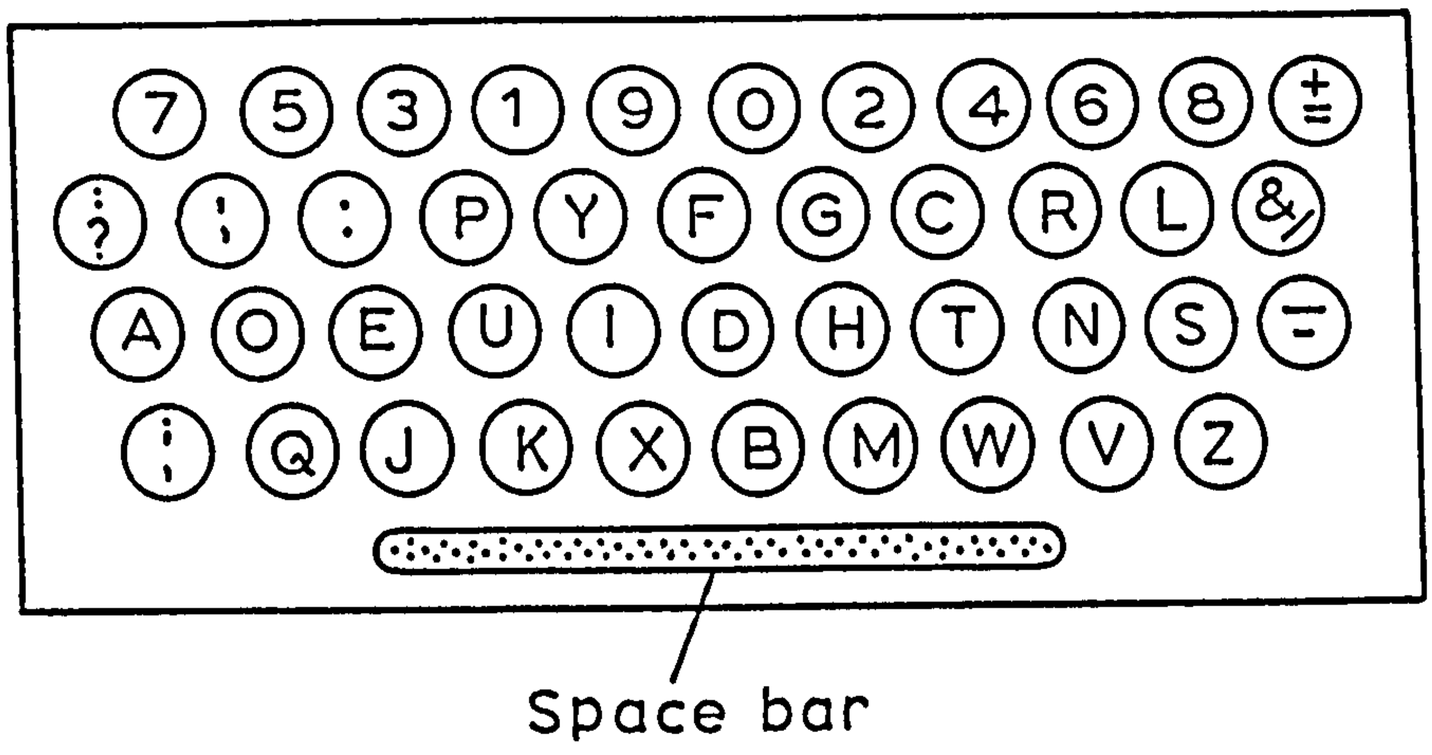


Figure 1.2 Dvorak Simplified Keyboard

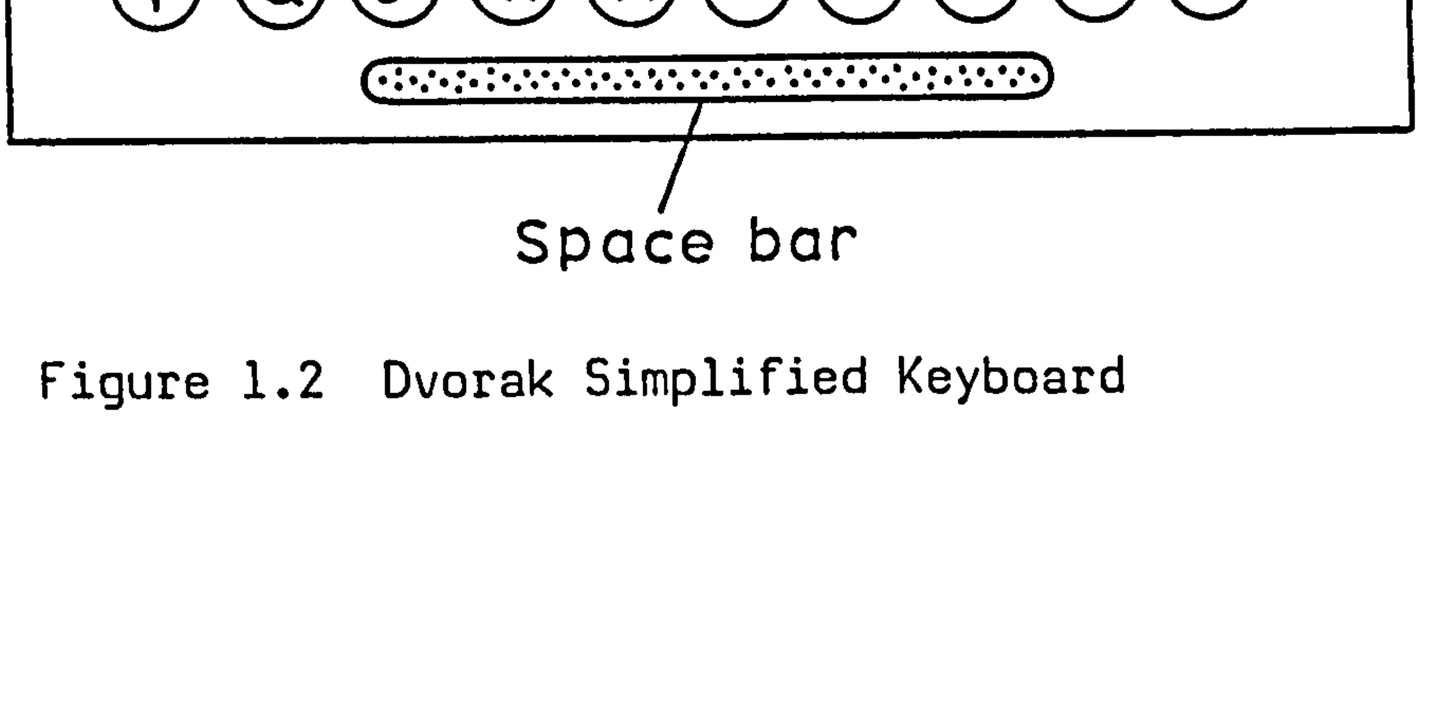


Figure 1.3 Alphabetic Keyboard

but nonetheless recommends the QWERTY as it is a "de facto" international standard.

Another example of a sequential keyboard is the Alphabetic Keyboard (figure 1.3) which was introduced as superior to the QWERTY for unskilled or occasional users. Michaels (1971) compared the Alphabetical and the QWERTY layouts for varying levels of user skill and found no performance difference. Even unskilled operators for whom the Alphabetic keyboard was intended performed better on the QWERTY. Klemmer (1971, p.3) claims that "there seems to be no potential population of users who perform better on the alphabetically ordered keyboard".

The Maltron Keyboard (figure 1.4) was first described by L. Malt at the Printing Industries Symposium in London, in 1977 (Hobday, 1981). According to her "... For high speed keying, it is necessary to: (a) balance the load between the two hands - making some allowance for right hand dominance; (b) balance the load between the fingers - allowing for individual finger capacity; (c) reduce finger movements to a minimum - by placing the most commonly used letters directly under the ten digits; (d) reduce difficult finger motions to a minimum - reduce hurdles and stretches and avoid use of the same finger twice in succession; (e) allow for fastest finger movements to be used most frequently - provide for [ipsi-]lateral as well as contra-lateral keying; (f) avoid long sequences on one hand - balance lateral with contra-lateral keying... For accurate keying and for ease of learning, letter

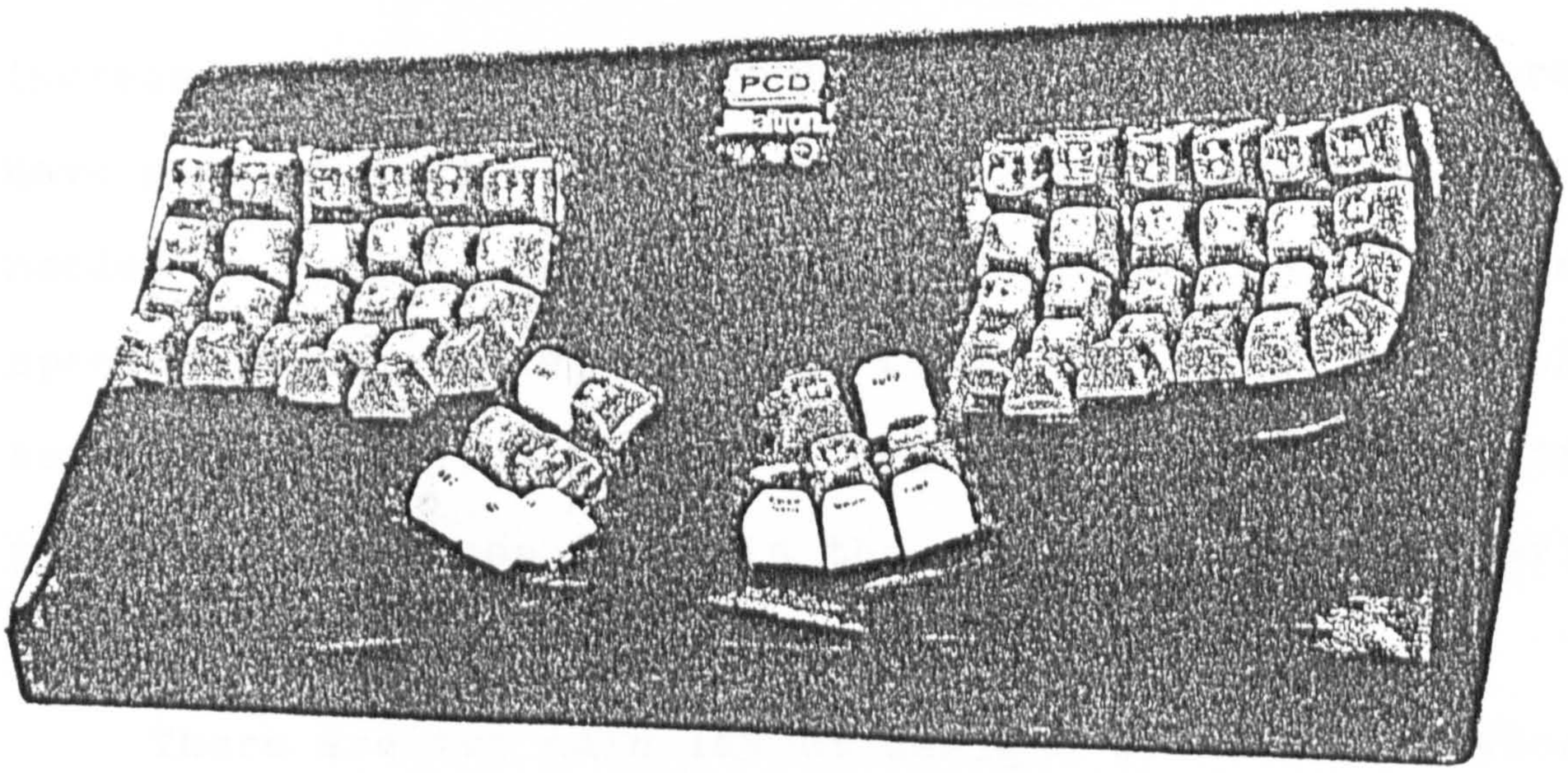


Figure 1.4 Maltron Keyboard

layout should take account of cybernetic requirements related to language..." (In Hobday, 1981, p.17). Although the manufacturers of this keyboard promote it as very superior to the QWERTY and the Dvorak, there has been no assessment of it by an independent researcher (at least to the present author's knowledge).

It can be concluded that irrespective of possible increases in the efficiency of performance, "market forces" have made it very difficult to justify the cost and effort needed to replace the "standard" QWERTY keyboard. For certain special keying tasks, however, (e.g. Letter Sorting Machines) new types of keyboards can be considered without marketing problems like those faced in the design of a new typewriter.

Numeric Keyboards

There are two main layout designs of numeric keyboards (often referred to as "keypads") both of which consist of three rows of three keys plus one extra key: (1) the push-button telephone layout (figure 1.5) with the keys 1,2,3 on the top row ("123" arrangement); and (2) the adding machine layout (figure 1.6) with the keys 7,8,9 on the top row ("789" arrangement). Seibel (1970) claims that the two arrangements are equally efficient when used by skilled operators, but occasional users tend to expect the "123" arrangement. Studies by Conrad and Hull (1968), Deininger (1960) and Minor and Revesman (1962) conclude that the "123" arrangement is preferable to the "789" one, particularly for low-skill operators.

The "789" arrangement has been used for another keyboard

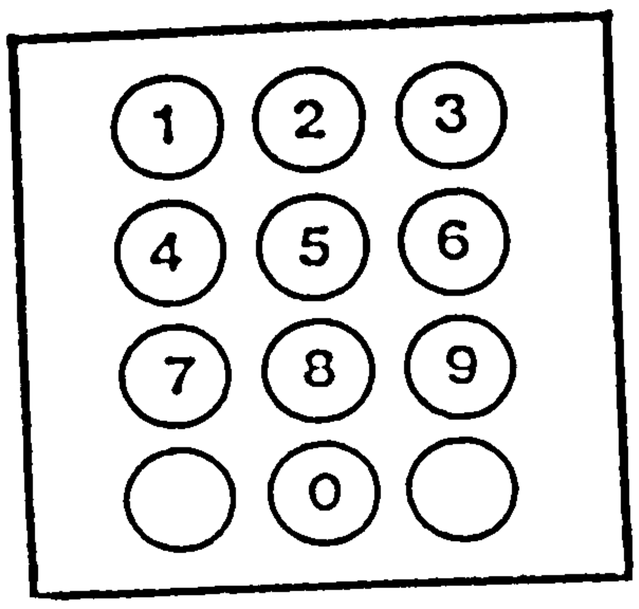


Figure 1.5 Telephone Layout

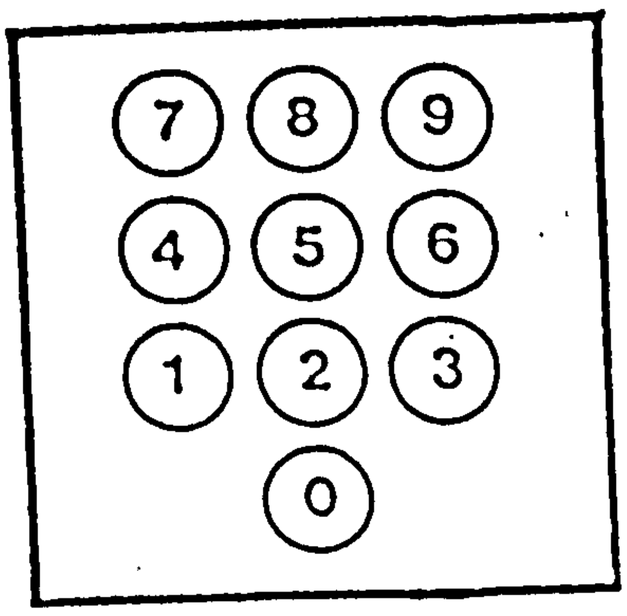


Figure 1.6 Adding Machine Layout

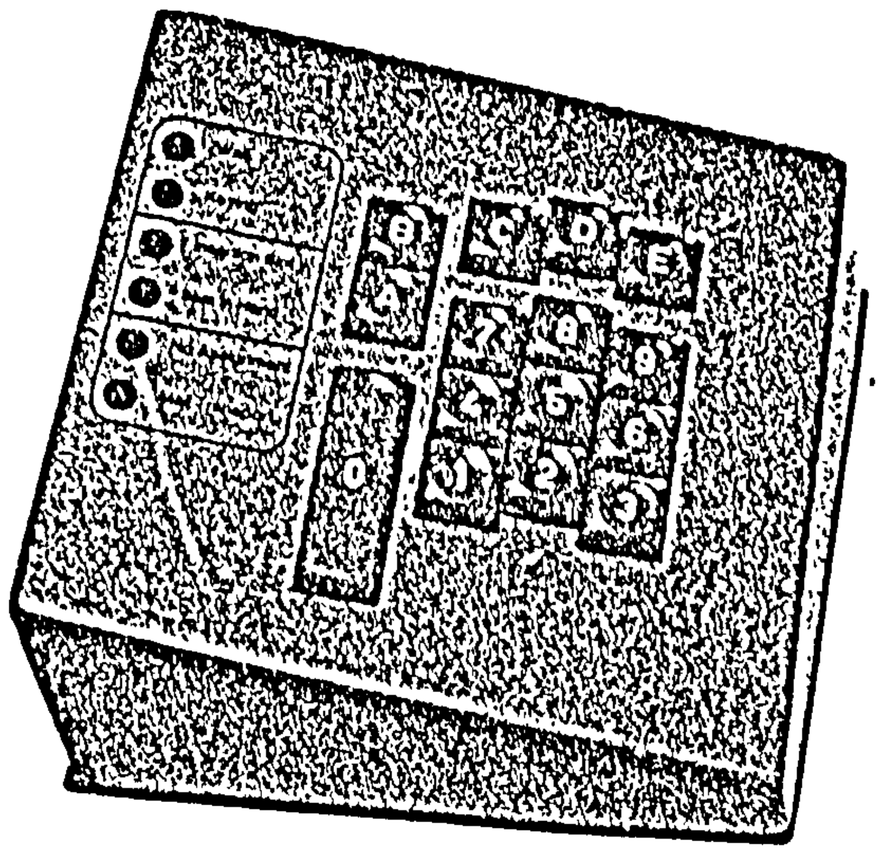


Figure 1.7 AEG-Telefunken Numerical Keyboard

designed by AEG-Telefunken for sorting of postal "flat packs" (figure 1.7). The keyboard has a total of fifteen keys: ten keys represent the digits 0 to 9 and five keys represent the letters A,B,C,D and E. The keyboard is operated by the right hand only (unless designed for left-handed operators) while the other hand is free to pick up the pack and place it on the sorting machine. This keyboard is used in Germany, the U.S. and shortly it is going to be used in the U.K.

1.4 History of Chord Keyboards

Chord keyboards require the depression of one or more keys at a time and, consequently, a small number of keys can allow a large number of key combinations. The basic argument for the use of chord keyboards is that because of the small number of keys the number of "reach" movements is reduced and the keying performance can become more efficient.

Several types of keyboards have been tried out since the 1950's, especially keyboards to be used for Letter Sorting Machines. These layout designs differ in the number of keys they have, in the way the keys are positioned, in the number of keys that must be pressed at a time, etc. Examples of such keyboards are those of Levy (1955) (figure 1.8), which was designed for the Toronto Post Office; Conrad (1960) (figure 1.9), which is still in use at the British Post Office but is to be replaced; Cornog, Hockman, and Craig (1963) (figure 1.10), which was designed for the Post Office Department in the U.S.A.

Levy (1955) was one of the first to experiment with a 10-

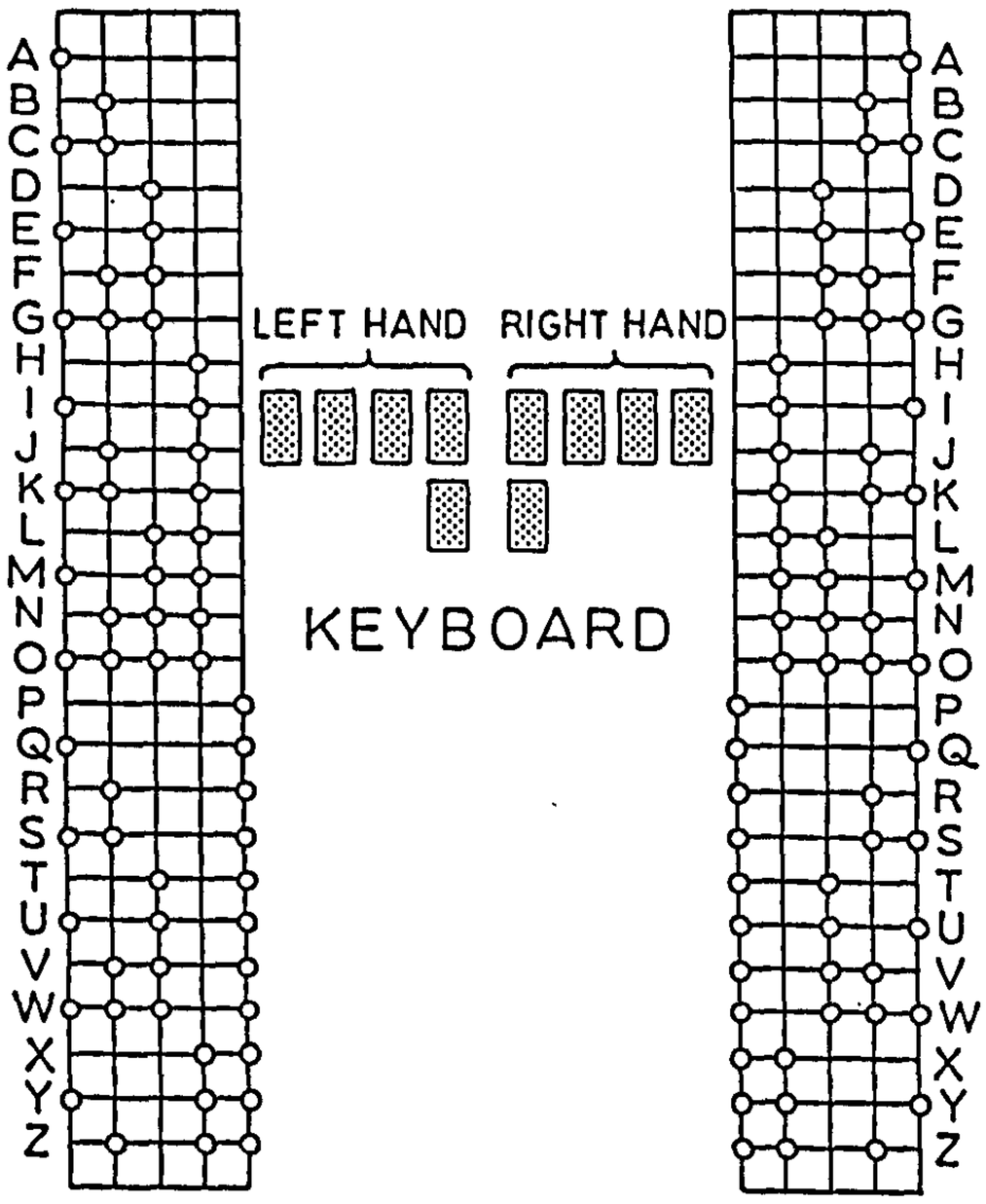


Figure 1.8 Levy Keyboard

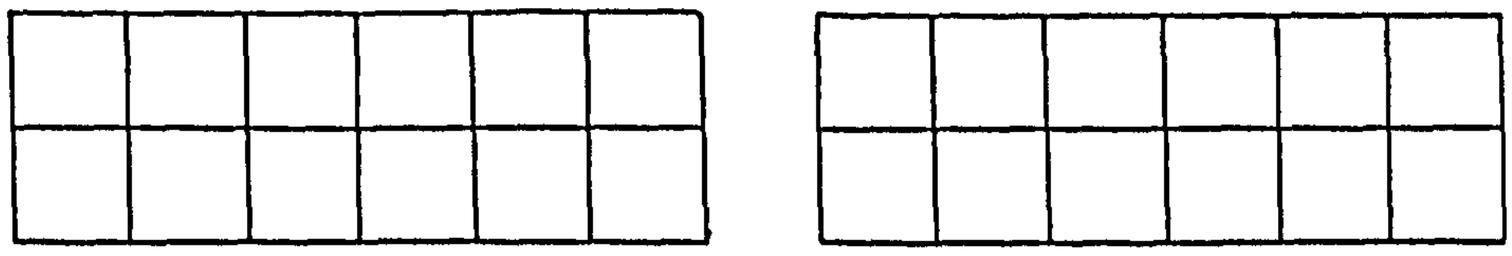


Figure 1.9 Conrad's 2x6 Chord Keyboard (1960)

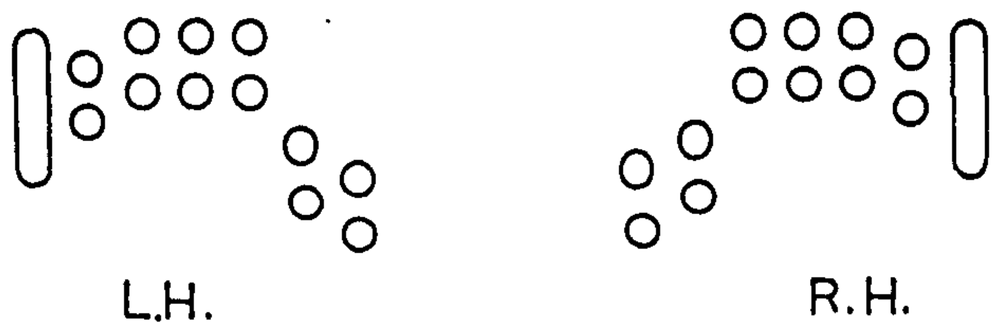


Figure 1.10 "Double-binary" Chord Keyboard
(Cornog, Hockman & Craig, 1963)

key "binary" keyboard. The Levy keyboard (figure 1.8) consisted of two sets of five keys and required a two-key combination for a letter or number. He trained three female subjects in the use of this keyboard with the training schedule being three hours a day for twenty-eight weeks. By the end of their training the subjects had reached a speed of 70 words per minute and were still improving. His control subject, a female typist, could type the same practice material at a speed of 95 words per minute. Somewhat surprisingly, he concluded that the keying speeds on the "binary" keyboard compared favorably with the speed on the typewriter. Furthermore, he concluded that the use of this keyboard combined with an effective method of encoding information would result in a higher speed than sorting letters manually. The limitations of his work are that he gives no accuracy data and that three experimental subjects plus one control are not sufficient to provide statistically reliable and conclusive comparisons.

Conrad (1960a,b) experimented with a 2x6 chord keyboard at the British Post Office (still in use for overseas mail sorting, figure 1.9). Operators had to simultaneously depress two keys, one by each hand, to forward a letter to one of 144 possible destination areas. Mechanical limitations imposed an upper limit of 110 letter sorts per minute. Following nine months of practice, the sorting rates had improved from about 35 sorts per minute for the manual system to approximately 60 sorts per minute when using the keyboard.

Cornog, Hockman and Craig (1963) evaluated a "double-binary" keyboard (figure 1.10) for use in a mail-sorting task. It involved one block of twelve keys for each hand, in a hand-configured layout, as well as a code-completion bar that lay on the outer edge of each block of keys. Each of the fingers operated two keys and the thumb operated four keys. Their training console consisted of the keyboard, the projection system that provided the stimuli, an arm-rest bar positioned between the operator and the keyboard, a foot operated switch to advance the film in the projector, and a panel of lights to check faulty strokes. After thirty-six weeks of training, their results suggested that such a type of keyboard could be efficiently used for a mail-sorting task.

In addition to this kind of applied research involving chord keyboards for Letter Sorting Machines, considerable work has been carried out on a number of experimental keyboards. Experimental studies looking at the efficiency of various chord keyboards have been reported by Klemmer (1958), Hillix and Coburn (1961), Ratz and Ritchie (1961), Seibel (1962), Leonard and Newman (1965). All of them argue in favour of chord keyboards.

Klemmer (1958), working for IBM, trained two subjects in the use of a ten-key chord keyboard (five keys for each hand arranged in a semi-circle). He allotted the most frequently used letters of the English alphabet (i.e. the letters A, D, E, H, L, N, O, R, S, T) to single-hand chord patterns and the remaining letters to two-hand chord patterns. He concluded that performance on this keyboard was similar to performance

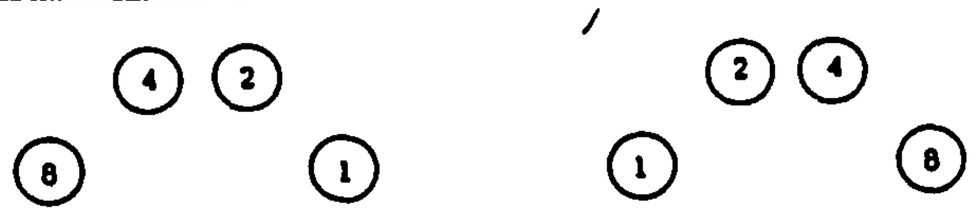
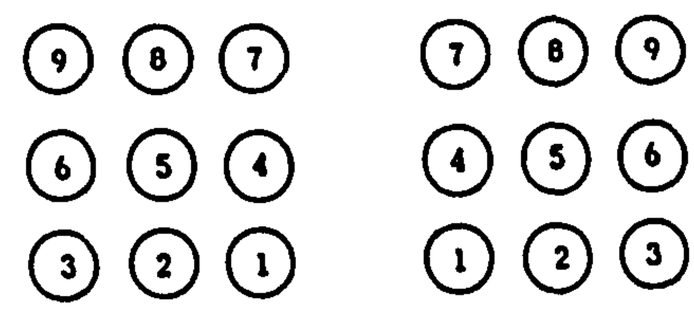
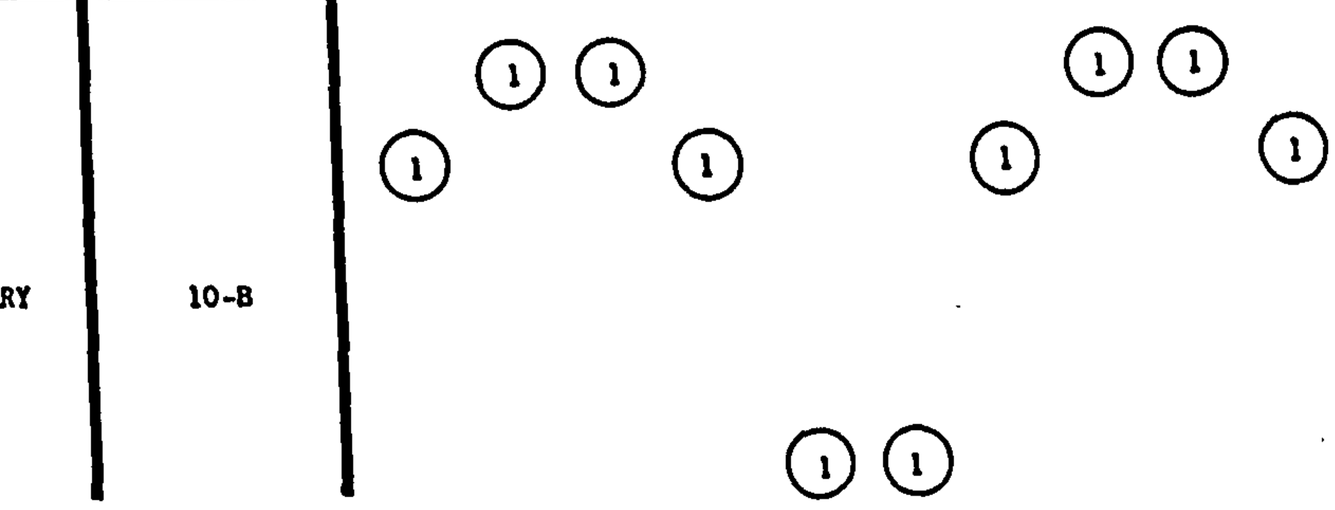
KEYSET NAME	ABBREVIATION	LAYOUT	STIMULUS
SIX-KEY BINARY	6-B		101110
SIX-KEY OCTAL	6-O		65
EIGHT-KEY BINARY	8-B		11100101
EIGHT-KEY 16-BASE	8-16		6F
EIGHT-KEY DECIMAL	8-D		94
EIGHTEEN-KEY DECIMAL	18-D		94
TEN-KEY BINARY	10-B		1011011100

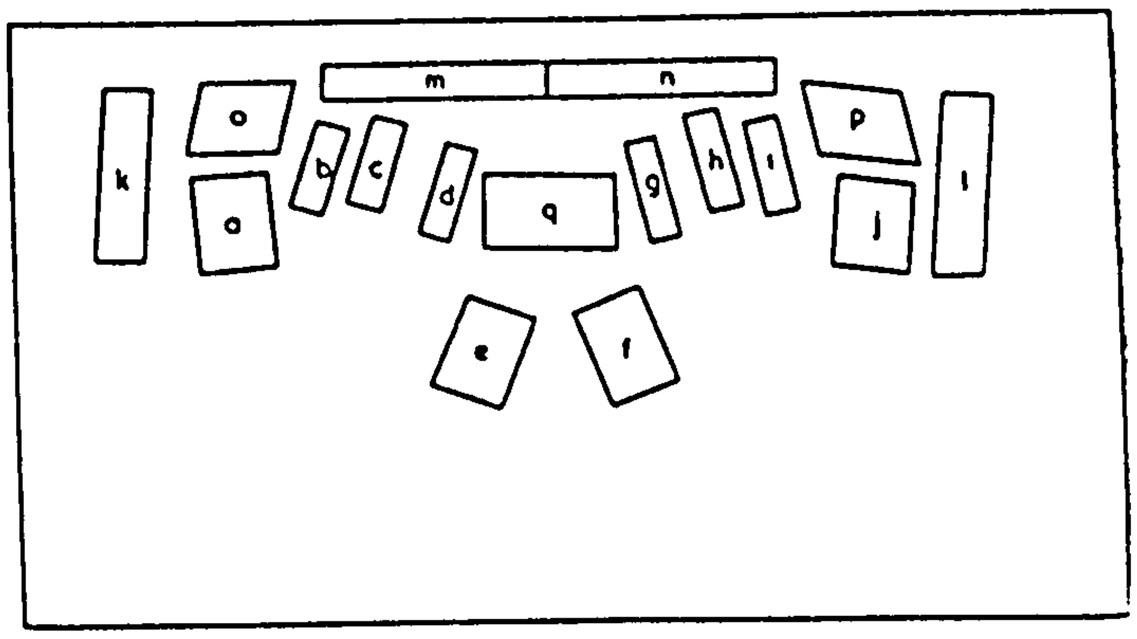
Figure 1.11 Keyset Units Used by Hillix and Coburn (1961)

on the conventional typewriter.

Hillix and Coburn (1961), working at the U.S. Navy Electronics Laboratory, compared seven different keyset designs (figure 1.11). They combined time and error measures in a single index to indicate the number of bits of information transmitted per second. One of their basic conclusions was that the design of a keyset must be based on "... a thorough analysis of the human keysetting processes involved in the particular situation for which the keyset is intended" (1961, p.1). Their suggestions for the "ideal" keyboard involve features like a touch operation facility, a hand-configured layout, provision for two-handed operation with maximum alteration facility, provision for correction of individual character errors and an operating procedure that is easy to learn.

1.5 Comparison between Sequential and Chord Keyboards

One of the very few comparisons between a sequential and non-sequential keyboard was a well-conducted study by Conrad and Longman (1965). They compared the conventional QWERTY keyboard with a chord keyboard (figure 1.12) which required simultaneous depression of two keys (one key by each hand) per character in an alpha-numeric task. Their hypothesis was that by minimizing the number of finger movements the chord keyboard would produce greater performance efficiency. Both of their groups (postmen between 30 and 40 years of age) received training of three and a half hours per day, five days a week for seven weeks. They found that the chord keyboard group



a + f = A	1 + a = 1
a + g = B	1 + b = 2
a + h = C	1 + c = 3
a + i = D	etc
a + j = E	k + f = 6
b + f = F	k + g = 7
b + g = G	k + h = 8
etc	k + i = 9
	k + j = 0

Figure 1.12 Chord Alphanumeric Keyboard (Conrad and Longman, 1965)

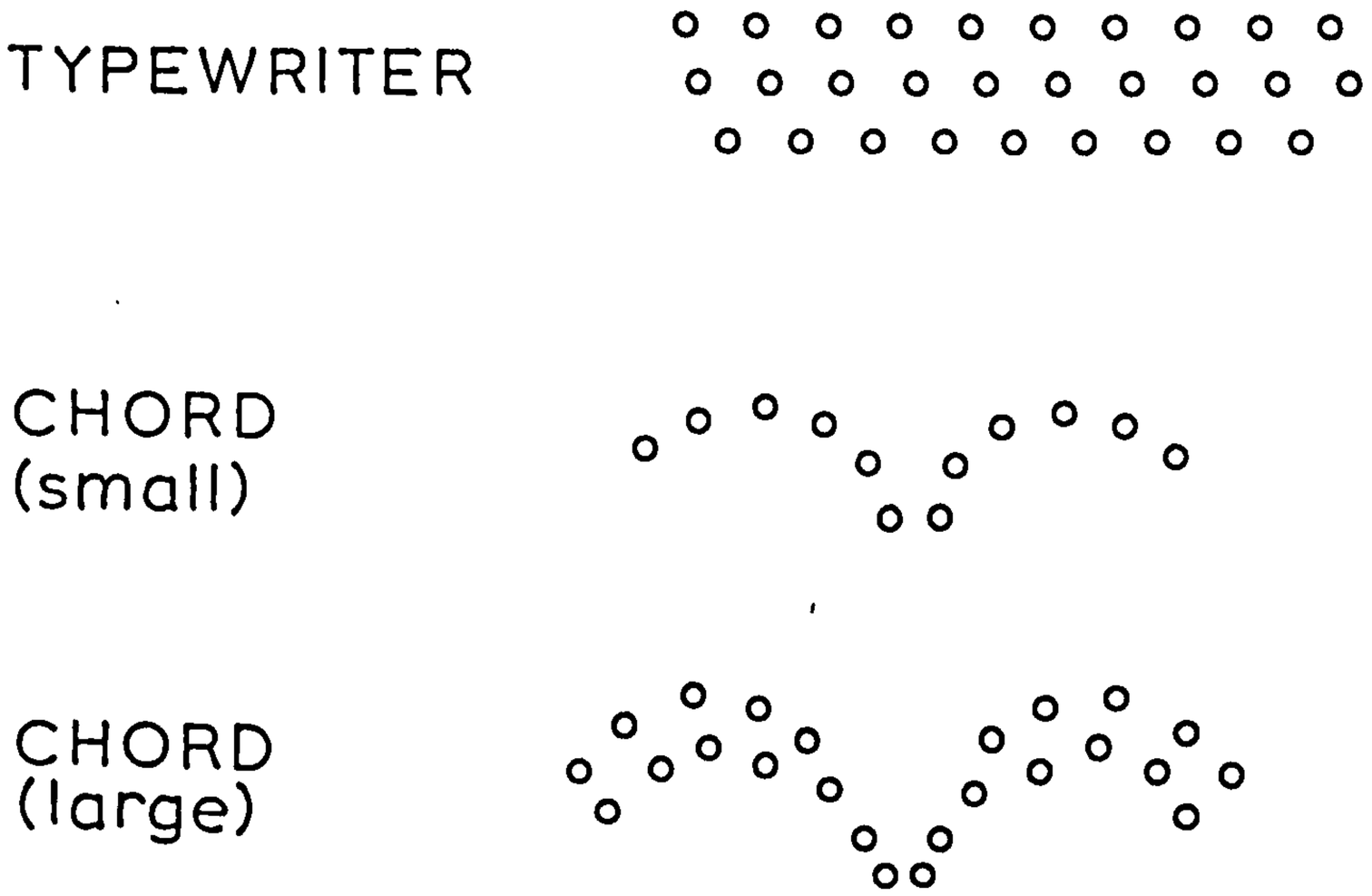


Figure 1.13 Keyboard Designs Used by Bowen and Guinness (1965)

reached an "operational" speed about two weeks earlier than the group that used the typewriter. Their error rate was similar despite the fact that the typewriter group could see their errors immediately, whereas the chord keyboard group were told about their errors by their instructor, sometimes a few days later.

Another study which compared sequential and chord keyboards was that conducted by Bowen and Guinness (1965) who were examining optimal keyboards for Letter Sorting Machines in the U.S. They used three keyboards: the ordinary QWERTY keyboard, a 12-key chord keyboard and a 24-key chord keyboard (figure 1.13). Furthermore, they employed two ways of encoding the information that had to be typed: "memory" coding, in which operators had to remember a combination of keys for each address and "extraction" coding, in which operators applied certain rules to "extract" a different code for every address. They concluded that the chord keyboard was preferable for mail sorting to the sequential keyboard.

1.6 Recent Research on Chord Keyboards

Stewart patented the ANTEL chord keyboard in 1973 (Martin, 1980). Basically, it is a 12-key keyboard which looks like a pocket calculator (figure 1.14). The ANTEL keyboard operates as a sequential one when one of the twelve keys is depressed (e.g. A) and as a chord when a certain key combination is depressed (e.g. a simultaneous depression of keys A and K produces the character F). The main advantages claimed for the ANTEL keyboard are the following:

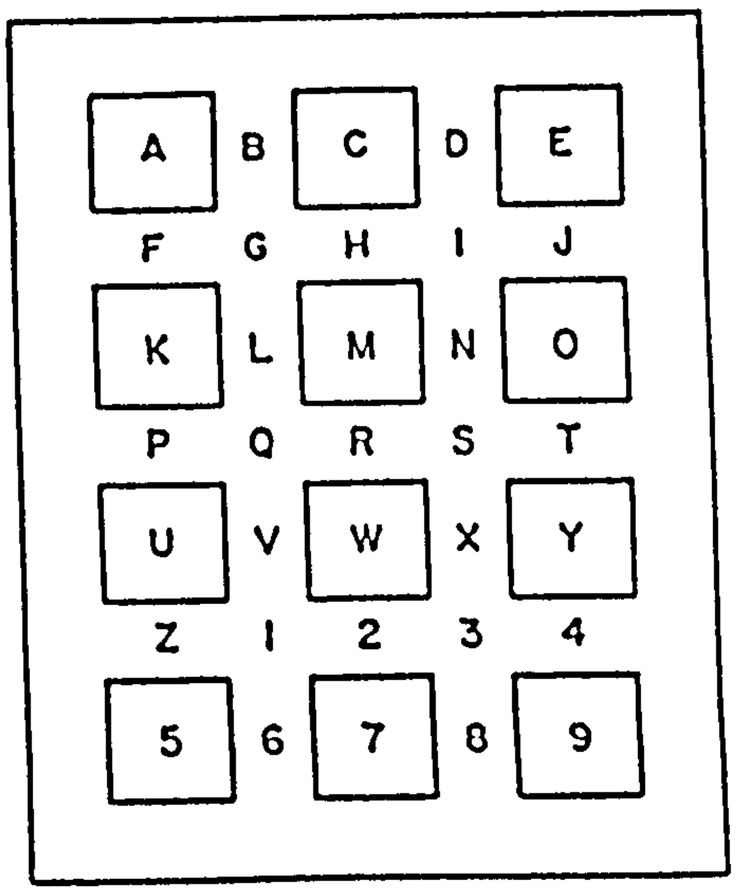


Figure 1.14 ANTEL Chord Keyboard

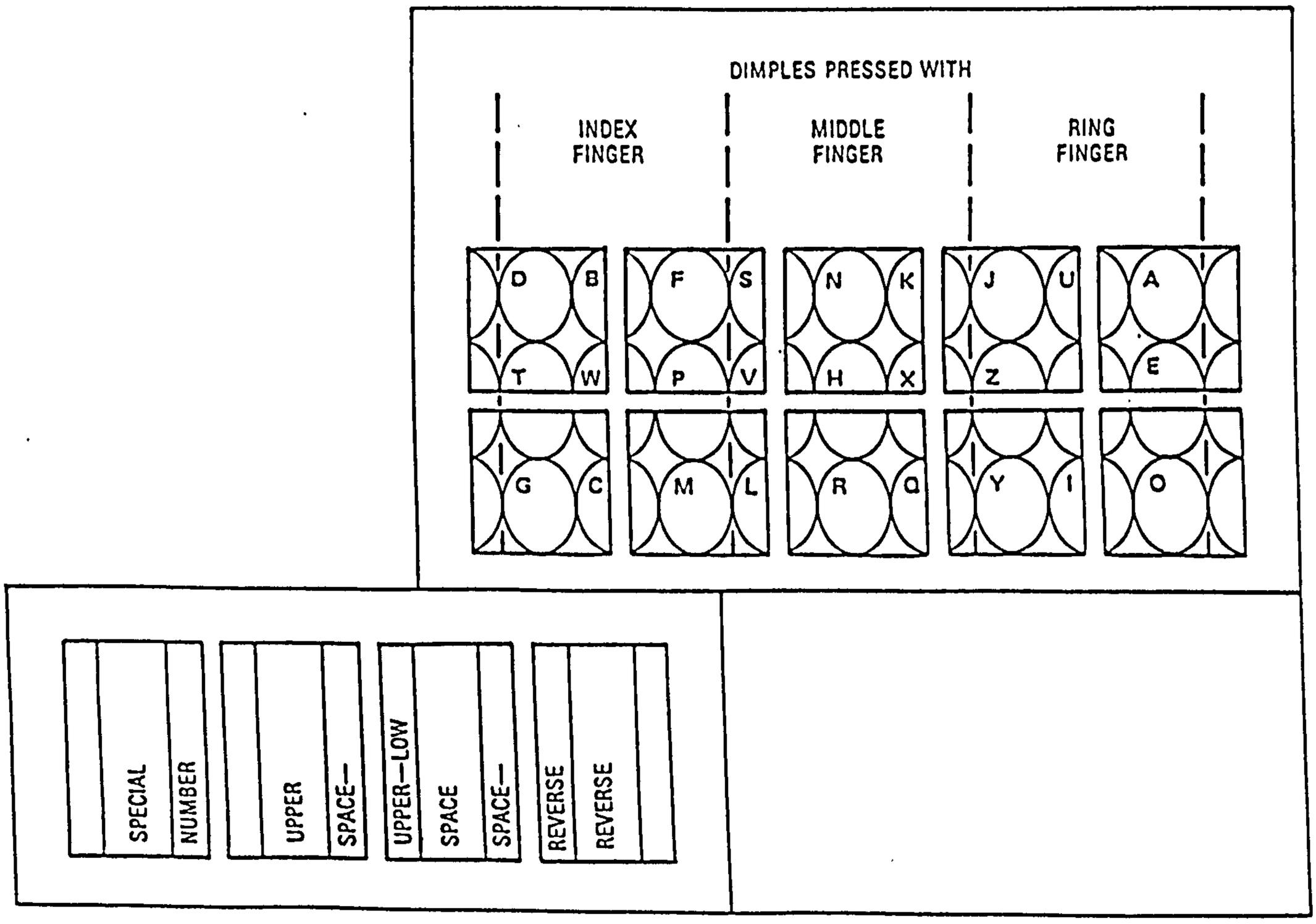


Figure 1.15 IBM Chord Keyboard

"(a) Only 12 keys are necessary to provide a full alphanumeric keyboard generating up to 39 characters.

(b) One-handed operation is possible.

(c) The compactness of the design results in a reduction of the number of hand movements without incurring higher mis-keying rates.

(d) No previous training is required.

(e) The simplicity of the chords avoids overloading the user's short term memory" (Martin, 1980, p. 14).

Rochester, Bequaert and Sharp (1978), working for IBM, developed another keyboard, similar in principle to the ANTEL. This keyboard consists of one bank of square keys arranged in a 2x5 array which are operated by the index, middle and ring fingers and a row of four rectangular keys operated by the thumb (figure 1.15). The finger keys have rounded depressions which are called "dimples". If the F dimple is depressed, the character F is typed. If the dimple S is depressed, two keys go down and the character S is typed. If the dimple V is depressed, four keys go down and the character V is typed.

The thumb keys form eight possible "troughs" (seven troughs produced by the depression of a single thumb key or the depression of adjacent thumb keys and one "no trough" produced by not depressing any thumb keys). By depressing various dimples and troughs it is possible to type single characters or sequences of characters. For example (Rochester et al., 1978, p. 57):

/I/n/ thi/s/ se/g/me/nt/ of/ te/xt/ the/ cho/rd/

bo/und/ar/i/es/ ha/ve/ be/en/ s/ho/wn/ by/ li/ne/s/.

Regarding the efficiency of its operation Rochester et al. (1978, p.61) report that "a number of students have learned to type at a gross typing speed of over 40 words per minute with low error rates" and that "... the initial experience of three programmer-engineers ... indicates that the user-defined-chord facility is extremely valuable to an on-line computer user and speeds up his work significantly".

Probably the most recent chord keyboard to be developed and marketed commercially is the Microwriter (Endfield, 1978, figure 1.16). It is a portable, battery-operated, chord keyboard which has five keys to be operated by the fingers and the thumb and a sixth key for entering capital letters. It also allows for functions such as back-space, insert, delete. The most frequent letters of the english alphabet are assigned to "easier" chord patterns and the less frequent to more "difficult" chord patterns.

Endfield claimed that it takes about 30 minutes of training for one to use this keyboard. His personal performance on the Microwriter after seven months of practice was 40 words/min.

1.7 Ergonomic Considerations in Keyboard Design

Human Factors Considerations

In the design or selection of a new keyboard (or any other data entry device), it is most important that certain specific ergonomic considerations are taken into account. As it has already been argued, "the optimal keyboard should allow

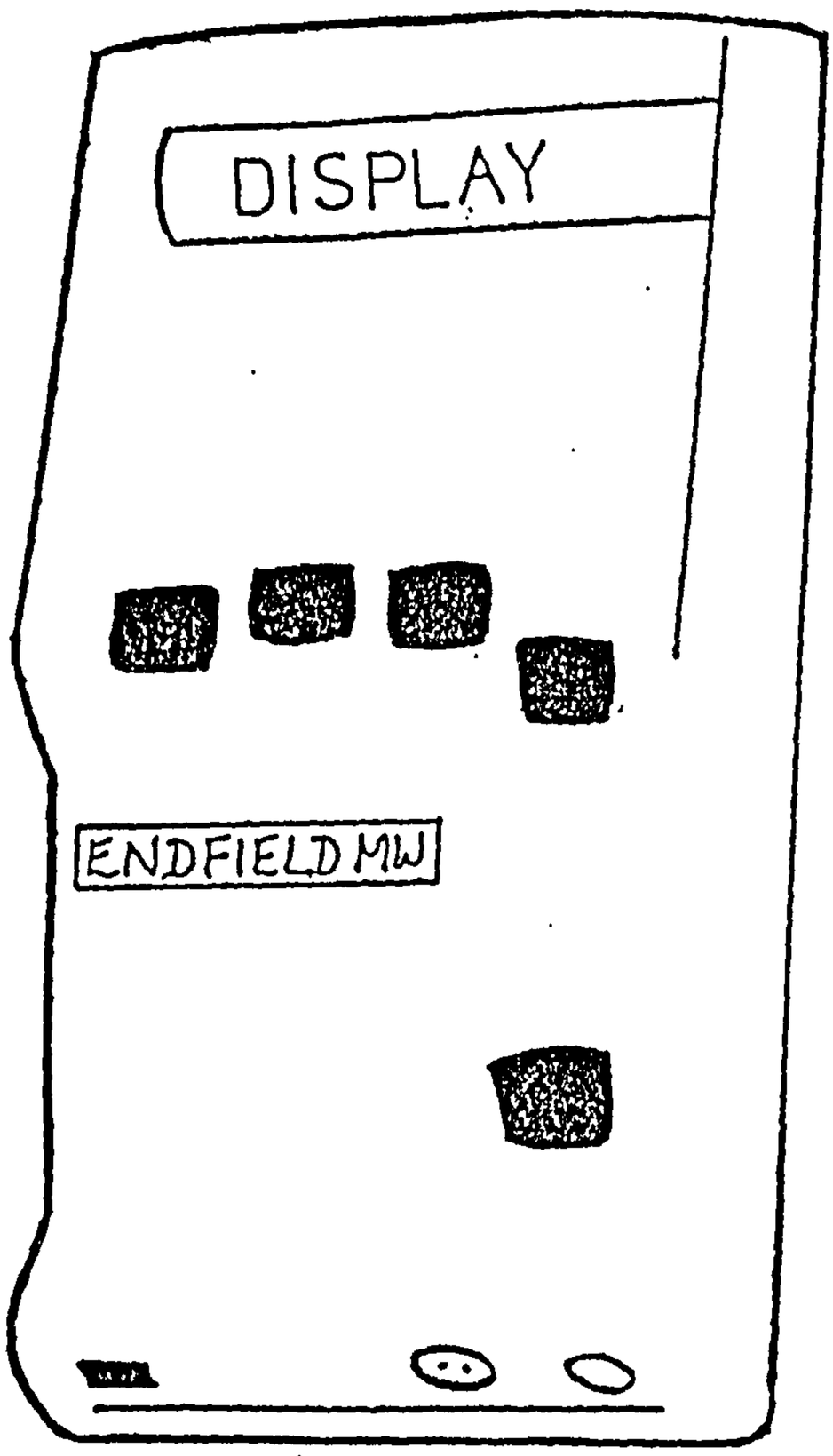


Figure 1.16 Microwriter (Endfield, 1978)

comfortable, high keying frequency and error free operation" (section 1.2). The term "comfortable" was used somewhat loosely in section 1.2. As "comfort" is a major factor in acceptability of a keyboard to operators it is discussed further here.

Kroemer (1972) summarised the results of a number of studies - by German researchers and published in German - on the effects of keyboard orientation on operator strain and level of performance. He explains that the posture of the arms which must be adopted for the operation of the standard keyboard, leads to "static muscle tension". Operators have to lift their elbows laterally and forward to relieve some of this tension. There is also more dynamic muscular work to move fingers between keys and depressing keys. All of these muscular activities, together with the "static muscle tension" create a considerable muscular strain which leads to poorer performance, as well as to pains and aches especially in the shoulders, arms, wrists and fingers. Although he does not present hard experimental evidence to show that performance is affected significantly in this sort of posture, his theoretical considerations are convincing. Kroemer (1972, p. 51) suggests that these postural constraints could be relieved by the following design recommendations: "(1) the keys should be arranged in a "hand-configured" grouping to simplify the motion patterns of the fingers; (2) the keyboard sections allotted to each hand should be physically separated to facilitate the positioning of the fingers; and (3) the keyboard sections allotted to each hand should be declined

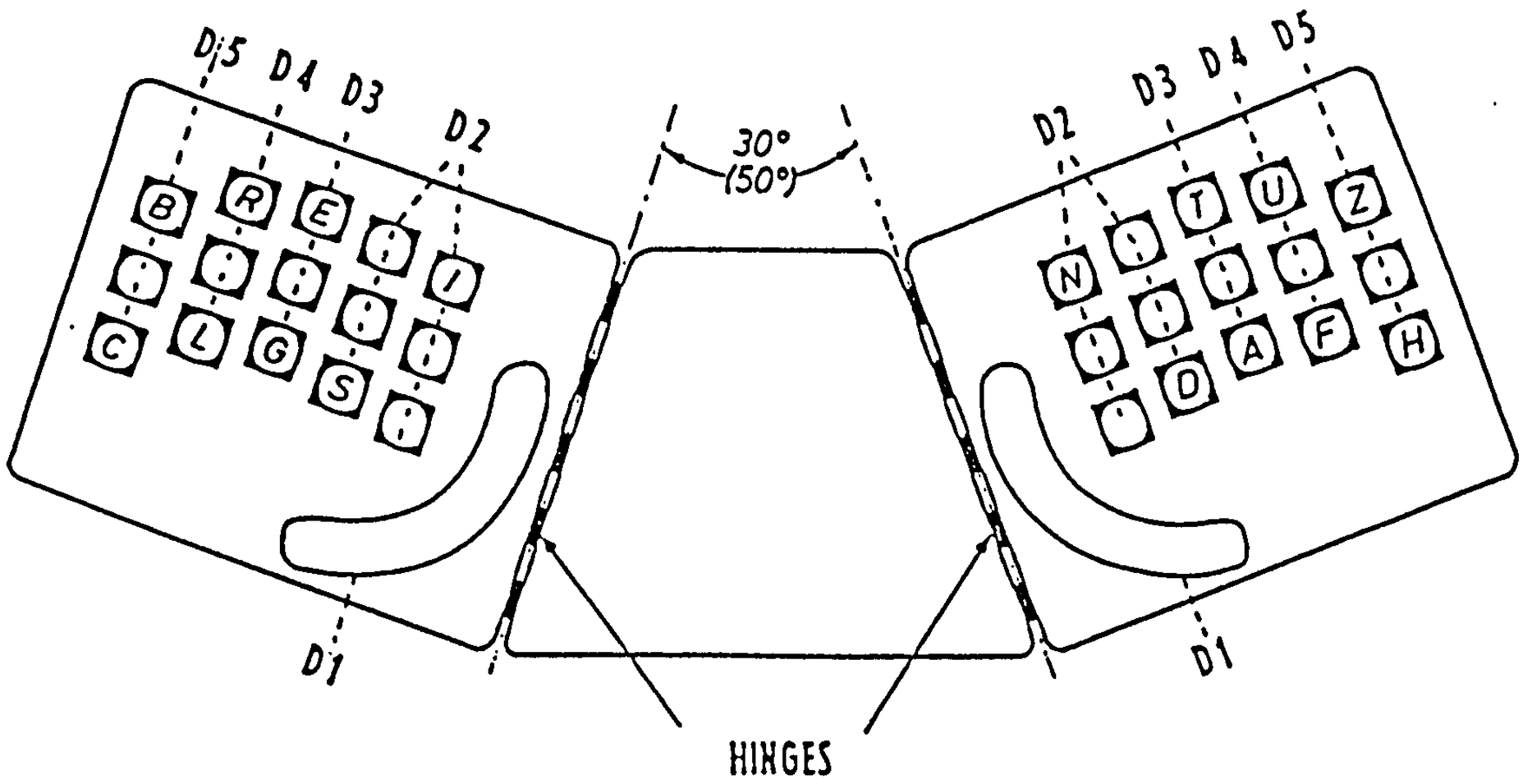


Figure 1.17 Keyboard Design Recommended by Kroemer (1972)

University
Library
Hull

laterally to reduce postural muscular strain of the operator" (figure 1.17).

Zipp, Haider, Halpern and Romhert (1983) investigated myoelectric activities of the muscles involved in typing and discussed the implications of their findings for keyboard design. They present a series of experiments and conclude that "An ergonomically designed keyboard should take into account the tolerable angles of the joints of the shoulders, arms and hands" (Zipp et al., 1983, p. 121). They recommend that, at least for full-time typists and data entry operators, the optimal keyboard should be split in two halves, with a lateral inclination of approximately 10 to 20 degrees for the right and the left halves.

The superiority of this type of inclined, split keyboard design (similar to that of the Maltron keyboard) has been discussed by Romhert and Luczak (1978) who used such a keyboard for the Postal Video Letter Coding systems, used at the German Central Post Office.

Human Operator Characteristics

Experimental evidence, so far, suggests the existence of certain factors affecting individual finger performance as well as individual key operation. Early experimentation by Gatewood (1920) on differences of performance between fingers, based on some 70,000 reactions resulted in the following conclusions:

(1) There are measurable differences between the reactions of fingers in terms of both speed and accuracy.

(2) The fingers of the right hand (for right-handed people) are faster and slightly more accurate than the fingers of the left hand.

(3) The reaction time of any finger varies according to which other finger it is used in conjunction with.

(4) Two-finger reactions (but not in a choice situation) are faster and more accurate than single-finger reactions.

(5) In two-finger combinations, RT is faster when the two fingers belong to different hands.

(6) Practice increases the speed and accuracy of all fingers.

(7) Individuals are different in terms of speed and accuracy. Trained subjects showed less variance between fingers than untrained subjects.

Although Gatewood's claims would not be fully accepted today, her study was very important in that it conveyed some basic ideas such as: there are differences (which are physiologically determined) in the reactions of individual fingers; individual operators are different in terms of speed and accuracy.

Dvorak, Merrick, Dealey and Ford (1936) measured the maximum tapping rate of each finger. They found that the index finger had the highest tapping rate, the middle finger was second, the ring finger third, and the little finger fourth. They also found that the tapping rate for the right hand was higher than that for the left hand (table 1.1).

Hand	Index	Middle	Ring	Little
Left	66	63	57	48
Right	70	69	62	56

Table 1.1 Maximum tapping rate (over a 15 secs period)

Haaland, Wingert and Olson (1963) measured the maximum finger-pushing force and found that the index finger was the strongest, followed by the middle finger, the ring finger and the little finger (table 1.2). Haaland (1962) reports that different fingers have different resistance to fatigue. The susceptibility to fatigue increased from the thumb to the little finger.

	Thumb	Index	Middle	Ring	Little
Mean	37	24	22	18	12
Range	30-43	17-31	17-26	12-22	6-19

Table 1.2 Maximum finger pushing force (lbs)

More recently, Hayes, Wilson and Schafer (1977) studied the relationship between reaction time and typewriting rate of skilled typists. They concluded that typewriting rate was related to three main factors: (1) the reaction time of the fingers, (2) the difference between the reaction time of the hands, and (3) the reaction time to individual keys on the typewriter.

Hayes and Halpin (1978) measured the reaction time of the fingers of 24 skilled typists in a typing task. They found a small but significant difference [$F(7,161)=2.87$, $p<.01$] in reaction time among fingers, across subjects. The rank order

of the mean finger reaction times, from slowest to fastest, was: L4, L3, L1, L2, R1, R2, R3, R4 (the fingers were named, starting from the little finger of the left hand to the little finger of the right hand, L4, L3, L2, L1, R1, R2, R3, R4 respectively). It seems at least surprising that the little finger of the right hand was the fastest! They also report that the left hand was significantly slower than the right hand and that there was a consistency of reaction times across the fingers of individual typists.

Specific Keyboard Characteristics

In the case of the conventional keyboards (i.e. flat keyboards like ordinary typewriters), it is recommended (Snyder, 1984) that keyboards are thin, less than than 30 mm from the base of the keyboard to the top of the second key row (A, S, D,...). However, Snyder adds, even slimmer keyboards should be preferred for ensuring maximum flexibility in workplace design, if all other ergonomic criteria can be fulfilled.

The size of the keytops should be a compromise between enough space for comfortable positioning of fingers on keys and total size of the keyboard. On one hand, keys should be big enough so that they are operated comfortably and, on the other, the size of the keyboard should not be too big. The tops of keys for touch-typewriters will probably be sufficiently big if they are square and 12-15 mm (Snyder, 1984). However, for other keyboards with fewer keys (e.g. keyboards for Letter Sorting Machines) keytops could be

bigger.

Alden et al. (1972) recommend that the force required for key displacement should be between 25.5 and 150.3 gr (0.9 to 5.3 oz). Furthermore, they recommend a key displacement (travel) of 1.3 to 6.4 mm (0.05 to 0.25 in.).

Bearing all these considerations in mind, the aim of the study described in the following two chapters was to provide data to assist the selection of a new keyboard for Foreign Destination Letter Sorting Machines at the British Post Office.

Chapter 2

PILOT EXAMINATION OF EXPERIMENTAL KEYBOARDS

Three new keyboard designs and the one still in use were examined so as to select a new keyboard for the Post Office foreign destination coding desks. The three new designs were the product of a cooperation between ergonomists at the University of Hull and Post Office engineers. The present experimental project examined, evaluated and compared the three new keyboard designs and the existing one. The aim of the project was to provide information needed to decide which one should replace the keyboard currently in use.

2.1 Description of the Four Keyboards

1. 2x6 Chord Keyboard

This keyboard, which is the one currently in use by the Post Office, consists of two rows of six keys for each hand (figure 2.1). It is operated by the simultaneous depression of two keys, one by each hand giving a total of 144 (12x12=144) combinations. The bottom row is the "home" row over which fingers are positioned. Fingers are placed over the four middle keys of the "home" row and subjects move them upwards to reach the keys of the top row. The little finger must also move to one edge of the key bank to depress two keys on that side whereas the index finger must also move to depress the two keys on the other side of the key bank.

2. 4x4 Chord Keyboard

It consists of four rows of four keys for each hand

2-by-6

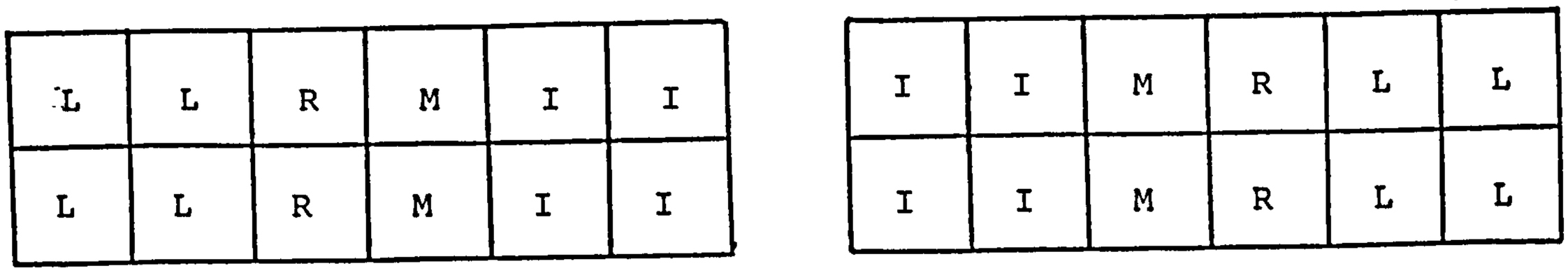


Figure 2.1 2x6 Chord Keyboard and Finger Position on Individual Keys (L=Little, R=Ring, M=Middle, I=Index)

4-by-4

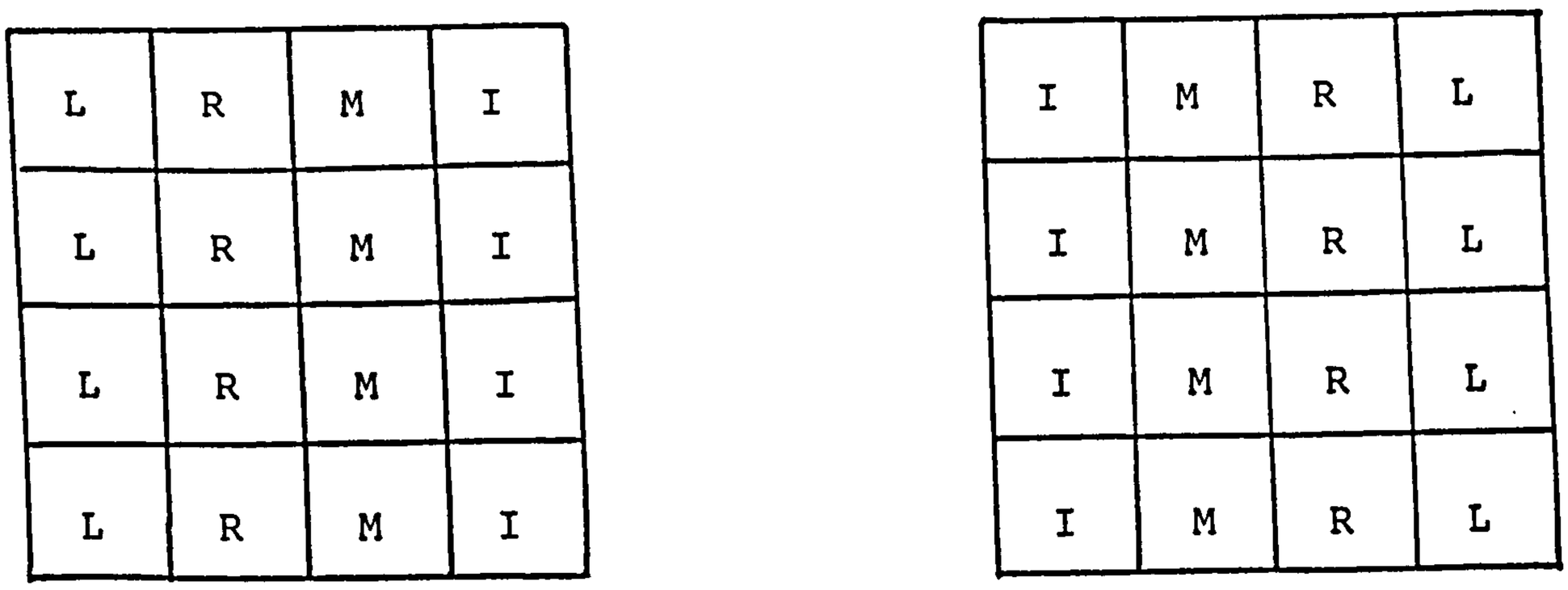


Figure 2.2 4x4 Chord Keyboard and Finger Position on Individual Keys (L=Little, R=Ring, M=Middle, I=Index)

Sequential and Chord

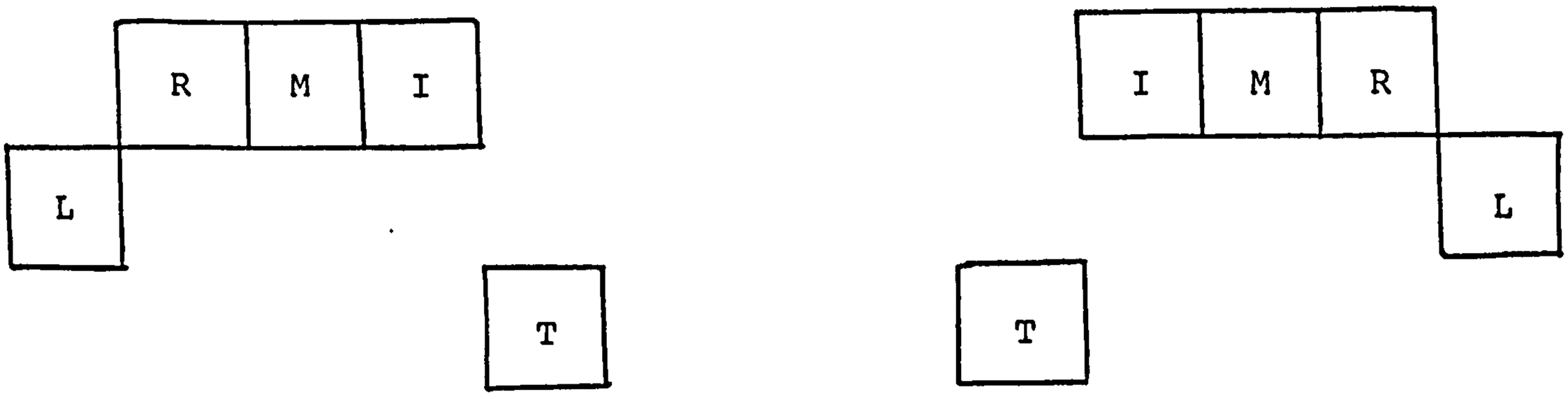


Figure 2.3 10-Key Sequential or 10-Key Chord Keyboard and Finger Position on Individual Keys (L=Little, R=Ring, M=Middle, I=Index, T=Thumb)

(figure 2.2). Its operation, like that of the 2x6, requires the simultaneous depression of two keys, one by each hand giving a total of 256 (16x16=256) combinations. The row above the bottom one is designated the "home" row. Fingers move upwards and downwards from the "home" row to reach the rest of the keys.

3. 10-Key Chord Keyboard

This has five keys for each hand arranged in a hand configured layout, so that all fingers and the thumbs can be used (figure 2.3). Its operation involves the simultaneous depression of from one to all ten keys. It gives a maximum of 1,023 different key-combinations ($2^{10}-1$).

4. 10-Key Sequential Keyboard

The layout of the keys in this keyboard is identical to that of the 10-key chord keyboard, but its operation is different. It requires the depression of keys in a sequence. Key sequences for this keyboard could consist of from 1 to an infinite number of keys - like an ordinary typewriter. In ordinary text the end of a word is signalled by a space or punctuation. In computer programs the end of a command is signalled by the depression of the RETURN key. In the 10-key sequential keyboard the end of a key sequence could be shown in two ways. First, it could be a "fixed length sequence" (e.g. a sequence of three keys per combination); second the number of keys per sequence may vary and an "enter" key could be depressed to signify the completion of the key sequence. In the second case the effective number of keys would be reduced to nine, since one key has to be used as an "enter" key. The

variable number of keys per combination has the advantage of allowing short keys sequences for frequently used codes but also the disadvantage of requiring an extra key depression.

2.2 The Need for a Pilot Examination

Prior to the comparison and evaluation of the four keyboard designs another aspect of the 10-key chord and the 10-key sequential keyboards had to be examined. Bearing in mind that the number of postal destinations has risen from 144 to over 230, and that the 4x4 chord keyboard had a maximum of 256 key combinations, it was decided that 256 key combinations should be used for all three of the new designs. However, as mentioned above, the 10-key chord keboard gives a total of 1,023 combinations and the 10-key sequential keyboard gives a theoretically infinite number of key combinations. It is well accepted in keyboard literature that some key combinations are "easier" than others, in the sense that some key combinations will be processed faster and more accurately than others (e.g. Ratz and Ritchie, 1961; Seibel, 1963, 1964a). In the cases of these two keyboards, there was the option of selecting the "easiest" key combinations. Therefore, two pilot experiments were carried out to select the 256 "easiest" chord patterns (for the 10-key chord keyboard) and the 256 "easiest" sequences (for the 10-key sequential keyboard).

2.3 Displays

Highly compatible stimulus displays were designed for each keyboard so that any performance differences could be

attributed to the effects of response generation and execution. For the chord keyboards each stimulus display was a "picture" of the keyboard on the screen. The keys of the keyboard were represented on the computer screen by square boxes. For each stimulus the boxes that corresponded to the keys that had to be pressed appeared "filled" (figure 2.4).

This way of displaying stimuli was not possible for the sequential keyboard, so a somewhat less satisfactory display was used. The stimulus that appeared on the screen consisted of a string of three digits that corresponded to the respective fingers (figure 2.5). The fingers were numbered from the little finger of the left hand to the little finger of the right hand: from 1, 2, 3 ... 9, 0.

2.4 Apparatus

Each one of the keyboards was connected to a BBC model B microcomputer, with 40-track single disk drive, and green-screen Sanyo monitor.

When a response was correct, a new stimulus appeared after a 500 msecs pause. (The occurence of an error was treated differently in different experiments and is discussed later when the experiments are presented.) Response times were measured from the moment the stimulus was presented until the response was completed.

It cannot be assumed that in the case of chord keyboards all keys which must be depressed are depressed exactly simultaneously. Therefore, measuring reaction time posed a problem to which two solutions exist. First, one may use a

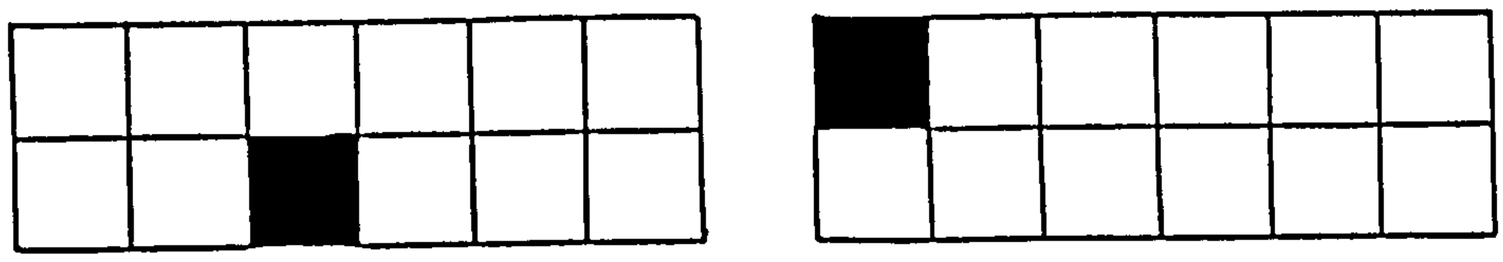


Figure 2.4 Examples of Stimuli for the Chord Keyboards

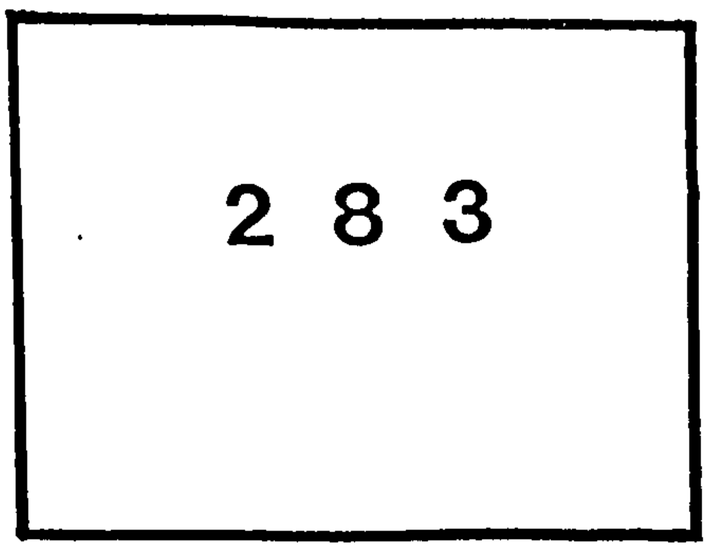


Figure 2.5 Example of Stimulus for the 10-Key Sequential Keyboard

time interval of say 50 msecs from the moment one of the keys of the chord pattern was depressed, within which all the rest of the keys of the chord pattern must be depressed. Second, one may use the release of keys as the signal that the response is completed.

The first solution, which was used by Conrad and Longman (1965), is complicated by the fact that some chord patterns are much "easier" than others and therefore the time interval which is selected may be too short for some chord patterns and too long for others. If the time interval is too long, the reaction time recorded for some "easy" stimuli is unnecessarily extended by the fact that the time interval has to elapse before the response is recorded. If the time interval is short, a number of errors may be recorded as a result of the fact that all keys are not depressed within the given time interval. However, these errors would be an artifactual component into the accuracy measures.

The second solution to the problem of recording reaction times when using chord keyboards was adopted by Ratz and Ritchie (1961). The problem with this solution, however, is that subjects may choose to depress keys one after the other, in a sequential fashion, without releasing any keys until the whole chord pattern is depressed. This implies that reaction time may be longer than when all keys are depresseed at once. However, if the instructions make it clear to subjects that they must depress all keys simultaneously, it is logical to assume that they will do otherwise only when they have "difficulties" with a particular chord pattern. Therefore it

is justified that their reaction time to those stimuli is longer.

Bearing all the foregoing in mind, the second solution to the problem of recording reaction times when using the chord keyboards was adopted.

This apparatus and these displays were used in all the experiments that followed.

2.5 Experiment 1

A PILOT INVESTIGATION OF THE 10-KEY CHORD KEYBOARD

This pilot experiment aimed, primarily, at collecting some performance data during the first stage of practice on the 10-key chord keyboard so as to assess the relative difficulty of various chord patterns. The latter was important for experiment 3 in which the four keyboards were compared and evaluated. The questions that were raised included: the rank order of various chord combinations in terms of difficulty as expressed in measures of reaction time and error rate; the examination of the difficulty of various chord combinations as a function of the number of fingers required for each chord pattern and also as a function of finger separation (a finger separation would be a chord pattern that involved the depression of non-adjacent keys); the establishment of possible differences in efficiency between the left and the right hand.

Ratz and Ritchie (1961) experimented with a 10-key chord keyboard, one key for each finger, in conjunction with a compatible 10-light display. They ranked all possible 31 one-handed chords (2^5-1) in terms of difficulty as measured by reaction time (i.e. the longer the RT the more difficult the chord is). Ratz and Ritchie do not explain what instructions they gave to their subjects, except that the task was "self-paced". They used six subjects, three of whom were trained on two-handed chords and the others on one-handed chords.

Subjects were given relatively little practice, and the authors' results are based on 2,660 responses. They found that the average RT for two-handed chords was 2.63 secs and the average RT for one-handed chords 1.20 secs. Surprisingly, with regard to speed, they report that "... little improvement took place after the second day" and with regard to accuracy, that "... because self-pacing was used, the number of erroneous responses made was negligible" (1961, p. 304).

Seibel (1962) replicated their experiment but found that improvement continued to take place after 4,000 to 11,000 trials. He gives both Discrimination Reaction Times and number of errors for each chord, and argues that subjects' performance, in terms of both speed and accuracy, continues to improve for much longer than Ratz and Ritchie suggest. After at least 50,000 trials by each subject, their RT ranged between 0.281 secs and 0.352 secs for the 31 chord patterns; the error rates ranged between 1.8% and 25.9%. Seibel's argument is supported by the results of many choice reaction time experiments (e.g. Mowbray and Rhoades, 1959; Cornog et al., 1963; Conrad and Longman, 1965). It seems that in the Ratz and Ritchie experiment subjects may have traded speed for accuracy and quite quickly established a relatively slow rate of response that produced a very small number of errors.

Furthermore, Seibel (1962) reports that there was a significant correlation beteween RTs and error rates (the higher the RTs the higher the error rates and vice versa). Seibel (1963) ran another experiment, using three subjects (himself and two research assistants). Subjects were trained

on all 1,023 chord patterns. Seibel (1963) reports that improvement was still continuing after 50,000 trials. The average RT for the full set of 1,023 chords was about 0.4 secs and for the one-handed chords it was 0.03 secs faster.

One major implication from both of the above experiments is that the various chords for the 10-key chord keyboard are not of equal difficulty. In the Ratz and Ritchie (1961) study, response times range from 0.94 secs for one-key chord patterns to 2.63 secs for multiple-key chord patterns and the number of errors was "negligible". In Seibel's (1962) rank order of the 31 possible one-hand chord patterns, RTs range from 281 msecs to 352 msecs and the percentage of errors for the various patterns ranges from 1.8% to 25.9%. In the light of this evidence it can be safely concluded that there are substantial differences in difficulty of various chord patterns. If both hands are used to operate this keyboard there are 1,023 possible combinations. Consequently, unless all 1,023 chord patterns are needed, there is the option of selecting the subset of the "easiest" chord patterns.

The research of Ratz and Ritchie (1961) and Seibel (1962, 1963) produced similar results in terms of the relative difficulty of chords (see below). However, their findings on the effect of practice on RT and error rates were very different. These inconsistencies most likely reflect methodological differences. Although there is no mention of the kind of instructions given to subjects it is quite apparent from the performance data that are reported that Ratz

and Ritchie's (1961) subjects had more lax criteria for speed and stricter criteria for accuracy than Seibel's (1962, 1963) subjects (i.e. different speed-accuracy tradeoff functions seem to have been adopted by the subjects). Furthermore, Seibel (1962, 1963) trained subjects for much longer than Ratz and Ritchie (1961) did. These differences make direct comparisons between their studies difficult if not inappropriate. Therefore new data were needed to answer the questions outlined above regarding the relative difficulty of various chords, the effect of the number of fingers involved in each stimulus and the effect of finger separation on speed and accuracy of perfomance.

METHOD

Subjects

Fifteen subjects, male and female students and employees from the University of Hull, were used. They had no prior experience with a 10-key chord keyboard.

Apparatus and Displays

The apparatus and displays that were used in this experiment have already been described.

Task Description, Stimuli and Feedback

The task involved the presentation of stimuli on the screen requiring the subjects to depress the relevant keys quickly and accurately. Subjects were given a full demonstration of the task before they started their sessions and were instructed to pay equal attention to both speed and

accuracy.

From the 1,023 possible key combinations only 288 were selected. The selection was based on Ratz and Ritchie's (1961) experiments. From all 31 possible one-handed chords, 16 were selected in this experiment, namely those that had an odd rank number in Ratz and Ritchie's rank order of difficulty classification (i.e. those ranked 1, 3, 5, 7, ..., 31). The reason for the elimination of 15 chord patterns was that this was a small-scale experiment and it was not feasible for all possible key combinations to be tested. The choice of the key combinations, however, was such that it allowed the examination of the whole range of chord difficulty. The 16 chords that were chosen were used for each hand separately and, in combination, for both hands simultaneously (i.e. 16 chords for the left hand + 16 chords for the right hand + 16x16 chords for both hands = 288 chords). Stimuli were presented once each in a random order.

If an error occurred there was auditory and visual feedback. The auditory feedback was an alerting bleep; the visual feedback was the display of the keys that were pressed correctly, those pressed by mistake, and those that should have been pressed and had not been (figure 2.6).

At the end of each session subjects received further feedback about their performance in the form of total number of correct responses, mean reaction time for correct responses, total number of errors, and mean reaction time for error responses.

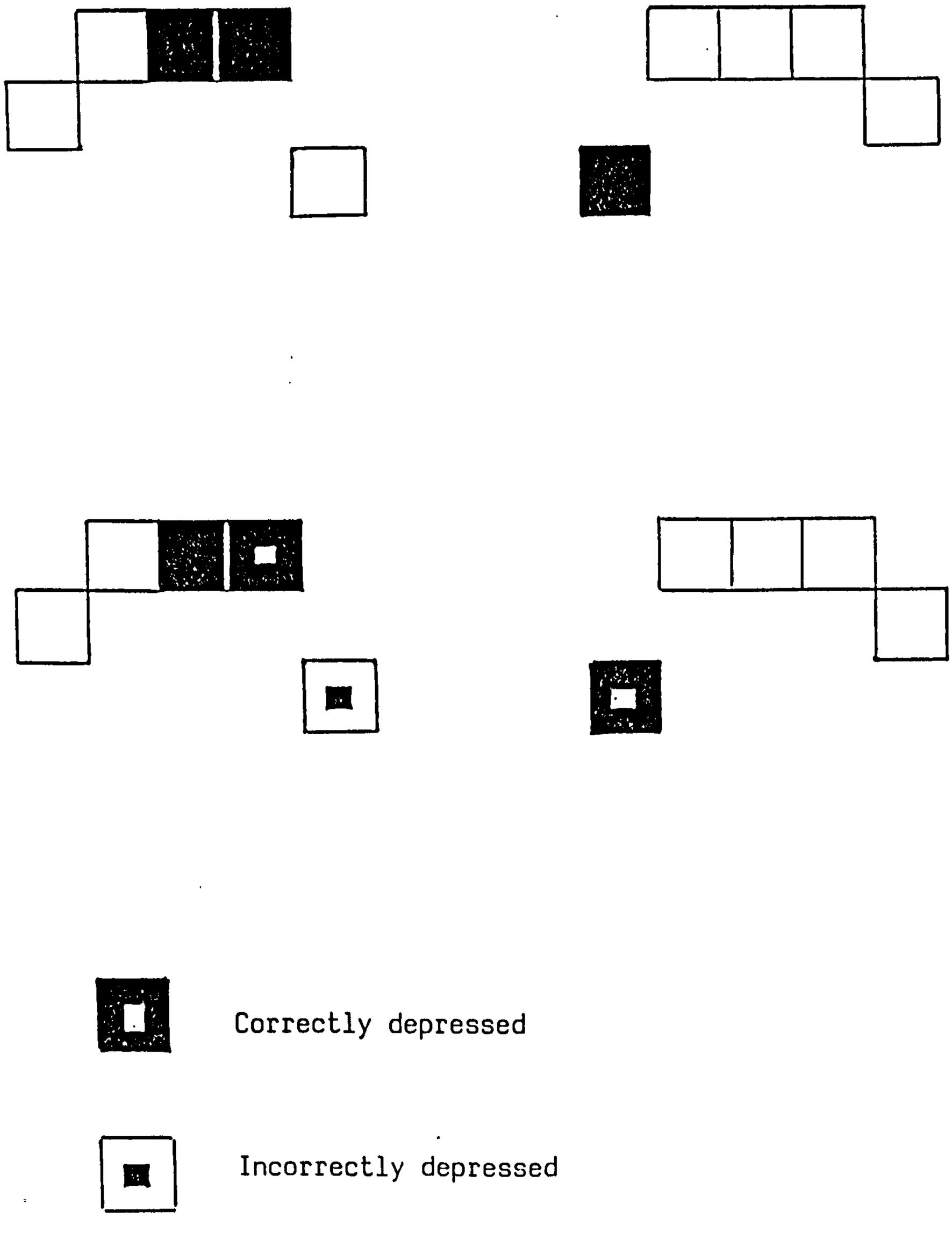


Figure 2.6 Type of Stimulus and Visual Feedback for the 10-Key Chord Keyboard

Sessions

Every subject had four sessions of 72 trials with a 5-minute break between sessions.

A full record of all stimuli, responses and reaction times was kept.

Statistical Analysis

For each hand, the rank order of the 16 one-hand chords (averaging across all chord patterns in the other hand) was produced and correlated with the rank orders obtained by Ratz and Ritchie (1961) and Seibel (1962). A Pearson correlation was carried out to examine the relationship between the performance (that is RT and error rate) of the left hand and that of the right hand. Another correlation matrix was produced to examine the relationship between the number of fingers (required for the depression of each chord pattern) and keying performance. A final correlation matrix was produced to examine the relationship between finger separation (a chord pattern with finger separation involved the depression of non-consecutive keys for at least one hand) with performance.

RESULTS

Effects of Chord-Pattern Difficulty

All sixteen one-handed chord patterns that were used are shown in table 2.1. The results shown for each chord pattern are averages across all trials in which these patterns

occured, either in one hand alone or in conjunction with another chord pattern in the other hand. The table presents the mean reaction times for each chord pattern, for all subjects, for both the left and right hand and the standard deviation for each mean. It also gives the percentage of errors for each chord. Moreover, it includes the rank order by difficulty from Ratz and Ritchie (1961) and Seibel (1962) as well as the mean reaction times and percentage of errors from Seibel's results. The chord patterns that are shown refer to the right hand. The keys that had to be pressed are indicated by "1" and the others by "0". For each chord pattern, the "0's" and "1's", correspond, from left to right, to the thumb, index finger, middle finger, ring-finger and little finger. The left hand patterns are a mirror image of those for the right hand. Finally, the table shows the number of fingers (NF) and whether finger separation was required for the depression of each chord pattern (SEP=1) or not (SEP=0).

The rank order of the 16 patterns that was obtained from this experiment is very similar to those of Ratz and Ritchie (1961) and Seibel (1962). Both RT and error rate are highly correlated with the rank orders of Ratz and Ritchie and Seibel (table 2.2).

CHORD	LEFT HAND			RIGHT HAND			R&R	SEIBEL			
Right Hand	RT	SD	ERR	RT	SD	ERR	RANK	RANK	RT	ERR	NF	SEP
00001	2.58	1.31	7.06	2.41	1.13	3.53	1	5	.294	5.9	1	0
00100	2.58	1.17	6.27	2.69	1.34	5.10	3	2	.285	2.4	1	0
01000	2.63	1.17	10.20	2.61	1.32	9.41	5	4	.292	5.0	1	0
10001	2.73	1.38	5.47	2.82	1.31	10.59	7	14	.316	5.6	2	1
11100	3.05	1.27	9.80	2.92	1.27	7.06	9	13	.315	5.3	3	0
00110	2.91	1.75	9.02	2.69	1.13	9.41	11	8	.306	10.3	2	0
11111	3.16	1.96	10.59	2.96	1.35	8.24	13	20	.325	7.4	5	0
11110	2.93	1.28	8.24	3.20	1.50	12.16	15	12	.314	4.1	4	0
10100	2.94	1.25	7.84	2.85	1.15	11.37	17	11	.312	5.0	2	1
00111	3.02	1.31	12.16	3.13	1.30	9.02	19	19	.321	7.6	3	0
11001	3.23	1.47	8.23	3.21	1.28	8.24	21	25	.335	11.5	3	1
01001	3.00	1.46	9.80	2.91	1.26	7.06	23	21	.326	12.4	2	1
00101	3.18	1.44	9.02	3.02	1.36	14.90	25	26	.343	13.2	2	1
11010	3.23	1.44	9.02	3.18	1.30	12.16	27	25	.328	13.2	3	1
11011	3.74	1.67	16.86	4.10	2.50	14.12	29	30	.351	25.9	4	1
01011	3.43	1.56	17.65	3.60	1.70	14.51	31	29	.349	20.9	3	1

Table 2.1 Mean Reaction Time (RT), Standard Deviation (SD) and percentage of errors (ERR) for each hand for each chord pattern. Also shown are the rank orders obtained by Ratz and Ritchie, 1961 (R&R) and Seibel, 1962, together with Seibel's RTs and error rates. (All RTs in secs)

	LEFT HAND		RIGHT HAND		R&R	SEIBEL				
	RT	ERR	RT	ERR	RANK	RANK	RT	ERR	NF	SEP
	1	2	3	4	5	6	7	8	9	10
2	.77									
3	.92	.78								
4	.66	.51	.68							
5	.88	.66	.81	.74						
6	.90	.62	.81	.63	.89					
7	.93	.66	.83	.71	.91	.98				
8	.87	.80	.82	.61	.82	.83	.83			
9	.70	.47	.66	.39	.49	.58	.61	.36		
10	.54	.20	.49	.57	.71	.71	.68	.60	.06	

Table 2.2 Correlation matrix among the variables
1: Mean RT for the left hand
2: Mean percentage of errors made by the left hand
3: Mean RT for the right hand
4: Mean percentage of errors made by the right hand
5: Rank order of stimuli by Ratz and Ritchie, 1961
6: Rank order of stimuli by Seibel, 1962
7: Mean RT from Seibel's (1962) data
8: Average error rate from Seibel's (1962) data
9: Number of fingers required for the appropriate response
10: Finger separation

The high correlations (0.77, 0.68) between the RT for each hand and the error rate shows that the smaller the RT the smaller the error rate and vice versa. This difference in speed suggests that there was no "between patterns" effect of speed-accuracy tradeoff.

Differences between Chords Involving One or Both Hands

Two analyses of variance were carried out to examine whether there was a difference in RT and/or error rates between the left and the right hand and also differences in RT and/or error rate between one-handed and two-handed chord patterns. The comparison between the left and the right hand

was based on the data for the one-handed chords only. No significant difference ($F<1$) between the left and the right hand was found in either RT or error rate data (table 2.3). There was a significant difference between one-handed and two-handed chords with responses to one-handed chord patterns being significantly faster [$F(1,14)=80.50$, $p<.01$] and more accurate [$F(1,14)=20.43$, $p<.01$] than responses to two-handed patterns (table 2.3).

	N	RT	SD	ERR
One-handed chords				
Left hand	240	1.82	.873	4.17
Right hand	240	1.77	.787	3.75
Two-handed chords	3,840	3.02	.023	9.82

Table 2.3 Comparisons between left-hand and right-hand chords as well as between one-handed and two-handed chords (N=number of trials, RT=reaction time, SD=standard deviation, ERR=percentage of errors; RT in secs)

Effects of "Number of Fingers" and "Finger Separation"

A factor that appears to contribute to the difficulty level of various chord patterns is the number of fingers involved in each such pattern. The reaction time and the error rate of both hands are highly correlated with the number of fingers needed for each chord (table 2.2). That is the smaller the number of fingers required for each pattern the shorter the reaction time, and the smaller the number of errors.

A relatively systematic increase in reaction time and error rate, for both hands, can be observed with the increase in number of fingers involved in a chord pattern (table 2.4).

The only exception to that is the five-finger chord pattern which produces the third most efficient measures of speed and accuracy. This observation about differences among chord patterns as a function of the number of fingers required for the depression is also supported by the high correlations between number of fingers and the rank orders from both Ratz and Ritchie, and Seibel (table 2.2).

No of Fingers	RT	% of Errors	Seibel RT	Seibel % of Err
1	2.59	6.93	.288	4.14
2	2.90	9.45	.315	8.89
3	3.20	10.78	.331	12.38
4	3.49	12.84	.331	13.12
5	3.06	9.41	.325	7.40

Table 2.4 Relationship between the number of fingers required for the depression of a stimulus and mean RT and percentage of errors from Experiment 1 and Seibel's (1962) data (RT in secs)

Another factor affecting the difficulty level of various chords is finger separation. For both hands, the correlations between reaction times and finger separation are high (table 2.2) which shows the tendency for chord patterns with no finger separation to produce shorter reaction times. A high correlation was found between the error rate of the right hand and finger separation whereas the correlation between the error rate of the left hand and finger separation is smaller (table 2.2).

It appears that the chord which involves a 3-finger separation (i.e. when just the little finger and the thumb have to be depressed) has, on average, the shortest reaction

time for both the right and the left hand. Overall though, it seems that there is no systematic relationship between degrees of finger separation and either RT or error rate (table 2.5).

Finger Separation	RT	% of Errors	Seibel RT	Seibel % Err
0	2.84	8.58	.305	6.36
1	3.33	12.75	.336	15.42
2	3.09	8.33	.324	10.24
3	2.77	8.04	.315	5.60

Table 2.5 Relationship between finger separation (the numbers denote the degree of finger separation) and mean RT and percentage of errors from the data of Experiment 1 and Seibel's (1962) data (RT in secs)

Prediction of Speed and Accuracy

Multiple regression analyses were carried out to formulate prediction equations for reaction time and accuracy. It was found that the variables "number of fingers" (NF), "finger separation" (SEP), and their interaction were good predictors of RT. The variables NF and NF*SEP were good predictors of percentage of errors (ERR). The prediction equations obtained were:

Predicted RT(secs)=2.492+0.1401(NF)-0.511(SEP)+0.323(NF*SEP)

(Standard error = 0.149 secs)

The multiple correlation (R=.92) was statistically significant ($F=49.53$, $df=3,28$, $p<.01$). This implies that 85% of all variance in RT can be accounted for by the "number of fingers", "finger separation" and their interaction. The contributions of each variable to the prediction equation were also significant ($F=28.2$ for NF, $F=9.68$ for SEP and $F=29.32$ for NF*SEP; df 1,28, $p<.01$ in each case). From the prediction

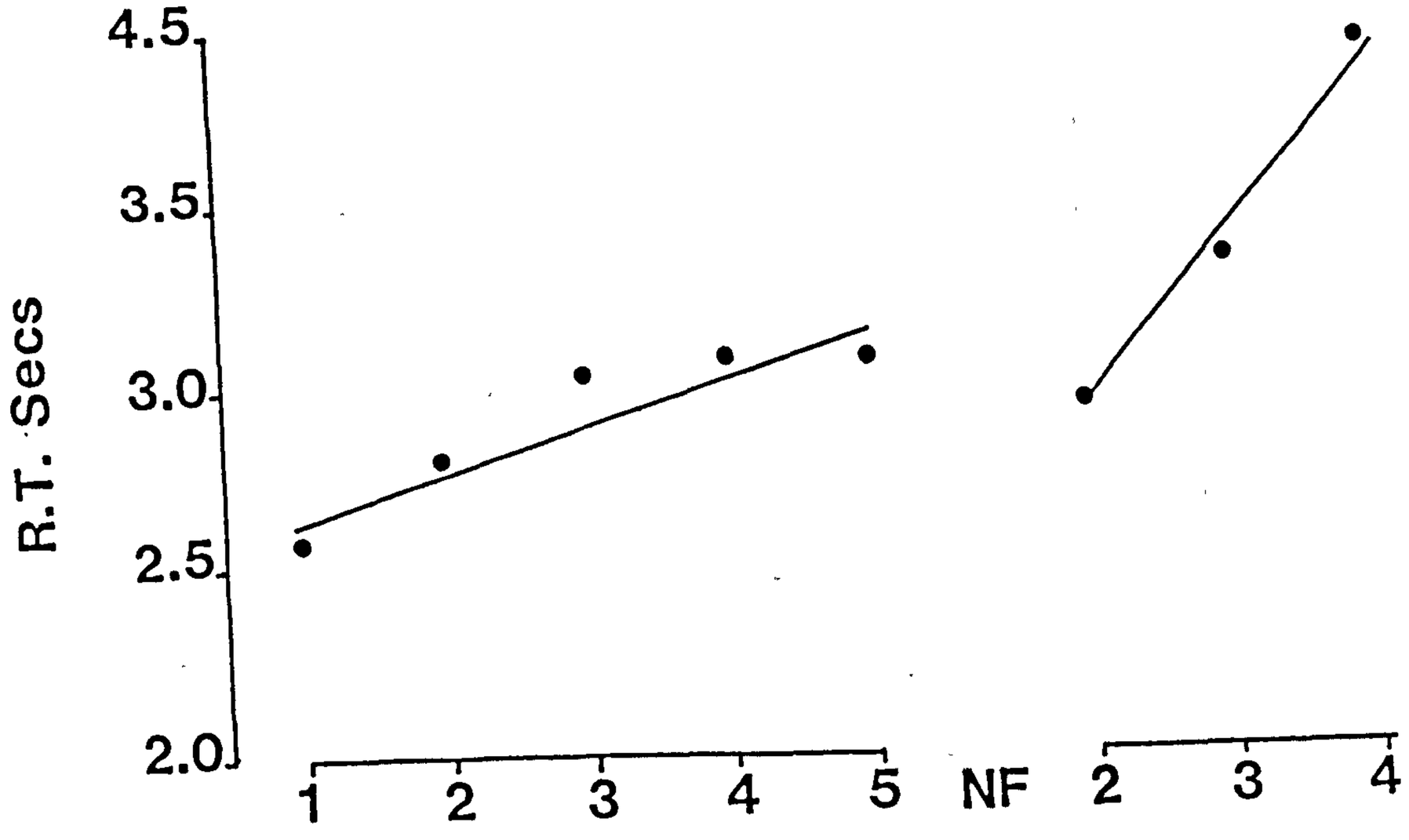


Figure 2.7 Obtained and predicted mean RTs for SEP=0 and SEP=1 as a function of the number of fingers involved in each chord (SEP= Finger separation)

equation, it can be concluded that the slope of the function relating NF to RT is steeper when SEP=1 than when SEP=0 (figure 2.7). For SEP=0, RT=2.492+0.1401(NF) (NF can range from 1 to 5) while for SEP=1, RT=1.981+0.463(NF) (NF can range from 2 to 4).

If the obtained regression lines are plotted for SEP=0 and for SEP=1 together with the actual RTs for the various values of NF, a difference in the two lines becomes apparent. The regression line is a much better fit when SEP=1 than when SEP=0 (figure 2.7). It seems that when SEP=1, the function is linear, whereas when SEP=0, the function is curvilinear.

The prediction equation for percentage of errors is not as good as that for RT ($F=8.34$, $df=2,29$, $p<.01$).

Predicted percentage of errors ERR =

6.031+1.01(NF)+1.01(NF*SEP)

(Standard error = 2.69)

The multiple correlation was R=.60 which implies that 36% of all variance in the percentage of errors can be accounted for by the "number of fingers" and the interaction of "number of fingers" and "finger separation". For SEP=0, ERR=6.031-1.01(NF), with NF ranging from 1 to 5, while for SEP=1, ERR=6.031-2.02(NF), with NF ranging from 2 to 4.

DISCUSSION

The rank order of difficulty of the 16 different chord patterns obtained from this experiment is similar to that of Ratz and Ritchie (1961) and Seibel (1962) and, therefore, can be considered relatively reliable. The data analysis shows that various chords are clearly more difficult than others since those chord patterns with a high RT also had a high error rate, a finding also supported by Seibel's (1962) results.

The average reaction times in the present experiment are much higher than those in Seibel's (1962) experiment. This, however, was expected since Seibel's subjects had been trained on this keyboard for much longer than the subjects in the present experiment. Ratz and Ritchie's (1961) claims that their subjects reached "asymptotic behavior" after the first day, were not confirmed. In the current experiment, subjects' performance would continue to improve for many sessions, as argued by Seibel (1962, 1963), Mowbray and Rhoades (1959), Cornog et al. (1963) and Conrad and Longman (1965).

No difference of performance was found between the left and the right hand. A significant difference was found between one- and two-handed chord patterns and it appears that, at this early stage in practice, two-handed chord patterns were treated almost as two separate one-handed chords.

Similarly to Ratz and Ritchie's (1961) as well as Seibel's (1962) findings, it was found that chord patterns that involved a small number of fingers and/or no finger separation produced, overall, faster and more accurate

responses. The multiple regression analyses suggest that the relationship between "number of fingers" and "finger separation" is not simple. It appears that decreases in speed and accuracy with increasing number of fingers per chord are greater when there is also finger separation. The impact of degrees of finger separation was rather vague and non-systematic, both in the present data as well as in Seibel's.

It can be safely concluded that the "easiest" chord patterns are those which involve a small number of fingers and no finger separation.

2.6 Experiment 2

A PILOT INVESTIGATION OF THE 10-KEY SEQUENTIAL KEYBOARD

The 10-key sequential keyboard allows a potentially infinite number of key combinations, as it operates similarly to an ordinary typewriter. However, only 256 combinations are needed for Experiment 3 (i.e. the experiment by means of which the four keyboard designs are going to be compared and evaluated). As discussed in the beginning of the chapter, this keyboard could process either fixed length sequences or sequences of variable length. When using fixed length sequences any one of the ten keys can be used in a given key sequence. When there is variable sequence length only nine keys can be used for processing key sequences because one has to be the "entry" key. The latter alternative has the advantage of being able to use short key sequences for the most frequent stimuli. Moreover, the effective number of keys is nine. However, it has the disadvantage of requiring the depression of an additional key (i.e. entry key) to signal the completion of a sequence. On the other hand, the obvious advantage of using a "fixed length" sequence is that no entry key has to be depressed.

If the information content and the average number of keys that have to be depressed are computed (table 2.6), it becomes apparent that, in terms of both of these considerations, constrained 3-digit sequences are more efficient than sequences of variable length.

	N of codes	Mean Bits	Keys/response
Using 10 keys and a fixed 3-digit sequence length			
Unconstrained	1000	9.77	3
Constrained	256	8.16	3
Using 9 keys plus an "enter" key and variable length sequences			
Unconstrained	1000	10.77	4.08
Constrained	256	9.28	3.61

Table 2.6 Information content and number of keys per response for constrained and unconstrained conditions for the 10-key sequential keyboard

This pilot experiment was carried out to obtain data to help in the choice of the "easiest" key sequences. It examined differences in speed and accuracy as a function of the fingers involved, as well as the ordinal position of each keystroke in the 3-digit sequence.

METHOD

Subjects

Four paid subjects, all students in the Psychology Department of the University of Hull, were used. All of them were right-handed. No-one had any previous experience either with ordinary typing or with a 10-key sequential keyboard.

Apparatus and Displays

The apparatus and the displays used in this experiment have already been discussed.

The microcomputer was programmed to record the actual key-strokes made, the inter-stimulus time and the time taken for each one of the three key-strokes to be made.

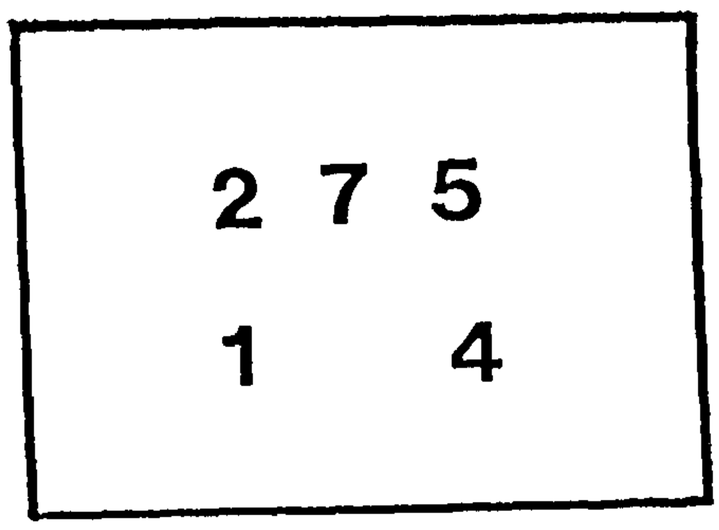


Figure 2.8 Type of visual feedback for the 10-Key Sequential keyboard (the keys 2, 7 and 5 should have been depressed but the keys 1, 7 and 4 were depressed instead)

Task Description, Stimuli, and Feedback

Each subject was seated in front of the apparatus with their fingers placed over the ten keys of the keyboard. The stimulus that appeared on the screen was a 3-digit number, selected at random. For the purpose of the experiment, three-digit number sequences were chosen as stimuli. The fingers were numbered from the little finger of the left hand to the little finger of the right hand as 1, 2, 3,..., 9, 0. Subjects were instructed to key each digit sequence as quickly and accurately as possible.

When an incorrect response was made, auditory and visual feedback were given. The auditory feedback was an alerting bleep. The visual feedback was the display of the digit(s) which had been incorrectly depressed shown below those that should have been depressed (figure 2.8).

At the end of each session subjects received further feedback in the form of: total number of correct responses; average reaction time for correct responses; total number of error responses; and average time for error responses.

Sessions

Subjects received four sessions, one in the morning and one in the afternoon of two consecutive days. Each session consisted of 105 trials.

Before each session, subjects had twenty "warm-up" practice trials which were not recorded.

1st Keystroke		BLOCK 1		BLOCK 2		BLOCK 3		BLOCK 4	
HAND	FINGER	RT	ERR	RT	ERR	RT	ERR	RT	ERR
L	Little	1.39	1.27	1.35	1.20	1.19	3.65	1.10	6.12
	Ring	1.57	5.12	1.38	5.40	1.32	3.77	1.23	2.92
	Middle	1.55	3.27	1.55	5.27	1.39	3.66	1.26	2.50
	Index	1.57	7.10	1.38	9.20	1.26	9.42	1.09	13.65
	Thumb	1.21	3.45	1.11	3.42	1.11	2.55	0.97	3.42
R	Little	1.36	1.75	1.46	2.87	1.21	5.12	1.09	2.72
	Ring	1.60	1.75	1.39	1.22	1.34	5.87	1.28	3.75
	Middle	1.55	1.90	1.45	4.92	1.25	0.65	1.20	6.67
	Index	1.22	2.85	1.15	2.40	1.07	2.17	0.99	1.35
	Thumb	1.31	1.47	1.23	3.20	1.14	1.45	1.00	2.40
Average		1.43	2.99	1.34	3.91	1.23	3.84	1.12	4.55

2nd Keystroke		BLOCK 1		BLOCK 2		BLOCK 3		BLOCK 4	
HAND	FINGER	RT	ERR	RT	ERR	RT	ERR	RT	ERR
L	Little	0.65	3.80	0.56	4.62	0.56	3.00	0.48	5.17
	Ring	0.67	10.50	0.67	7.52	0.66	4.25	0.54	6.60
	Middle	0.75	7.95	0.80	7.12	0.67	5.35	0.60	7.20
	Index	0.82	3.55	0.76	7.87	0.62	4.20	0.66	6.20
	Thumb	0.69	0.67	0.65	1.10	0.58	2.72	0.58	4.30
R	Little	0.61	4.20	0.59	8.47	0.51	6.80	0.54	2.30
	Ring	0.81	7.60	0.69	6.42	0.55	9.35	0.60	6.97
	Middle	0.76	8.37	0.60	8.50	0.59	6.85	0.44	11.92
	Index	0.67	3.00	0.64	1.95	0.53	4.65	0.47	3.52
	Thumb	0.64	4.82	0.58	3.87	0.52	1.07	0.46	2.85
Average		0.71	5.45	0.65	5.75	0.58	4.72	0.53	5.70

3rd Keystroke		BLOCK 1		BLOCK 2		BLOCK 3		BLOCK 4	
HAND	FINGER	RT	ERR	RT	ERR	RT	ERR	RT	ERR
L	Little	0.46	6.35	0.43	3.47	0.39	4.57	0.36	4.40
	Ring	0.52	5.30	0.47	8.85	0.42	7.55	0.40	6.15
	Middle	0.62	7.47	0.54	9.90	0.50	8.60	0.50	7.40
	Index	0.56	9.72	0.54	3.67	0.46	7.45	0.43	5.07
	Thumb	0.51	3.45	0.42	2.72	0.40	1.32	0.37	7.35
R	Little	0.47	8.17	0.44	4.42	0.41	10.52	0.34	4.87
	Ring	0.60	10.02	0.49	7.60	0.42	9.05	0.42	9.65
	Middle	0.62	6.62	0.51	8.55	0.43	8.22	0.39	7.50
	Index	0.50	6.10	0.44	3.57	0.38	6.45	0.35	7.10
	Thumb	0.49	3.85	0.42	7.15	0.41	4.32	0.35	7.82
Average		0.54	6.71	0.47	5.99	0.42	6.81	0.39	6.73

Table 2.7 Mean RTs (secs) and error rates for all fingers as a function of practice and ordinal position of the keystroke on the digit sequence

RESULTS

RT and error data for each finger, each key-stroke and each practice block are shown in table 2.7.

The Effects of Practice and Key-stroke Sequence

RT varied systematically as a function of the ordinal position of keys in the key sequence. The first keystroke was slower than the second and the second slower than the third. This effect is very obvious in figure 2.9 where RT averaged across subjects is plotted as a function of the ordinal position of keys, for each one of the four blocks of practice (B1 to B4). The number of errors increased with every keystroke (i.e. overall, the first keystroke is more accurate than the second and the second more accurate than the third, figure 2.10).

Reaction time also decreased systematically with practice (figure 2.9). The improvement is practically equal for all key strokes, hence the almost parallel nature of the four functions. Accuracy, on the other hand, remained relatively stable from one block of practice to the other (figure 2.10). The lack of any general improvement in accuracy as a result of practice may imply either that accuracy does not improve with practice as fast as speed (i.e. it might need longer practice for accuracy to improve) or that subjects adopted differential speed-accuracy tradeoffs from one block of trials to another (i.e. they paid more attention to improving speed while maintaining accuracy).

Differences between Keys

The analyses of the differences in the RTs of keystrokes revealed that there is a tendency for the little finger and the thumb of each hand to be faster than the other fingers. An exception to that is the index finger of the right hand which is about as fast as the thumb of the same hand. In other words, there is a clear "end-anchoring" effect for the reaction times of individual fingers (figure 2.11). This effect is most pronounced for the first keystroke.

Performance Differences between Hands

No substantial difference in RTs between the left and the right hand can be observed (figure 2.11). Although the RTs for the left hand are consistently longer than those of the right hand, the difference is very small and mainly caused by the index finger of the right hand being faster than that of the left hand. There is no obvious systematic difference in accuracy between the two hands (figure 2.12).

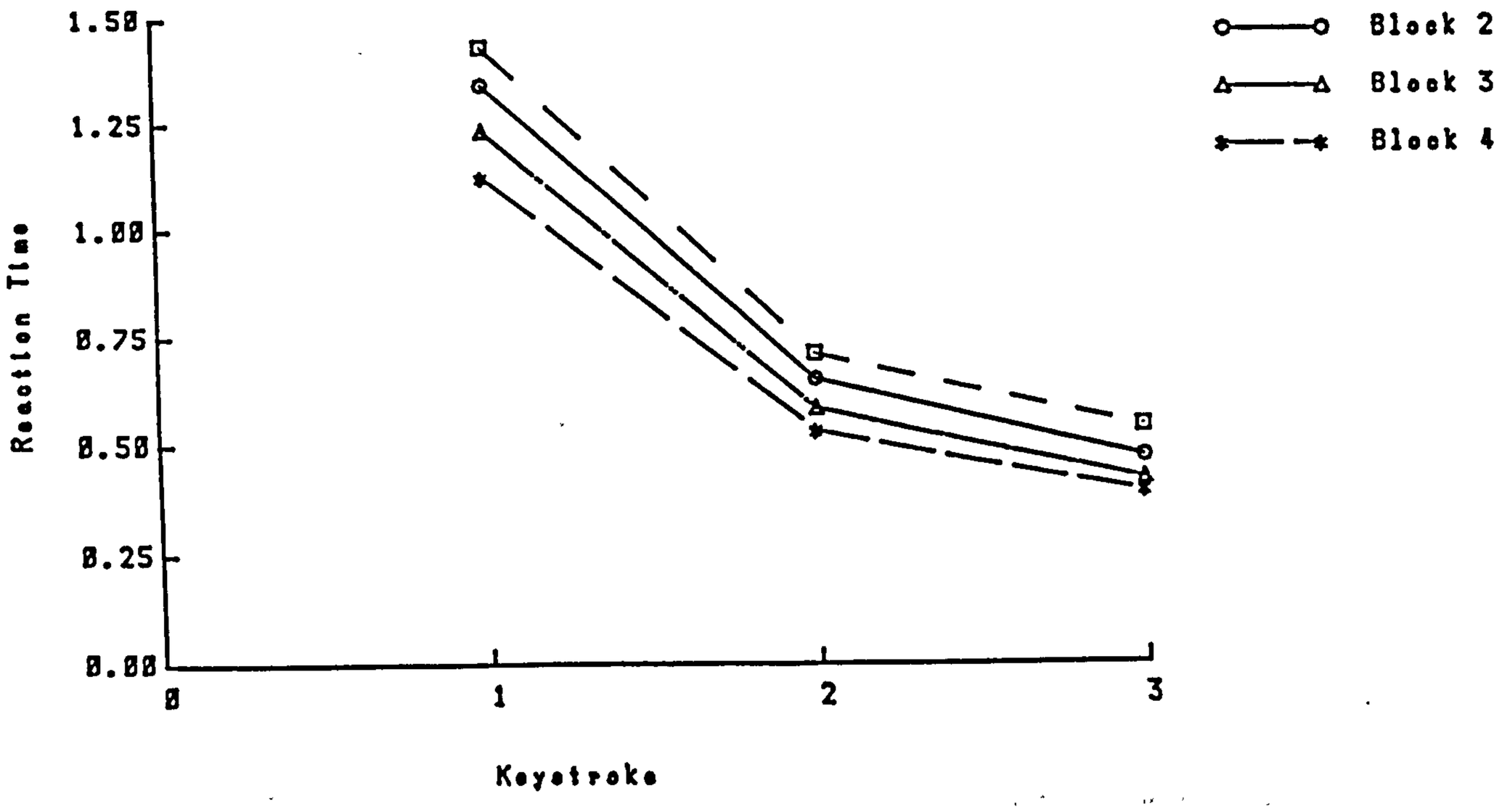


Figure 2.9 Mean RT for each keystroke as a function of practice blocks

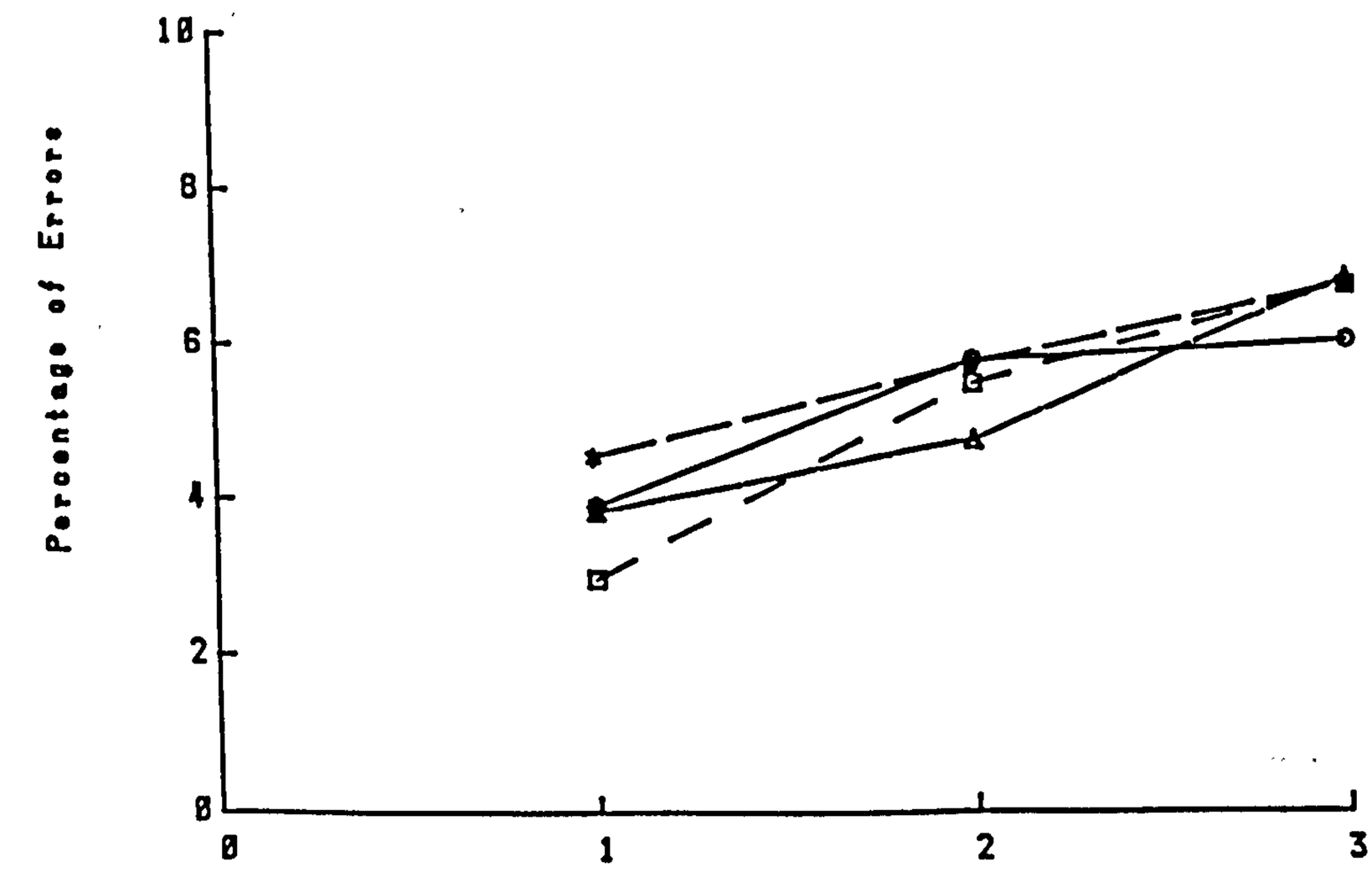


Figure 2.10 Percentage of errors for each keystroke as a function of practice blocks

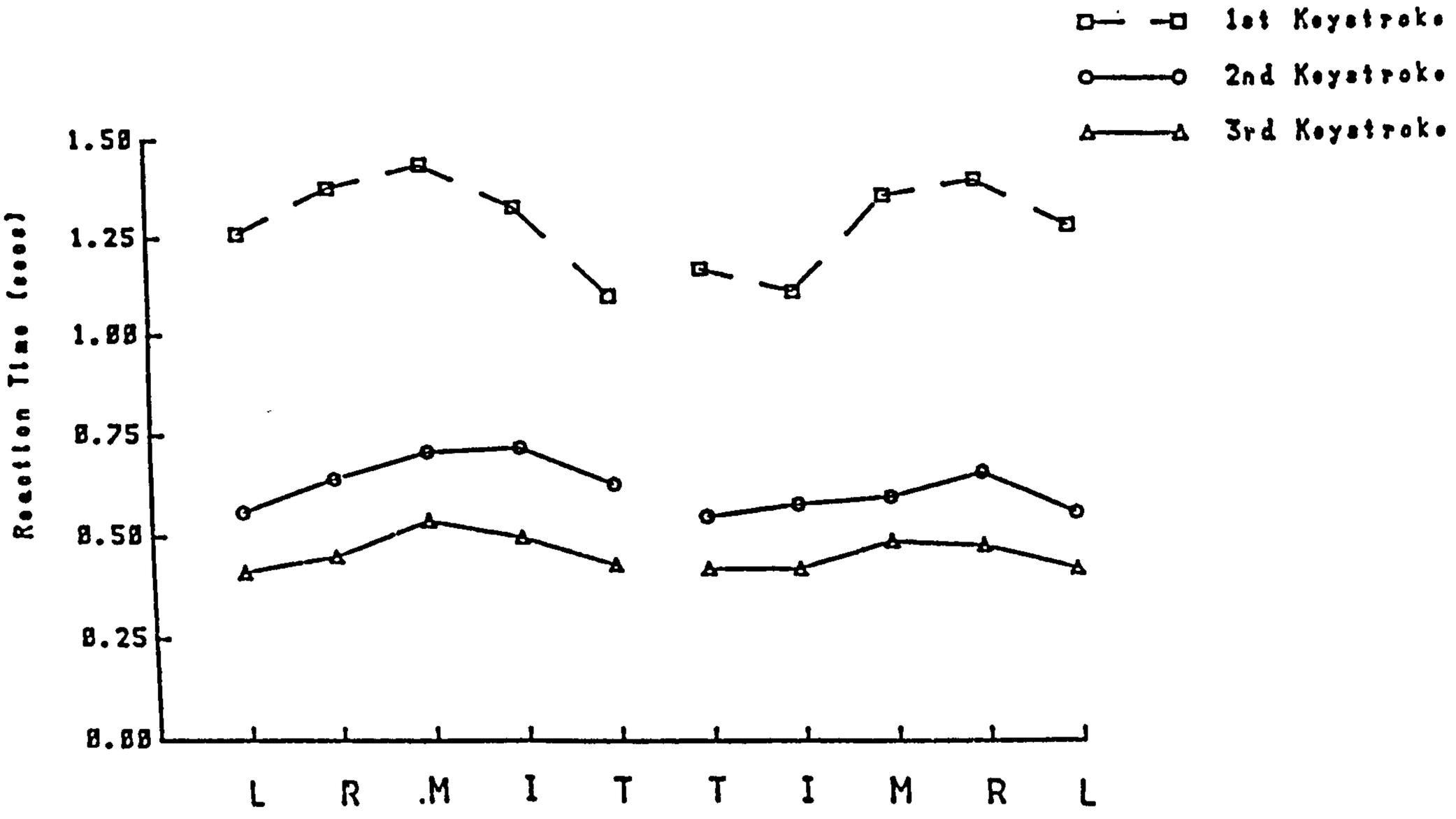


Figure 2.11 Mean RT for each finger and each keystroke averaged across practice (L=Little, R=Ring, M=Middle, I=Index, T=Thumb)

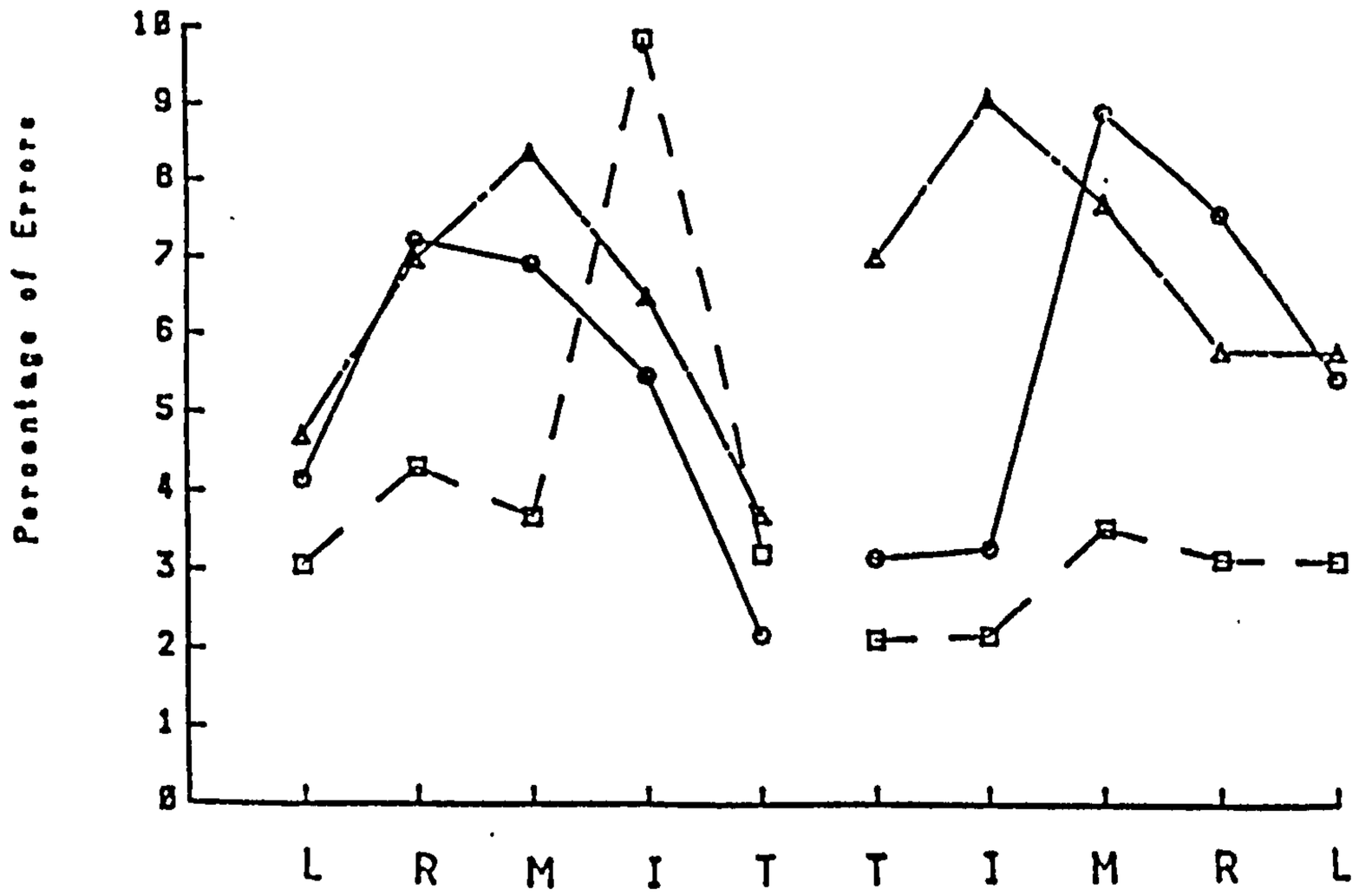


Figure 2.12 Mean percentage of errors for each finger and each keystroke averaged across practice

DISCUSSION

The results suggest that there is an improvement in speed across the four sessions of practice while accuracy remains relatively constant. It is most likely that improvement in accuracy would occur with more training.

There are two possible ways in which subjects could have responded to key sequences. First, subjects could attend to each digit individually, in other words, look at one digit and respond to it, then look at the second digit and respond and then look at the third digit and respond. In this case, RTs for all three keystrokes would be equal or slightly decreasing from one keystroke to the next because of "previewing" effects. In other words, the second and the third keystrokes could be just slightly faster than the first and the second respectively because subjects would have already previewed the second and the third keystrokes while attending to the first and the second respectively. However, it was found that RTs for the second and third keystrokes were much faster than the first one and this difference in speed is too big to be considered as only the result of "previewing" effects.

The second way in which subjects could respond was that they looked at the whole digit sequence and then depressed each key. This alternative seems consistent with the observed differences in speed from the first keystroke to the second and from the second to the third. Moreover, it explains the slight decrease in accuracy which was observed among the keystrokes. As subjects perceived the whole digit sequence, the information was stored in the short-term memory which

faded slightly from one keystroke to the next and therefore the probability of an error increased.

The results also suggest that some keys are easier to respond to than others (i.e. the fingers towards the middle are generally slower than the peripheral fingers). These differences, however, do not seem to be caused by differential latencies related to individual fingers. If the differences were reflections of finger "strength" (Haaland et al., 1963) or "ease of articulation" (Dvorak et al., 1936), it would be expected that speed and accuracy would generally improve from the little to the thumb and from the left hand to the right. Since the results of this experiment do not suggest anything of this nature, it must be concluded that the difference in difficulty of various keys are related to the spatial layout of keys, with the end keys of each set of five being faster than the middle ones. In other words, this "end anchoring" suggests a differential effect of spatial uncertainty on the choice of a response.

In summary, these results show that there are differential effects associated with responses to three-digit sequences. Although it is clear that some key sequences are more "difficult" than others (for example those which include keys in the middle of each five-key pad), it is unlikely that this information would be of great value for the selection of key sequences for Experiment 3. The reason for this is that there is no easy way to convey such information to subjects. It has been clearly demonstrated (e.g. Fitts and Switzer,

1962) that unless the population from which stimuli are sampled is very well-defined, subjects cannot make effective use of any constraints imposed on the population.

Chapter 3

EVALUATION AND COMPARISON OF FOUR KEYBOARD DESIGNS

Experiment 3

3.1 Introduction

The present experiment was carried out to evaluate and compare the four different keyboard designs which have already been described, so that the most efficient one could be selected for the foreign destination coding desks at the British Post Office. The criteria of keyboard efficiency were: speed and accuracy of keyboard operation as well as the time needed for operators to be trained in the use of the new keyboard. Therefore, the concentration of the present experiment was on speed and accuracy measures of response execution as affected by the differences in keyboard design and operation.

The procedure followed in this experiment, at least for the 10-key chord and the 10-key sequential keyboards, was based on the findings of the two pilot experiments discussed in Chapter 2. The results of Experiment 1 were in agreement with the findings of Ratz and Ritchie (1961) and Seibel (1962, 1963) and provided some useful guidelines about the selection of stimuli for the present experiment. They revealed that the "easiest" chord patterns were those that involved a small number of keys and no finger separation (except for the easiest finger separation pattern, the little finger and the thumb which was also included). For the present experiment therefore the chord patterns that involved one- or two-finger

separation were excluded. It was well-established by Experiment 1 that one-handed chord patterns produced much shorter reaction times than two-handed ones. For the present comparison however it would be inappropriate to use any one-handed patterns since the other three keyboards would necessarily require the use of both hands for all stimuli.

For the 10-key sequential keyboard, the findings of Experiment 2 made clear that factors affecting the "choice" of a response were more important than those affecting the actual "response execution". The fastest key combinations were those that required the least "preparation" time. Consequently, for the present experiment it was decided that the set of stimuli which would be selected should reduce the amount of time required for the "preparation" of each digit sequence. This was done by constraining the set of possible digits for the first keystroke to the digits "1", "2" and "3" while all ten digits ("0" to "9") were used for the second and the third keystrokes. The specific stimuli that were selected are discussed in the Method below.

METHOD

Subjects

Sixteen paid subjects, male and female students at the University of Hull, were used. They were divided in four equal sized groups, one group for each of the four keyboards. The subjects were matched across groups on the basis of their performance on an 8-choice reaction time pre-test.

All subjects were told from the beginning what the

project involved and were asked to sign agreement forms committing themselves to the requirements of the experiment which included participation five days a week for six weeks. They were given financial incentives to maintain the level of motivation which was needed for this kind of task. An incentive scheme was devised and depending on their performance, subjects received between £115 and £195 at the end of the experiment.

Stimuli

For all but the 2x6 chord keyboard, a set of 256 different stimuli were used. Each trial consisted of the presentation of one stimulus randomly selected from the set of stimuli. The stimuli used for each keyboard were as follows.

2x6 Chord Keyboard: 144 stimuli representing all possible combinations.

4x4 Chord Keyboard: All 256 possible key combinations.

10-Key Chord Keyboard: The key combinations that were selected were the fifteen chord patterns with no finger separation and the chord pattern 10001 (16x16=256), which was the "easiest" finger separation chord pattern.

10-Key Sequential Keyboard: 256 3-digit numbers starting from 100. These stimuli were chosen so that the first digit of each sequence was restricted to three alternatives (1, 2, or 3). This procedure minimized the average response information for the 256 sequences.

The way the stimuli were presented to subjects has already been discussed in Chapter 2.

Information Transmission Capacity of the Four Keyboards

As already mentioned, the number of alternative responses varied among the four keyboards. It was important therefore to equate, as far as possible, the information load. Given the way the stimuli were selected, the information capacity of each keyboard was:

The 2x6 Chord Keyboard had 144 codes (an average of 7.17 Bits per response) and required the depression of two keys per response.

The 4x4 Chord Keyboard had 256 codes (an average of 8 Bits per response) and required the depression of two keys per response.

The 10-Key Chord Keyboard had 256 codes (an average of 8 Bits per response) and required the depression of an average 4.63 keys per response.

The 10-Key Sequential Keyboard had 256 codes, each requiring the depression of 3 keys per response. Because of the unequal key frequencies, this represented an average information of 8.16 Bits per response.

Pre-test

Subjects were pretested with an 8-choice reaction time test on the BBC micro-computer,and the results were used to assign subjects to four matched groups. They were asked to place their fingers of the left hand over the keys 1,2,3 and 4 and the fingers of their right hand over the keys 7,8,9 and 0 at top row of the computer keyboard. Each trial consisted of the presentation of one digit at a time from the set 1,2,3,4,7,8,9,0 on the screen of the computer monitor. They

were instructed to depress the keys that corresponded to the digits on the screen as quickly and accurately as possible. When they made an error they were alerted by a bleep (i.e. auditory feedback). They were given two sessions of 100 trials each, one session immediately after the other.

Subjects' performance was assessed using an adjusted score (total time/number of correct responses). Subjects were assigned their numerical "label" on the basis of the adjusted scores' rank-order (table 3.1) and they were divided into four matched groups for assignment to the four keyboards (table 3.2).

SUBJECT	RT	N of Ers	ASC
S1	75.70	0	75.70
S2	73.46	2	75.76
S3	78.53	5	83.63
S4	89.71	2	91.00
S5	88.73	7	96.04
S6	97.94	3	101.43
S7	100.34	4	104.71
S8	109.05	2	111.16
S9	108.95	4	112.69
S10	109.33	11	123.70
S11	124.68	13	137.13
S12	132.48	7	138.70
S13	140.57	5	144.07
S14	124.07	32	144.91
S15	143.80	3	146.96
S16	145.45	6	159.50

Table 3.1 Subjects' mean reaction time (RT), number of errors (N of Ers) and adjusted scores (ASC) on the pretest RT in csecs

2X6	4X4	SEQ.	CHORD
S1	S2	S3	S4
S8	S5	S6	S7
S11	S12	S9	S10
S14	S15	S16	S13

Table 3.2 Assignment of subjects to matched groups on the basis of their ASC on the pretest

Procedure

Subjects received training, by the same instructor, five days a week for six weeks. During the first two weeks the training consisted of one session of five sets of 105 trials each day (i.e. one session or 525 trials per day). For the rest of the time there were two such sessions each day (i.e. 1050 trials per day), with a ten minute break in between (table 3.3). At the beginning of each session subjects had a "warm-up" practice of 20 trials that was not recorded.

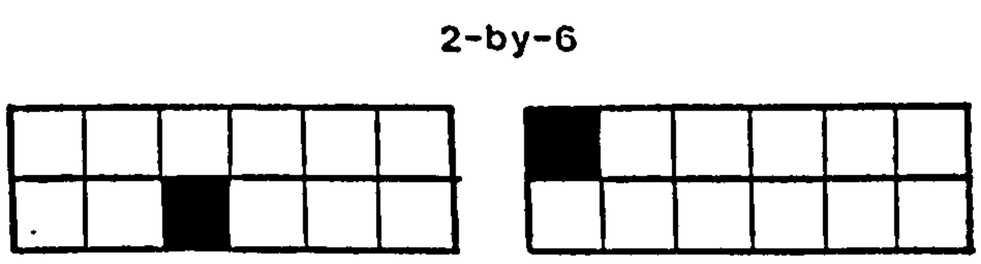
Week	Sessions/day	Trials/day	Trials/week	VF	AF
1 & 2	1	525	2,625	yes	yes
3 & 4	2	1,050	5,250	no	yes
5 & 6	2	1,050	5,250	no	no

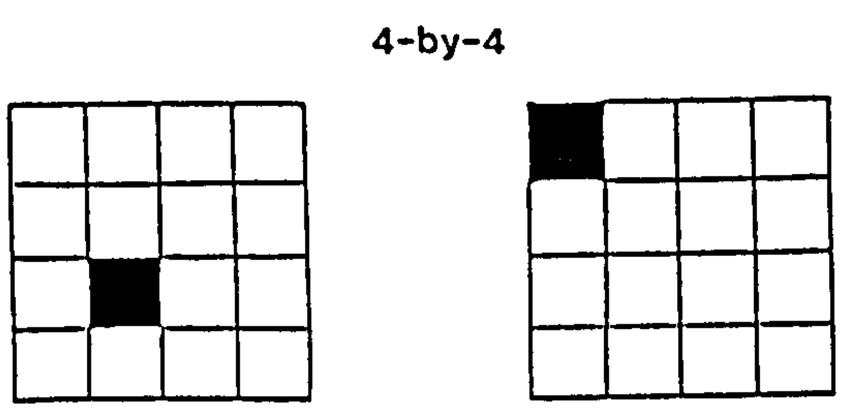
Total number of sessions per subject: 10
Total number of trials per subject: 26,250
(VF=Visual Feedback, AF=Auditory Feedback)

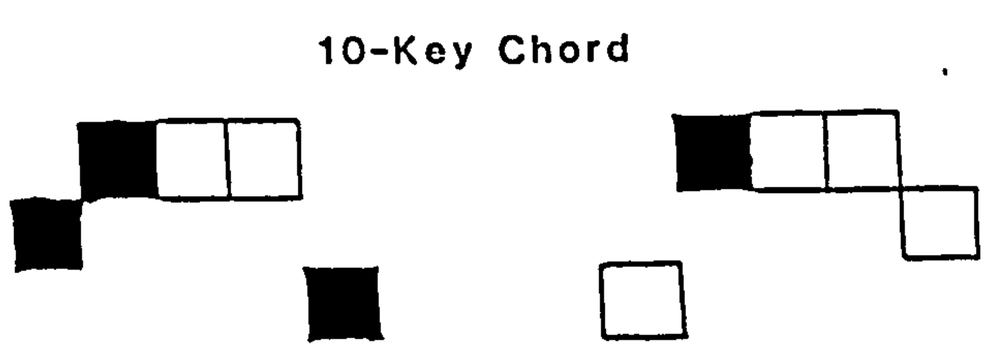
Table 3.3 Experimental Design

During the first and the second week, when an error was made both auditory and visual feedback was provided. The auditory feedback was a bleep; the visual feedback was an indication on the screen of the keys that had been pressed correctly, those pressed by mistake, and those that should have been pressed but had been omitted (figure 3.1). In the case of the 10-Key Sequential keyboard the visual feedback indicated the digits that had been incorrectly pressed under those that should have been pressed (figure 3.1).

During the third and the fourth week, the visual feedback was removed and, when an error occured, subjects would only







10-Key Sequential

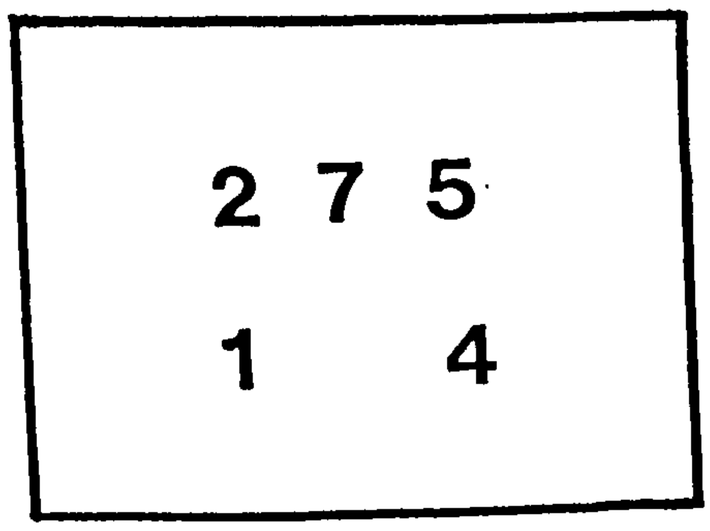


Figure 3.1 Examples of stimuli and visual feedback for each keyboard group

get a bleep for feedback.

During the fifth and the sixth week there was neither auditory nor visual feedback.

Throughout the six weeks, at the end of each set of 105 trials, a summary of each subject's performance was displayed. This consisted of the number of correct responses, the average reaction time of correct responses, the number of error responses, and the average reaction time of error responses. At the end of each session (i.e. 525 trials) they were given another summary of their performance level similar to that given at the end of each block. Furthermore, at the beginning of each week, each subject was shown, individually, graphs of his/her average reaction time and number of errors for each session of the previous week.

3.3 RESULTS

Three measures were examined in the analysis of the data: the mean reaction time (RT); the percentage of errors (ERR); and an adjusted score (ASC) which was derived by dividing the total reaction time by the number of correct responses and represents the mean time between correct responses. This final measure was useful in that it took account of both RT and ERR at the same time.

Pre-test

As a consequence of the way subjects were assigned to the four groups the average RT, ERR and ASC on the pre-test for the four groups are very similar (table 3.4). Analyses of variance on all three measures showed that indeed there were

no significant differences between the groups (F<1).

	G R O U P			
	2x6	4x4	10-key Chord	10-key Sequential
RT	100.6	108.75	109.03	109.18
ERR	12%	5%	5%	6%
ASC	117.23	114.36	114.31	115.87

Table 3.4 Mean reaction time in csecs (RT), percentage of errors (ERR) and adjusted scores (ASC) for the four groups on the pretest

To examine the predictive value of the pre-test, a correlation matrix was produced for the pre-test ASCs and the ASCs during the ten practice sessions. Because these correlations would be confounded by the effect of each keyboard on RT and error rate, the residual ASCs were used in the correlational analysis (where the residual score was the difference between a subject's average adjusted score during a session and the mean adjusted score of the group that the subject belonged to for the same session): see table 3.5. All correlations were very high (the lowest was 0.44 and the highest 0.85), which confirmed the utility of the pre-test that was used.

Session	1	2	3	4	5	6	7	8	9	10
Pretest	.85	.80	.75	.67	.59	.62	.47	.44	.48	.50

Table 3.5 Correlation coefficients between residual adjusted scores on the pretest (across subjects) and residual adjusted scores during the ten training sessions

Practice Effects on the Performance on the Four Keyboards

For the examination of the effects of training, data were grouped in ten blocks of sessions (each session consisting of 525 trials per subject and each block consisting of five sessions). RT for each of the four keyboards decreased across the ten blocks of sessions (figure 3.2). The 10-key Sequential Group had the slowest reaction times. The 2x6 Group started at about the same level as the 10-key Sequential Group, but they improved more during the second and the third blocks of sessions. From the fourth block onwards, the improvement in speed of the two groups was almost parallel. The 4x4 Group produced the shortest RTs of all groups, during the first block. In the second and the third blocks of sessions, the 10-key Chord Group reached the same level of speed and from then on became fastest.

	2x6			4x4			10-key seq.			10-key chord		
Block	ASC	RT	ERR	ASC	RT	ERR	ASC	RT	ERR	ASC	RT	ERR
AF+VF:												
1	2.05	1.99	3.44	1.79	1.71	3.69	2.13	2.00	6.50	1.92	1.79	6.06
2	1.46	1.42	2.36	1.29	1.27	1.94	1.72	1.63	5.36	1.33	1.26	5.20
AF only:												
3	1.34	1.30	2.81	1.17	1.14	3.05	1.48	1.41	5.08	1.19	1.13	4.42
4	1.24	1.20	3.63	1.11	1.06	3.92	1.39	1.31	5.72	1.09	1.03	5.35
5	1.20	1.16	3.45	1.04	1.01	3.32	1.32	1.25	5.15	1.01	0.98	2.67
6	1.16	1.12	4.13	1.02	0.98	3.39	1.28	1.21	5.87	0.98	0.93	5.55
No feedback:												
7	1.13	1.09	3.40	1.01	0.98	3.24	1.25	1.18	5.88	0.99	0.97	2.24
8	1.08	1.04	3.55	0.97	0.93	4.57	1.21	1.13	6.34	0.94	0.91	3.51
9	1.04	1.00	3.84	0.94	0.91	3.89	1.18	1.08	8.37	0.89	0.88	1.87
10	1.00	0.95	4.96	0.91	0.88	3.47	1.18	1.07	9.13	0.86	0.83	3.59

Table 3.6 Mean adjusted scores (ASC), reaction time in secs (RT) and percentage of errors (ERR) for each group, averaged across subjects (AF=Auditory Feedback, VF=Visual Feedback)

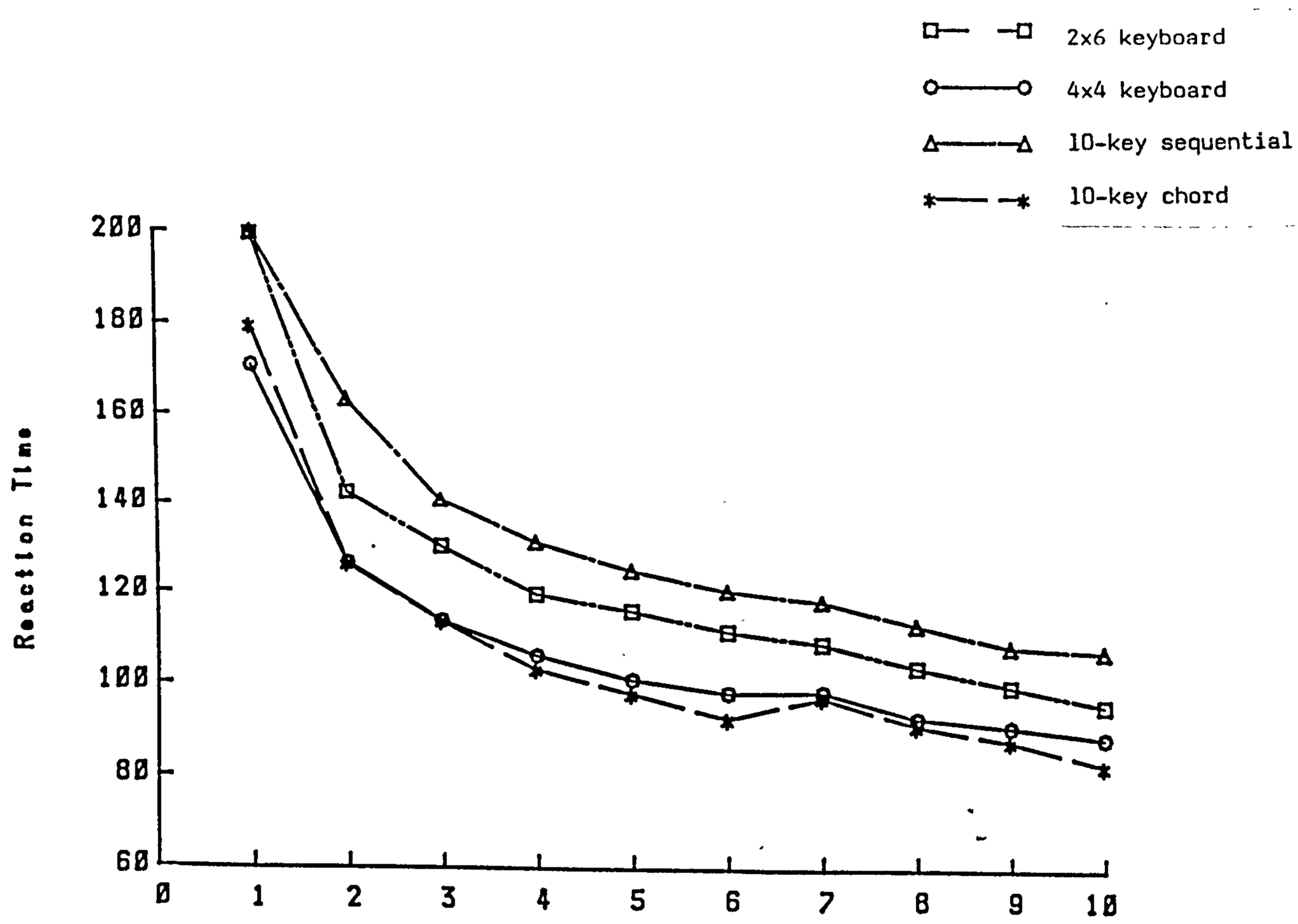


Figure 3.2 Mean RT (csecs) as a function of practice for each of the four groups

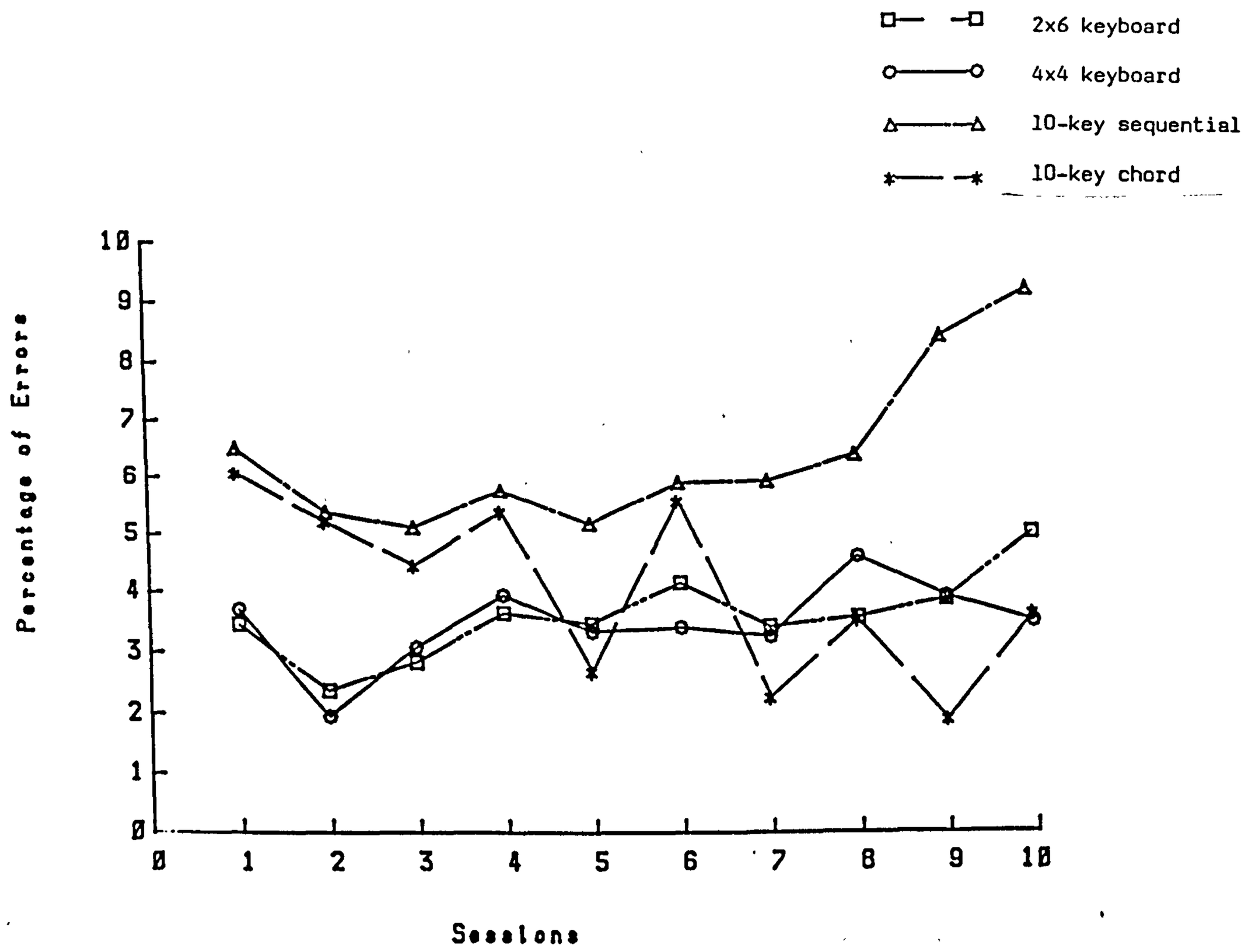


Figure 3.3 Mean percentage of errors as a function of practice for each of the four groups

When all feedback was removed in the seventh block of sessions, all groups' improvement in speed was reduced. The 10-key Chord Group, actually, became slower for these sessions. Their improvement however continued steadily from that point on. All four groups continued to improve in terms of speed (table 3.6) up to the end of the ten blocks.

The relationship between practice and the percentage of errors for each of the four groups is shown in figure 3.3. The 10-key Sequential Group made more errors than all other groups throughout the ten practice blocks. The gradual removal of feedback caused this group to make an increasingly bigger number of errors until the end of training. The 2x6 and the 4x4 Groups had average error rates below 5%. The number of errors by these two groups rose slightly during the sessions when feedback was first removed. Their error rates, however, were rather steady. The group that showed the biggest improvement in accuracy was the one that used the 10-key chord keyboard. Their error rate was about 6.1% in the first set of trials and, with some fluctuations between sessions, they reached a final level of about 3% (table 3.6).

The intercorrelations (tables 3.7, 3.8 and 3.9) between RT, ERR and ASC for the ten blocks of sessions (each block consisting of five sessions or 2,625 trials) are consistently high, indicating that there is consistency in subjects' performance from one session to the next. It also seems that their initial performance (especially speed performance) is a good predictor of their final performance. This relationship between initial and final performance is even clearer when

examining the intercorrelations of the residual scores for the ten blocks of practice. These intercorrelations (table 3.10) are consistently very high, ranging from 0.76 to 0.98.

	1	2	3	4	5	6	7	8	9
1									
2	.88								
3	.86	.96							
4	.79	.94	.97						
5	.73	.91	.95	.99					
6	.74	.92	.94	.96	.98				
7	.69	.85	.90	.92	.93	.92			
8	.67	.84	.58	.91	.92	.91	.99		
9	.70	.88	.88	.91	.93	.94	.96	.96	
10	.65	.85	.85	.89	.90	.91	.97	.98	.98

Table 3.7 Intercorrelations of mean RTs during the ten blocks of sessions

	1	2	3	4	5	6	7	8	9
1									
2	.74								
3	.72	.81							
4	.65	.69	.89						
5	.43	.47	.80	.81					
6	.50	.73	.90	.83	.77				
7	.23	.24	.63	.67	.90	.66			
8	.38	.38	.69	.77	.87	.71	.91		
9	.22	.20	.53	.49	.84	.51	.91	.73	
10	.22	.26	.45	.54	.75	.50	.83	.72	.87

Table 3.8 Intercorrelations of mean percentage of errors during the ten blocks of sessions

	1	2	3	4	5	6	7	8	9
1									
2	.87								
3	.87	.95							
4	.83	.94	.98						
5	.75	.91	.96	.98					
6	.79	.92	.96	.98	.99				
7	.73	.88	.93	.94	.94	.95			
8	.73	.88	.91	.94	.93	.95	.99		
9	.72	.90	.91	.93	.94	.96	.97	.98	
10	.68	.88	.88	.91	.92	.94	.96	.98	.99

Table 3.9 Intercorrelations of adjusted scores (ASC) during the ten blocks of sessions

	1	2	3	4	5	6	7	8	9
1									
2	.92								
3	.89	.93							
4	.87	.90	.96						
5	.79	.87	.95	.98					
6	.86	.90	.96	.98	.97				
7	.76	.80	.90	.89	.89	.89			
8	.78	.80	.86	.89	.87	.90	.97		
9	.81	.87	.89	.90	.89	.92	.95	.96	
10	.81	.84	.87	.89	.86	.90	.96	.98	.98

Table 3.10 Intercorrelations of the residuals of the adjusted scores during the ten blocks of sessions

Final Performance of the Four Groups

In this experiment, there were three different conditions in terms of amount of feedback provided: (1) auditory and visual feedback, (2) auditory feedback, and (3) no feedback. The third condition was the same as that under which operators work at the Post Office. Consequently, it was the most important of the three in indicating final performance. Analyses of variance were carried out on the data from the last two weeks (i.e. the last four blocks of practice or the

last 20 sessions) to examine the effects of certain factors on RT and error rate.

First, a repeated measures analysis of variance was carried out on RT measures to examine the effect of practice (n=10 sessions per week) and the factor "weeks" (n=2, 5th and 6th week of practice) on the speed of the four groups. It was found that there was a significant difference between groups [$F(3,12)=3.35844$, $p<.05$], which suggests that the speed of operation is a function of the keyboard used. The group that used the 10-key Chord keyboard was fastest, followed by the groups that used the 4x4 Chord, the 2x6 Chord and finally the 10-key Sequential Keyboards.

The speed of all groups during the last week was significantly higher than their speed the previous week [$F(1,108)=103.044$, $p<.001$]. Their speed also increased significantly across practice sessions within each week [$F(9,108)=28.2003$, $p<.001$]. Moreover, there was a significant interaction between the factor "weeks" and practice sessions [$F(9,108)=3.0932$, $p<.01$]. Examination of the means suggests that the improvement during the fifth week was bigger than the improvement during the sixth week.

The effects of the same three factors on accuracy were examined by a second repeated measures analysis of variance on the error data. It was found that there was a significant increase in error rate from the penultimate to the final week of practice [$F(1,12)=4.3933$, $p<.05$] and within the last ten sessions [$F(9,108)=3.12539$, $p<.01$]. This increase was basically due to the large increase of errors that the 10-key

Sequential Group produced. The latter is confirmed by the signifigant interaction between groups and "weeks" [$F(3,108)=3.0859$, $p<.05$], which suggests that the change in the error rate was not the same across groups. If the means are examined it becomes obvious that this resulted from the fact that there was a substantial increase in errors by the 10-key Sequential Group, a small increase for the 2x6 whereas there was a decrease for the other keyboards (4x4 and 10-key Chord).

The results of the above two analyses of variance suggest that the 10-key chord keyboard was best in terms of improvement of both RT and percentage of errors. However, there is a discrepancy as far as improvement of overall performance on the other three keyboards is concerned. While speed improved with practice during the last two weeks, for two of the groups, accuracy decreased during the same time, especially for the 10-key Sequential Group. Although precise explanations for these results cannot be given at this stage, it is very likely that the improvement in speed occurred at the expense of accuracy, or it could be that these groups were not ready for the removal of all feedback.

Since there was this discrepancy, a third repeated measures analysis of variance on the adjusted scores data was carried out to consider the effects of the same factors on the overall performance as measured by ASC.

The results of this analysis show that there was a significant difference in ASC among the four groups

[F(3,12)=5.8777, p<.01], with the 10-key Chord Group being the best and the 10-key Sequential Group being the worst. Also, there was significant improvement in ASC from one week to the other [F(1,12)=104.556, p<.001] and a significant improvement across the ten practice sessions within each week [F(9,108)=16.4113, p<.01]. These results suggest that there was a difference in the efficiency of the four groups and that, despite the different speed-accuracy tradeoff functions encountered in the groups, overall performance improved during the last two weeks.

Having compared the four keyboards so far, it would be interesting to look into the performance of each individual group and examine various individual aspects of the four keyboards.

(1) 2x6 Chord Keyboard

The reaction times and the percentage of errors for every row and column of this keyboard, during the last 20 sessions, (each consisting of 525 trials per person), averaged across all keys in the other hand, were examined (figures 3.4a and 3.4b). No real difference was found between the two rows of keys for either RT or percentage of errors. The first and the last keys of each row for both banks of keys seem to be easier to respond to, in terms of both speed and accuracy. This "end-anchoring" effect is slightly more pronounced for speed than for accuracy. No difference was found between the two hands for either speed or accuracy.

(2) 4x4 Chord Keyboard

Similarly to the results of the 2x6 chord keyboard, no

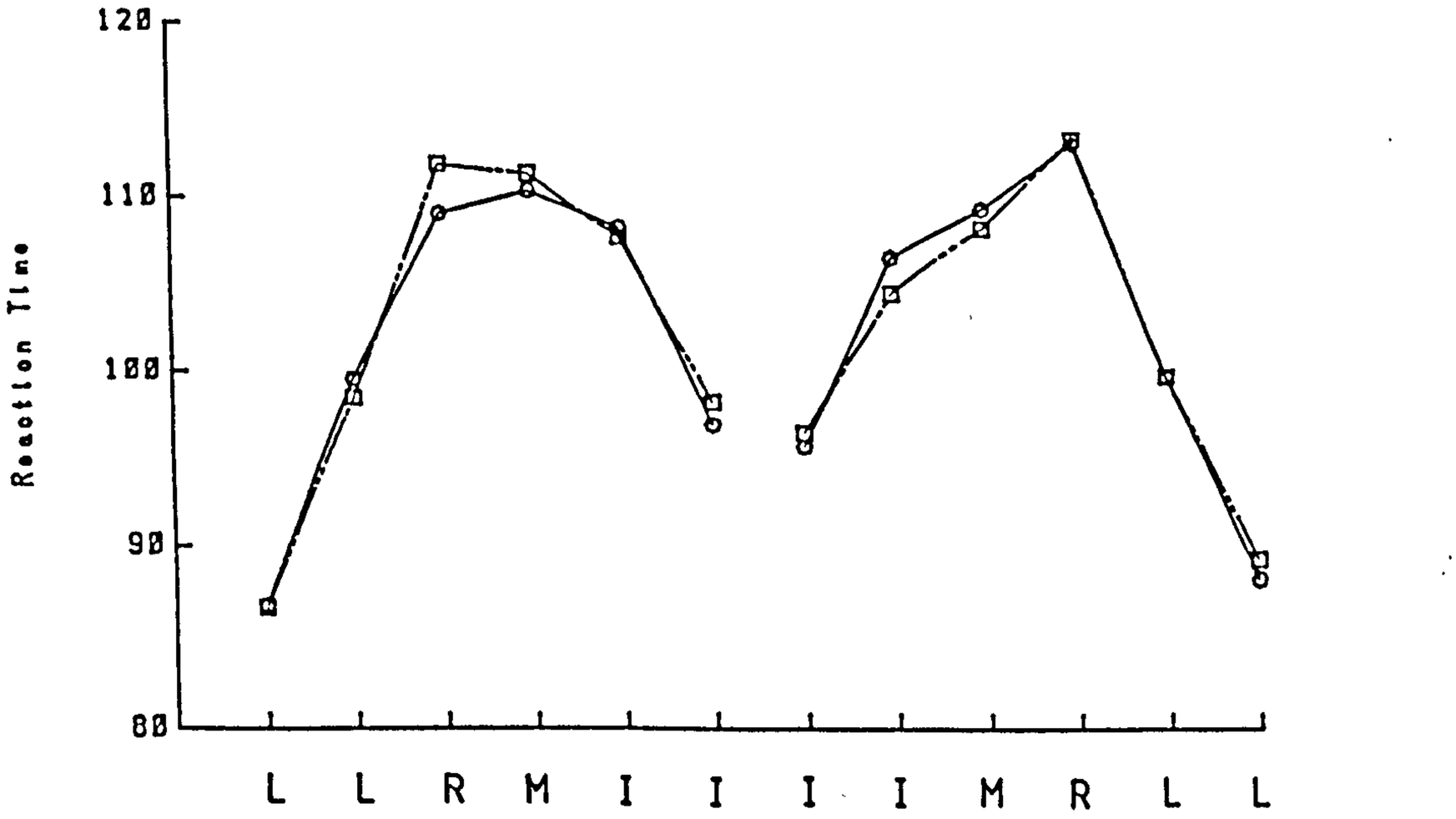


Figure 3.4a Mean reaction time (csecs) for each key of the 2x6 keyboard during the last twenty training sessions

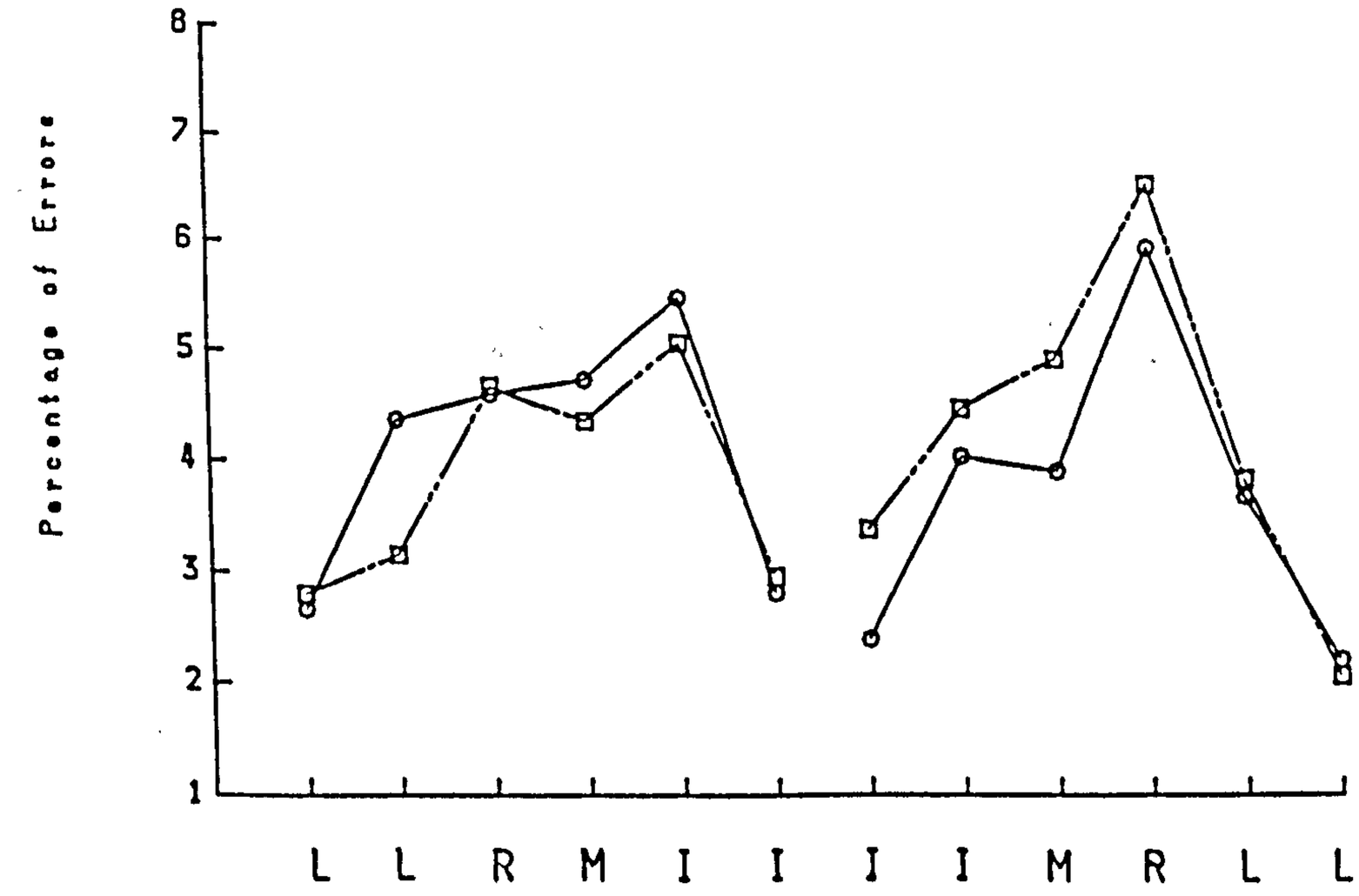


Figure 3.4b Mean percentage of errors for each key of the 2x6 keyboard during the last twenty training sessions

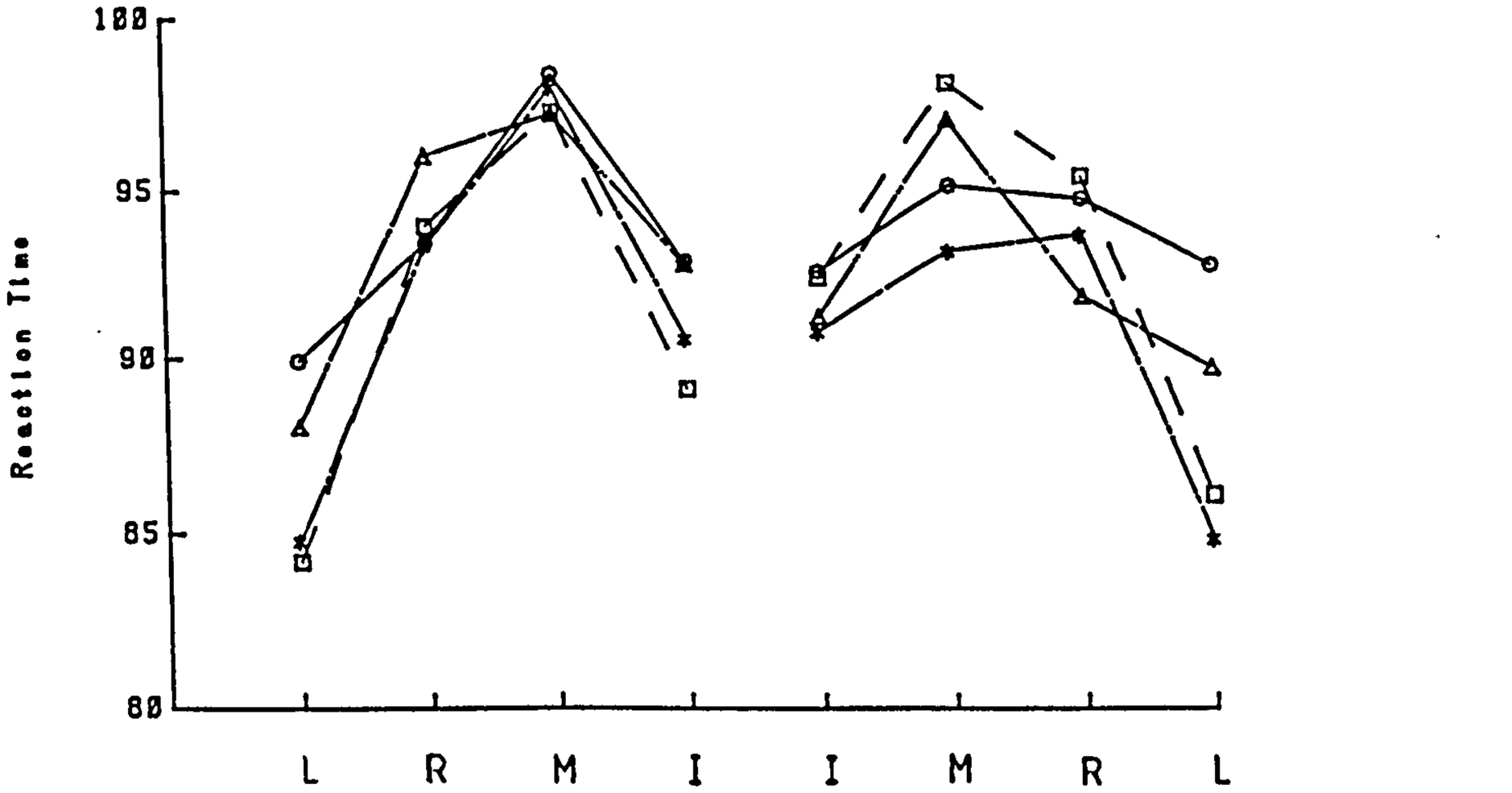


Figure 3,5a Mean reaction time (csecs for each key of the 4x4 keyboard during the last twenty training sessions

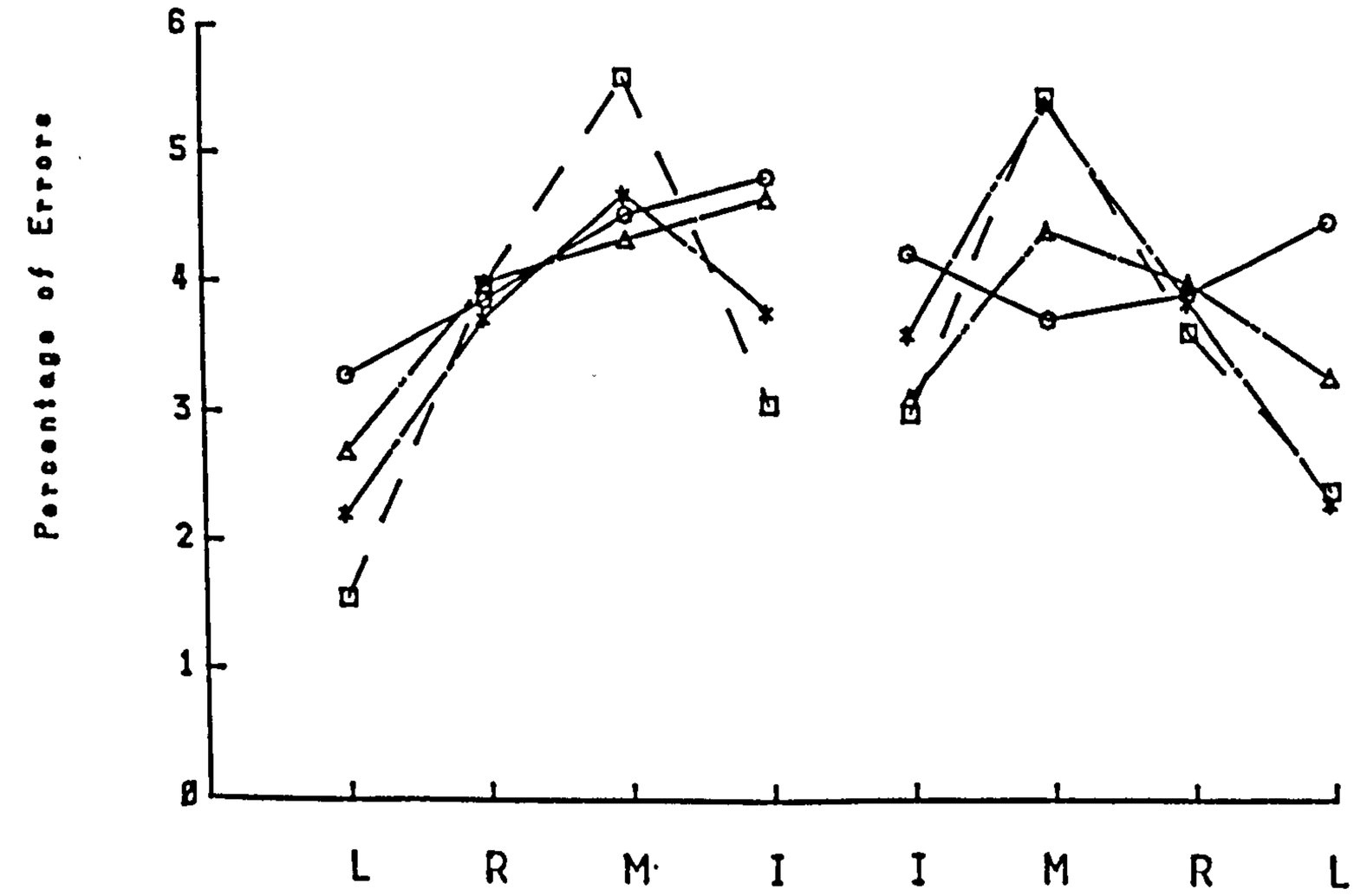


Figure 3.5b Mean percentage of errors for each key of the 4x4 keyboard during the last twenty training sessions

difference was found among the four rows of keys for either hand. The left and the right hands are very similar in the patterns of both speed and accuracy.

The "end-anchoring" effect found in the RT and ERR functions of 2x6 Group was also encountered in the functions of the 4x4 Group (figures 3.5a and 3.5b). The middle keys of this keyboard, also, seem to require longer reaction times and, generally, produce more errors than the keys at the edge columns.

(3) 10-Key Sequential Keyboard

The RT data that were collected from the 10-key Sequential Group consisted of the time that subjects spent on the whole 3-digit sequence. The responses to individual digits of every digit sequence were not recorded. The RTs for each individual key in each position were calculated by taking the means of all the sequences where each key appeared in the same position. (Figure 3.6a shows the functions of these RTs during the last twenty sessions, each session consisting of 525 trials.) The error data were collected in exactly the same way as the RT data. The functions for percentage of errors for each key at every one of the three possible positions during the last twenty sessions of practice (figure 3.6b) are similar to the RT functions. The "end-anchoring" effect which was found in Experiment 2 seems to have changed. In Experiment 2 this effect was apparent in the responses of each hand. In the curves of figures 3.6a and 3.6b though, this effect has shifted from each hand to the whole range of keys, with the

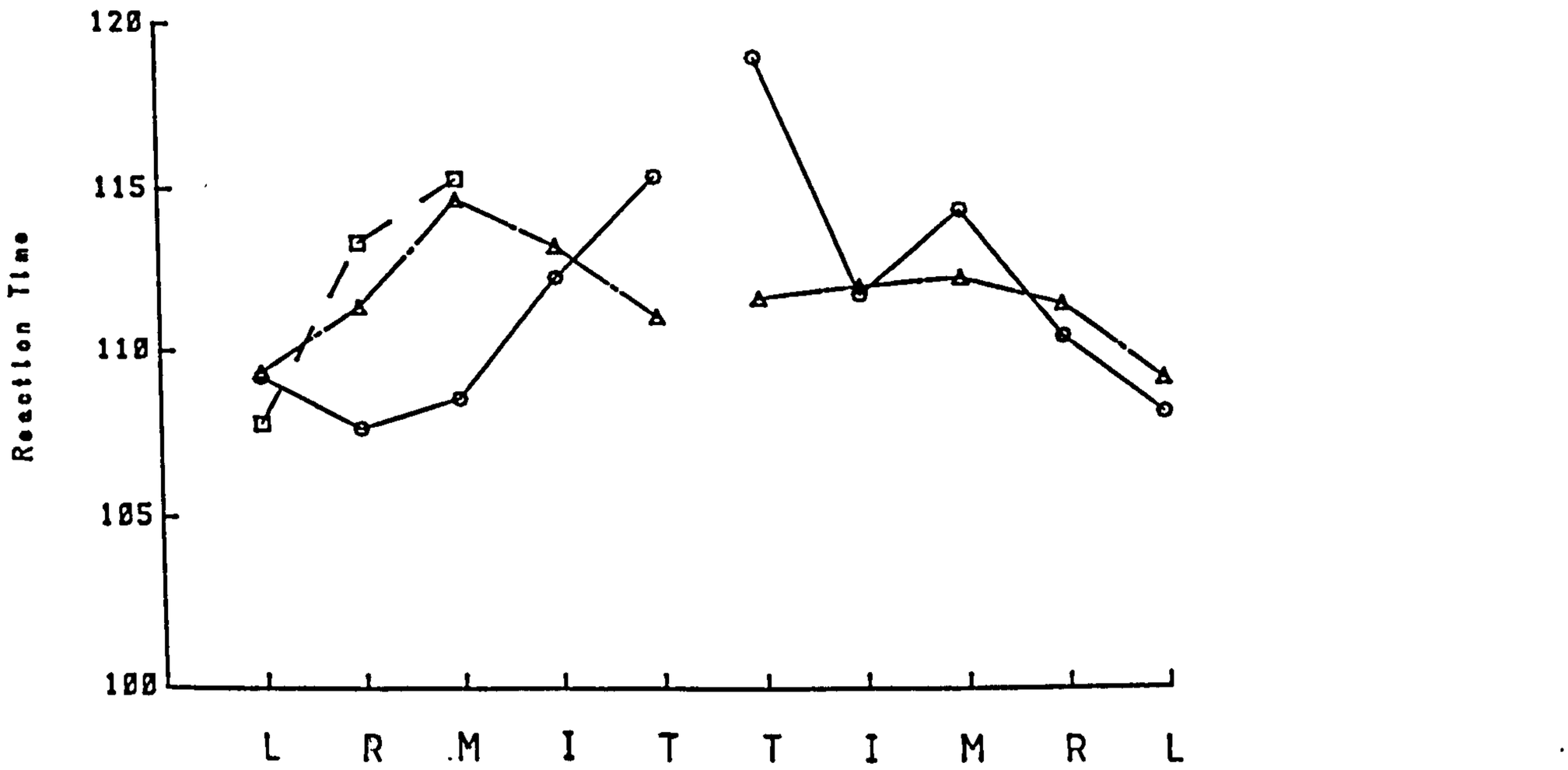


Figure 3.6a Mean reaction time (csecs) for each keystroke, for each key of the 10-key sequential keyboard during the last twenty training sessions

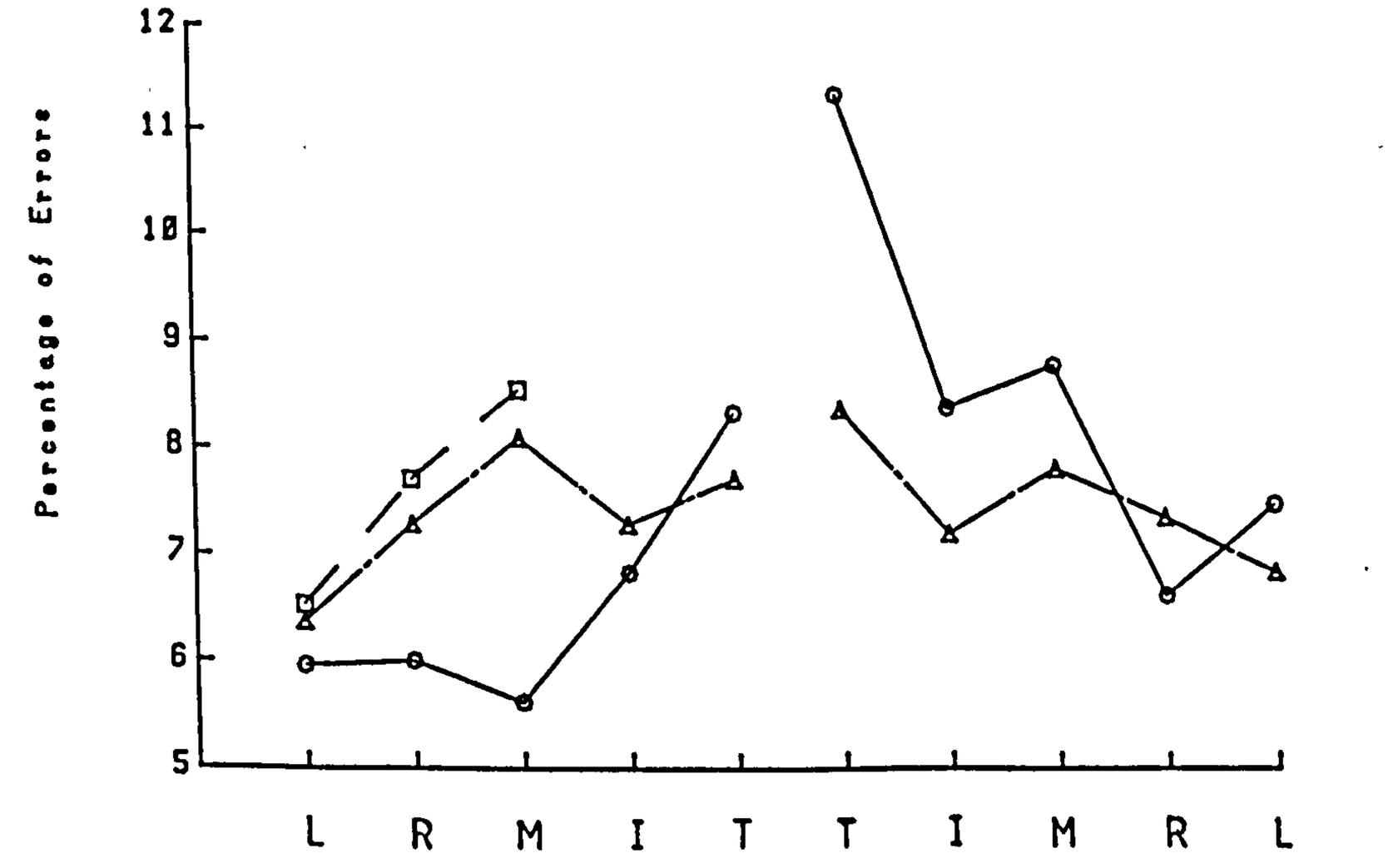


Figure 3.6b Mean percentage of errors for each keystroke, for each key of the 10-key sequential keyboard during the last twenty training sessions

small fingers of both hands producing the fastest and most accurate responses.

Further analysis was carried out to look into the effect of the following three factors on RT and percentage of error. The factors were: (1) the repetition of the same key two consecutive times in one stimulus (e.g. 116 or 277), (2) the presentation of adjacent keys in a stimulus (e.g. 128 or 278), and (3) reverse finger movements in a digit sequence (a reverse finger movement would require a reversal of direction in the digit sequence, e.g. 284 or 412).

Variable		RT	ERR
Keystrokes 1 & 2		r=.431 **	r=.173 **
Repetition	R12=1	103.02	5.61
No repetition	R12=0	112.72	7.66
Keystrokes 2 & 3		r=.269 **	r=.235 **
Repetition	R23=1	105.69	4.76
No repetition	R23=0	112.72	7.73
Keystrokes 1 & 2		r=.011 ns	r=.126 *
Adjacent	A12=1	111.25	6.34
Non-adjacent	A12=0	111.74	7.69
Keystrokes 2 & 3		r=.023 ns	r=.054 ns
Adjacent	A23=1	111.76	7.94
Non-adjacent	A23=0	111.62	7.32
Keystrokes 1 to 3		r=.277 **	r=.101 ns
Reversals	REV=1	113.56	7.84
Sequential	REV=0	109.75	7.02

ns = not significant
* = significant, p<.05
** = significant, p<.01

Table 3.11 Differences in mean RTs and percentage of errors between key sequences as a function of key repetition, adjacency and reversals (point-biserial correlations, r, have 254 degrees of freedom)

It was expected that the repetition of keys and the presentation of adjacent keys would produce faster and more accurate responses whereas the need for reverse finger movements would result to slower and less accurate responses. These kinds of effects were expected because of the level of difficulty in the "preparation" of the response. The results of point-biserial correlations (table 3.11) support the hypotheses about the effect of repetition and reverse finger movements. The presentation of adjacent keys produces slightly faster and more accurate responses for keystrokes 1 and 2, but this is not the case for keystrokes 2 and 3.

A multiple regression analysis was carried out on the dummy variables R12 (repetition of keys 1 and 2), R23 (repetition of keys 2 and 3), A12 (adjacent keys 1 and 2), A23 (adjacent keys 2 and 3) and REV (reverse finger movements in a digit sequence). It was found that the variable REV's covariation with RT and error rate was accounted for by the variables R12 and R23 which are significant predictors of RT and error rate. A12 was a significant predictor of error rate only. The prediction equations for RT and ERR were:

Predicted RT (secs) =

1.139-0.105(R12)-0.067(R23)-0.018(A12)

where R=0.516 (F=30.47, df=3,252, p<.001) and

Standard Error = 0.065

Predicted Error Rate =

8.365-3.01(R23)-2.43(R12)-1.58(A12)

where R=0.33 (F=10.47, df=3,252, p<.001) and

Standard Error = 3.67

(4) 10-Key Chord Keyboard

The RT and error rate results, during the last 20 sessions of practice, of all sixteen right-hand chord patterns (averaged across all patterns in the left hand and vice versa for the left-hand patterns - see appendix for a complete list of all 256 stimuli, rank ordered and with their RTs and error rates) together with Ratz and Ritchie's (1961) rank order and Seibel's (1962) RTs and error rates are presented in table 3.12a. As shown from the correlation matrix (table 3.12b), even after four weeks of practice on this keyboard, the RTs of both hands correlate highly with Seibel's RTs for one-hand chord patterns, as well as with Ratz and Ritchie's rank order. Moreover, the errors made by the left hand correlate highly with those of the right hand, which implies that the kind of errors made by both hands is quite similar.

The reaction time and error rate results for each finger were computed by taking the mean RT and percentage of errors of all chord patterns that involved the use of the specific finger under consideration. Once more the "end-anchoring" effect is present (figures 3.7a and 3.7b), especially pronounced on the RT functions. The little finger and the thumb of both hands seem to respond faster and more accurately.

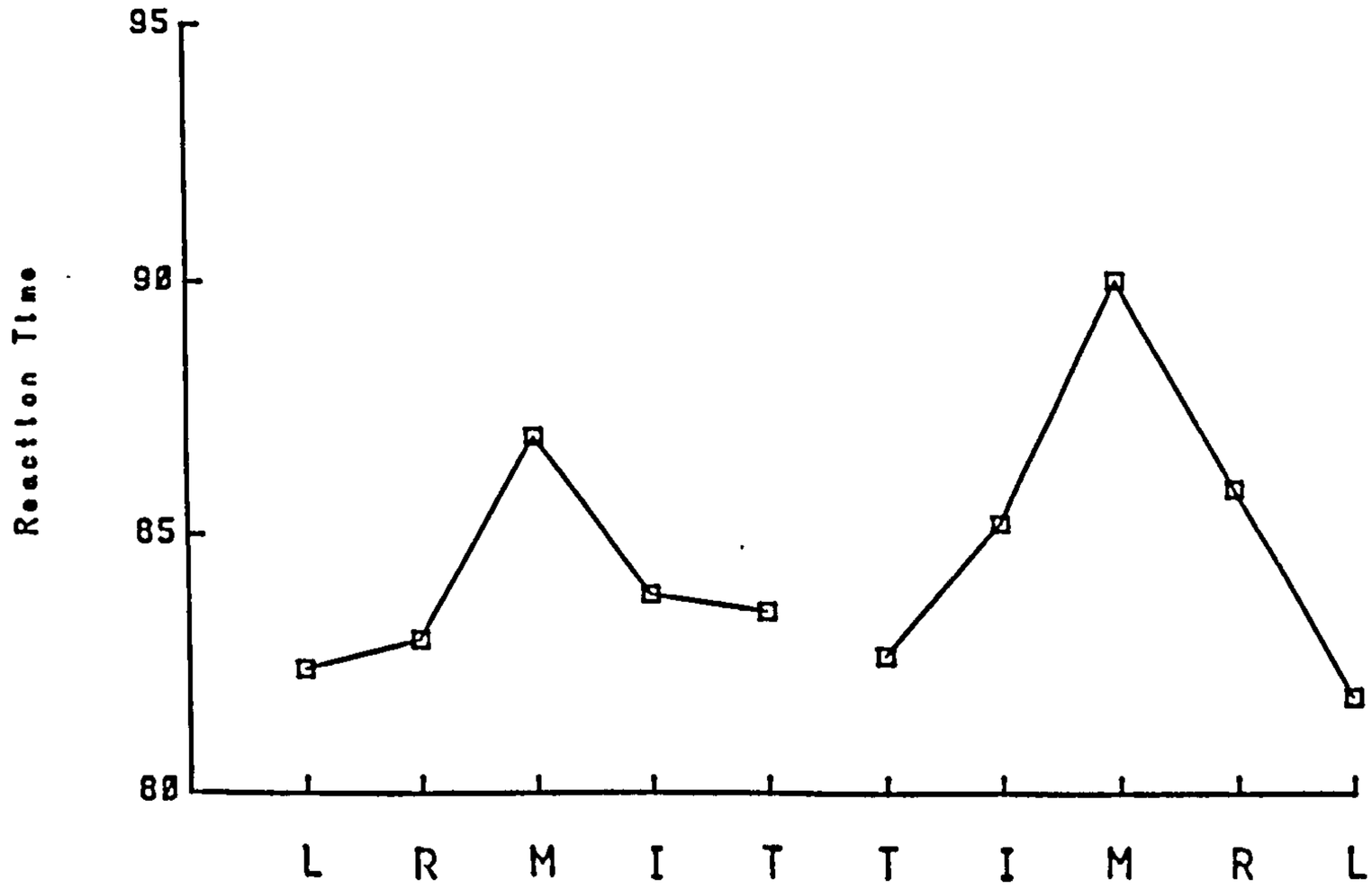


Figure 3.7a Mean reaction time (csecs) for each finger (averaged across all chord patterns) of the 10-key chord keyboard during the last twenty training sessions

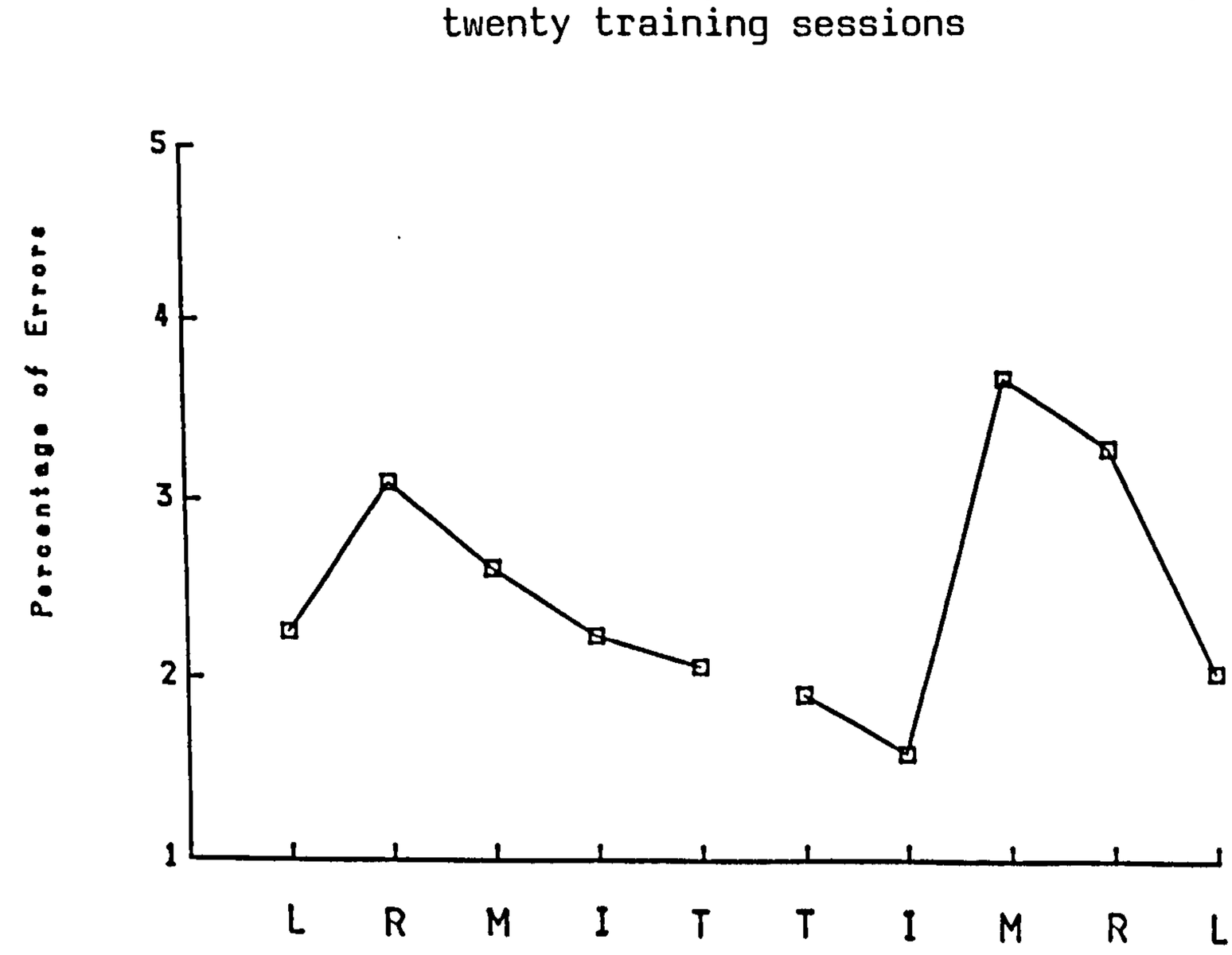


Figure 3.7b Mean percentage of errors for each finger (averaged across all chord patterns) of the 10-key chord keyboard during the last twenty training sessions

	LEFT HAND			RIGHT HAND			SEIBEL		R&R	
	1	2	3	4	5	6	7	8	9	10
CHORD	RT	SD	ERR	RT	SD	ERR	RT	ERR	Rank	NF
10000	.82	.14	2.26	.82	.15	2.04	.289	1.8	2	1
01000	.83	.15	3.10	.86	.14	3.29	.292	5.0	5	1
11000	.94	.17	4.68	.86	.18	3.08	.310	6.2	6	2
00100	.87	.16	2.62	.90	.16	3.68	.285	2.4	3	1
01100	.91	.15	2.51	.89	.15	2.71	.306	8.8	8	2
11100	1.03	.19	5.88	.95	.16	4.25	.315	5.3	9	3
00010	.84	.14	2.24	.85	.15	1.58	.281	5.9	4	1
00110	.88	.14	2.90	.93	.16	3.09	.306	10.3	11	2
01110	.88	.14	2.38	.89	.15	2.19	.311	9.1	10	3
11110	.91	.14	2.80	.94	.16	3.85	.314	4.1	15	4
00001	.84	.13	2.07	.83	.13	1.91	.294	5.6	1	1
10001	.89	.14	2.31	.87	.17	2.13	.315	5.6	7	2
00011	.91	.17	3.29	.93	.16	2.70	.316	11.5	12	2
00111	.99	.19	3.23	1.01	.17	4.56	.321	7.6	16	3
01111	.95	.16	2.97	.96	.17	2.97	.317	4.4	14	4
11111	.97	.16	3.60	.97	.17	4.74	.325	7.4	13	5

Table 3.12a Right-hand chord pattern (left-hand chord patterns are exact mirror images of the right-hand ones), mean RT (secs), standard deviation (SD) and percentage of errors (ERR), for each hand and each chord pattern, plus RT and ERR from Seibel's data, the rank order obtained by Ratz and Ritchie, and the number of fingers required for the depression of each stimulus

	1	2	3	4	5	6	7	8	9	10
1										
2	.84									
3	.77	.75								
4	.80	.67	.41							
5	.71	.58	.51	.64						
6	.73	.67	.60	.80	.53					
7	.78	.47	.44	.35	.71	.54				
8	.20	.16	.07	.31	.13	.00	.45			
9	.67	.42	.26	.90	.63	.59	.84	.44		
10	.71	.31	.34	.77	.60	.61	.83	.18	.83	

Table 3.12b Intercorrelations of all of the above variables

In the present experiment, the 16 chord patterns that were used differed in the number of fingers involved in each hand. From the results of Experiment 1, it was expected that the number of fingers (NF) involved in a chord pattern would affect performance. Indeed, it was found in this experiment as well that both RT and error rate, for both hands, correlate highly with the number of fingers involved in the chord pattern (table 3.12b).

Since the correlations between NF and RT and NF and error rate were so high, a curvilinear multiple regression was carried out to examine whether RT and error rate can be predicted by the number of fingers involved in a chord pattern (figures 3.8 and 3.9). The variables used in this analysis were the number of fingers used by the left hand (NFL), the number of fingers used by the right hand (NFR), the interaction of the two (NFL*NFR), and the square values of NFL and NFR (NFL2 and NFR2). The prediction equations that were produced were:

Predicted RT (secs) =

0.5919+0.1232(NFL)+0.0881(NFR)-0.0134(NFL2)-0.0074(NFR2)-0.0066(NFL*NFR)

The multiple correlation was quite high [R=0.76 (F=68.48, df=5,250, p<.001), Standard Error = 0.055]

Predicted Error Rate =

0.0943+1.222(NFL)+0.083(NFR2)-0.177(NFL2)

The level of prediction for error rate was lower than that for RT [R=0.306, F=8.68, df=3,252, p<.01, Standard Error = 2.171].

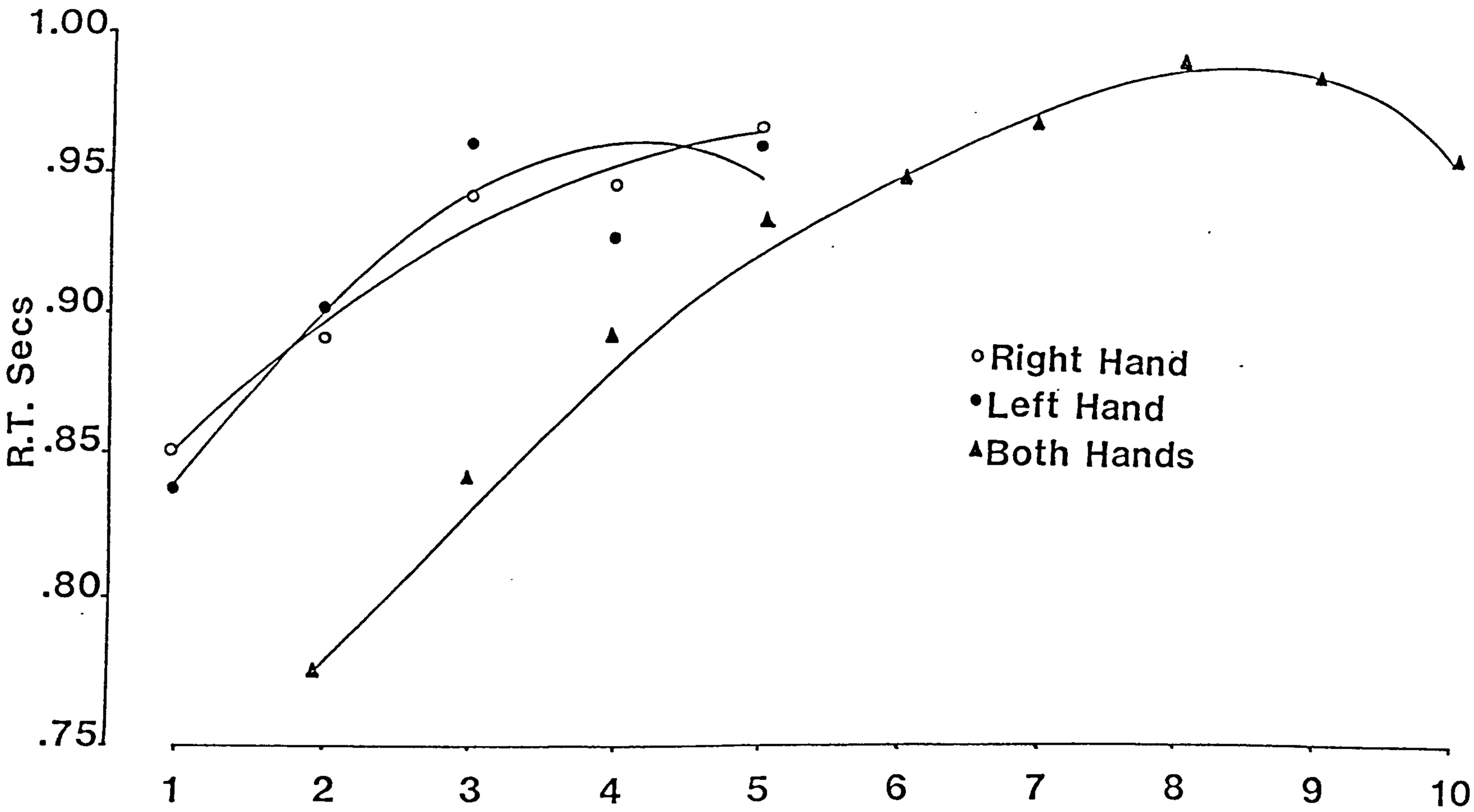


Figure 3.8 Mean reaction time as a function of the number of fingers, for the 10-key chord keyboard (obtained and predicted values from the last twenty training sessions: values for the left hand are averaged across right-hand chords and vice versa)

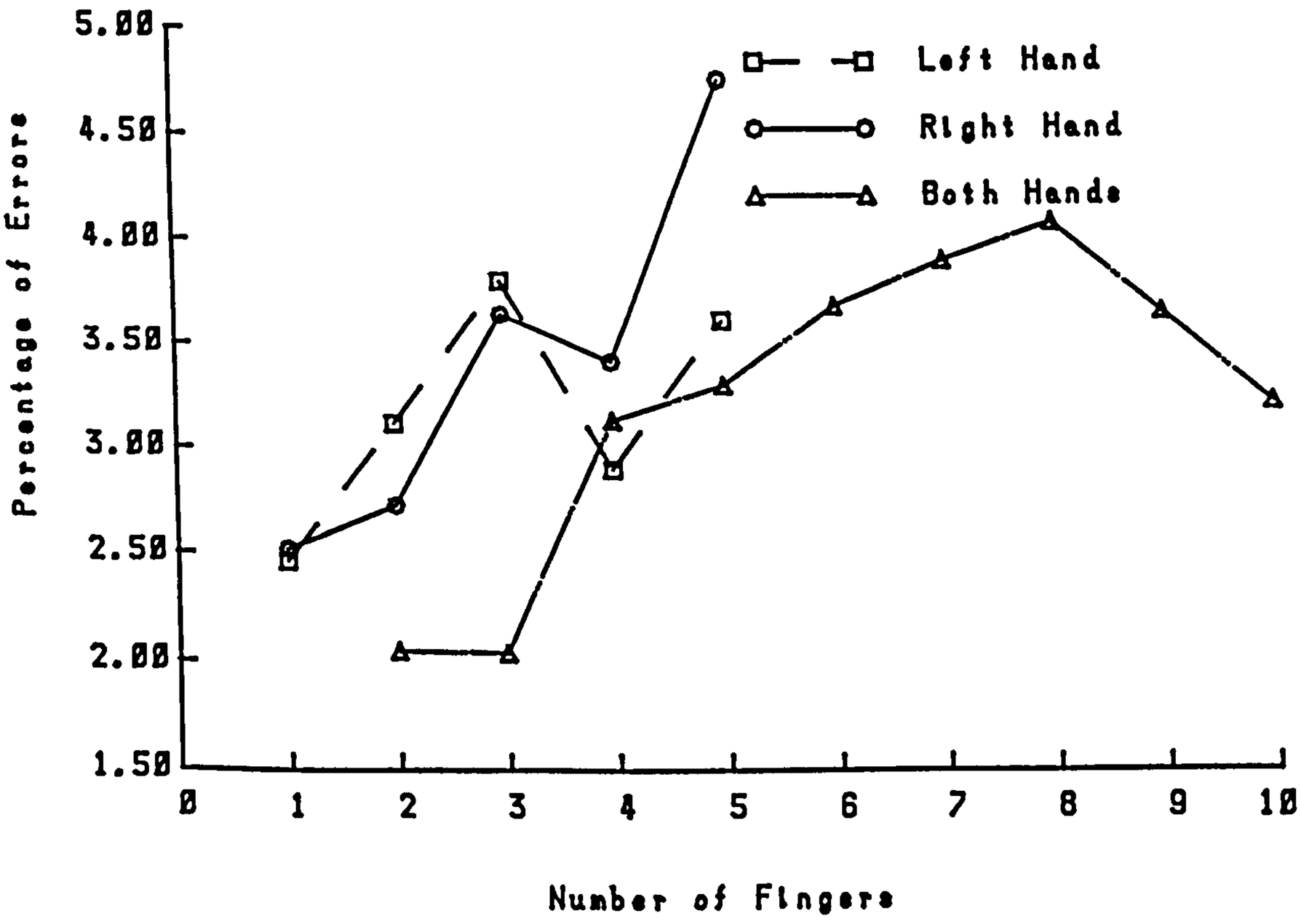


Figure 3.9 Mean percenntage of errors as a function of the number of fingers for the 10-key chord keyboard (obtained values from the last twenty training sessions)

3.4 CONCLUSIONS AND RECOMMENDATIONS

The results showed quite clearly that the 10-key chord keyboard was the most efficient of the four, in terms of speed, accuracy and rate of improvement. The 4x4 keyboard was almost as efficient in terms of speed as the 10-key chord, but produced a less stable reduction in error rate. The 10-key sequential keyboard proved to be the least efficient of the four keyboards in speed, accuracy and improvement rate. It has to be recognised that subjects' performance on the 10-key sequential keyboard must have been reduced as a result of the lower (by comparison to the other three keyboards) display-control compatibility. However, it is expected that further training on these keyboards would result in a bigger difference in efficiency between the chord keyboards and the sequential one, as the time taken for the "execution" of a key sequence, itself, is necessarily longer than that for a chord pattern.

The 2x6 keyboard was less efficient than the 4x4 and the 10-key chord keyboards, despite the fact that there were only 144 stimuli to be learnt whereas there were 256 for the others. The main disadvantage of this keyboard is that it has rows of six keys to be operated by four fingers.

In addition to the superiority of the 10-key chord keyboard in terms of its operation, a practical advantage of it is that it can cope with 1023 possible key combinations. Thus, if the Post Office's demands require an increase in the number of destinations this keyboard would not have to be replaced.

Consequently, the 10-key chord keyboard was recommended as a substitute for the 2x6 keyboard for the foreign destination coding desks at the Post Office. However, on the basis of the ergonomic considerations discussed in chapter 1, further recommendations were made to the Post Office. More specifically, it was recommended that the design be modified so that the two sets of keys for each hand are physically separated and laterally inclined to reduce muscular strain on the operator. This design modification would result in a more comfortable position for the wrists, forearms, upper arms and shoulders. Furthermore, it was recommended that pads are incorporated in the design for the hands to rest during keyboard operation.

The Post Office adopted the present recommendations and the final keyboard design (figure 3.10) incorporated these suggestions.

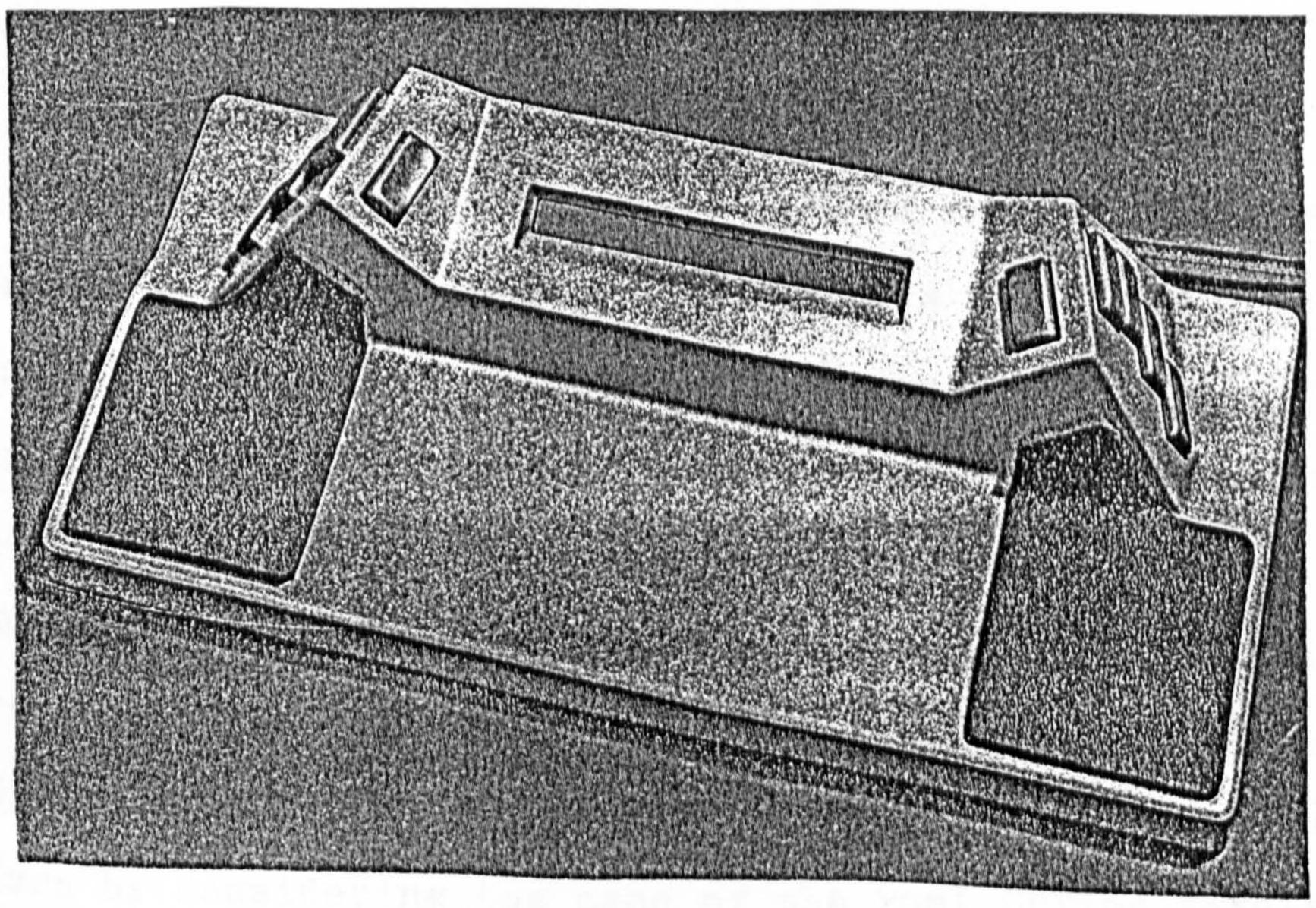


Figure 3.10 The 10-key chord keyboard design adopteed by the Post Office

Chapter 4

DESIGN OF AN EFFICIENT TRAINING METHOD FOR OPERATORS OF THE 10-KEY CHORD KEYBOARD

Experiment 4

4.1 Introduction

The implementation of a human-machine system requires careful consideration of both of its main components - the machine and the human. The increasing complexity of these systems has made it necessary to recognize that, to the extent that humans take action which influences the system, the human element is very important. Even the best designed and best constructed machines, unless operated by skillful operators, will not usually realize the potential system efficiency. This is particularly important for training in industries where individual performance efficiency is valued most. Operators need to be trained to operate machines efficiently and there is almost no controversy about it.

The design of a training programme is a complicated procedure where a number of factors have to be considered. It has to be recognised that training is only one consideration for a human-machine system to work effectively; the others are the machine itself, the human operators themselves and the possible interactions between the three. In practice, it is often the case, that some of these considerations and solutions cannot be implemented. A relevant example can be given by considering the case of the Post Office Foreign

Destination Coding Desk operators. Although it is recognised that keyboard operation involves perceptual-motor skills which vary among different individuals, for organizational reasons they cannot adopt any selection procedures. The job of Foreign Destination Coding Desk operators is carried out by postmen who after a number of years' work have been promoted. This implies that, under the present conditions, all postmen have the opportunity to be promoted to the post irrespective of their perceptual-motor skills. Therefore, the Post Office have to "make up" for it by introducing efficient equipment and by providing effective training to their staff. The problem for the Post Office, however, is similar to that for all organizations and industries: "optimum" training, in other words, minimum training time and cost for maximum training efficiency.

A relatively simple but quite comprehensive systems approach to training has been presented by Eckstrand (1964, figure 4.1). According to this approach, the first step is to define the objectives of the particular training programme. Once the training objectives have been defined, it is possible to develop criterion measures which can test whether people have reached these objectives at the end of their training. It is important to mention that if these objectives have not been reached by the end of the training procedures, it is the system which has failed and not the trainees. Following the definition of the training objectives, the content of the training must be specified so that the training objectives are obtained. Once it has been decided what should be included in

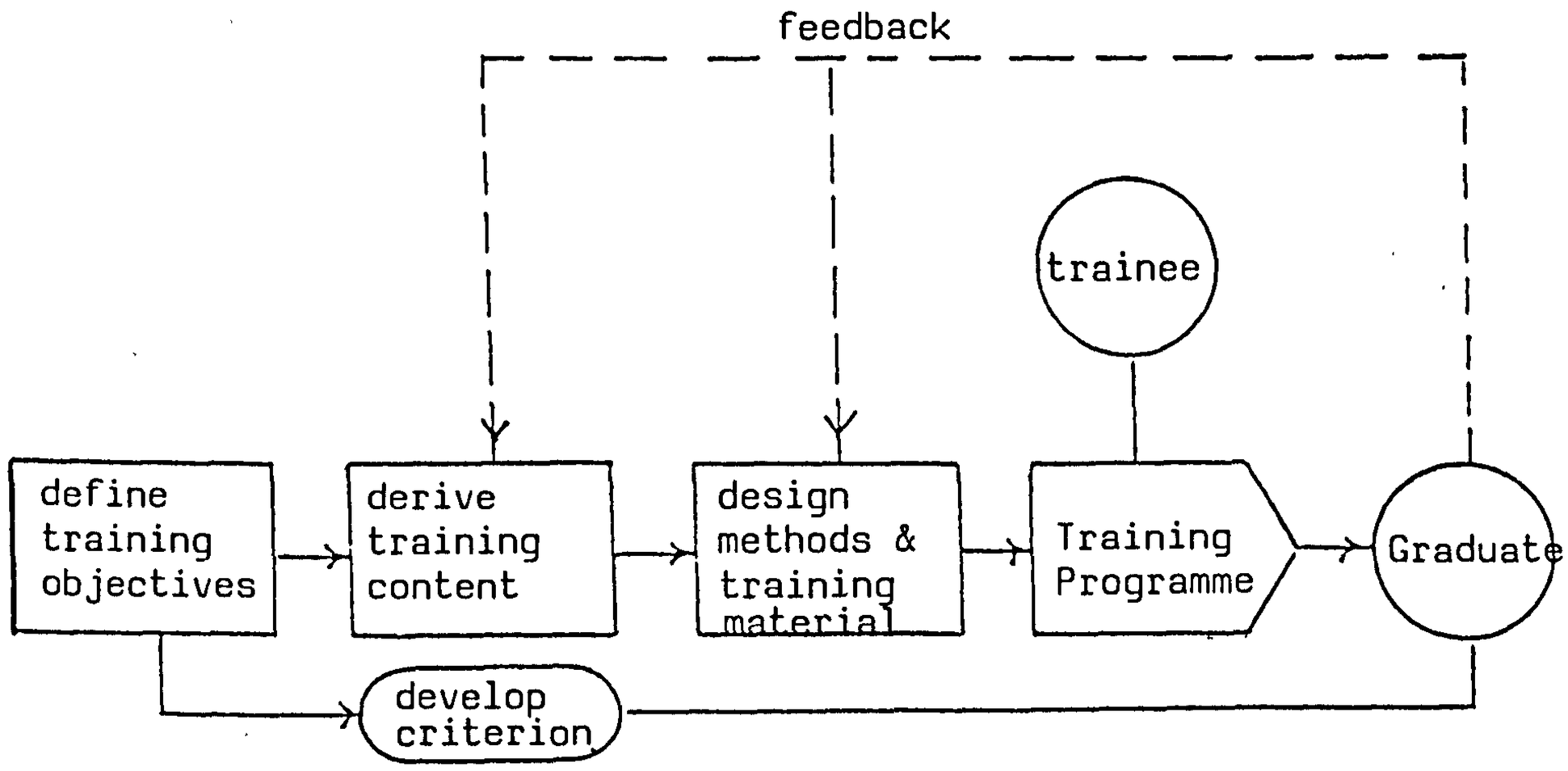


Figure 4.1 Eckstrand's (1964) training system

the training, the specific training methods, material, equipment, etc. have to be chosen. Thereupon, the training programme is implemented. After their training, individuals are tested against the criterion measures. If there are any discrepancies, modifications are made either at the training content or the training methods and material.

In the specific case of the Post Office, the problem to be investigated was related to finding the training procedure for the 10-key chord layout that would result in reaching a criterion level of performance in the minimum time. As already discussed in Chapter 1, the work of the operators at the Letter Coding Desks involves the sorting of letters by reading the address on the envelope (display encoding), relating the destination on the envelope to an internal representation of the letter code for that destination (display-response mapping), retrieving the appropriate motor-response programme and preparing the response (response preparation) and pressing the key combination that will activate the machine to forward the letter to the appropriate box. In seeking the best possible training method for this job, several issues have to be considered and a number of decisions be made.

The main objectives which had to be obtained by training operators were speed and accuracy of performance on this particular task. Therefore, training operators should involve teaching them: (1) one code (i.e. key combination) for each geographical destination; (2) how to depress the chosen key combination efficiently (i.e. quickly and accurately).

In devising the training procedures and the material for training, the first decision that had to be made was whether operators should be trained on the whole task (i.e. read the address on the envelope and depress the correct key combination for that destination) or separately on each component of the whole task (i.e. the associations between geographical destinations and codes and the efficient entry of a code on the keyboard). In other words, the first decision was related to part- versus whole-training.

Part Versus Whole Task Training

Naylor (1962) examined thoroughly all part- versus whole-task training studies since 1930. He found that the results of each research study were directly related to the particular characteristics of the specific task on which people had to be trained. Naylor used two basic terms to describe task difficulty. First, he used "task organization" to refer to the extent to which the subtasks are interrelated, that is, are dependent upon or influence the other. Second, he used "task complexity" to refer to the difficulty of each of the separate task components viewed individually. Naylor proposed that the total difficulty of a task was the product of task organization and task complexity.

Based on his literature review he suggested two training principles with regard to part- versus whole- training:
(1) When a task has a relatively high organization, then with increases in task complexity, whole-task training should be relatively more efficient than part-task training.

(2) When a task has a relatively low organization, then as task complexity increases, part-task training becomes superior to whole-task training.

Naylor's (1962) hypothesis which is supported by other studies as well (e.g. Bilodeau, 1954, 1955, 1957; Briggs & Naylor, 1962; Naylor & Briggs, 1963) was adopted in the design of the Post Office training system. The task that operators have to perform at the Foreign Destination Coding Desks is one of low organization. Although the efficiency of final performance is a function of the efficiency with which each subtask is performed (i.e. reading the address on the letter, relating this address to a "code", retrieving the appropriate motor programme to process this "code", and executing it), each subtask is relatively independent of the others. That is, the various subtasks are different in the kind of skills needed for each one to be performed efficiently (e.g. an operator may read the address and relate it to the appropriate "code" quickly but be slow in depressing the appropriate keys on the keyboard). Thus, task organization is "low". Consequently, according to Naylor's (1962) principles, part-task training should be more efficient in this case.

The present study was carried out to investigate the relative merits of alternative methods of designing the first part of the training system: the part dealing with "keyboard familiarization". This was to involve training people to operate the keyboard under high compatibility display-control conditions.

Computer-Based Training

It had been decided that the training of operators to use the 10-key chord keyboard would be computer-based. The main reason for this decision was "technical". For the keyboard to operate, it has to be connected to some kind of machinery, which can be either the actual Foreign Destination Coding Desks or a microcomputer. The microcomputer was, obviously, the logical alternative, given that the number of Foreign Destination Coding Desks at the Post Office is limited, permanently positioned in certain Letter Sorting centres around the country and in continuous use for letter sorting. However, the choice of computer-based training (CBT) was made with full awareness of its adavantages as well as limitations.

Advantages of CBT

CBT offers potential advantages (Dean and Whitlock, 1983; Shoemaker and Holt, 1965) over traditional types of training in a number of cases:

(1) When a small number of people need to be trained over a long period of time, unless CBT is used, a number of instructors have to spend a lot of time to train very few people.

(2) When trainees cannot be released from their job to attend a training course, CBT offers the opportunity for them to be trained for a few hours a day, at or near to their place of work.

(3) When many people need to be trained, CBT saves time and "boredom" to instructors who, otherwise, would be teaching

the same material again and again.

(4) When trainees differ in their initial levels of skill but need to be of equal levels by the end of their training, CBT can make it possible for every trainee to progress at his/her own speed until he/she has reached the same level of competence that is required. Although it is not always the case that a CBT system allows trainees to pace their own training, it is probably easier and cheaper to devise a CBT system that allows for self-pacing than to use self-pacing in more conventional types of training.

(5) Large organizations with many branches have the problem of whether to centralise training (which involves costs of sending trainees to one place) or send instructors to different branches. A CBT system could be a cheaper and more convenient solution.

(6) When the task (or skill) to be taught is too dangerous to be taught "on-the-job" or when the consequences of an error are too serious for an organization, CBT can provide simulation of varying complexity without any risk.

According to Dean and Whitlock (1983), the general advantages for the trainee are the following.

(1) It makes more effective use of trainee time.

(2) The training is available when the trainee is ready for it.

(3) The training material is presented consistently.

(4) It reduces travel time and expenses.

(5) It makes it possible for special short-term requirements

to be met (e.g. in cases of changes in policy, legislation, etc. a large number of the working force of an organization may need retraining fairly rapidly).

(6) The trainee does not need to be away from the place of work.

(7) Practical training can be made more effective.

(8) CBT may be made available to trainees at any time, day or night.

(9) Training can be carried out from home.

According to the same authors the benefits of CBT for the training department are the following.

(1) It reduces the involvement of instructors in a specific course.

(2) Amendments can be incorporated relatively quickly and cheaply.

(3) The training department can monitor more easily and more acurately the performance of trainees.

(4) CBT reduces space requirements in the training department.

(5) CBT can provide the training department with statistics about trainees' progress and therefore make evaluation of training easier and improvement of training more feasible.

Limitations of CBT

Without detracting from the importance or the validity of the foregoing it has to be added that "computer-based training is not a panacea" (Dean and Whitlock, 1983, p. 144). In some cases CBT may be inappropriate and/or too expensive by comparison to more traditional methods of training. There are

three basic problems with using CBT:

(1) Inconclusive Evidence: The evidence on the effectiveness of CBT and its superiority over traditional training is limited and inconclusive. According to three reviews of industrial training literature (Campbell, 1971; Goldstein, 1980; Hinrichs, 1976), relatively few organisations have implemented CBT. The reasons are not obvious but it is suggested that they are financial, situational and administrative. All of these three authors criticise much of the existing literature on CBT as non-theoretical, non-empirical, with inadequate emphasis on the definition and evaluation of training programmes and lacking in work on training effectiveness, comparisons between training techniques and use of experimental design.

Kamouri (1983-84, p.289) claims that "the research to date has been inconclusive in determining whether computer assisted training is more effective than traditional methods". Following her own research, she accepts that most dissatisfaction with CBT arises from users of early and relatively unsophisticated Programmed Instruction (PI) designs, which paid very little attention to cognitive factors in learning.

Nash, Muczyk and Vettori (1971) examined various studies on PI and concluded that: (1) PI almost always reduces training time to about two thirds of traditional training techniques (also reported by McCormick and Ilgen, 1980); (2) from the few studies carried out on retention, there is no clear superiority of PI techniques; (3) the application of PI

has more positive effects in industry than in academic settings; and (4) PI is superior for high and low ability groups than average groups. They suggested that the development of training technology which involves sophisticated feedback mechanisms should result in much clearer superiority of PI techniques.

In summary, the evidence on the effectiveness is inconclusive because of two reasons. First because relatively few studies have been conducted to evaluate the effectiveness of CBT programmes and second because some of the studies which have been conducted present conflicting or inconclusive results.

(2) CBT Can Be Expensive: Although in a number of cases CBT is cost-effective and therefore justified, it has to be recognized that in other cases it is not. CBT requires an initial expense which, often, is far above the cost of a new conventional training programme. First, the implementation of a CBT programme may necessitate the purchase of microcomputers, or even minicomputers which are much more expensive. Furthermore, CBT programmes have to be designed by specialists. The organization may have such specialists in their permanent workforce, if there is a continuous need for the development of CBT programmes. It is obvious that the organization must have the necessary financial resources to implement any such scheme as well as a "realistic" need for it to be cost-effective. If the organizations do not have such specialists in their workforce they have to engage an outside

contractor. In this case, as well, the implementation of CBT may be quite costly, especially if it is taken into account that the CBT programmes may take a long time to be designed and tested before they are implemented. Therefore, CBT is justified if traditional training is almost equally expensive, or if the CBT programme could be used more than just once or twice.

(3) Minimum Human Interaction: It is inevitable that CBT minimises the human interaction between trainees or between instructors and trainees. Although this, by itself, is not a limitation, it has to be born in mind. A CBT programme, no matter how sophisticated it is, has a degree of rigidity. Some old PI programmes were too rigid while today many CBT programmes are quite flexible and adaptive. Nevertheless, the degree to which they can be flexible and/or adaptive is limited.

It must be recognised that in some cases CBT cannot effectively replace the instructor (e.g. in schools, etc.). Moreover, sometimes there are benefits to the organization when a residential training course is organised and their employees from different backgrounds and locations meet informally as well as in groups. They might bring knowledge from their individual work experience to the group, in this way helping the other trainees and maybe the instructor as well. This sort of interaction with colleagues may provide fresh views and experiences and potentially increase motivation and interest or even generate new perspective on the job.

The Future of CBT

The introduction of high technology in today's offices, the decreasing prices of computers and software and the increasing familiarization of people with computers, are all factors likely to increase the use of CBT. Furhermore, CBT is now based on more sophisticated learning models than the original Skinnerian types of PI. The flexibility of microcomputers today has enhanced the possibilities of fully automated training programmes which are designed to meet particular training requirements and aim at teaching trainees with minimum use of instructors' time.

Adaptive Training

The idea of adaptive training stemmed from the need to account for trainees' individual differences in a training programme. Pask is one of the pioneers in the development of automated adaptive training techniques. He (1960) claims that a training system should be like a "conversational game" which keeps the trainees' attention and interest by adopting a partly competitive and partly cooperative strategy. He views an adaptive training machine as a closed loop system which interacts with the trainee. The system "learns" about the trainees' difficulties and proficiencies and modifies its training routine to match their training needs.

Pask devised a model of learning to enable an effective form of training to be designed. His model has two restrictions:

(1) it entails certain assumptions about trainees, and

(2) it applies only to training of "structured skills". He defines a "structured skill" as a skill which "can be reduced to a collection of component subskills which are related to each other in a specifiable fashion and for which systematic simplification or 'cuing' procedures are known" (Lewis and Pask, 1965, p. 216).

The assumption this model makes about trainees is that "their behaviour is a specially restricted kind of self-organizing system" (Lewis and Pask, 1965, p. 217). They express this assumption in the following terms.

(1) Trainees necessarily and continually learn about something, however trivial; therefore, a training system is concerned about teaching material relevant to the subject matter.

(2) Although learning may be regarded as goal-directed adaptation, a complete specification of a training system must entail the possibility of both creating and satisfying goals.

(3) Given that trainees have a finite decision-making capacity, the instructional situation has to be defined within the field of attention that trainees can contemplate. This implies a learning situation which reduces trainees' uncertainty about choosing the appropriate response to a given stimulus.

(4) Given the postulate that trainees need to maintain a certain rate of learning two more constraints have to be satisfied: (a) the training material must be intelligible,

and (b) it must provide enough variety for trainees to maintain his/her interest, motivation, etc to learn about.

Lewis and Pask (1965) argue that if these assumptions are satisfied, it can be shown that the learning and teaching processes have to be imaged at several levels of discourse depending on the specific skill being taught. As a "structured skill" consists of component subskills, the stimuli of the learning situation will belong to the subsets of these subskills. Furthermore when a "structured skill" is taught, it is possible to specify the procedures for every subskill which simplify the problems for trainees. These simplifying procedures which partially solve problems for trainees are viewed as cooperative actions.

An additional complication that Lewis and Pask note is that the various subskills of a "structured skill" are very often related. In other words, there might be positive or negative transfer from one subskill to another or there may be a precedence relationship among subskills, so that some subskills have to be taught before proceding to the next ones.

On the basis of this model, Pask viewed adaptive training systems as "a mechanised simulation of the environment" (Pask, 1964, p. 83). He said (p. 84) that the main parts of the simulation are:

(1) a display of problems (or for perceptual skills, problem sequences) selected from several alternative categories;

(2) a separate display with facilities for the controlled introduction and removal of cue information when this method of simplification is adopted;

(3) a multiple-choice response board on which the student can indicate his selection of a response alternative (for perceptual motor skills, this part of the device is replaced by a real-life response mechanism such as a typewriter keyboard);

(4) an additional display for "reinforcement" or "encouragement" variables indicating immediate and long-term success.

A schematic representation of such an adaptive teaching system is shown in figure 4.2.

Pask (1964) designed such an adaptive training system, SAKI, Self Adaptive Keyboard Trainer, to instruct the skill of card punching. This device measures both the latency and accuracy of trainees' responses. In this system, the machine "learns" about the trainee by a series of "probing experiments". In other words, the system presents the trainee with different categories of problems which are modified by "cue information". At the end of these experiments, the system has recorded performance measures for both latency and accuracy of responses (that is, the category of problems associated with each experiment). From these average performance measures the system has "learnt" about the trainee's initial difficulties and this serves as the basis of the adaptive mechanism. From that point onwards, this adaptive mechanism decides about the strategy by which sequences of problems associated with the various categories will be presented. As a result of this strategy, trainees learn and

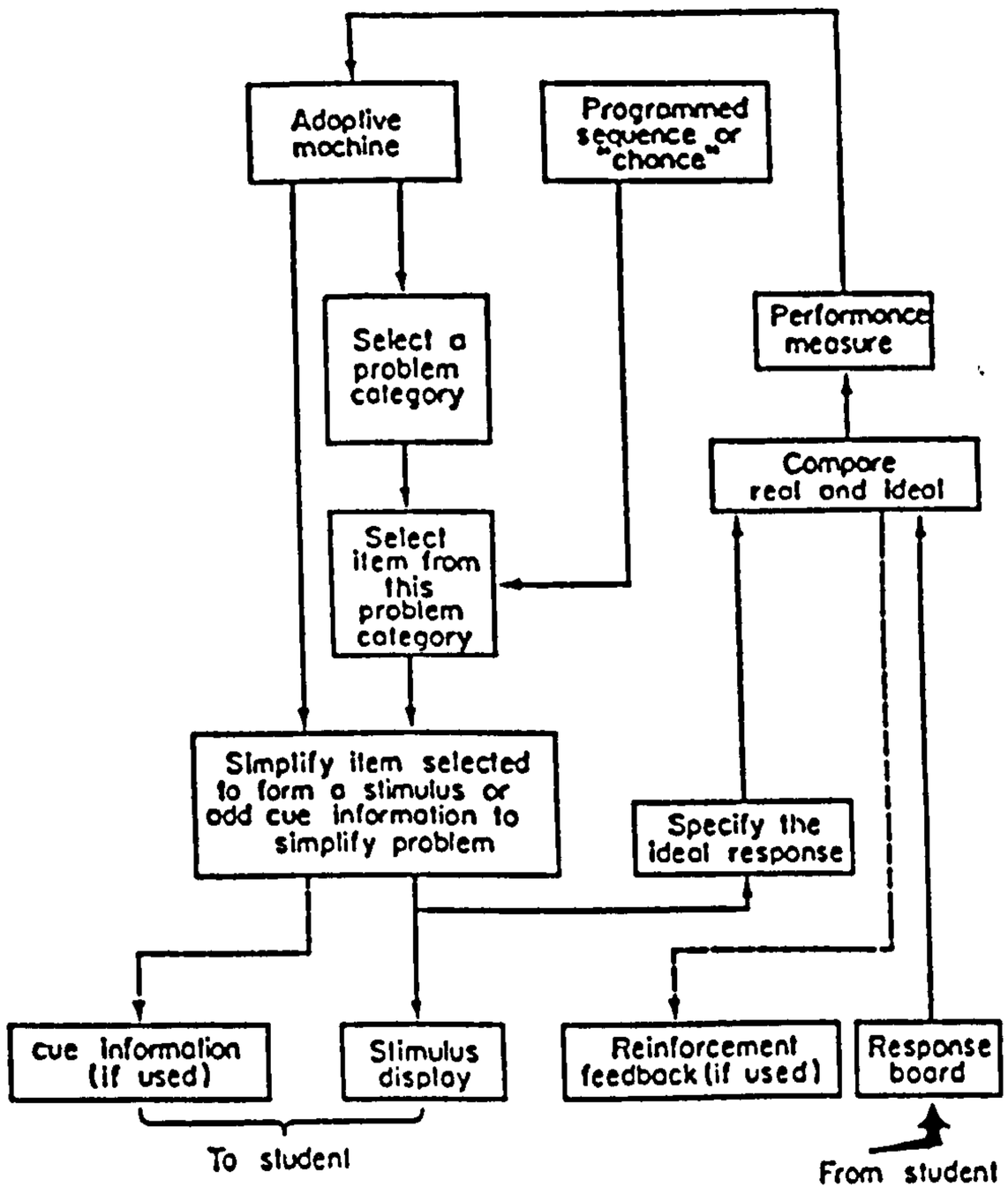


Figure 4.2 Pask's (1964) Adaptive Teaching System

the adaptive mechanism is continually modified to take account of trainees' improvement. Pask (1964) argues in favour of a strategy that maintains trainee' performance stable by increasing (or decreasing) the level of difficulty of the task.

Pask's strategy stems from cybernetic theory. In these terms, it produces "a stable interaction by maximising information trasfer between participants" (Pask, 1964, p. 91).

Kelley (1969) studied and argued in favour of adaptive training systems. He pointed out that such a system is a closed-loop feedback system and that its basic requirements are that "performance be continuously or repetitively measured in some way and that the measurement be employed to make appropriate changes in the stimulus, problem, or task" (Kelley, 1969, p. 547). He explained that adaptive training machines must have three elements: (1) a means for performance measurement, (2) an adaptive variable, which is an adjustable characteristic of the problem or task which affects the level of difficulty of the system, and (3) an adaptive logic which automatically changes the adaptive variable as a function of the performance measurement.

Kelley's adaptive training technique is based on the equation

$$dy/dt=ax-b$$

where y is the adaptive variable

x is the performance measurement

t is time

a and b are constants

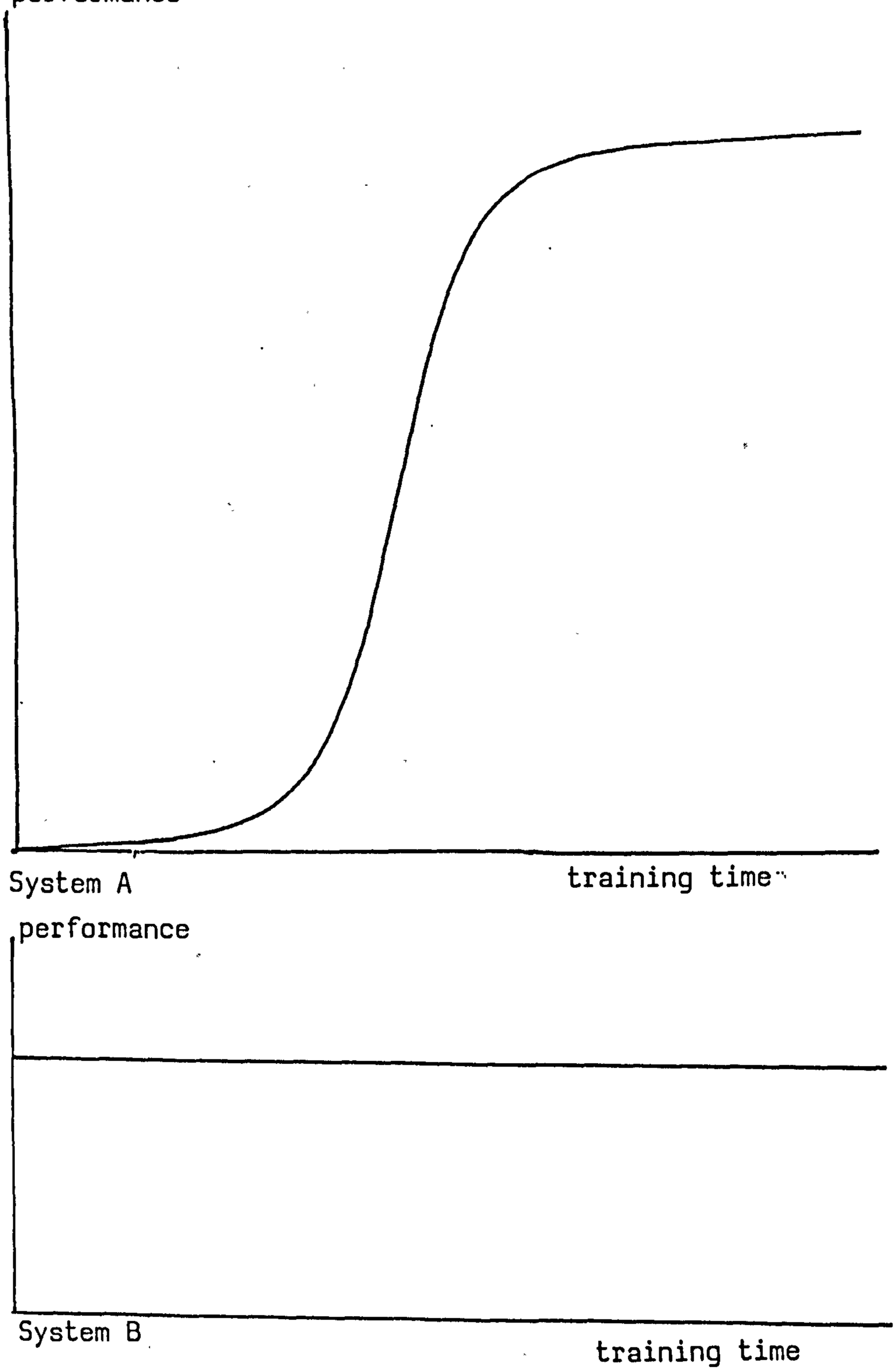


Figure 4.3 Conventional training system (system A) and adaptive training system developed by Kelley (1969) (system B)

In this kind of system, the rate of change of the adaptive variable, expressed as its time derivative, changes with the performance measurement. The performance measurement remains stable and it is only the adaptive variable that changes progressively during adaptation (figure 4.3).

In Kelley's adaptive training system, any aspect of the task which is being trained can serve as an adaptive variable, provided that it can alter the difficulty of the task. One or more adaptive variables may be used either independently or in combination. The way in which an adaptive variable (or the combination of two or more adaptive variables) is tied to performance measurement is the adaptive logic of the system. The adaptive logic may be expressed either as a mathematical equation involving the adaptive variable and the performance measurement or as a set of rules which adjust the level of difficulty of the system.

Examining both Pask's and Kelley's adaptive training systems one can find both similarities and differences. The main similarities are the following.

(1) Both agree that an adaptive training system is a closed loop feedback system where the task changes as a function of performance efficiency.

(2) Both find that adaptive training is more efficient than fixed training on the grounds that effective learning only takes place when the level of difficulty is appropriate for each individual trainee.

(3) Both agree that in an adaptive training system it is

preferable to keep performance measurement stable by varying the level of difficulty.

They seem to differ in two main points.

(1) Pask's system is based on a theoretical model which deals with more conceptual elements of learning. SAKI is just one application of this model. Kelley, on the other hand, has tried to establish an optimum technique for training tracking skills and his work was rather concentrated on the final effectiveness of the adaptive training technique.

(2) In Pask's system, the first step is an assessment of individual trainees' difficulties by a series of "probing experiments" and once this assessment is done then the machine generates the appropriate strategy which suits each trainee. In other words, training starts from easier or more difficult levels depending on individual difficulties. In Kelley's system, all trainees are initially treated the same way and individual differences only affect the rate of adaptation.

Adaptive Training On Chord Keying

The first three experiments revealed substantial individual differences in performance. Therefore, it was decided that the training system for Post Office operators should be designed so as to "adapt" itself to trainees' different capabilities. Phrased differently, the system should be designed so as, from trial to trial, to present each trainee with stimuli "suitable" for the level of skill the

particular trainee has achieved.

The optimization of such a training system depends, to a large extent, on the way stimuli are presented to trainees. In the present case two factors defined item-selection. The first one related to the size of the set from which stimuli were sampled while the second to the way individual stimuli were selected (which might in turn be constrained by the first factor).

(1) Control of Size of Current Item Set

This factor involved two modes: the Fixed mode and the Cumulative mode. For the Fixed mode all items are available for presentation. For any practice trial, every item in the population may be presented. For the Cumulative mode only a subset of the item population is available originally and as performance improves the size of the subset increases until it becomes equal to the population. Increments and decrements of the size of the item subset are controlled by predefined upper and lower criteria of performance (speed and/or accuracy).

(2) Selection of Stimuli from Item Set

This factor also involved two modes: the Random mode and the Adaptive mode. For the Random mode items are presented randomly (with replacement) from the current item set. In the Adaptive mode, items are presented so that the trainee has more practice with the stimuli he/she finds most difficult. An item subset is selected from the current item set and the items are presented. The items which produced correct responses are dropped from this item subset and replaced by

new ones while the items which resulted in error are kept in the subset.

The two levels of each of these factors were completely crossed to produce four training conditions.

(1) Fixed-Random: The current set is always the whole population of items and items are sampled from that at random.

(2) Fixed-Adaptive: As for (1) the full population of items is used but items are sampled adaptively.

(3) Cumulative-Random: Here the current set is initially a small subset of the population of items and stimuli are sampled randomly from that subset.

(4) Cumulative-Adaptive: As for (3) the current set is initially a subset of the population of items but stimuli are sampled adaptively.

The aims of this experiment were twofold. First, to see whether or not there was any clear superiority of the Cumulative over Fixed training modes. Second, to see if the Adaptive modes were superior to Random ones. In addition, the experiment would provide information on the relative merits of using a combination of the factors (e.g. Fixed-Random versus Fixed-Adaptive, Fixed-Adaptive versus Cumulative-Random, etc.). It was hoped that the results of this experiment would provide information necessary for selecting the most appropriate technique for the Post Office training programme.

4.2 Method

Subjects

Forty paid subjects, male and female students from the University of Hull, volunteered to participate in this experiment. Their ages ranged from nineteen to thirty-seven years. All of them were right-handed and none had any previous experience with a 10-key chord keyboard. In addition to a basic £3 payment, a reward of £10 was offered as an extra incentive for the subject whose performance incorporated the combination of highest speed, accuracy and rate of improvement.

Apparatus

The 10-key chord keyboard (already described in Chapter 2) was connected to a BBC model B micro-computer, with a 40-track disk drive and a Sanyo green-screen monitor.

Displays and Stimuli

The same highly compatible stimulus displays that had been used for the 10-key chord keyboard in Experiment 3 were used again in this experiment.

Left-hand Chords	Right-hand Chords
0 0 0 0 1	1 0 0 0 0
0 1 0 0 0	0 0 0 1 0
0 0 0 1 1	1 1 0 0 0
0 1 1 0 0	0 0 1 1 0
1 0 0 1 0	0 1 0 0 1
1 0 1 0 0	0 0 1 0 1
1 0 1 1 0	0 1 1 0 1
1 1 0 1 0	0 1 0 1 1

Table 4.1 Eight one-hand chord patterns, for the left and right hand

Eighty stimuli were selected for the present experiment. The full range of chord pattern difficulty was used, based on Experiment 1 and the experimental results of Ratz and Ritchie (1961), and Seibel (1962). These were all possible combinations of the eight one-handed chord patterns shown on table 4.1. More specifically the eighty stimuli were divided into five difficulty levels:

(1) 16 one-handed ones including both no finger separation and finger separation (a finger separation would require the depression of non-adjacent fingers);
(2) 16 two-handed stimuli with no finger separation;
(3) 16 two-handed stimuli where the right hand pattern involved finger separation and the right hand did not;
(4) 16 two-handed stimuli where the left hand pattern involved finger separation and the left hand did not; and
(5) 16 two-handed stimuli with both hand patterns involving finger separation.

Pretest

Subjects were first pretested with the 8-choice reaction time pretest on the BBC micro-computer (as described in Chapter 2). As the pretest had been found to be of good predictive value in Experiment 3, it was used again to assign subjects to four matched groups. On the basis of their adjusted scores (ASC: total time divided by the number of correct responses) subjects were divided into four groups so that the mean ASC of the four groups were equivalent.

Training Conditions

Four training conditions were tested. They differed in the way stimuli were selected from the total set of eighty stimuli which subjects had to learn.

(1) Fixed-Random Condition (FR): In this condition each stimulus was selected randomly from the full population of eighty stimuli. The only constraint on the sampling of stimuli for the FR condition was that the same stimulus could not be presented twice in a row. This constraint applied also to the other three conditions.

(2) Fixed-Adaptive Condition (FA): In this condition, an "adaptive subset", consisting of 10 items was set up, by initially sampling at random from the full population of eighty stimuli. The ten stimuli were presented over ten consecutive trials. After the presentation of each subset of ten trials, all stimuli associated with an error response were kept in the subset. All correct responses were rank ordered in terms of speed and the five stimuli that had produced the fastest responses were dropped from the set and replaced by five new ones randomly selected from the eighty stimuli. If fewer than six stimuli had produced correct responses, then, all those that had produced correct responses were dropped and replaced. In other words those stimuli with which a subject had difficulties were presented more often than the other. Therefore, there was control over the selection of individual items.

(3) Cumulative-Random Condition (CR): For this condition the

order of difficulty for the eighty stimuli was determined on the basis of the findings of Experiment 1 and 3. The way the eighty stimuli were classified in terms of difficulty has been discussed above.

An initial "current set" of the 16 easiest items was established. In each trial, the stimuli were randomly sampled from this "current set". The initial Adjusted Score for each subject was set to be 2 secs. After every ten trials the Adjusted Score (ASC) was re-computed and on the basis of this new ASC an "increment" decision was made. The new ASC was computed as follows:

for each ten trials

SCORE = total time/number correct

New ASC = (old ASC + SCORE)/ 2

If the new ASC was larger than 2.5 secs, the four most recently added items were dropped from the "current set"; if it was between 1.5 secs and 2.5 secs the "current set" remained the same; and, if the new ASC was less than 1.5 secs, the "current set" was increased by four new items. These were selected either from the current diffuclty level (if any unsampled items remained) or from the next level of difficulty. In this condition therefore, there was control over the set of stimuli from which items were selected for presentation to subjects but individual items were randomly selected from that set.

(4) Cumulative-Adaptive Condition (CA): In this condition, the "current set" initially contained the 16 easiest stimuli. From these an "adaptive subset" of ten was drawn. After ten trials,

the New ASC was calculated and the "current set" was adjusted accordingly. Moreover, the "adaptive subset" was adjusted as for condition FA. In this condition therefore, there was control over both the set from which stimuli were selected and the individual items that were presented. But the two methods operated independently.

Design

Following the pretest, subjects were assigned to four groups that were equivalent in terms of their average ASC (table 4.2).

Group	ASC on the pretest (in csecs)
Fixed-Random	116.774
Fixed-Adaptive	116.802
Cumulative-Random	116.792
Cumulative-Adaptive	116.879

Table 4.2 Average Adaptive Scores of the Four Training Groups on the Pretest

Each group of subjects was trained under one of the four training conditions described above. Their training consisted of five sessions of 320 trials each. They had three training sessions in the morning and two in the afternoon of the same day. Between sessions they had a five-minute break. The interval between the morning and the afternoon sessions was about four hours. Following the fifth training session there was a final test session on condition FR for all four groups. The total number of trials for each subject was 1600 training trials and 320 "test" trials.

Procedure

First, subjects were given the pretest a few days before the actual experiment and were assigned to four matched groups.

Before the first training session, subjects were given instructions and a demonstration of the actual task. They were told about the number of sessions they had to do and were given specific instructions on the operation of the keyboard. The demonstration consisted of the presentation of ten stimuli and the aim of it was to ensure that subjects had understood the task before they started the sessions. They were also instructed to keep their fingers over the keys of the keyboard at all times during a session and respond as quickly and accurately as possible. This latter somewhat ambiguous instruction was given on purpose. Subjects were expected to pace themselves so that they responded optimally. If the emphasis on either speed or accuracy had been set by the experimenter, it would be likely that subjects would have been slower in adjusting their own performance to the demands of the task. The task was self-paced, in that a new stimulus appeared only after the subject had responded to the last stimulus.

During all training and test sessions subjects received feedback about their performance. Every time they made an error, there was an alerting bleep and visual feedback on the screen which showed them which keys had been pressed correctly, which incorrectly and which had been omitted

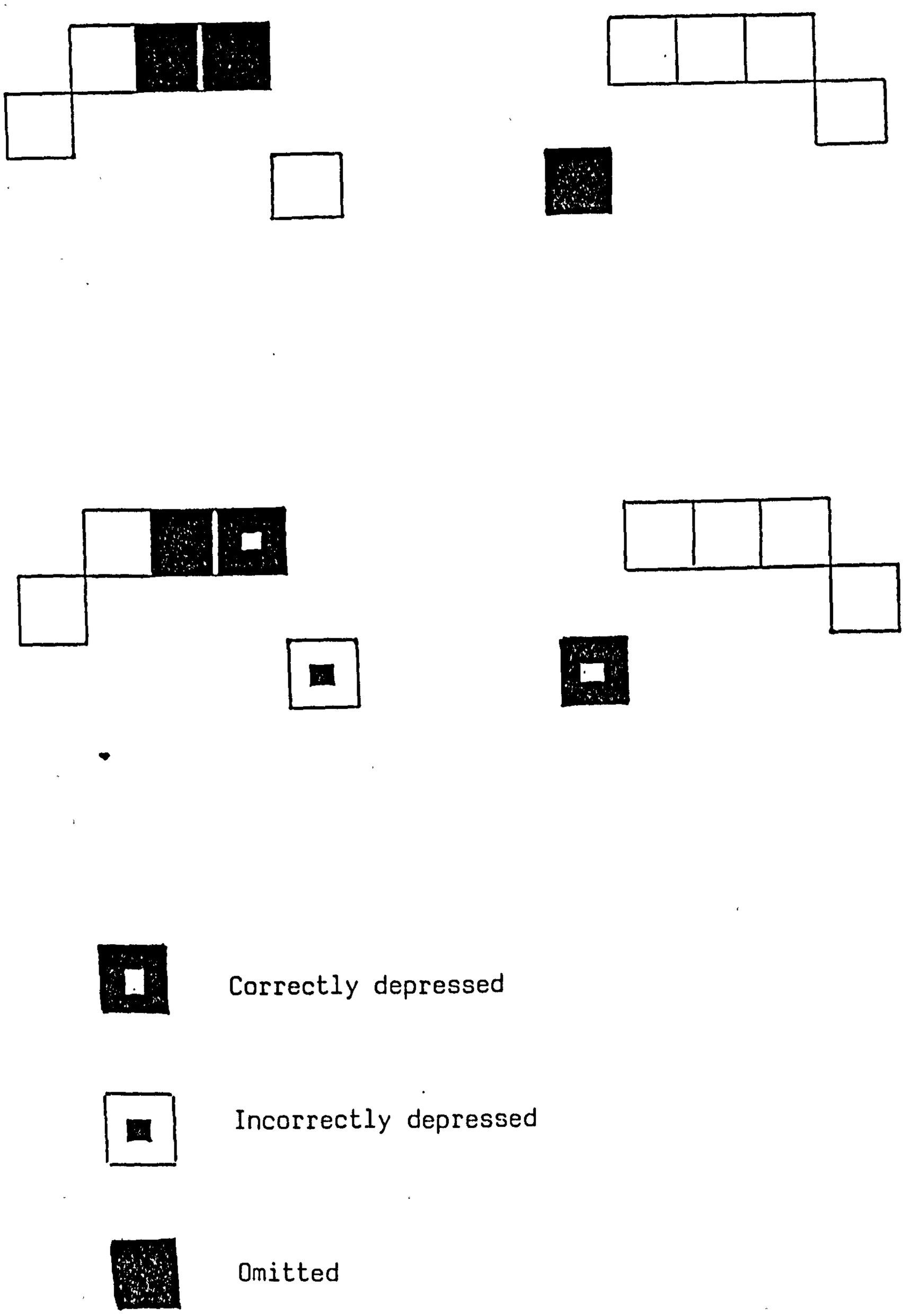


Figure 4.4. Example of stimulus and visual feedback

(figure 4.4). At the end of each session, they received further feedback about their performance in the form of total number of correct responses, average time for correct responses, total number of wrong responses and average time for wrong responses.

4.3 Results

Three measures were used in the analysis of the data: (1) the reaction time (RT), (2) the percentage of errors (ERR), and (3) a combined score of speed and accuracy (ASC) which was calculated by dividing the total reaction time by the number of correct responses.

Rates of Learning of the Four Groups

To examine the effects of training on each group the RT, ERR, and ASC measures were computed for each session of 320 trials. The mean RTs, ERRs and ASCs of all groups, for all six sessions (five training and one test session), are shown in figures 4.5, 4.6 and 4.7 (see also the relevant tables in appendix 3).

The CR condition group maintained, throughout the six sessions, the lowest ASCs and RTs. Their ERR rate was reduced from 11.2% in the first session to 8.1% in the test session. Two subjects of the CR group were not trained on all 80 stimuli (appendix 3), since their rate of progress was not fast enough to reach the criterion level that would allow them to be trained on the more difficult patterns. Consequently, some of the stimuli they received during the test session were novel.

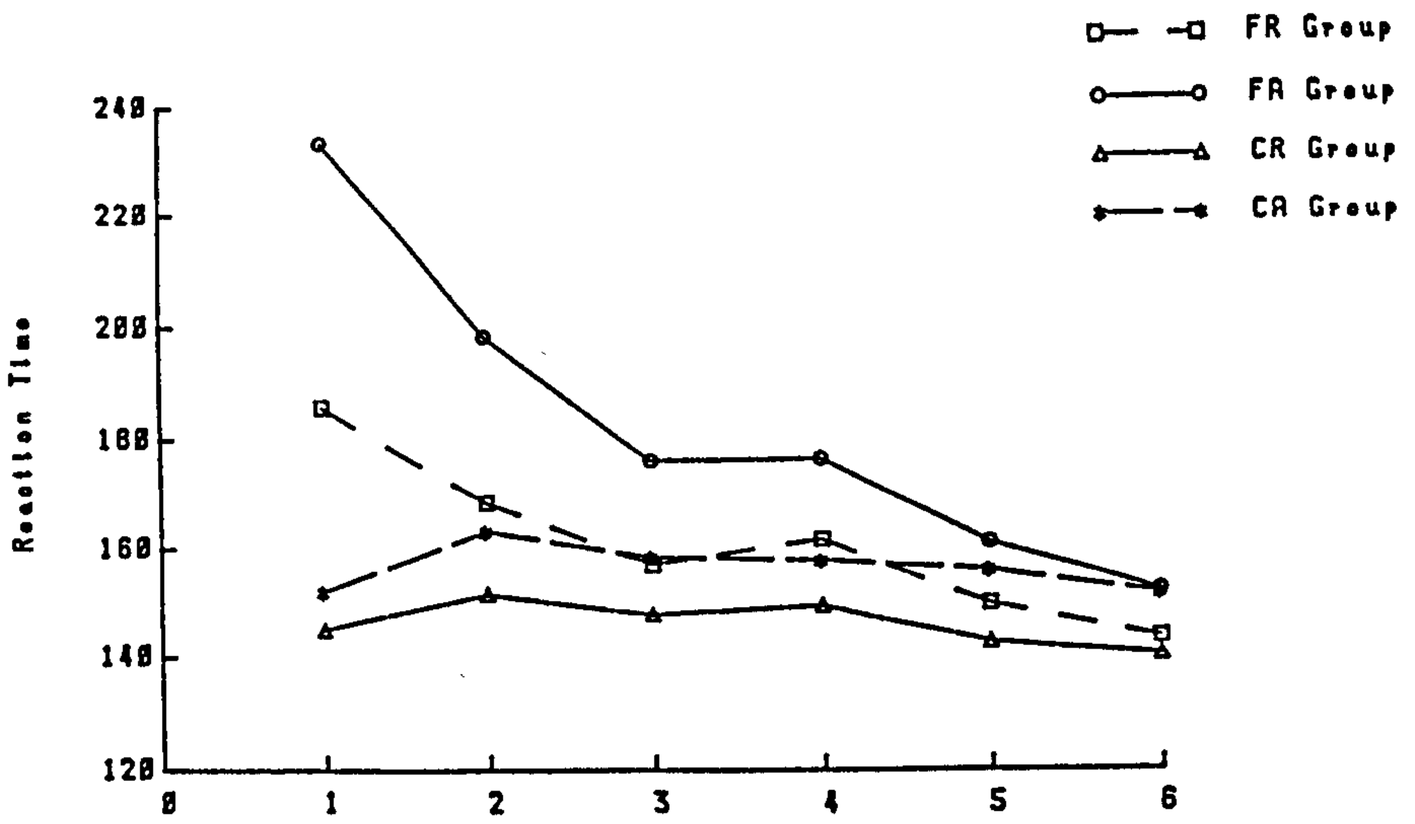


Figure 4.5 Mean RT of each training group as a function of practice

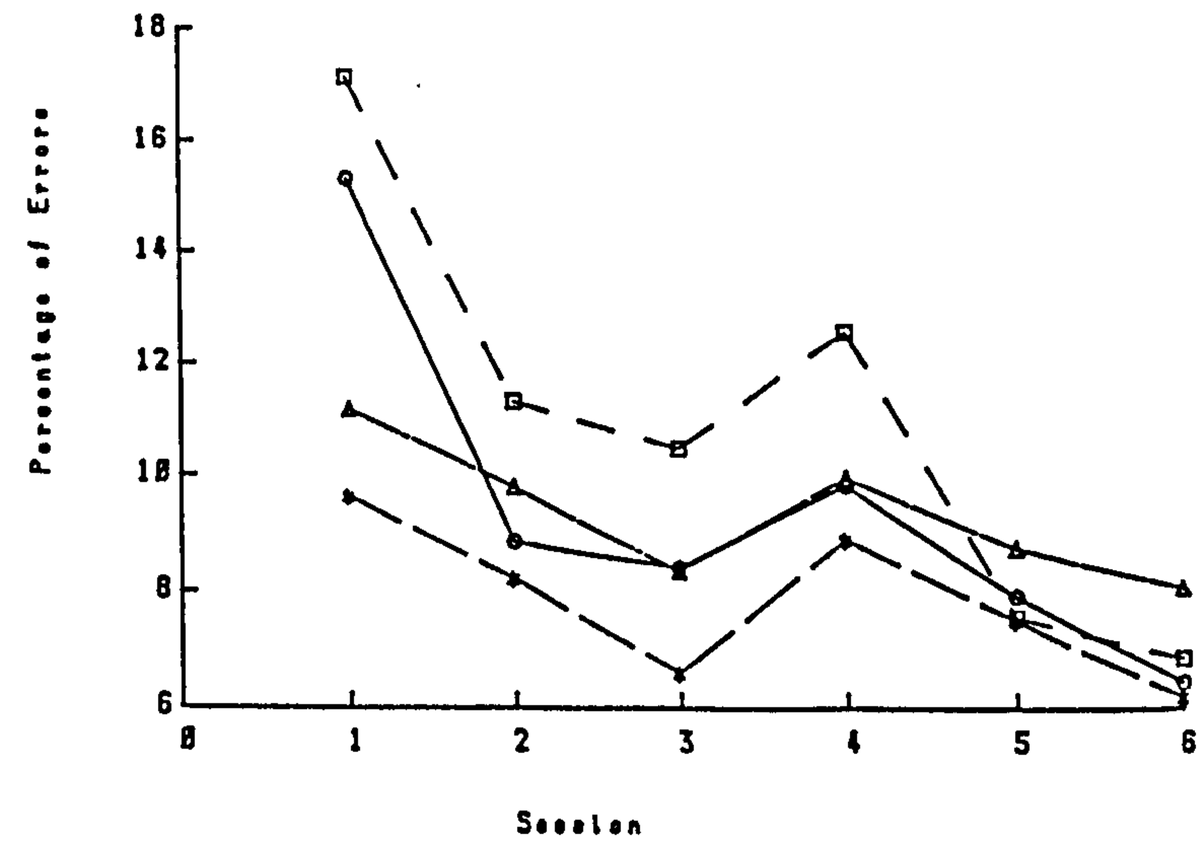


Figure 4.6 Mean percentage of errors of each training group as a function of practice

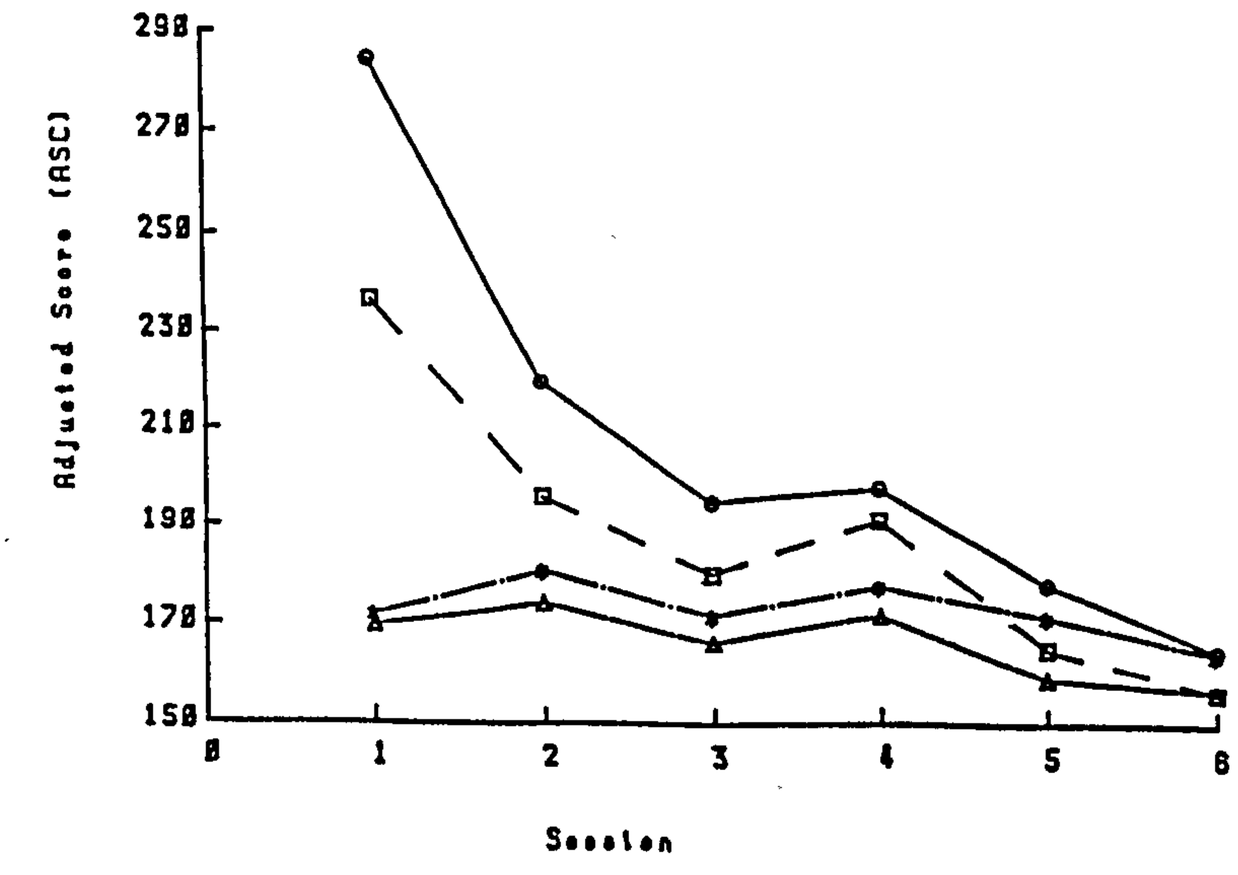


Figure 4.7 Means adjusted score (ASC) of each training group as a function of practice

The CA condition group started with the second best ASCs and RTs but in the test session they had dropped to the third place. Their ERRs were kept at the lowest rate of all groups, between 9.6% in the first session and 6.2% in the test session. Only three subjects of the CA group were trained on all 80 chord patterns. Seven of the subjects did not progress fast enough to be trained on all chord patterns (appendix 4). As a result of this, as for the two subjects of the CR group, these subjects received stimuli during the test session which they had not practised earlier.

The FR condition group started with the third best ASCs and by the test session they had reached the same level as the CR group. Both their RTs and ERRs showed a systematic improvement throughout the six sessions.

Finally, the FA condition group had the worst RTs and ASCs throughout the training and test sessions. However, their ERRs were better than both both the FR and the CR groups in the end.

The description of the data from the four groups, so far, has not shown that any one training condition resulted in substantially more efficient final performance. It has been pointed out, however, that the two Cumulative conditions (i.e. when the size of the current item set was controlled) produced better speed and the Adaptive conditions (i.e. when there was control over the selection of stimuli from the item set) slightly better accuracy. Furthermore, it appears that the Cumulative groups took a shorter time to complete their training than the Fixed groups. The analyses that follow aim,

mainly, at examining the statistical significance of these differences.

Analyses of variance on the RT, ERR, and ASC measures during the final test session revealed no statistical difference among the four groups ($F<1$). To reduce within group error variance, analyses of covariance were carried out using the pretest ASCs as the covariate (the pretest ASCs correlated highly with RT, ERR and ASC for all six sessions - table 4.3). Still, no significant differences were found among the four groups ($F<1$).

	Pretest ASC		
	RT	ERR	ASC
Session			
1	.41	.40	.47
2	.51	.52	.63
3	.58	.46	.66
4	.62	.34	.65
5	.63	.40	.66
6	.63	.41	.67

Table 4.3 Correlations between pretest ASCs and RT, ERR and ASC for the six sessions (N=40)

Given the similarity of the four groups in their final performance, the training methods had to be evaluated on the basis of other criteria. An obvious such criterion was training time. Since all groups had received the same number of training sessions, it was justified to examine the time taken by each group to complete their training (table 4.4). A two-way analysis of variance on total training time taken by each group showed that the groups that were trained under the

Cumulative conditions took significantly less time to reach the final level of performance [$F(1,36)=2.23$, $p<.05$].

Training Group	Total Training Time (mins)	SD
Fixed-Random	44.71	9.93
Fixed-Adaptive	51.02	11.28
Cumulative-Random	40.32	5.35
Cumulative-Adaptive	42.60	6.67

Table 4.4 Total training time over five training sessions and standard deviations for each group (n=10), averaged across subjects

Since there was a significant difference in total training time among groups, the learning functions had to be analysed further during the first stages of training, when the group differences were biggest. The aim of this further analysis was the examination of the effects of the various experimental parameters on the learning functions.

Examination of the learning functions during the first three training sessions

(1) Examination of the RT functions

A three way analysis of covariance was carried out on the RT measure. The ASCs of the pretest were used as the covariates and the factors under consideration were: Adaptive vs Random, Cumulative vs Fixed and practice. The Cumulative groups were significantly faster than the Fixed [$F(1,35)=13.92$, $p<.01$) and the Adaptive groups were significantly slower than the Random [$F(1,35)=5.491$, $p<.05$].

Overall, all groups' RT improved significantly with practice [$F(2,72)=45.27$, $p<.01$].

The effect of practice was not the same for all factors as the analysis revealed significant interactions between practice and the other two factors. The Adaptive groups, although overall slower than the Random, improved their speed at a faster rate than they did [$F(2,72)=4.802$, $p<.05$], as a result of practice (figure 4.8). The significant interaction between the factors practice and Cumulative vs Fixed [$F(2,72)=74.168$, $p<.01$] indicates the difference in the way the RT functions of the Cumulative and Fixed groups occurred under the effect of practice (figure 4.9). From the first session to the second, the Fixed groups became faster while the Cumulative groups became slightly slower; from the second session to the third the Fixed groups improved their speed at a higher rate than the Cumulative. Finally there was a significant 3-way interaction between practice and the Adaptive vs Random and the Cumulative vs Fixed factors [$F(2,72)=8.202$, $p<.01$].

As expected the Cumulative groups had a rather constant RT. The training system was designed so that, initially, these groups received stimuli sampled from a small set. Therefore they were faster than the Fixed groups in the first stages of training with their RT remaining basically constant as the set from which stimuli were sampled increased.

(2) Examination of the ERR functions

A second three way analysis of covariance was carried out on the ERR data this time, using the same factors and

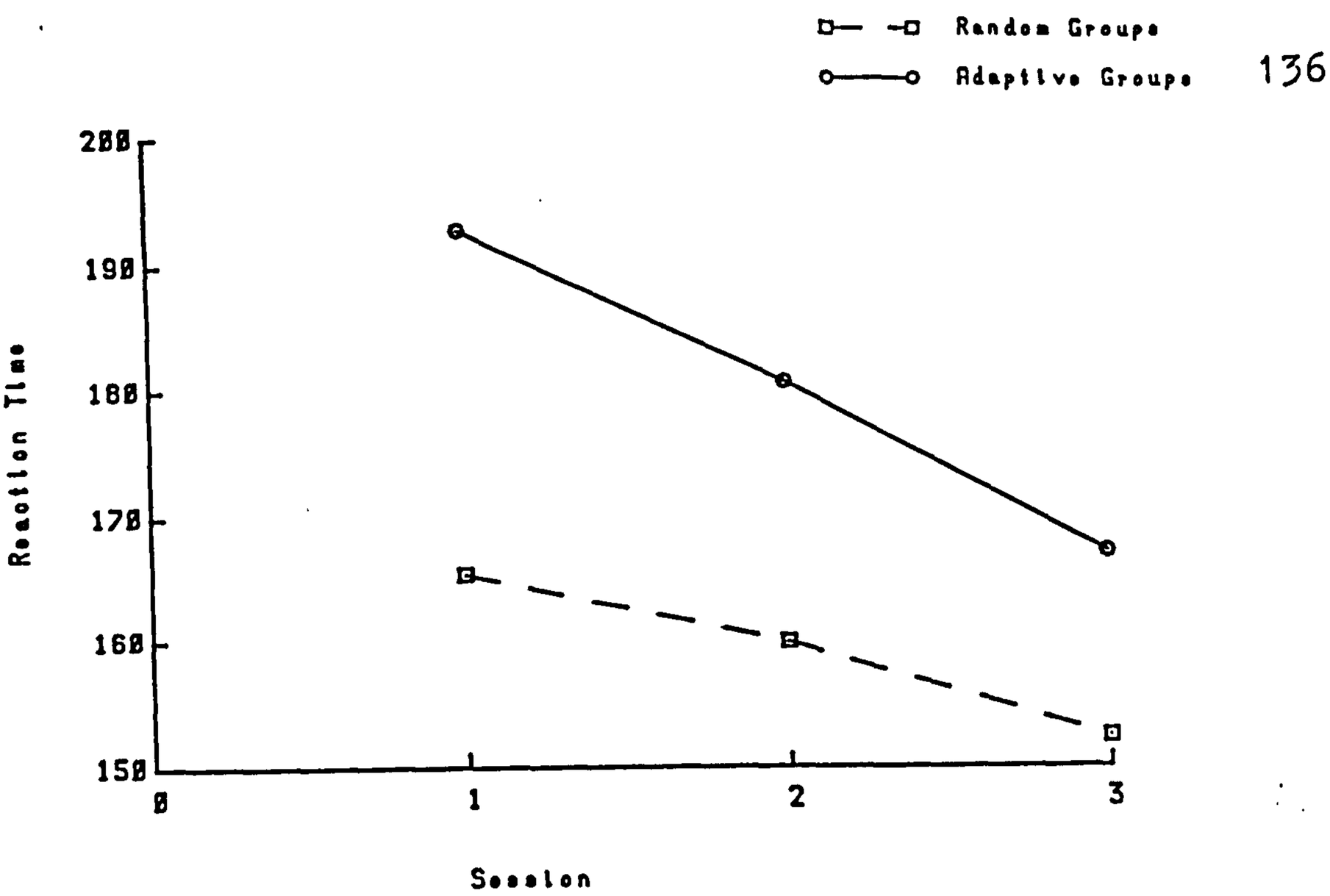


Figure 4.8 Mean RT of the Adaptive and Random groups during thr first three training sessions

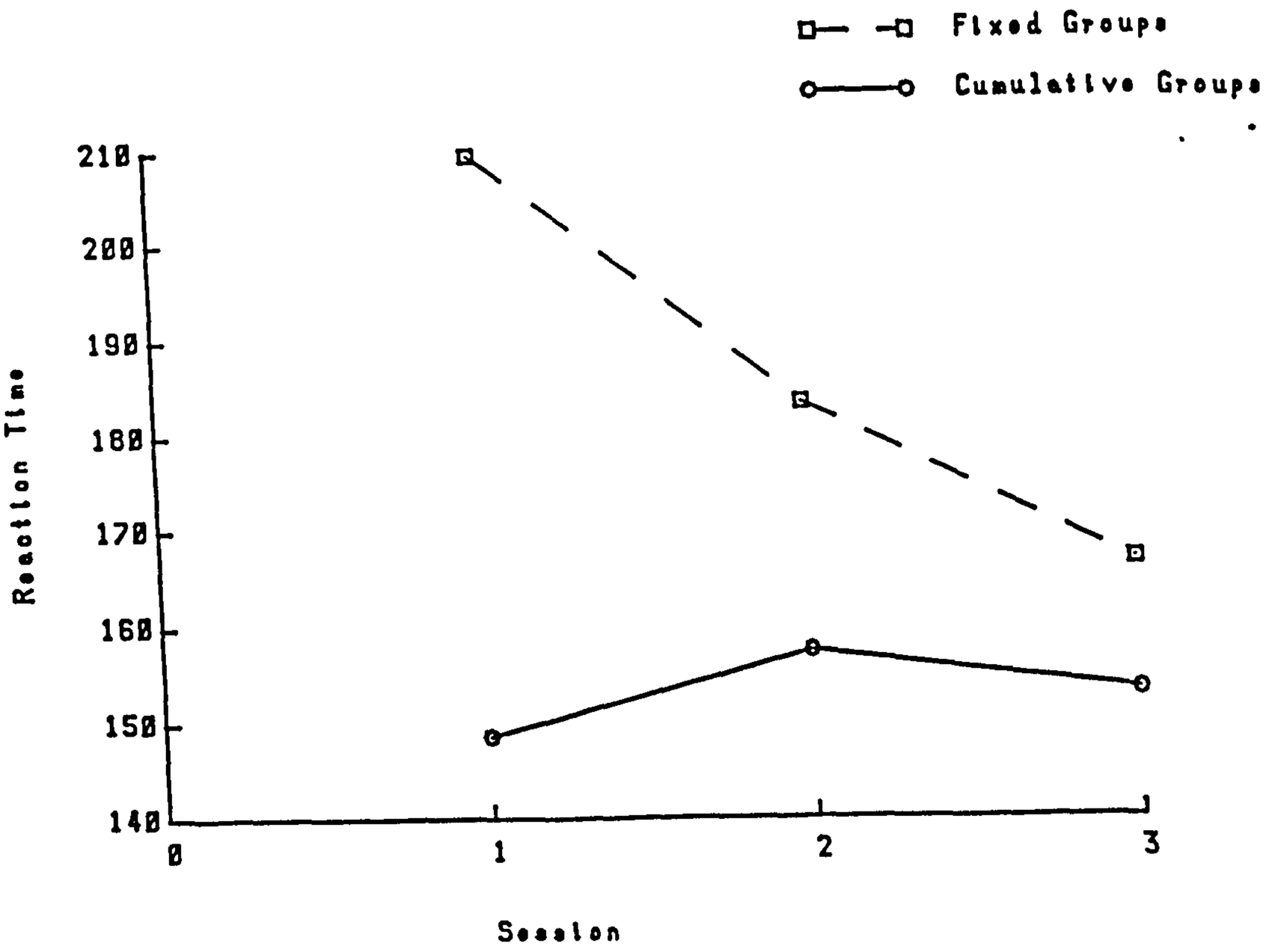


Figure 4.9 Mean RT of the Cumulative and Fixed groups during the first three training sessions

covariates that are described above, for the covariance analysis on the RT measure. The average accuracy of the Cumulative groups was significantly better than that of the Fixed groups [$F(1,35)=4.211$, $p<.05$]. No significant difference in accuracy was found between the Adaptive and the Random groups. Overall, the mean accuracy of all groups improved with practice [$F(2,72)=27.506$, $p<.01$].

The significant interaction between practice and the Cumulative vs Fixed factor [$F(2,72)=6.398$, $p<.01$] indicates that the effect of practice was not the same on the Cumulative and Fixed groups (figure 4.10). The Cumulative groups improved much less than the Fixed from the first session to the second, whereas the Fixed groups improved less than the Cumulative from the second session to the third. It appears that because the Cumulative groups started their training with fewer stimuli and proceeded with an increasing number of stimuli, they were more accurate than the Fixed groups which started their training with the full set of stimuli.

(3) Examination of the ASC functions

A third three-way analysis of covariance was carried out on the ASC measure which represents the time between correct responses and accounts, therefore, for both speed and accuracy. The mean ASC of the Cumulative groups during the first three training sessions was significantly better than that of the Fixed groups [$F(1,35)=17.092$, $p<.01$], which indicates that the Cumulative groups performed more efficiently than the Fixed groups. The mean ASCs of groups

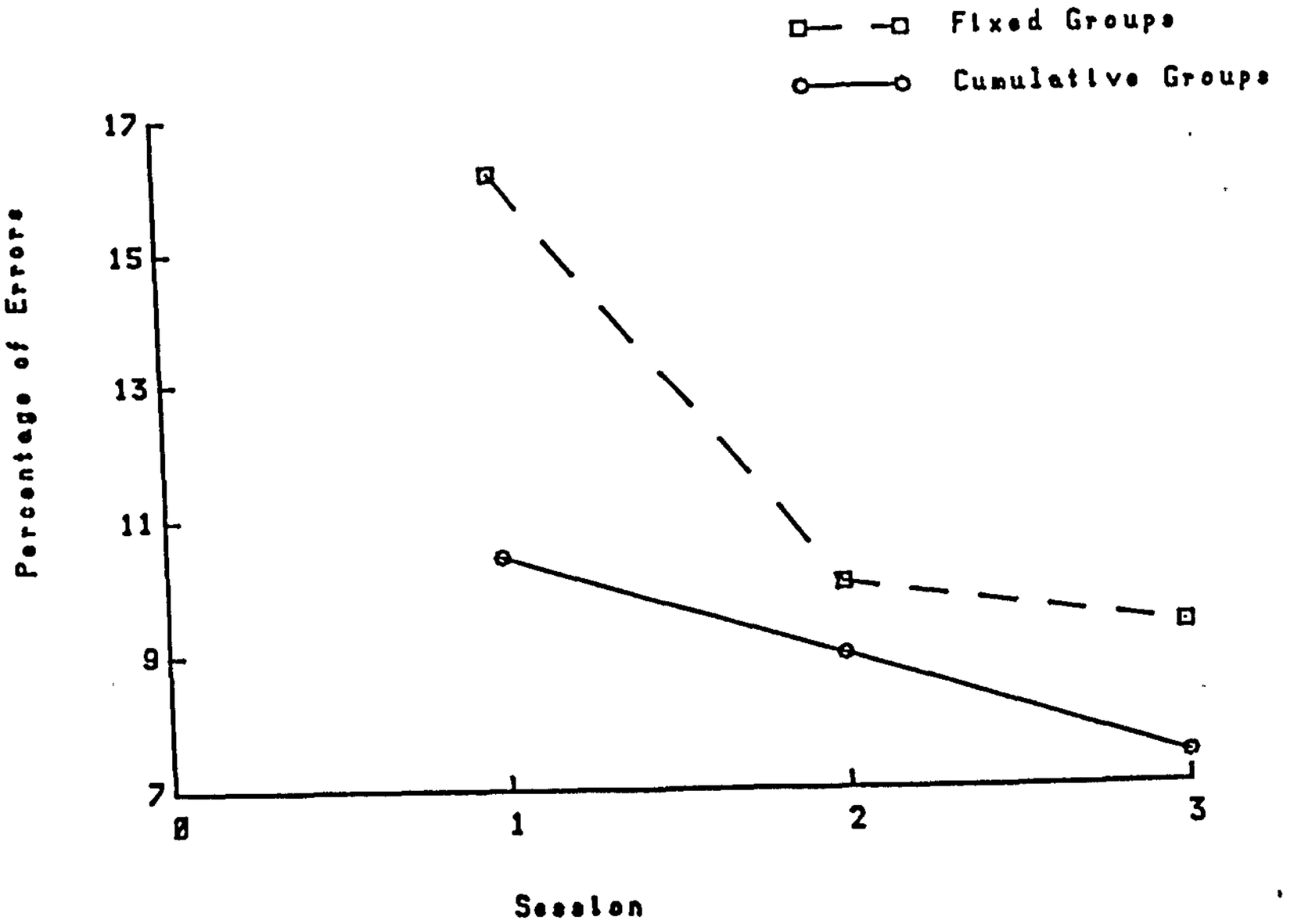


Figure 4.10 Mean percentage of errors of the Fixed and Cumulative groups during the first three training sessions

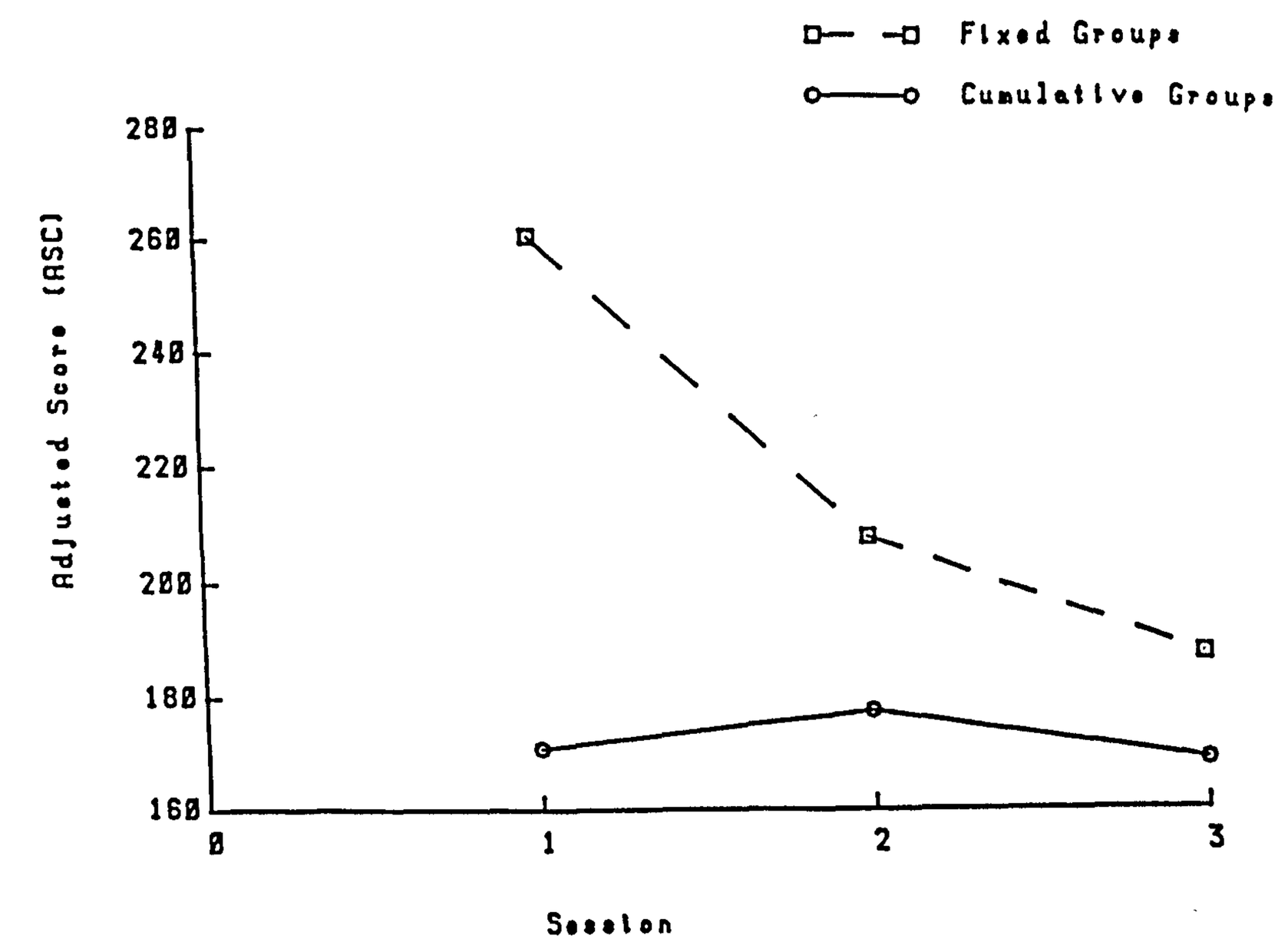


Figure 4.11 Mean adjusted score (ASC) of the Fixed and Cumulative groups during the first three training sessions

improved significantly with practice [$F(2,72)=65.854$, $p<.01$].

The significant interaction between practice and the Cumulative vs Fixed factor [$F(2,72)=66.628$, $p<.01$] indicates that the effect of practice was different for the Cumulative and Fixed groups (figure 4.11). From the first session to the second, the Fixed groups were improving a lot, while the Cumulative groups were becoming slightly worse; from the second session to the third, while both groups were improving, the rate of improvement of the Fixed groups was higher than that of the Cumulative. Finally a significant 3-way interaction [$F(2,72)=4.099$, $p<.05$] was found between all three factors (i.e. practice, Cumulative vs Fixed and Adaptive vs Random) which suggests that the effect of practice was not the same under all conditions. The Fixed groups improved more than the Cumulative groups and the degree of improvement depended on whether they were trained under the Adaptive or Random conditions (table 4.5).

	Training Group			
Session	FR	FA	CR	CA
1	185.71	145.04	233.17	152.24
2	168.34	151.68	198.34	163.08
3	157.05	147.85	175.96	158.57

Table 4.5 Mean ASCs for each training group (averaged across subjects) for each one of the first three training sessions (n=10)

The effect of "chord" difficulty on RT, ERR and ASC

Chord difficulty affected all four groups in a very similar way. The difference in the effect on performance of the five levels of stimulus difficulty was very clear (figures 4.12, 4.13, 4.14). The first level was, obviously, the easiest and performance in terms of all three measures was the best on this level of difficulty. For the second level of difficulty, performance was relatively worse, especially in terms of the RT and ASC measures. For the other three levels of difficulty, RT, ERR and ASC increased quite steeply. Obviously, the chord patterns that required the use of both hands and involved finger separation were much more difficult and responses to these patterns both with regard to speed and to accuracy were worst.

Three analyses of covariance were carried out on the three measures (RT, ERR and ASC) for the test session (e.g. the sixth session in which stimuli were presented in a Fixed Random mode). The factors used in the analyses were "chord" (that is, level of chord difficulty), Adaptive vs Random and Cumulative vs Fixed. The pretest ASCs were used as covariates. These analyses revealed significant differences in the difficulty level of the five levels of chord difficulty for all three measures (RT, ERR and ASC). For the RT measure, $F(4,144)=175.221$, $p<.01$; for ERR, $F(4,144)=30.842$, $p<.01$; for ASC, $F(4,144)=143.425$, $p<.01$.

No significant effect of any of the other two factors was found ($F<1$). This result was quite interesting, given the fact that not all subjects of the Cumulative groups had been

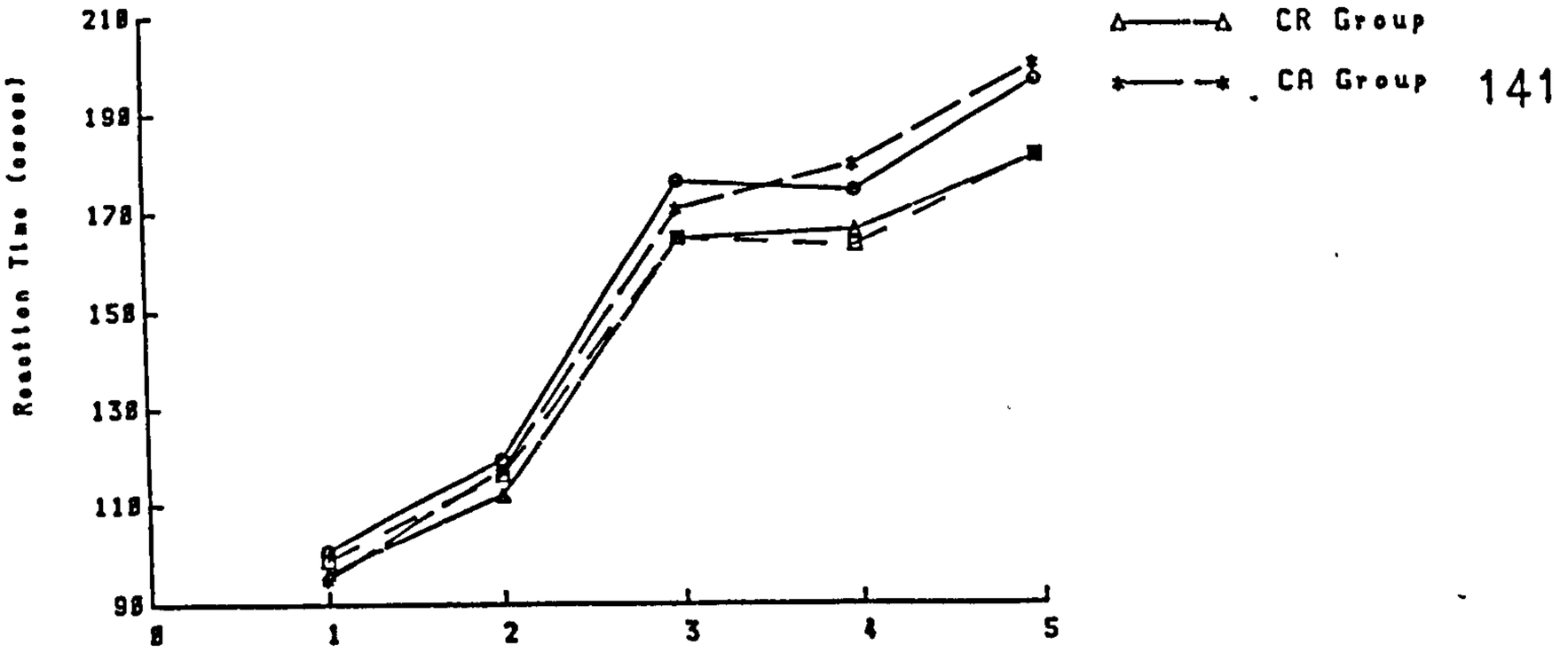


Figure 4.12 Mean RT of each training group as a function of the level of stimulus difficulty

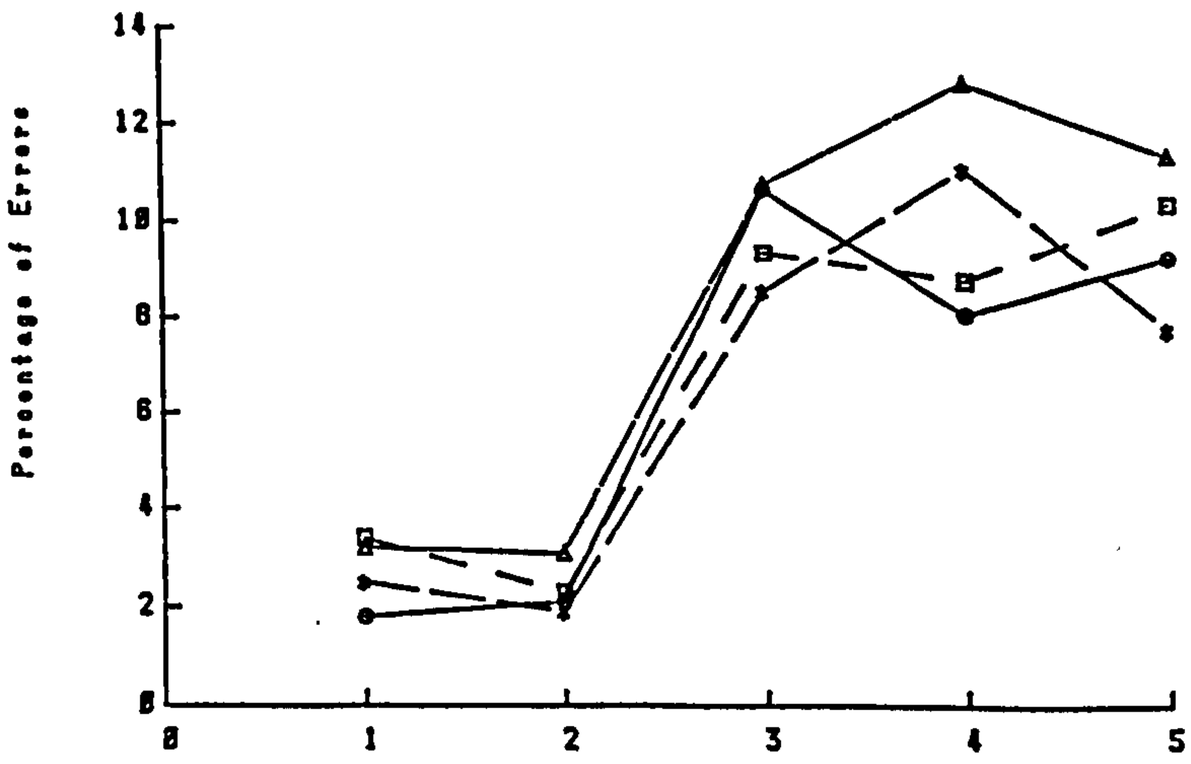


Figure 4.13 Mean percentage of errors of each training group as a function of the level of stimulus difficulty

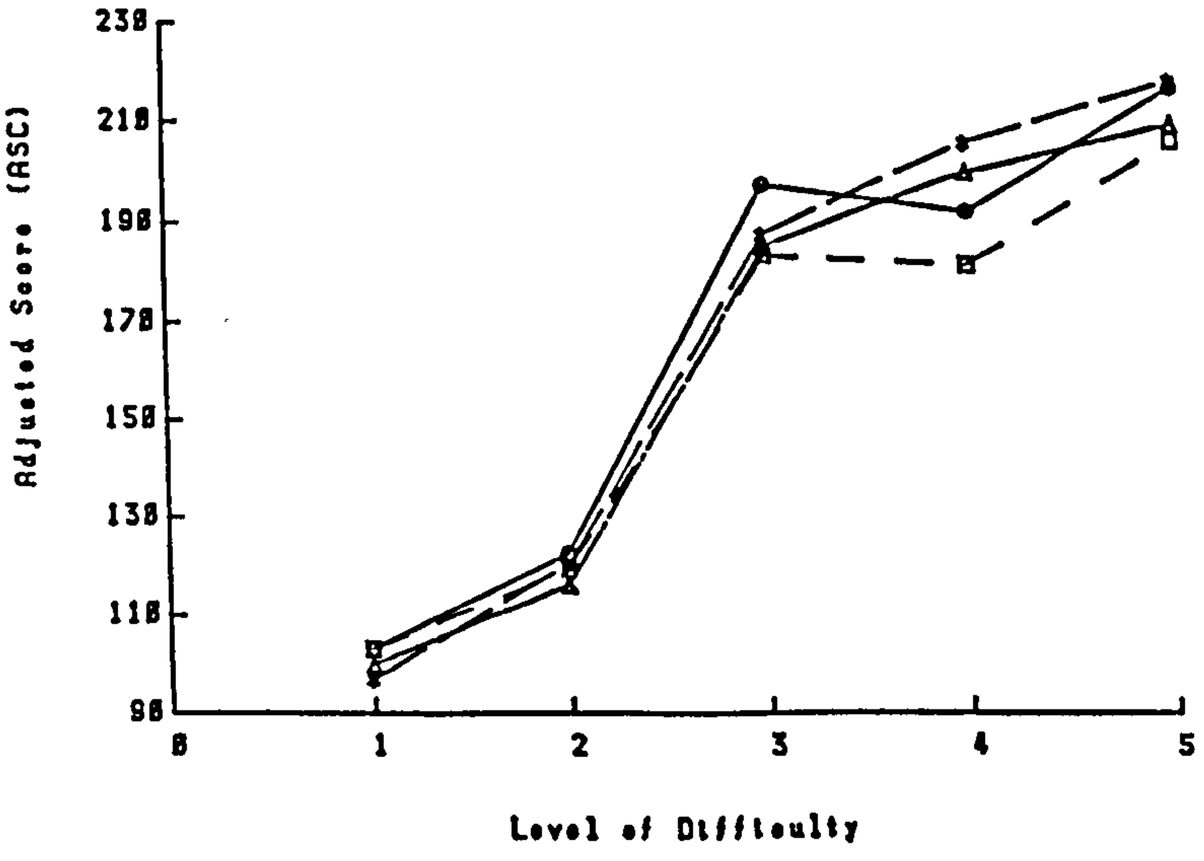


Figure 4.14 Mean adjusted score (ASC) of each training group as a function of the level of stimulus difficulty

trained on all five levels of chord difficulty. One would expect that the Fixed groups would perform better than the Cumulative groups in the fifth level of chord difficulty since not all subjects of the Cumulative groups had been trained on all five levels of chord difficulty. Since this sort of difference was not found from the analysis, it must be assumed that there was positive transfer of training from stimuli of one level of chord difficulty to the other.

4.4 Discussion and Conclusions

The four training methods which were examined in this experiment, were evaluated in terms of the following criteria:

(1) Speed of performance by the end of training;

(2) Accuracy of performance by the end of training;

(3) Total training time.

The results revealed that all groups were almost equal in terms of both speed and accuracy during the test session at the end of the training programme. It was also found that during the first training sessions, the Cumulative groups were faster but less accurate than the Adaptive groups. These differences however decreased during the last two training sessions and practically disappeared in the test session.

This difference between the Cumulative and Adaptive modes of training reflects the difference in the theoretical position from which the two methods have stemmed. In the Adaptive modes, which are based on Pask's system of adaptive training, more emphasis is placed - by the system - on accuracy of performance. This is also apparent by the fact

that the CA group did not progress as fast as the CR group and, overall, seven out of ten subjects subjects did not have any training on the full population of eighty stimuli as opposed to only two out of ten for the CR group (appendix 4).

For this stage of operators' training (i.e. keyboard familiarization) Kelley's adaptive system seems to be more effective. Given that the level of speed and level of accuracy at the test session were practically equal for all groups, the CR mode of training may be considered superior in that the CR group completed their training faster than the other groups. The results of the experiment provided no evidence to support Pask's view that subjects should be treated differentially in the beginning of their training. As has been discussed in beginning of the chapter, Pask argued that the first step of an adaptive system should be to assess the trainee's difficulties by a series of "probing experiments" and to use this information to generate the appropriate strategy which suits the trainee. This may be appropriate in cases where subjects may have prior knowledge of some aspects of the task to be taught before training begins. In the keyboard familiarization task, however, there is no reason to assume that subjects may have knowledge of some chord patterns which should be taken into account in the first stages of training. The Adaptive types of training would be more useful in the other stages of Post Office operators' training where knowledge of area codes, geography, etc. are important aspects of the task of processing letters.

In terms of total training time the Cumulative groups showed a rather clear superiority over the Fixed groups. This, also, has to be interpreted within the theoretical framework from which the training procedures stemmed. In the Cumulative modes, subjects were "encouraged" by the training system itself to perform fast, hence they completed the required number of training and test sessions faster than the subjects of the other groups. The fact that their accuracy level did not decrease as a result of the higher speed they achieved provides further evidence of the superiority of the Cumulative methods for this kind of task.

In conclusion, when all criteria are taken into account, the Cumulative methods appear to be more efficient training methods than the Fixed ones. For the keyboard familiarization stage which basically involves training of motor skills, the Cumulative Random method was shown to meet the training criteria more closely than the Cumulative Adaptive method.

PART B

Chapter 5

EXAMINATION OF SPEED-ACCURACY RELATIONSHIPS

The results of Experiment 4 raised questions about the relationship between speed and accuracy. Given that both speed and accuracy are important for the efficient operation of the 10-key chord keyboard, the issue was studied in more detail.

The problem of accuracy and its relation to speed in CRT tasks has been relatively neglected, although the topic of choice reaction time has been under study for over a century. This however is hardly surprising considering the framework within which most theoretical models have been developed. A number of criticisms against some of the main assumptions and positions adopted by various CRT models will be presented in an attempt to examine why no model can sufficiently explain the learning that occurs in a CRT task: The term "learning" is used to refer to the development of increasing speed and accuracy with practice in a choice response task.

5.1 Limitations of CRT Models

Many models emphasise the effects of variables such as number of alternative stimuli, S-R compatibility, stimulus discriminability, probability of certain stimuli being presented, stimulus repetition to account for changes in RT. For example, it has been found that response times are affected by variables such as the number of alternative stimuli (e.g. Hick 1952; Hyman, 1953; Morin and Forrin 1962; Welford, 1960, 1968); S-R compatibility (e.g. Conrad, 1962; Fitts and Deininger, 1954; Fitts and Seeger, 1953; Leonard,

1959); stimulus discriminability (Crossman, 1955; Sternberg, 1964b; Thurmond and Alluisi, 1963); probability of a certain stimulus being presented (e.g. Favreau, 1964; Fitts, Peterson and Wolpe, 1963; LaBerge and Tweedy, 1964; Lamb and Kaufman, 1965); stimulus repetition effects (e.g. Bertelson, 1961, 1963; Kirby, 1976; Kornblum, 1969, 1973; Sternberg, Kroll, and Wright, 1978). However, the variance in RT which these variables account for is relatively small when compared with the variance accounted for by practice (Bertelson, 1965; Rabbitt, 1968a). Moreover, there has been good evidence that the effect of these variables tends to decrease very substantially, and may even be entirely abolished, as a result of prolonged practice (Crossman, 1956; Mowbray and Rhoades, 1959). Yet many CRT models concentrate on these sources of variance to explain choice reaction time. Such models include those of Hick, 1952; Hyman, 1953; Falmagne, 1965; Kornblum, 1973; McGill, 1963.

Another major problem with most CRT models is that they describe a system which is "steady" in itself. The operator is generally considered as a passive part of the system and the only changes come about from the manipulation of the variables mentioned above. However, it is going to be argued that operators do affect the system by adjusting their own performance according to the demands and/or limitations of the system. Moreover, it should be added that changes in the system may also occur as a result of the particular task (e.g. adaptive training, where the presentation of stimuli is

dependent on the subject's performance).

Rabbitt (1981, p. 154) expresses the opinion that "... the trouble can be traced back to the uncritical, tacit adoption of the assumption of sequential, independent, successive processing stages made by F.C. Donders in 1868 to justify his decomposition of measured mean CRTs into additive components".

Donders (1868, in Koster, 1969) distinguished between three types of reactions (A, B, and C reactions), where: the A reaction involved a single stimulus which required a single response; the B reaction consisted of two stimuli and two responses and a 1:1 mapping between them; and the C reaction which involved two stimuli but only one response which was required for one stimulus but not for the other. Donders assumed that a subject requires a certain measurable time interval to identify any stimulus which is presented to him/her as part of the task he/she has to perform. He also supposed that until the identification of the stimulus is fully made the subject cannot begin the process of selecting a response. Donders used this basic assumption to develop the subtraction method. This assumption has been adopted by various researchers and has been the starting point of a number of CRT models.

Welford (1960, 1968, 1976) assumed that, in a CRT task, information processing involves three different central mechanisms in addition to the sensory and effector organs: a perceptual mechanism responsible for the categorization of the stimulus, a translation mechanism responsible for the

selection of the appropriate response and an effector mechanism responsible for the execution of the response. Other researchers (Morin and Forrin, 1963; Nickerson and Feehrer, 1964) working essentially on Donders' first assumption tried to isolate either response categorization or response selection by substituting many:1 (many stimuli and one response) and 1:many (one stimulus and many responses) mappings for the usual 1:1 mapping. The work of Hick (1952) and Hyman (1953) also advanced the subtraction logic. The equation:

$$CRT = a + bH$$

(where H refers to the bits of information transmitted) which was developed to measure CRT as a function of information transmitted, in actual terms promoted the idea that CRT is simple RT with other stages inserted (Smith, 1968).

Donders' subtraction logic was extended even further. Posner and Mitchell (1967) present an experimental paradigm for detecting different levels of processing which they call chronometric analysis of classification. They present a series of experiments in which the stimuli were pairs of letters to which subjects had to respond "same" or "different". They used three levels of instruction, physical identity (e.g. AA), name identity (e.g. Aa) and rule identity (e.g. both vowels). By using the subtractive method to analyse matches at different levels, they tried to find levels of processing which were dependent primarily upon the solely physical attributes of stimuli and levels which were dependent on more detailed

analyses like naming. They varied the level of instructions (given to subjects) before subjects made their classification, allowing this way the same stimulus-response combination (e.g. AB-different) to occur with instructions at quite different levels. They put emphasis not on the RT measures themselves but on their relevance for understanding the operations and mechanisms involved in perceptual matching, naming and classifying. They claim that "this technique can be used to bridge the gap between the very early stage of perception and the complex classifications underlying learned concepts" (pp. 406-7).

Donders had argued that stimulus categorization and response selection are additive. Christie and Luce (1956) and Sternberg (1966, 1969) conceptualized the stimulus categorization stage as composed of a number of tests or comparisons made in a serial order. Other researchers (e.g. Stanovich and Pachella, 1977; Taylor, 1976) argued the serial processing viewpoint but talked about stages overlapping or even being parallel. A further step concerning this controversy of viewpoints was taken by the mathematical theorems of Townsend (1972, 1974, 1976) who provided mathematical proof that all parallel processing models had an equivalent serial processing model when considered on the basis of RT. The reverse however is not always true, i.e. a serial processing model does not necessarily have an equivalent parallel processing model.

These "confusions" about processing stages do no more than give support to a basic and fundamental argument against

these models where the rigidity and simplicity of the basic assumptions prevents them from handling aspects of CRT data related to changes of perfomance within or between subjects. Although the idea of processing stages is intuitively attractive, it should not be taken any further than as a loose conceptualization of the CRT task situation in which after a stimulus is presented the subject chooses a response and executes it. There is no doubt that from the moment the stimulus is presented until the subject initiates a response, there is a lot of "activity" within the subject (whether described as cognitive, chemical, neural, etc.). However, it is simplistic to expect that because there is such "activity" going on, one can make a "logical" guess about which stages are involved, then measure the time taken by each one of these stages and conclude that the CRT is the total time taken by all these processing stages. This "logical" guess may be extremely sophisticated and based on extensive knowledge of the subject matter (although "such sophisticated knowledge is rarely available" Pachella, 1974, p. 47). However, there is no way to provide empirical proof of the nature or exact identity of these processing stages. An additional criticism to this sort of approach refers to the other two assumptions it makes. First it assumes that, through all levels of practice, the same processing stages exist. Second, it assumes these processing stages remain in the same order throughout all levels of practice.

Rabbitt (1981), too, has strongly opposed the kind of

models which he calls LISPs (Linear Independent Sequential Process). Within these models, through various techniques, mean RT is partitioned to give estimates for each one of a number of hypothetical stages like signal encoding, response selection and response execution. Regarding the assumptions made about the existence of processing stages, he claims that there is no rationale for deciding that particular errors are failures at one stage or another of this hypothetical chain of events. He argues that "if any one process in a linear sequence fails, input to all succeeding processing stages will be corrupt and identification of any unique point of failure in the chain will be uncertain" (1981, p. 156).

Therefore, it is argued that those models which base their theoretical arguments on the existence of specific processing stages have to be considered as inadequate to explain the learning process of a CRT task. Such models are, for example, those developed by Audley, 1960, 1973; Audley, Caudrey, Howell and Powell, 1975; Bertelson, 1963; Bower, 1959; Falmagne, 1965; LaBerge, 1962; Welford, 1968.

Another criticism against many of the existing CRT models is their failure to consider accuracy actively. In fact a basic condition for some of these models (e.g. Sternberg, 1969, 1975) is that errors are minimal and identical for all experimental conditions. This condition should be satisfied even if subjects in one condition must receive more training than in other conditions. In one of Sternberg's (1969) experiments by which he was testing his "additive-factor" method, he used payoffs that weighed accuracy more heavily

than speed specifically to keep errors to 1 or 2 per cent. In another experiment (Sternberg, 1969) he based his analyses only on correct responses. This approach, however, totally disregards any effects other than accuracy which might be affected by practice. Therefore a lot of information about the learning process is lost.

According to Rabbitt (1981), another limitation of all CRT models is that they do not consider trial-to-trial variations in RT either within or between subjects and/or tasks. That is, they consider neither how the performance of different subjects changes with practice nor how performance on easy as opposed to difficult stimuli changes with practice. Therefore, he says, these models describe a system which improves average performance over practice but does not discuss how or to what extent the system alters the maximum capabilities over time.

Moreover, Rabbitt (1981) argues that LISP models consider each transaction of signal identification and response production as an independent process. They do not take into account that subjects may have an overall supervisory control so as to anticipate events and maximize sensitivity of signal detection and eventually facilitate response production.

The final limitation of LISPs that Rabbitt discusses refers to "failures to describe control of information encoding and response production" (Rabbitt, 1981, p. 158). He argues that these models fail to consider that subjects may already have relevant information in long-term memory which

allows them to recognize stimuli easily and provides them with the appropriate motor programs for the emission of the response.

5.2 The Study of Speed-Accuracy Tradeoff

Wickelgren (1977) argues very strongly for the need to study speed-accuracy tradeoff relationships. He accepts that traditional RT experiments are acceptable if one wants to assess the relative difficulty of two tasks. However, he advocates the introduction of speed-accuracy tradeoff methodology to replace the traditional reaction time research in the study of information processing. He says that "...the speed-accuracy tradeoff method is so superior to the traditional reaction time method that many psychologists interested in studying the dynamics of information processing in perceptual, memory, performance, psycholinguistic, and other cognitive tasks, ought, in many instances, to do speed-accuracy tradeoff studies instead of reaction time studies" (p. 68).

Wickelgren discusses six methods of obtaining a speed-accuracy tradeoff (SAT) function.

(1) Instructions: Subjects must set a speed-accuracy criterion as a result of instructions given to them which emphasise either speed or accuracy. This approach has been used by various psychologists (e.g. Hale, 1969; Hick, 1952; Howell and Kreidler, 1963; 1964) but not for obtaining a speed-accuracy tradeoff function. Wickelgren however claims that this method can be used to obtain an SAT function.

(2) Payoffs: In this method subjects adjust their speed-accuracy criterion on the basis of explicit payoffs they receive for every correct response or for every error they make. For example, Swensson (1972) and Swensson and Edwards (1971) have used the method where subjects received a bonus [D-k(RT)] for every correct response and a charge [-k(RT)] for every error. By varying D and k in different sets of trials, subjects were induced to perform at different speeds and make different percentages of errors.

(3) Deadlines: This method involves the use of a deadline within which subjects must respond to a stimulus. In other words this method imposes an upper limit on reaction time. Pachella and his associates (Pachella, Fisher and Karsh, 1968; Pachella and Fisher, 1969; Pachella and Fisher, 1972) have used a pure deadline technique to study speed-accuracy tradeoff functions. They instructed subjects to respond faster than a deadline of 0.7 secs after the onset of a stimulus in one set of trials, or faster than 0.4 secs after the onset of a stimulus in another set of trials and so on. They gave subjects feedback about response times and subjects soon learned to respond at a variety of times usually below the deadline. Links (1971) used a mixed deadline technique which involved three deadlines. He instructed subjects to respond to each stimulus within the deadline that was given to them prior to the onset of the stimulus. Another deadline technique which has been used in a number of experiments (e.g. Pachella and Pew, 1968; Pew 1969; Yellott, 1971) combines the deadline and the payoff methods using a 2x2 payoff matrix. The matrix is

defined by correct versus error responses and by faster versus slower than the deadline.

(4) Time Bands (Time Windows): In this method subjects have to respond within a time band (or time window) following the onset of a stimulus. They must respond faster than a certain upper limit and slower than a lower limit. This method does not seem to have been used specifically to obtain SAT functions but Snodgrass, Luce and Galanter (1967) report that subjects can successfully time their responses to fall within a specified time band. A limitation of this method and, to a lesser extent, of the deadline method is that subjects who could possibly perform faster than the lower limit of a time band (or much faster than the deadline in the deadline method) do not do so.

(5) Response Signals: In this method subjects have to respond to a stimulus not as soon as the stimulus is presented but after they receive an additional stimulus which commands reaction. Schouten and Bekker (1967) employed this method, which they called "forced reaction time" method, to study the relationship between speed and accuracy. They used visual stimuli to which subjects had to respond and a series of three additional auditory signals (tone pips) which gave subjects the command to respond to the visual stimulus. The three tone pips were 20 msecs long each with intervals of 75 msecs between them. Subjects were instructed to respond to the visual stimulus in coincidence with the third pip. Schouten and Bekker (1967) report that subjects were able to adjust

their responses to coincide with the third pip and they did it with a high degree of accuracy. This method allowed the experimenters to vary the interval between the onset of the stimulus and the first auditory signal. This manipulation rendered it possible for Schouten and Bekker to achieve responses to a variety of different reaction times and study the relationship between speed and accuracy.

(6) Partitioning Reaction Time: In contrast to all of the above methods where the experimenter manipulates reaction time through one means or another, in this method reaction time is not manipulated. Schouten and Bekker (1967) used this method as well and called it the "free reaction time" method. After reaction times have been collected, they are partitioned in intervals, e.g. 200 and 220 msecs, 220 and 240 msecs, 240 and 260 msecs, etc. Accuracy in each interval is then studied. This method has also been used by Rabbitt and Vyas (1970) and in a series of studies by Lappin and his associates (Lappin and Disch 1972a, 1972b, 1973; Harm and Lappin, 1973).

Wickelgren (1977) admits that each one of these methods has various limitations, one of which is that they require the collection of at least five times as much data as a standard reaction time experiment. However, he claims that they can obtain a speed-accuracy tradeoff function and he suggests that more research of this kind is carried out.

Schmidt and Scheirer (1977) agree with the potential contribution of the examination of speed-accuracy relationships to the study of CRT tasks. They even extend Wickelgren's argument by suggesting that "the conditions that

lead to speed-accuracy tradeoffs are of interest in themselves" (p. 324). However, they disagree with Wickelgren's strong claims against RT research. They see the study of speed-accuracy tradeoff functions as complementing rather than replacing reaction time experiments.

Kantowitz (1978) agreed with Wickelgren that SAT methodology should be used more extensively to evaluate cognitive functions. However, he criticised Wickelegren for neglecting the issue of choosing the appropriate accuracy measure. Although Wickelgren (1977) had listed a few such measures (e.g. probability-correct, odds, probability-correct/probability-error, d', information transmitted), he had not discussed them in any detail and did not offer any guidelines about how they should be used. But in 1978, Wickelgren did accept the above criticism and agreed with Kantowitz (1978) that the choice of a specific accuracy measure is of great importance.

Rabbitt (1981) also stresses the importance of speed-error tradeoff functions (SETOFs) in CRT tasks. He claims that subjects respond increasingly faster until an error occurs. When an error occurs then the reaction times to subsequent trials are typically slow. Then, subjects speed up again until they make another error. He suggests that subjects trade speed for accuracy or vice versa and proposes a "tracking" model for maintaining optimum levels of speed and accuracy.

Rabbitt (1981) suggests that errors are studied in more detail and recommends that RTs are studied as a function of

accuracy on the previous response. By such an analysis of the data, he claims, the learning process is studied in more depth and more information becomes available about what and how subjects learn a particular CRT task.

Although Rabbitt's latter claim may not be necessarily fully accepted, it is believed that this kind of analysis can make the tradeoff function more explicit and, as such, it is important. Following his suggestions, all the data from Experiment 4 were analysed further. The emphasis of this new analysis was not on the four training methods but on the speed-accuracy tradeoff function. It was decided to investigate how errors on a given trial related to accuracy on the previous trial and how speed changed as a function of accuracy both of the current as well as the previous response. Also it was decided to examine how speed and accuracy on any given trial were affected by the level of practice and stimulus difficulty.

Chapter 6

SPEED-ACCURACY RELATIONSHIPS: A NEW ANALYSIS OF THE DATA OF EXPERIMENT 4

6.1 METHOD OF ANALYSIS

While differences between modes of training were not of specific concern in this analysis, it was recognised that such differences could have effects on speed-accuracy tradeoff functions. Hence, detailed results are presented separately for each mode, followed by an examination of their common and their distinctive functions.

The same analyses were carried out for all four modes of training.

First, two repeated measures analyses of variance were carried out for each one of the four groups of subjects. In the first ANOVA, the independent variables (factors) were the six practice sessions and the five levels of stimulus difficulty whereas the dependent variable was the mean RT. In the second ANOVA the independent variables were the same as in the first one and the dependent variable was the percentage of errors.

Second, for each mode, a 19,200 x 9 array was produced, where the following nine variables were examined in relation to the total 19,200 trials of practice.

(A) Three nested measures of practice level were included in the array, namely: SESSION (n=320 trials), BLOCK (n=80 trials) and TRIAL.

(B) The difficulty level of the current trial and difficulty

level of the preceding trial were also included. There were five levels of stimulus difficulty: (1) one-handed stimuli, (2) two-handed stimuli requiring the depression of adjacent keys only, (3) two-handed stimuli requiring the depression of non-adjacent keys by the right hand, (4) two-handed stimuli requiring the depression of non-adjacent keys by the left hand, and (5) two-handed stimuli requiring the depression of non-adjacent keys by both hands.

(C) The proportion of accurate current responses and the proportion of accurate preceding responses were two additional variables which were included in the array.

(D) Finally the residual value of the reaction time of the current response as well as the residual value of the reaction time of the preceding response were also included. Residual values rather than the raw RTs were used so as to remove all between subjects' variance. The residual value of the RT on trial i performed by subject j was the difference between the actual RT for trial i (RTij) and the mean RT for subject j (RTj), i.e. RESij = RTij - RTj

Having created such a matrix for each mode, three analyses of variance were carried out, to examine the effects of the various independent variables as well as the effects of the interactions between these variables on the residual value of RT of the current response (dependent variable). In the first analysis, the independent variables were accuracy of current response, difficulty level of the current response and practice session. In the second analysis of variance, the independent variables were the accuracy of the preceding

response, the accuracy of the current response and the difficulty level of the current response. Finally in the third analysis the independent variables were accuracy of the preceding response, accuracy of the current response and practice session.

Following these ANOVAs a stepwise multiple regression analysis was carried out. The dependent variable was the residual of the current response and the independent variables (predictor variables) were: the three measures of practice (session, block, trial); the difficulty level of the current trial; the difficulty level of the preceding trial; the accuracy of the preceding response; and, the residual of the preceding response.

The effect of accuracy of the preceding response on the accuracy of the current response was examined by various crosstabulations. First, a crosstabulation of the variables accuracy of the current response and accuracy of the preceding response was made. Second, a set of five crosstabulations was carried out examining the relationship between accuracy of the current response and accuracy of the preceding response but taking into account the different levels of difficulty. Third, another set of crosstabulations was carried out with the same two variables but accounting for the different levels of practice.

Finally, another stepwise multiple regession analysis was carried out with the dependent variable being the accuracy of the current response and the independent variables being:

session, block and trial; difficulty level of the current trial; difficulty level of the preceding trial; accuracy of the preceding trial; and, residual of the preceding trial.

Before the results for each training mode are presented, a reminder of the differences between the four training modes is presented.

(1) FR (Fixed-Random): In this condition each stimulus was selected randomly from the full set of eighty stimuli.

(2) FA (Fixed-Adaptive): In this condition, stimuli with which subjects had difficulties were presented more often than others. In other words, in this training mode, the selection of individual stimuli was not independent.

(3) CR (Cumulative-Random): In this condition, there was a "current set" of stimuli from which stimuli were presented. The number of items of the "current set" increased on the basis of subjects' accuracy and speed. In other words, there was control over the set of stimuli from which stimuli were presented to subjects but individual items were randomly selected from that set.

(4) CA (Cumulative-Adaptive): In this condition, there was control over both the size of the "current set" from which stimuli were presented as well as over the individual items which were selected for presentation to subjects.

6.2 Results of the Fixed-Random Mode of Training

Speed Changes with Practice, Stimulus Difficulty and Accuracy

The effects of practice, as measured by changes between "sessions", and "difficulty level" on RT were examined by a repeated measures analysis of variance using subjects' mean RTs. It was found that in general RT decreased significantly with "session" [$F(5,45)=25.33$, $p<.01$]. In the fourth session RT increased but this may be accounted for by the fact that subjects had their first three training sessions in the morning and the next three in the afternoon. This relatively long break between the morning and afternoon sessions apparently caused a decrease in speed during the fourth session, however subjects' speed increased quite substantially after the end of the fourth session. RT increased significantly with the level of difficulty [$F(4,36)=67.46$, $p<.01$]. The biggest increase in RT because of difficulty level was between the two least difficult and the three most difficult levels (all the latter involved the depression of non-adjacent keys). The significant interaction between "session" and "difficulty level" [$F(20,180)=7.14$, $p<.01$)] showed that subjects' RT to difficult stimuli decreased more than their RT to easy stimuli as a function of practice (figure 6.1).

Another analysis of variance was carried out to examine how speed was affected by the level of practice ("session"), level of stimulus difficulty and accuracy. The data used for this analysis were the mean residual values of RT for each

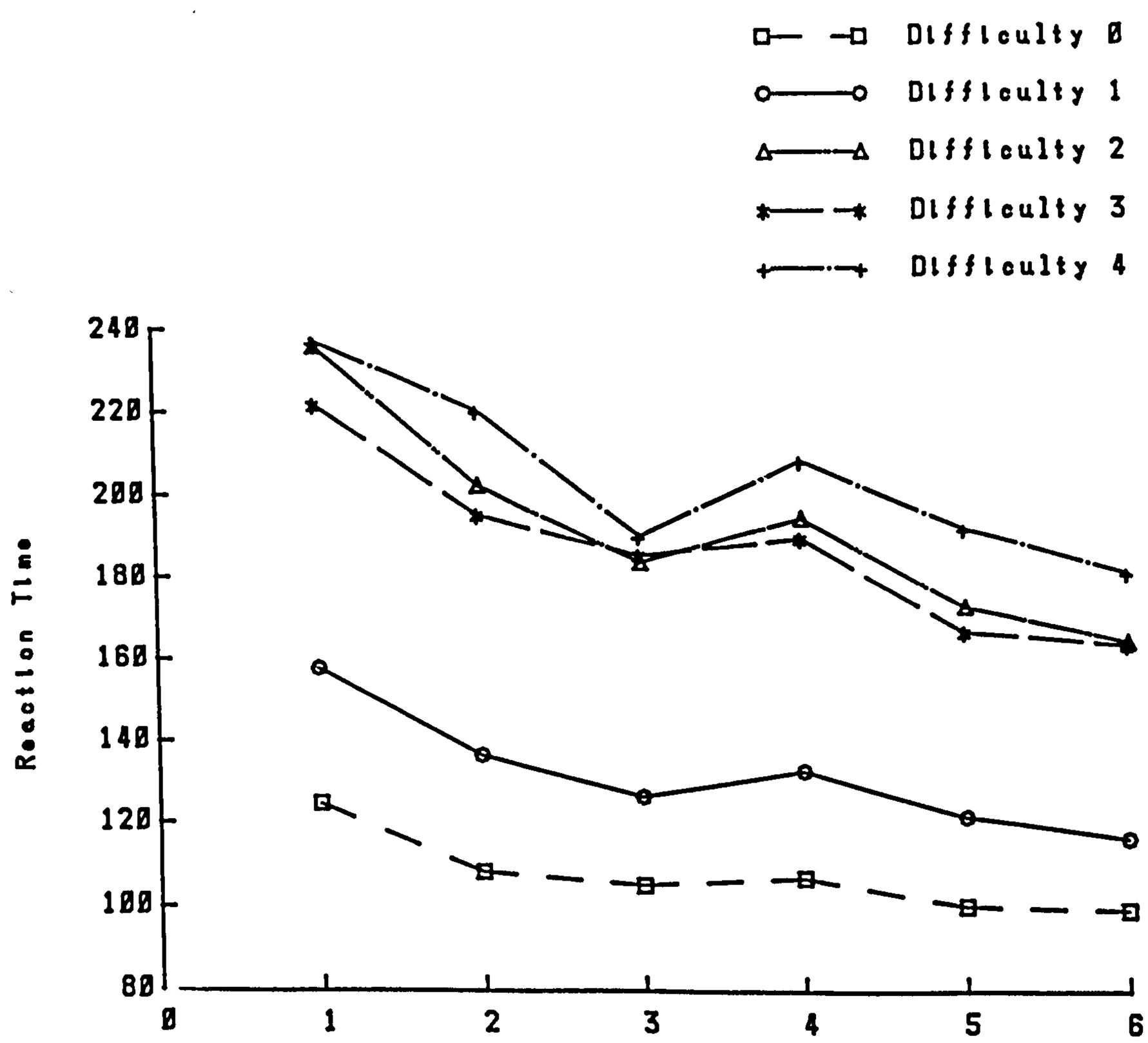


Figure 6.1 Mean RT of the current response as a function of the level of stimulus difficulty and the level of practice (FR group)

trial (i.e. all between subjects variance had been removed).

As expected from the repeated measures analysis, it was found that, overall, speed increased significantly with practice [$F(5,19140)=375.49$, $p<.0001$]. Similarly to the previous analysis speed did decrease slightly between the third and fourth sessions but it improved afterwards. Speed was affected significantly by the level of stimulus difficulty [$F(4,19140)=2587.91$, $p<.0001$] with subjects responding much faster to "easy" stimuli than to more "difficult" ones. The significant interaction between "session" and difficulty level of the current stimulus [$F(20,19140)=10.03$, $p<.0001$] implied that the effect of practice on speed was bigger for difficult stimuli than it was for easy stimuli. In other words, as already found from the repeated-measures analysis, speed improved more with practice for "difficult" stimuli than it did for "easy" stimuli.

The inclusion of accuracy as a variable in this analysis revealed an interesting result: irrespective of the level of practice, error responses were significantly slower than correct responses [$F(1,19140)=29.86$, $p<.0001$]. The difference between the residuals of error and correct responses varied significantly as a function of the level of difficulty, as expressed by the significant interaction between "accuracy" and "level of stimulus difficulty" [$F(4,19140)=3.24$, $p<.01$]. The effect of this interaction however was not systematic (figure 6.2).

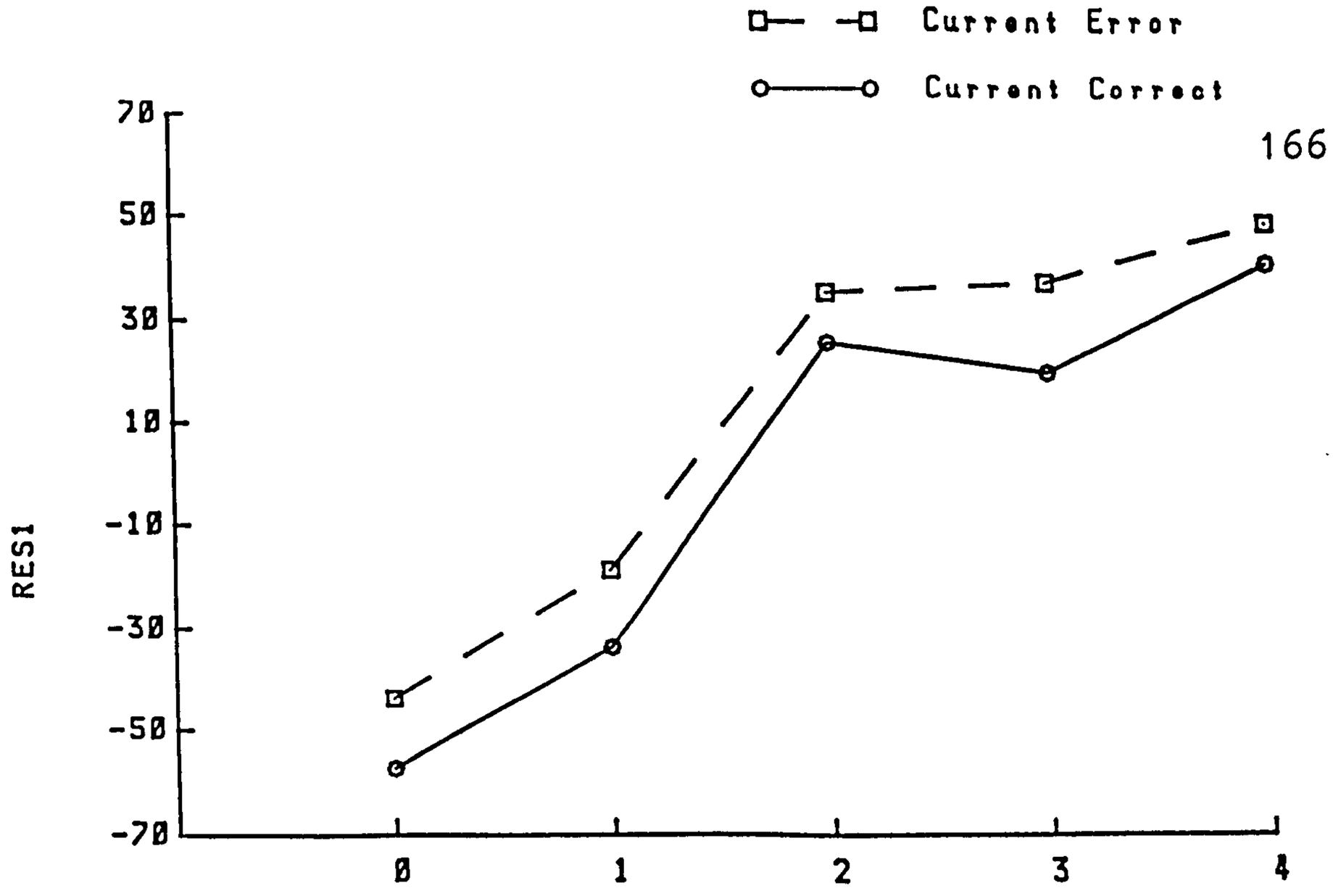


Figure 6.2 Mean residual RT of the current response as a function of the level of stimulus difficulty and the accuracy of the current response

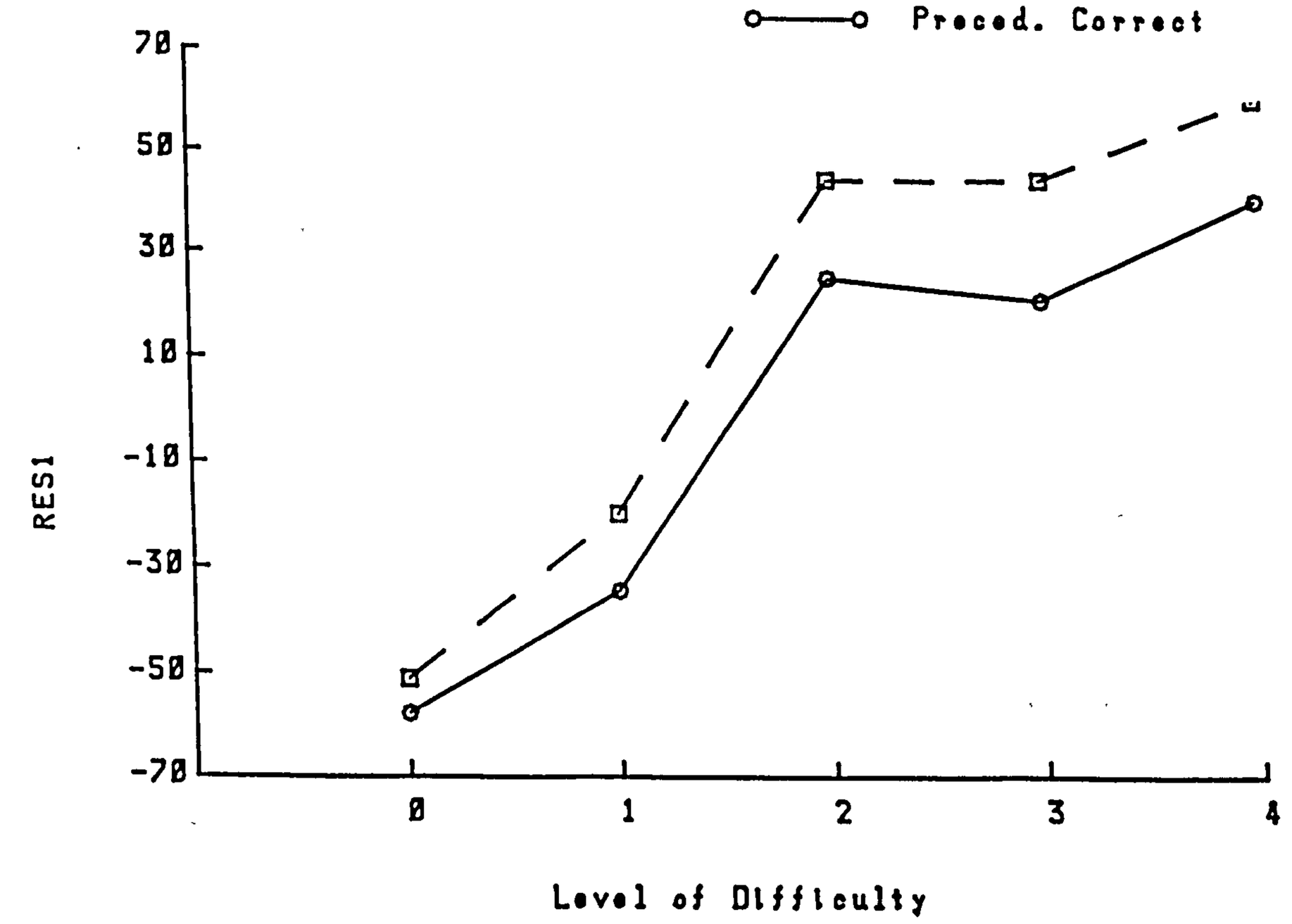


Figure 6.3 Mean residual RT of the current response as a function of the current stimulus difficulty and the accuracy of the preceding response

		DIFFICULTY				
	ACC1	0	1	2	3	4
SESSION 1						
	0	-17.39 (37)	2.95 (71)	74.51 (156)	71.62 (145)	69.60 (139)
	1	-41.31 (599)	-7.90 (656)	69.41 (441)	51.95 (450)	72.65 (506)
SESSION 2						
	0	-31.15 (17)	-23.89 (34)	34.72 (111)	30.75 (109)	76.45 (91)
	1	-57.22 (536)	-28.57 (714)	37.14 (517)	30.27 (537)	51.87 (534)
SESSION 3						
	0	-50.17 (28)	-24.41 (42)	22.12 (97)	25.30 (74)	32.95 (94)
	1	-59.96 (585)	-39.37 (659)	17.89 (515)	19.27 (490)	34.95 (616)
SESSION 4						
	0	-63.45 (28)	-20.13 (48)	30.35 (108)	40.47 (108)	44.12 (109)
	1	-57.35 (601)	-32.37 (658)	29.52 (523)	21.23 (478)	43.95 (539)
SESSION 5						
	0	-50.01 (21)	-43.49 (26)	4.94 (63)	.80 (63)	27.05 (69)
	1	-64.97 (526)	-42.61 (708)	8.48 (548)	1.97 (574)	27.50 (602)
SESSION 6						
	0	-59.44 (22)	-42.50 (17)	-3.62 (58)	6.26 (50)	14.95 (72)
	1	-64.85 (593)	-48.43 (678)	.44 (572)	-1.34 (513)	16.80 (625)

Table 6.1 Mean residual RT of the current response as function of accuracy, level of difficulty and level of practice of the current response

The effect of difficulty on the speed of correct and error responses varied significantly as a result of practice [$F(20,19140)=1.79$, $p<.05$]. This effect also did not seem systematic. The three-way interaction of the three variables showed that the effect of the interaction of difficulty and speed of correct and error responses was more pronounced at the early stages of practice (table 6.1).

The Effect of Accuracy of the Preceding Response on the Speed of the Current Response

It was found from the above analysis of variance that error responses were slower than correct responses and that speed decreased as the difficulty level increased. Further analysis was carried out to examine, for each trial, the effect of accuracy of the preceding response on the speed of the current one. Furthermore, the combined effects, on speed, of difficulty level and accuracy of both preceding and current trials were examined by another analysis of variance. This analysis looked into the effects of accuracy of the preceding response, accuracy of the current response and the difficulty level of the current stimulus on the residual RT values for the current stimulus.

When the preceding response was an error, the speed of the current response was significantly reduced [$F(1,18940)=171.68$, $p<.0001$]. In other words, the occurence of an error led to a decrease in speed on the next trial. Speed also decreased when the difficulty level increased (as had already been found from the previous analysis of variance).

However, the effect of the difficulty level on speed was more pronounced when the preceding response was an error. In other words, the combination of an error on the preceding response and a "difficult" current stimulus had the greatest effect on decreasing the speed of the current response. This was revealed by the significant interaction between "accuracy of the preceding response" and "difficulty level of the current trial" [$F(4,18940)=5.62$, $p<.0001$] (figure 6.3).

No significant interaction was found between "accuracy of the preceding response" and "accuracy of the current response". This meant that the effect of accuracy of the current response on speed was independent of the accuracy on the preceding response. Although the interaction of these two accuracy measures was not significant, together they did interact significantly with difficulty level (as shown from the three-way interaction between the three variables [$F(4,18940)=4.41$, $p<.005$]).

The effect of the three-way interaction did not seem to be systematic (figure 6.4). This however was, most likely, due to the large standard error of the mean residuals of those responses in the category "current errors preceded by errors" (table 6.2). The number of error responses preceded by errors was almost six times smaller than the number of errors preceded by correct responses and the number of correct responses preceded by errors and just over fifty times smaller than the number of correct responses preceded by correct responses.

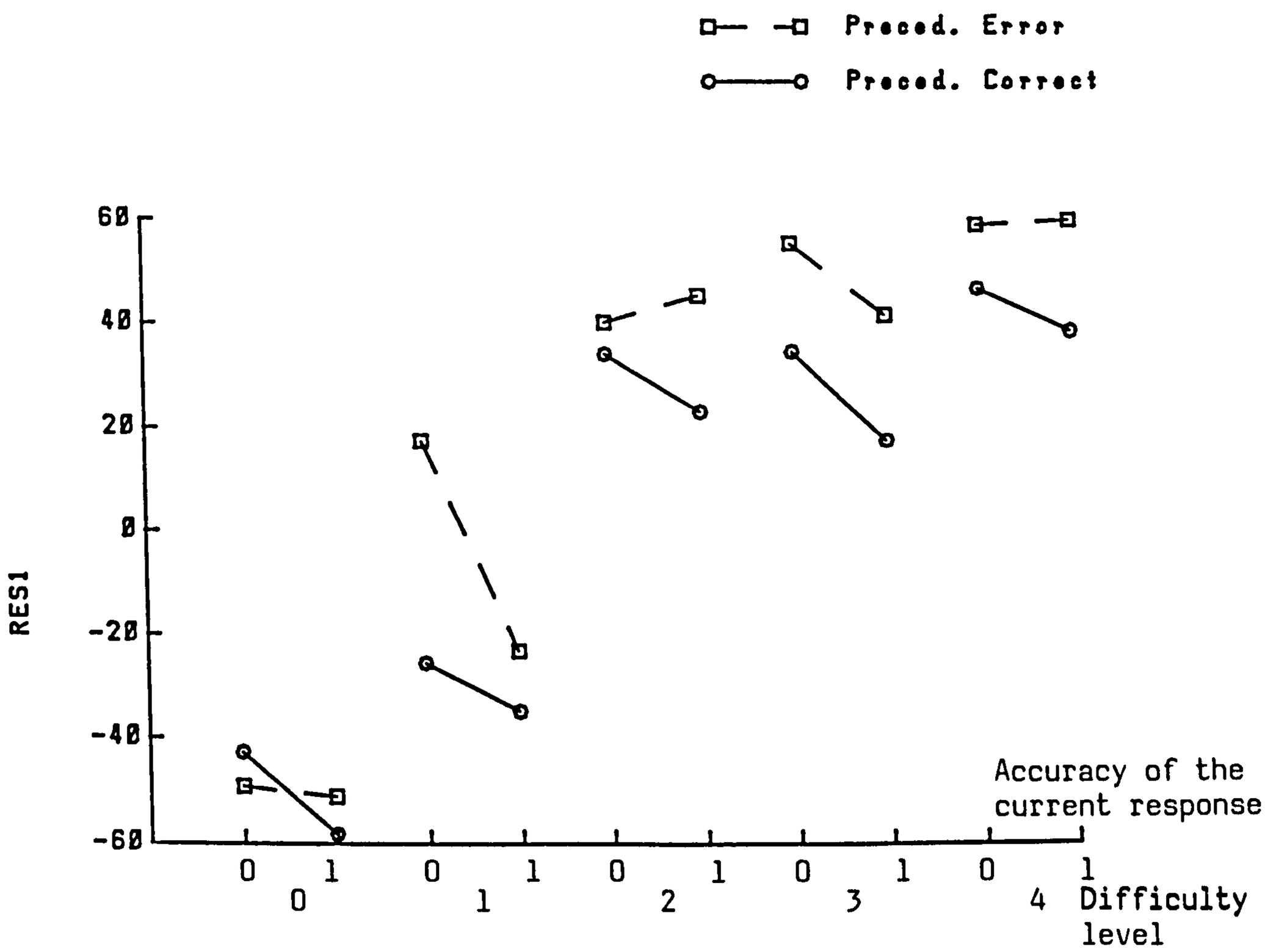


Figure 6.4 Mean residual RT of the current response as a function of the current stimulus difficulty, the accuracy of the preceding response and the accuracy of the current response

			PRECEDING RESPONSE								
			ERROR				CORRECT				
		DIFF1	MEAN	SD	N	SE	DIFF1	MEAN	SD	N	SE
C	ERROR	0	-49.22	73.51	28	13.89	0	-42.67	53.44	125	4.78
U		1	17.21	75.84	39	12.14	1	-25.72	52.01	199	3.69
R		2	39.68	69.78	93	7.24	2	33.72	79.50	489	3.59
R		3	54.81	83.38	59	10.85	3	34.14	66.42	487	3.01
E		4	58.23	85.45	82	9.44	4	45.78	74.80	483	3.40
N		MEAN	36.52	83.40		4.81		25.11	75.60		1.79
T		TOTAL			301					1,783	
R	CORRECT	0	-51.14	45.48	378	2.34	0	-58.25	39.72	3,036	0.72
E		1	-23.32	50.07	440	2.39	1	-35.09	42.07	3,556	0.71
S		2	44.78	69.99	341	3.79	2	22.76	55.03	2,730	1.05
P		3	40.90	59.54	283	3.54	3	17.10	49.16	2,732	0.94
O		4	58.88	62.10	337	3.38	4	37.60	59.55	3,043	1.08
N		MEAN	9.61	72.05		1.71		-5.19	61.73		0.50
S		TOTAL			1,779					15,097	
E											

Table 6.2 Residual RTs for current response as a function of the stimulus difficulty (DIFF1), accuracy of the response and accuracy of the previous response

The decrease in the speed of the current response as a result of an error in the preceding response was the same irrespective of the level of practice. This can be concluded from the non-significant interaction between accuracy on the preceding response and level of practice ($p>.2$).

Factors Predicting Speed

Having found a number of variables which were related to the speed of the current response, a stepwise multiple regression analysis was carried out to see which of those factors best explained the variance in speed. The dependent variable was the residual value of the reaction time of the current response and the predictor variables were: three

nested measures of practice level ["SESSION" (N=320), "BLOCK" (N=80) and "TRIAL"], difficulty level of the current trial (DIFF1), difficulty level of the preceding trial (DIFF0), speed (residual RT) of the preceding response (RES0) and accuracy of the preceding response (ACC0). The accuracy of the current response was excluded from the regression analysis on purpose, despite the high correlation between it and the dependent variable, as both speed and accuracy of the current response may only be known after the response is made (while all the other variables are determined prior to each response). Hence it would be inappropriate to use one dependent variable as a predictor of another.

It was found that 36 per cent of the variance of speed was explained by the difficulty level of the current trial, the practice "session" and the speed of the preceding response. Overall a multiple correlation of R=.60 [$F(3,18956)=3507.03$, $p<.0001$] was obtained, with the prediction equation being:

Predicted Residual RT (csecs)

= 25.27(DIFF1)-7.79(SESSION)+.12(RES0)-23.68

(mean Residual RT = -0.29, n=18,960)

Although all of the other variables did enter the equation at statistically significant levels, their contribution to the total variance was less than one per cent.

Accuracy Changes with Practice and Difficulty Level

A repeated measures analysis of variance was carried out on subjects' mean proportion of correct responses to examine

how accuracy varied as a function of practice "sessions" and "level of stimulus difficulty". The pattern of accuracy variations due to practice and stimulus difficulty was found to be similar to that for speed. More specifically, it was found that subjects' proportion of errors decreased significantly with "session" [$F(5,45)=16.49$, $p<.001$] and increased with increased stimulus difficulty [$F(4,36)=16.55$, $p<.001$]. The significant interaction between the variables "session" and "level of stimulus difficulty" [$F(20,180)=2.98$, $p<.01$] meant that the proportion of errors decreased more as a function of practice for "difficult" stimuli than for "easy" stimuli (figure 6.5).

The Effect of Accuracy of the Preceding Response on the Accuracy of the Current Response

Having found that the accuracy of the preceding response affected quite substantially the speed of the current response, the next question to be examined was whether the accuracy of the preceding response affected the accuracy of the current response. First, all responses were crosstabulated in a two-by-two matrix, the variables being accuracy of the current response and accuracy of the preceding response. This produced four categories of responses: errors preceded by errors, errors preceded by correct responses, correct responses preceded by errors and correct responses preceded by correct responses (table 6.3). (The values in each cell are: the actual number of responses, the number of responses expected if the variables were independent and the percentage

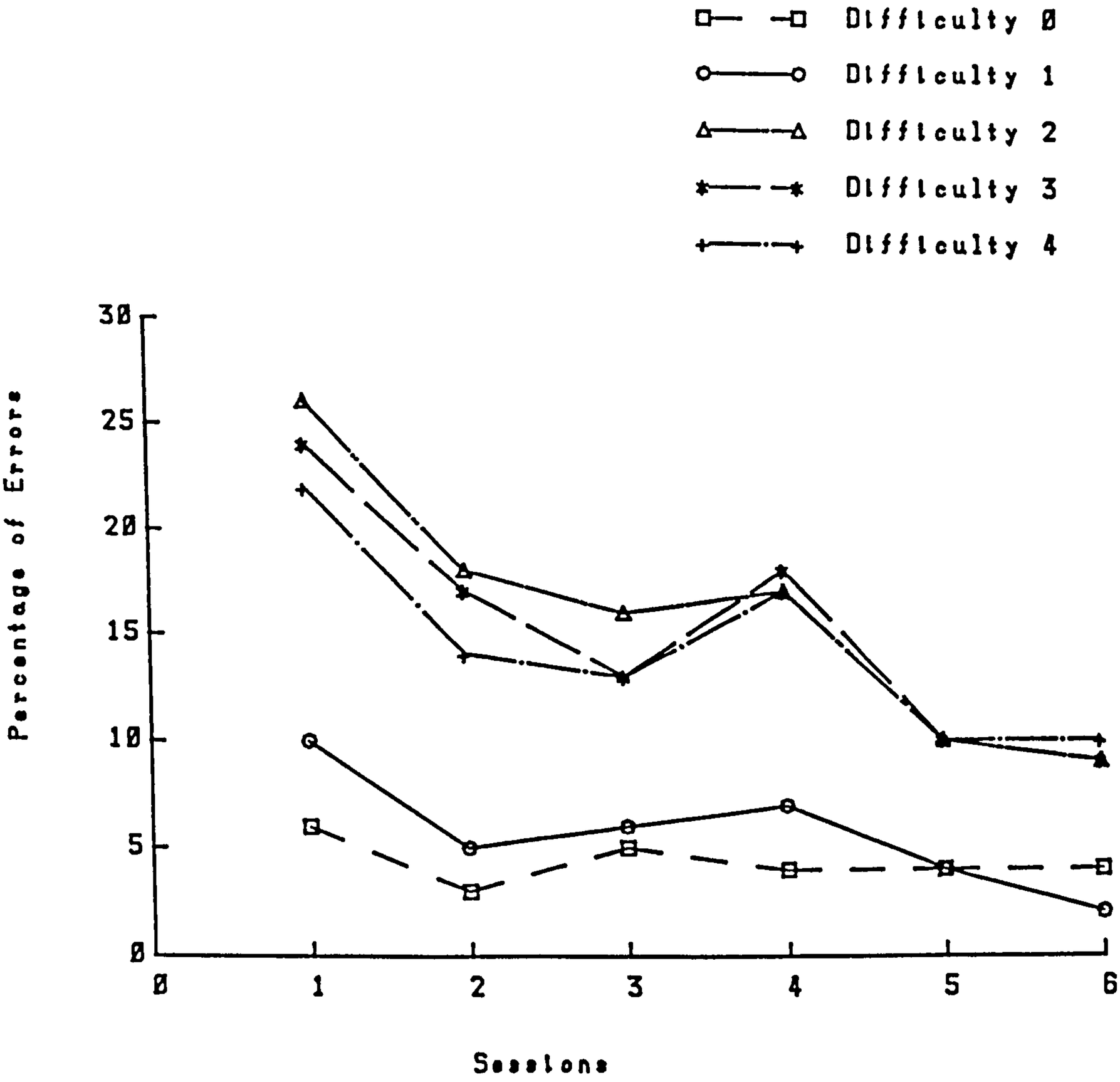


Figure 6.5 Mean percentage of errors as a function of the level of stimulus difficulty and the level of practice

of total responses.)

	Previous Error	Previous Correct
Current Error	Obs. 301 Exp. 228.6 1.6%	1,783 1,855.4 9.4%
Current Correct	1,779 1,851.4 9.4%	15,097 15,025 79.6%

x^2 = 28.52, df: 1, p<.0001
phi = 0.0391

Table 6.3 Errors on current trials as a function of the accuracy of the preceding response

The null hypothesis, that the two variables (namely accuracy of the preceding response and accuracy of the current response) were independent was tested using chi-square and the phi coefficient. The result of the chi-square test was statistically significant (x^2=28.52, p<.0001) which implied that the two variables were not independent. However, it must be taken into account that the phi coefficient was very small (phi = 0.03905). The fact that this low level of association was statistically significant was due to the large number of cases (N=18,960).

Secondly a set of five crosstabulations was carried out with the same two variables but controlling for level of difficulty. The same tests were carried out and, as expected, similar results were obtained. For each difficulty level, if there was a statistically significant relationship it was due

to the large number of cases and not to a substantial phi coefficient value (maximum value of phi was 0.055: see apendix 4).

Finally, a third set of six crosstabulations of the same two variables was carried out but this time controlling for each one of the six practice sessions. In this analysis, only the results for the first and the third sessions were found to be statistically significant. Again, though, the statistical significance did not indicate a substantial level of relationship (maximum phi value was 0.06). For the other sessions, the results were not significant (appendix 4).

In conclusion, although the null hypothesis that the accuracy of the preceding response and the accuracy of the current response were independent could not be wholly rejected, the occurence of an error on a given trial did not substantially alter the probability of a correct response on the next trial.

Factors Predicting Accuracy

A multiple regression analysis was carried out to examine the factors which explained the variance in the accuracy of the current response. As for speed the predictor variables were: three measures of practice (SESSION, BLOCK and TRIAL), the difficulty level of the current trial (DIFF1), the speed of the preceding response (RESO) and the accuracy of the preceding response (ACCO).

It was found that only a small proportion of the total variance could be accounted for by the independent variables.

Only three independent variables entered the equation with $p<.05$: the difficulty level of the current trial, the practice "session" and the accuracy of the preceding response. The prediction equation was:

Predicted ACC1 = 0.86-0.03(DIFF1)+0.02(SESSION)+0.03(ACC0)

The multiple correlation cofficient was 0.17 [$F(3,18956)=184.08$, $p<.0001$] which meant that only 2.89% of the variance could be predicted by the independent variables. This was not a good level of prediction, but the small multiple correlation coefficient could be partly due to the dichotomization into two unequal proportions: in this case, 89% correct responses and 11% errors. It has been established that when the dichotomization departs substantially from the mean (i.e. 50%), the variance is reduced and therefore its prediction is very difficult since the statistical power of tests is reduced (e.g. Cohen, 1983). The present dichotomization departed quite substantially from the mean and therefore the variance of the data had been reduced. Cohen (1983) gives the following formula with which one may find the maximum obtainable correlation for a given dichotomization:

$e=h/SQR[p*(1-p)]$

where e, the ratio of the point-biserial r to the biserial r, is the maximum value of r obtainable; h is the ordinate of the standard normal curve which cuts off p (the proportion of the cases in one of the two intervals). In the present case e was computed to be 0.60. The square of e gives the maximum

variance which could ideally be predicted by the multiple regression and in this case this maximum value is 36%. This is still more than ten times greater than the 2.89% of the variance which was predicted by the multiple regression. To summarize, the variables "difficulty level", "session", and "accuracy of the preceding response" explained a minimal proportion of the total variance most of which remained unaccounted for.

Summary of Speed and Accuracy Variations

In summary, the main speed and accuracy changes of the group who received training by the FR method were the following:

(1) Speed increased with practice and decreased with increasing levels of stimulus difficulty. Accuracy also increased with practice and decreased with level of difficulty.

(2) The decrease of speed as a function of practice was more pronounced for "difficult" stimuli than for "easy" stimuli.

(3) Error responses were slower than correct responses and this was true irrespective of the level of practice. However, the effect of accuracy on speed varied as a function of stimulus difficulty.

(4) The occurence of an error on the preceding response slowed down the speed of the current response and this effect was true across all levels of practice. However, the effect of the accuracy of the preceding response varied as a

function of stimulus difficulty. In other words, the decrease of speed because of stimulus difficulty was more pronounced if the preceding response was an error.

(5) A response following an error had a slightly increased probability of being an error itself.

(6) It was possible to explain 36% of the variance in speed as a function of the level of stimulus difficulty, the level of practice and the speed of the preceding response.

(7) It was not possible to explain and make a good prediction of variance in accuracy.

6.3 Results of the Fixed-Adaptive Mode of Training

Speed Changes with Practice, Stimulus Difficulty and Accuracy

The effects of practice "sessions" and "difficulty level" were examined by a repeated measures analysis of variance. The results were very similar to those of the FR group. It was found that: in general, RT decreased with increased levels of practice [$F(5,45)=25.68, p<.001$]; RT also increased as the level of stimulus difficulty increased [$F(4,36)=37.90, p<.001$]; the significant interaction of practice "sessions" and "stimulus difficulty" [$F(20,180)=8.58, p<.001$] showed that subjects' RT to difficult stimuli decreased more than their RT to easy stimuli as a function of practice (figure 6.6).

Another analysis of variance was carried out to examine the effects of practice ("session"), level of stimulus difficulty and accuracy on speed. As in the analysis for the

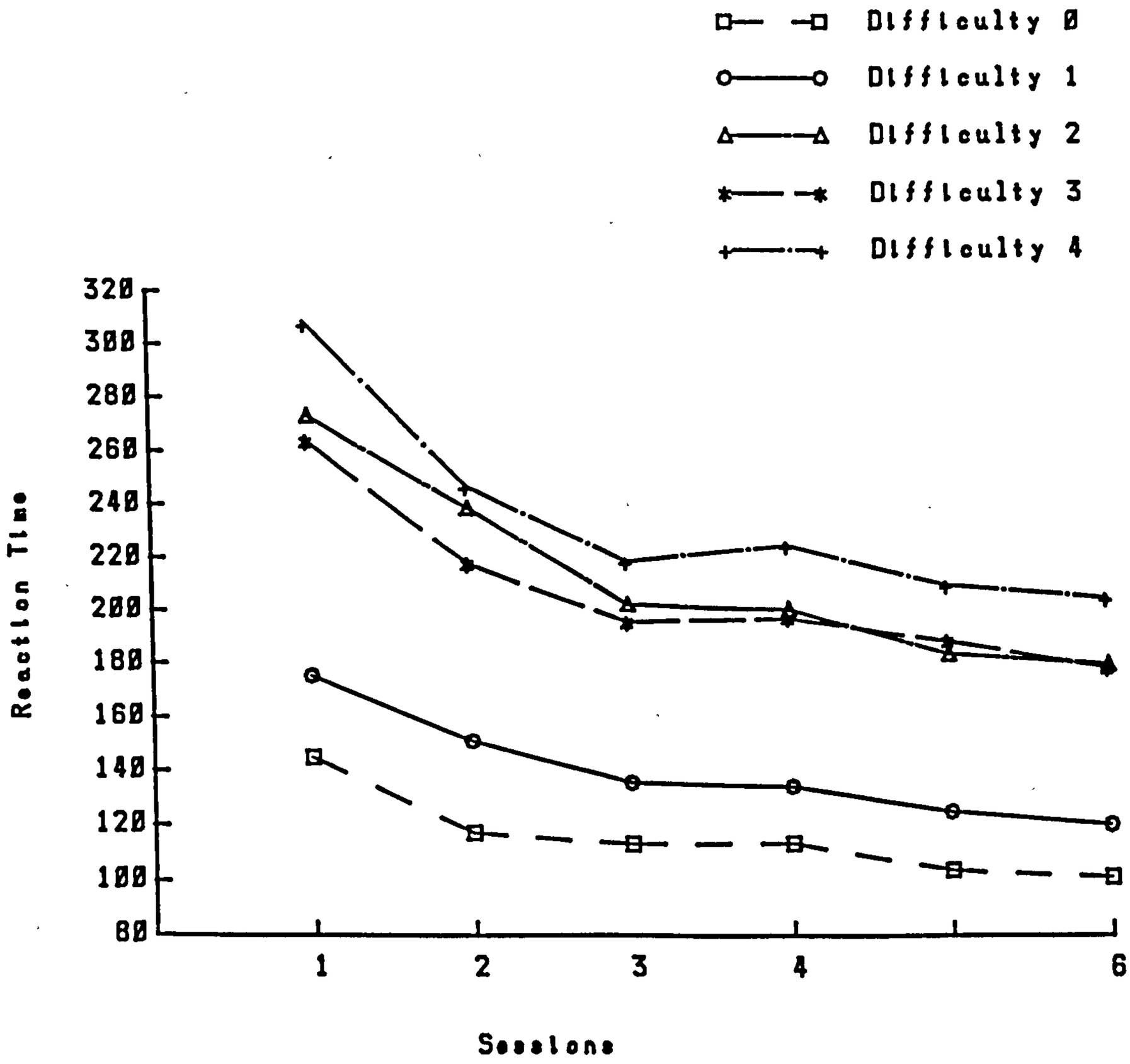


Figure 6.6 Mean RT of the current response as a function of the level of stimulus difficulty and the level of practice

FR group, the data used were the mean residual values of RT for each trial (i.e. all between subjects variance had been removed). Similarly to the results of the FR group and the repeated measures analysis, it was found that: speed increased with practice [$F(5,19140)=598.94$, $p<.0001$] and decreased with increasing stimulus difficulty [$F(4,19140)=1915.61$, $p<.0001$]. The significant interaction between "stimulus difficulty" and "session" [$F(20,19140)=15.95$, $p<.0001$] implied that the increase in speed because of practice varied as a function of stimulus difficulty, as already shown from the repeated measures analysis of variance above.

The effect of accuracy on the mean residual RT values was only marginally significant for this group ($0.1>p>0.05$), unlike the FR group, although the mean residual RT for error responses was markedly bigger than that for correct responses. The low level of significance arose from the use of the least-squares method of analysis of variance and the matter will be discussed again a little later when the effect of accuracy is considered in combination with other variables.

The difference between the residual RTs of error and correct responses varied significantly as a function of the level of difficulty [$F(4,19140)=4.26$, $p<.005$] (figure 6.7) and also as a function of practice sessions [$F(5,19140)=5.07$, $p<.0001$] (figure 6.8). The effects of these interactions however were not systematic and are difficult to interpret mainly due to the large standard errors of the residual RTs of the error responses (table 6.4).

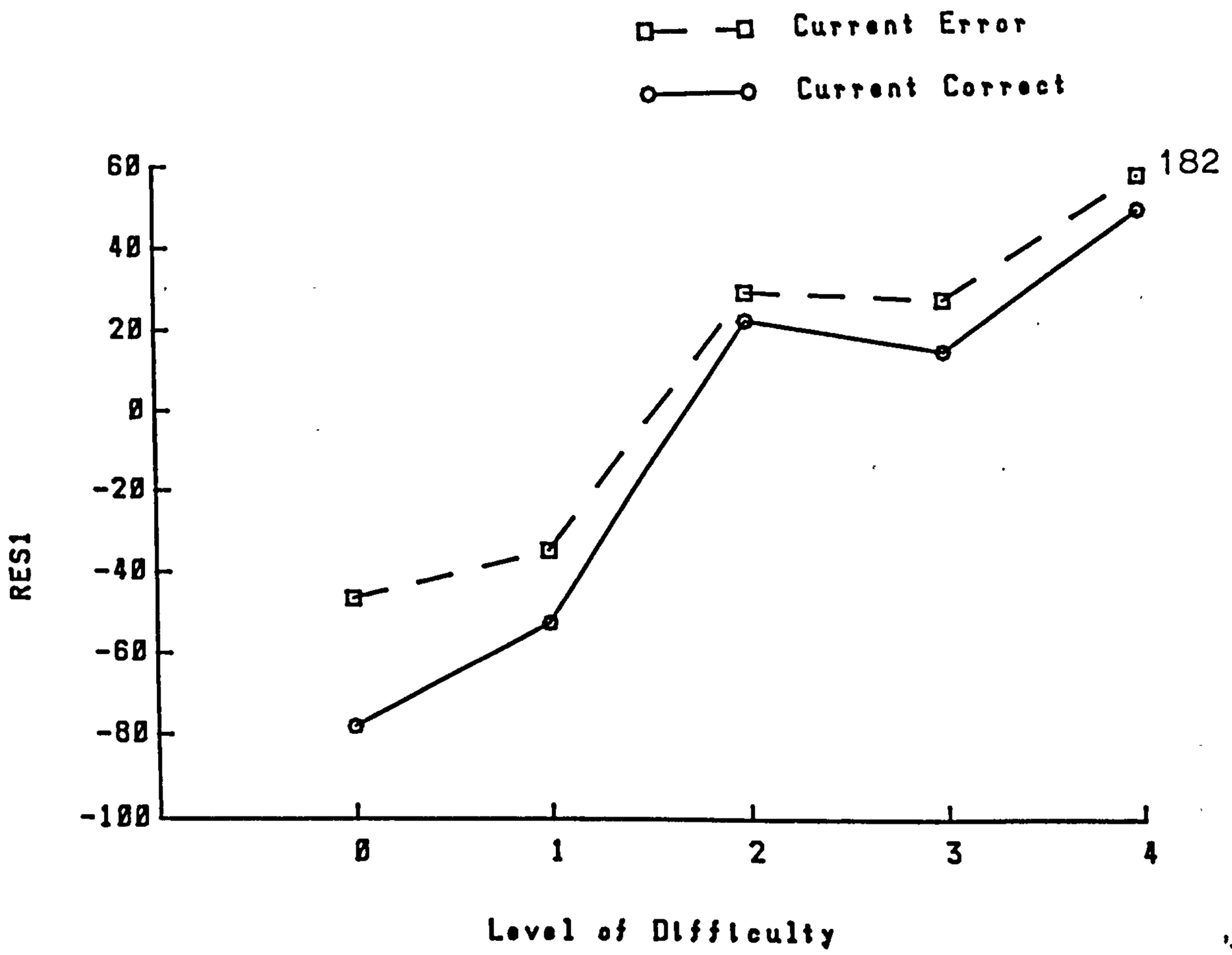


Figure 6.7 Mean residual RT of the current response as a function of the level of stimulus difficulty and the accuracy of the current response

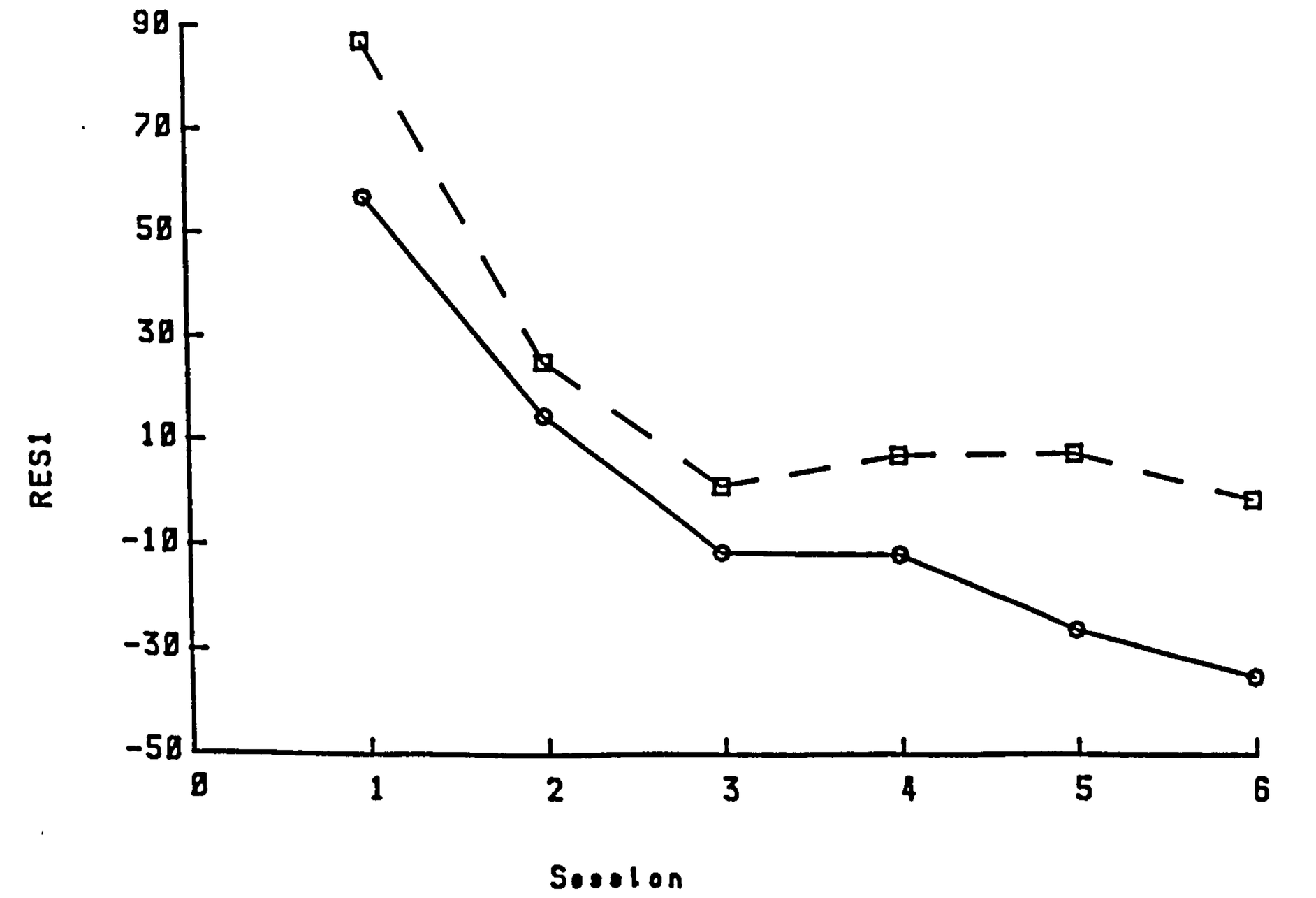


Figure 6.8 Mean residual RT of the current response as a function of the level of practice and the accuracy of the current response

	CURRENT ERROR				CURRENT CORRECT			
DIFF1	MEAN	SD	N	SE	MEAN	SD	N	SE
0	-46.41	106.17	99	10.67	-77.74	55.36	2,813	1.04
1	-34.54	57.91	116	5.38	-51.99	52.29	3,345	0.90
2	29.71	91.96	583	3.81	22.76	77.32	3,665	1.28
3	27.81	75.94	470	3.50	15.29	74.57	3,532	1.25
4	58.48	103.45	547	4.42	49.97	90.66	4,030	1.43
MEAN	29.63	95.28		2.24	- 3.09	86.83		0.66
TOTAL			1,815				17,385	

Table 6.4 Residual RTs for responses as a function of current stimulus difficulty and accuracy

The Effect of Accuracy of the Preceding Response on the Speed of the Current Response

Further analysis was carried out to examine, for each trial, the effect of accuracy of the preceding response on the speed of the current response. Furthermore, as in the analysis of the data of the FR group, the combined effects of difficulty level and accuracy of both the preceding and current responses on speed were examined by another analysis of variance. More specifically this analysis of variance looked into the effects of "accuracy of the preceding response", "accuracy of the current response" and the "difficulty level of the current trial" on the residual RT value of the current response.

When the preceding response was an error, the speed of the current response was significantly reduced [$F(1,18940)=199.31$, $p<.0001$]. An increase in the level of difficulty resulted in a decrease in speed (as already discussed). This speed decrease however was more pronounced

when the preceding response had been an error. In other words, the combination of an error in the preceding response and a "difficult" current stimulus had a great effect in reducing the speed of the current response. This latter effect was revealed by the significant interaction between "accuracy of the preceding response" and "difficulty level of the current trial" [$F(4,18940)=7.97$, $p<.0001$] (figure 6.9).

This analysis also revealed that error responses were significantly slower than correct responses [$F(1,18940)=33.68$, $p<.0001$]. (The main effect of "accuracy of the current response" was different in the present analysis of variance from what it was in the previous one. This point will be discussed and explained later.)

A significant interaction was found between the variables "accuracy of the preceding response" and "accuracy of the current response" [$F(1,18940)=4.25$, $p<.05$]. The effect of "accuracy of the preceding response" on the speed of the current response was greater than the effect of the "accuracy of the current response" (figure 6.10). In other words, the speed of the current response slowed down more when the preceding response was an error rather than it did when the current response itself was an error.

A three-way interaction among the three variables of this analysis of variance was found [$F(4,18940)=6.67$, $p<.0001$]. The effect of the interaction between "accuracy of the preceding response" and "accuracy of the current response" was, generally, more pronounced when the level of stimulus

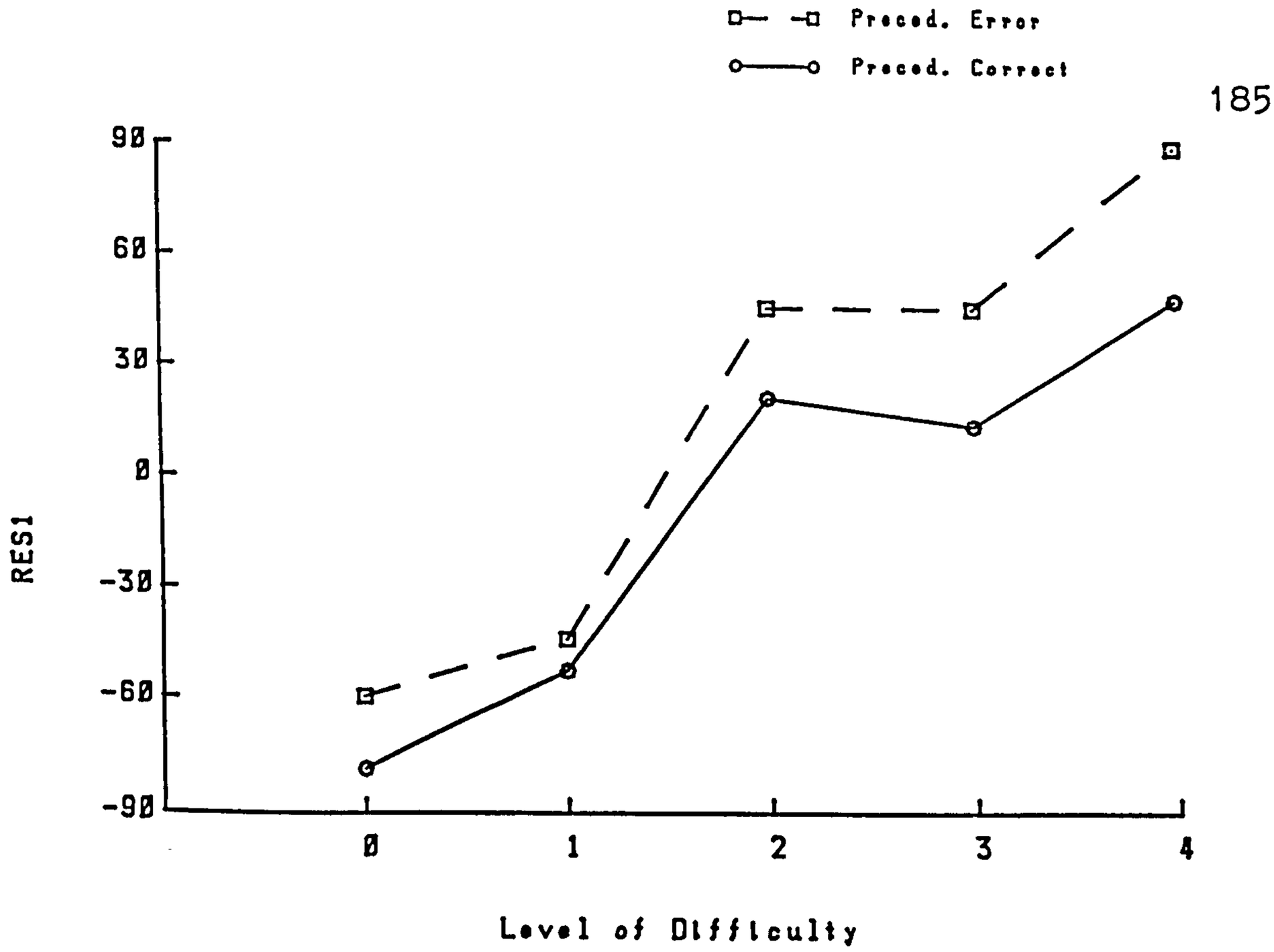


Figure 6.9 Mean residual RT of the current response as a function of the level of stimulus difficulty and the accuracy of the preceding response

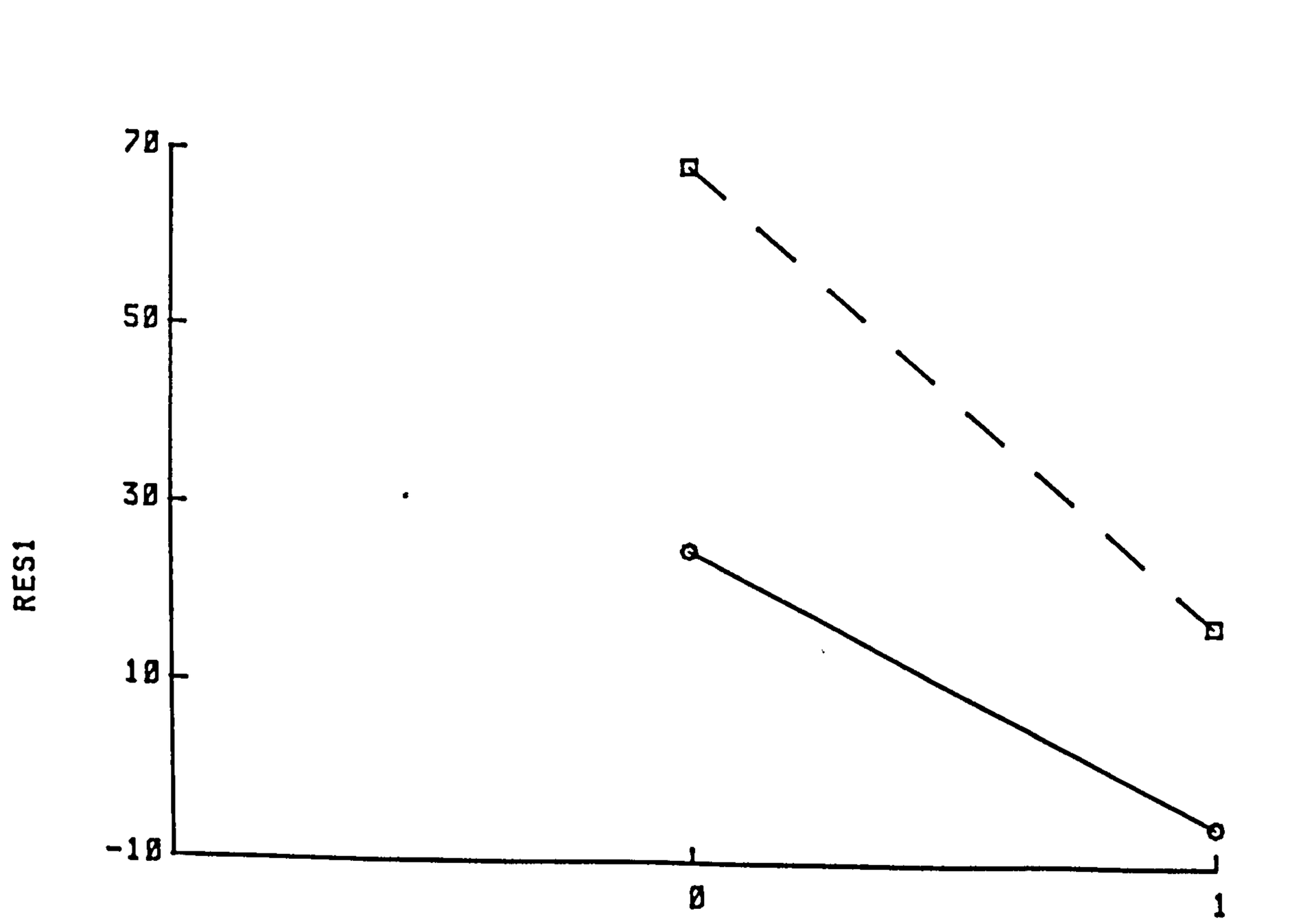


Figure 6.10 Mean residual RT of the current response as a function of the accuracy of the current and the preceding response

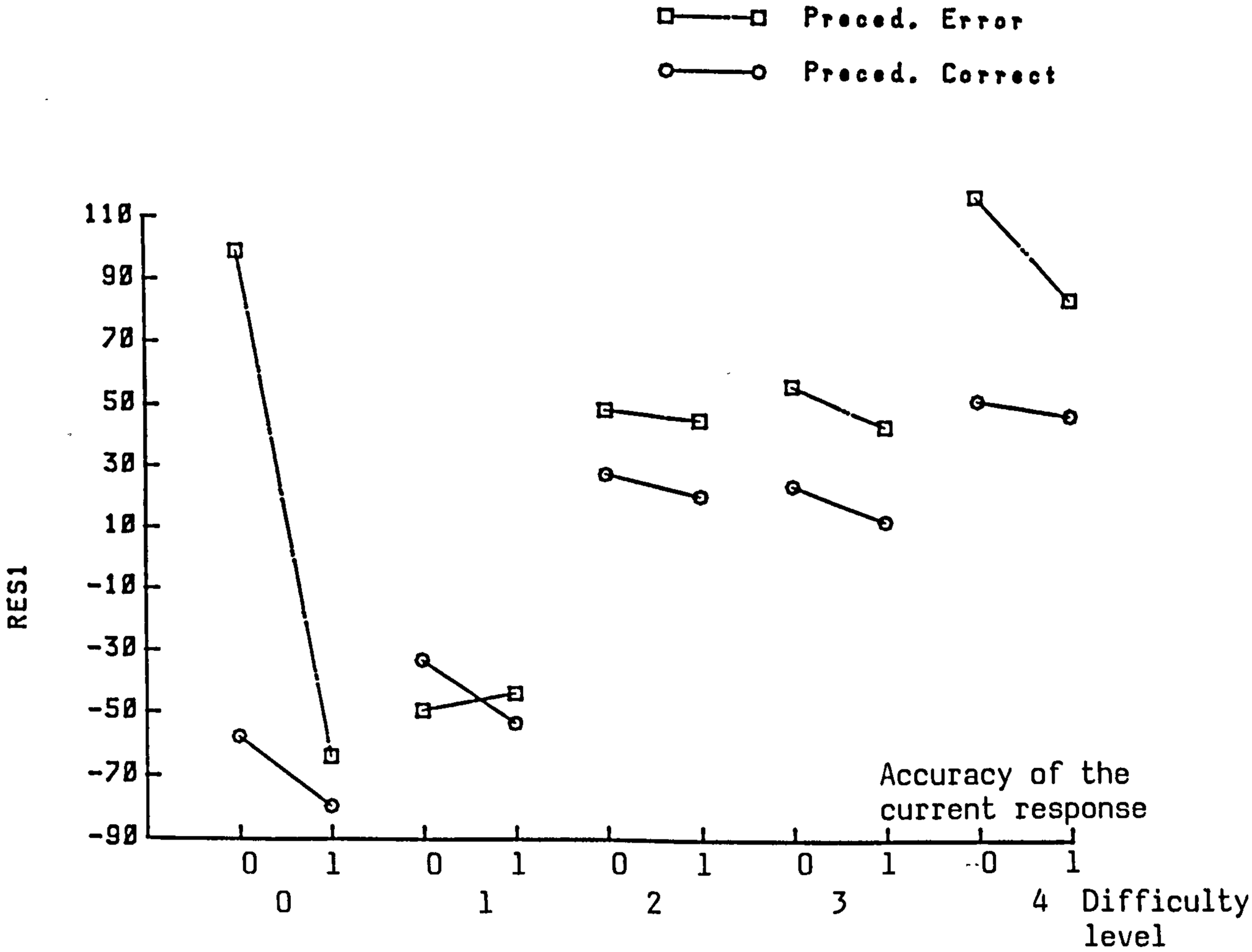


Figure 6.11 Mean residual RT of the current response as a function of the level of stimulus difficulty and the accuracy of the current and the preceding response

difficulty increased (figure 6.11). In the case where a current error was preceded by an error, the effect of difficulty level did not appear systematic. This however was most likely due to the fact that there were relatively very few such cases and therefore the standard errors of these means were too big to provide reliable estimates (table 6.5).

		PRECEDING				RESPONSE				
		DIFF1	MEAN	SD	N	SE	DIFF1	MEAN	SD	N	SE
C	ERROR	0	98.87	293.11	7	110.79	0	-64.00	62.91	266	3.86
U		1	-49.31	61.14	11	18.44	1	-43.84	61.11	327	3.38
R		2	48.29	98.06	51	13.73	2	44.89	92.81	339	5.04
R		3	56.02	92.53	54	12.59	3	43.04	110.97	330	6.11
E		4	116.96	126.65	57	16.78	4	83.41	111.63	352	5.95
N		MEAN	68.36	123.00		9.17		16.98	107.39		2.67
T		TOTAL			180					1,614	
R	CORRECT	0	-57.72	69.09	91	7.24	0	-79.47	54.16	2,508	1.08
E		1	-33.17	58.17	103	5.73	1	-53.50	50.95	2,962	0.94
S		2	27.11	91.22	527	3.97	2	19.95	74.88	3,281	1.31
P		3	23.59	73.25	407	3.63	3	12.02	69.01	3,166	1.23
O		4	51.25	98.00	487	4.44	4	86.35	86.35	3,634	1.43
N		MEAN	24.88	90.80		2.26		-5.52	83.73		0.67
S		TOTAL			1,615					15,551	
E											

TABLE 6.5 Residual RTs for current response as a function of stimulus difficulty, accuracy of the response and accuracy of the preceding response

The effect of accuracy of the preceding response on the speed of the current response varied significantly as a function of practice [$F(4,18940)=3.44$, $p<.005$]. This effect which did not appear to be systematic (figure 6.12) was revealed by another analysis of variance on the residuals, the variables being "accuracy of the preceding response", "accuracy of the current response" and "sessions".

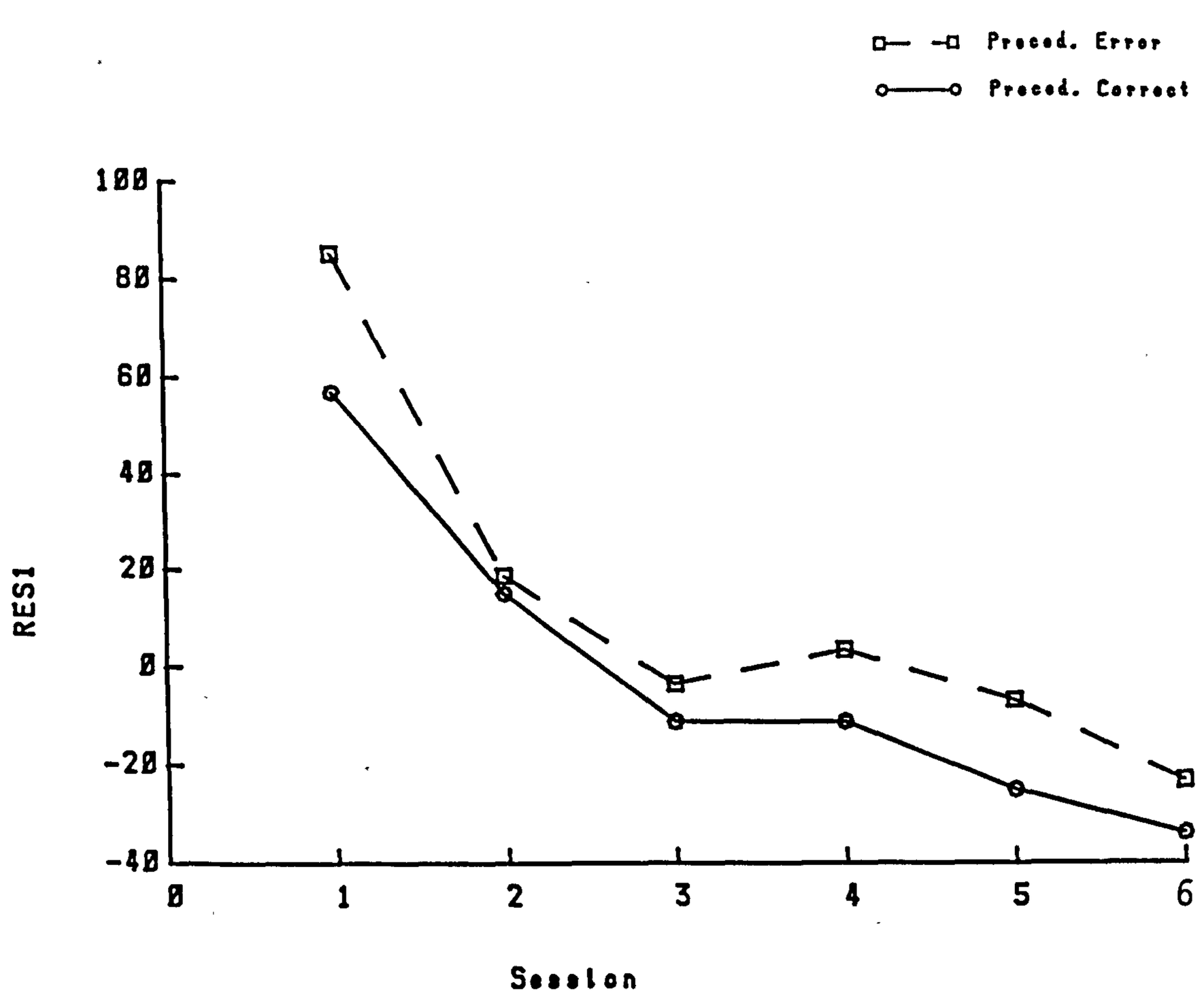


Figure 6.12 Mean residual RT of the current response as a function of practice and the accuracy of the preceding response

Another interesting result of this analysis of variance was related to the effect of "accuracy of the current response" on speed. When examining the main effects of the three variables above on the residuals, it was found that speed was slower when the current response was an error than it was when the current response was correct [$F(1,18940)=130.54$, $p<.0001$]. Since the main effect of the variable "accuracy of the current response" has appeared rather ambiguous so far it has to be discussed and explained further.

The Main Effect of Accuracy of the Current Response on Speed

Three analyses of variance on the residual values of RT and involving "accuracy of the current response" as a variable have been discussed so far. The first one involved the variables "accuracy of the current response", "level of stimulus difficulty" and practice "session". The second analysis involved the variables "accuracy of the preceding response", "accuracy of the current response" and "level of stimulus difficulty". The third analysis involved the variables "accuracy of the preceding response", "accuracy of the current response" and practice "session". In the first analysis the main effect of the variable "accuracy of the current response" was not statistically significant whereas in the other two analyses it was highly significant. The explanation of this "discrepancy" can be found in the way the analysis of variance was done.

The number of cell frequencies was unequal and therefore

ordinary analysis of variance would not have been appropriate. The unequal frequencies result in correlation among the classification variables and where two or more factors are correlated it is not obvious which factor influences the criterion and to what extent. For this reason, the default SPSS least squares method of analysis of variance was used. There are several least squares methods which are used to estimate the independent effect of each variable adjusted for relationships to other classification variables (Overall and Speigel, 1969, p. 311). The SPSS default method uses the Classic Experimental Approach which takes into account the experimental design hierarchy of main effects and interactions. The method adjusts each effect for all other effects at an equal or lower level and ignores higher order effects.

In our particular case, in the first analysis when the main effect of "accuracy of the current response" was adjusted for the effects of the other two variables of the ANOVA, it became non-significant. In the other two analyses however, the main effect of "accuracy of the current response" after the adjustment was significant. The difference among the three analyses was that they involved different variables and therefore different adjustments.

The conclusion which can be drawn at this stage is that an error response is generally slower than a correct response. However, this effect is not statistically significant if speed is studied as a function of the variables "accuracy of the

current response", "level of stimulus difficulty" and practice "session" (due to the partialling of covariation between these variables).

Factors Predicting Speed

Similarly to the analyses of the data for the FR group, a stepwise multiple correlation analysis was carried out to examine which factors best explained the variance in speed. The dependent and the predictor variables used in this analysis were the same as those used for the analysis of the FR group. [The dependent variable was the residual value of the reaction time of the current response and the predictor variables were: three measures of practice level ("SESSION" (N=320), "BLOCK" (N=80) and "TRIAL"), difficulty level of the current trial (DIFF1), difficulty level of the preceding trial (DIFF0), speed of the preceding response (RES0) and accuracy of the preceding response (ACC0).]

It was found that 34 per cent of the variance of speed could be explained by the difficulty level of the current trial, the practice "session", and the speed of the preceding response. The obtained multiple correlation was R=.59 [$F(3,18956)=3301.24$, $p<.0001$] and the prediction equation was:

Predicted Residual RT (csecs)

= 30.18(DIFF1)-12.30(SESSION)+0.14(RES0)-23.76

(Mean Residual RT = 0)

The variables "difficulty level of the preceding trial", "accuracy of the preceding response" and "block" also entered the equation at statistically significant levels, but their

contribution to the total variance was less than one per cent.

Accuracy Changes with Practice and Difficulty Level

A repeated measures analysis of variance was carried out on subjects' mean proportion of correct responses looking into the effects of practice "session" and "level of stimulus difficulty" on accuracy. It was found that the proportion of errors decreased significantly with practice [$F(5,45)=7.21$, $p<.001$] and increased significantly with increased stimilus difficulty [$F(4,36)=20.71$, $p<.001$]. There was no significant interaction between the two variables ($p>.05$) which suggests that, for the FA group, the effect of practice on accuracy was quite independent from the effect of stimulus difficulty.

The Effect of Accuracy of the Preceding Response on the Accuracy of the Current Response

Whether the accuracy of the preceding response affected the accuracy of the current response was next to be examined. Similarly to the data analysis of the FR group, all responses were crosstabulated in a two-by-two matrix, the variables being "accuracy of the current response" and "accuracy of the preceding response" (table 6.6).

The null hypothesis was that the two variables were independent, in other words, that the accuracy of the current response was independent from the accuracy of the preceding response. This hypothesis was tested using chi-square and the phi coefficient. Neither the obtained chi-square value nor the phi coefficient was statistically significant ($p>.1$). Therefore the null hypothesis could not be rejected. It is therefore likely that the two variables are independent.

	Previous Error	Previous Correct
Current Error	Obs. 180 Exp. 169.8 0.9%	1,615 1,625.2 8.5%
Current Correct	1,614 1,624.2 8.5%	15,551 15,541 82.0%

x^2 = 0.67, df: 1, not significant
Phi = 0.0062

Table 6.6 Errors on current stimuli as a function of accuracy on the preceding response

Two sets of crosstabulations were also carried out with the same two variables but one set of crosstabulations controlling for level of stimulus difficulty and the other controlling for level of practice. The same two statistical tests were carried out and, as expected, the same results were obtained. The chi-square values and the phi coefficients were non-significant across all levels of stimulus difficulty and all practice sessions (Appendix 4).

In conclusion, it can be confidently stated that for the FA group, the accuracy of the current response was not affected by the accuracy of the preceding response. In other words, the occurence of an error on a given trial did not affect the probability of a correct response on the next trial.

Factors Predicting Accuracy

A multiple regression analysis was carried out to examine

the factors which explained the variance in the accuracy of the current response. As for speed the predictor variables were: three measures of practice ("SESSION", "BLOCK" and "TRIAL"), the difficulty level of the current trial (DIFF1), the speed of the precesing response (RESO) and the accuracy of the preceding response (ACCO).

It was found that only a very small proportion (even smaller than that of the FR group) could be accounted for by the independent variables. Only three variables entered the prediction equation with $p<.05$: the difficulty level of the current trial and two practice measures, practice "session" and "trial".

Predicted ACC1 = 0.93-0.02(DIFF1)+0.01(SESSION)-0.00002(TRIAL)

The multiple correlation coefficient was R=0.13 [$F(1,18956)=113.77$, $p<.0001$] which meant that only 1.75% of the variance could be predicted by the independent variables.

This was a very poor prediction. Given that the dependent variable was a dichotomized variable rather than a continuous one, Cohen's formula which has already been discussed was used again. It was found that the maximum variance which could ideally be predicted was 32%. This is almost twenty times more than the variance predicted by the above multiple regression analysis. In conclusion, the variables "difficulty level of the current trial", "session" and "trial" explained 1.75% of the variance with most of the variance remaining unexplained.

Summary of Speed and Accuracy Variations

In summary, the main speed and accuracy changes of the group which received training by the FA method were similar to those of the FR group and were as follows.

(1) Speed increased with practice and decreased with increased levels of stimulus difficulty. Accuracy also improved with practice and was diminished by increased stimulus difficulty.
(2) The changes in speed because of practice varied as a function of stimulus difficulty.
(3) Error responses were generally slower than correct responses across the various levels of practice and stimulus difficulty.
(4) The occurence of an error slowed down the speed of the next response even if that response was correct. The effect of accuracy of the preceding response varied as a function of the level of practice and the level of stimulus difficulty.
(5) The accuracy of a given response was independent from the accuracy of the preceding response.
(6) 34% of the variance in speed was explained by the variables "difficulty of the current response", practice "session" and "speed of the preceding response".
(7) It was not possible to explain the variance in accuracy in any satisfactory way.

6.4 Results of the Cumulative-Random Mode of Training

Speed Changes with Practice, Stimulus Difficulty and Accuracy

Similar to the analysis of the data of the FR and FA groups, a repeated measures analysis of variance was carried out to examine the effects of practice "sessions" and "stimulus difficulty" on RT. It was found that RT generally decreased with increased levels of practice [$F(5,45)=27.90$, $p<.001$] and increased as the level of stimulus difficulty increased [$F(4,36)=91.61$, $p<.001$]. The significant interaction between practice "sessions" and "stimulus difficulty" [$F(20,180)=3.54$, $p<.001$] revealed that the increase in RT due to "stimulus difficulty" varied as a function of practice (figure 6.13).

Another analysis of variance was carried out to examine the effects of practice ("session"), level of "stimulus difficulty" and "accuracy" on speed. As in the previous analyses of the groups FR and FA, the data used were the mean residual values of RT for each trial (which meant that all between subjects variance had been removed). As expected, it was found that, in general, speed increased with practice [$F(5,19140)=222.82$, $p<.0001$] and decreased with an increase in the level of stimulus difficulty [$F(4,19140)=2972.26$, $p<.0001$]. The increase in speed related to practice varied as a function of stimulus difficulty but the effect did not seem systematic. This was implied by the significant interaction between the variables "session" and "stimulus difficulty" [$F(20,19140)=8.27$, $p<.0001$].

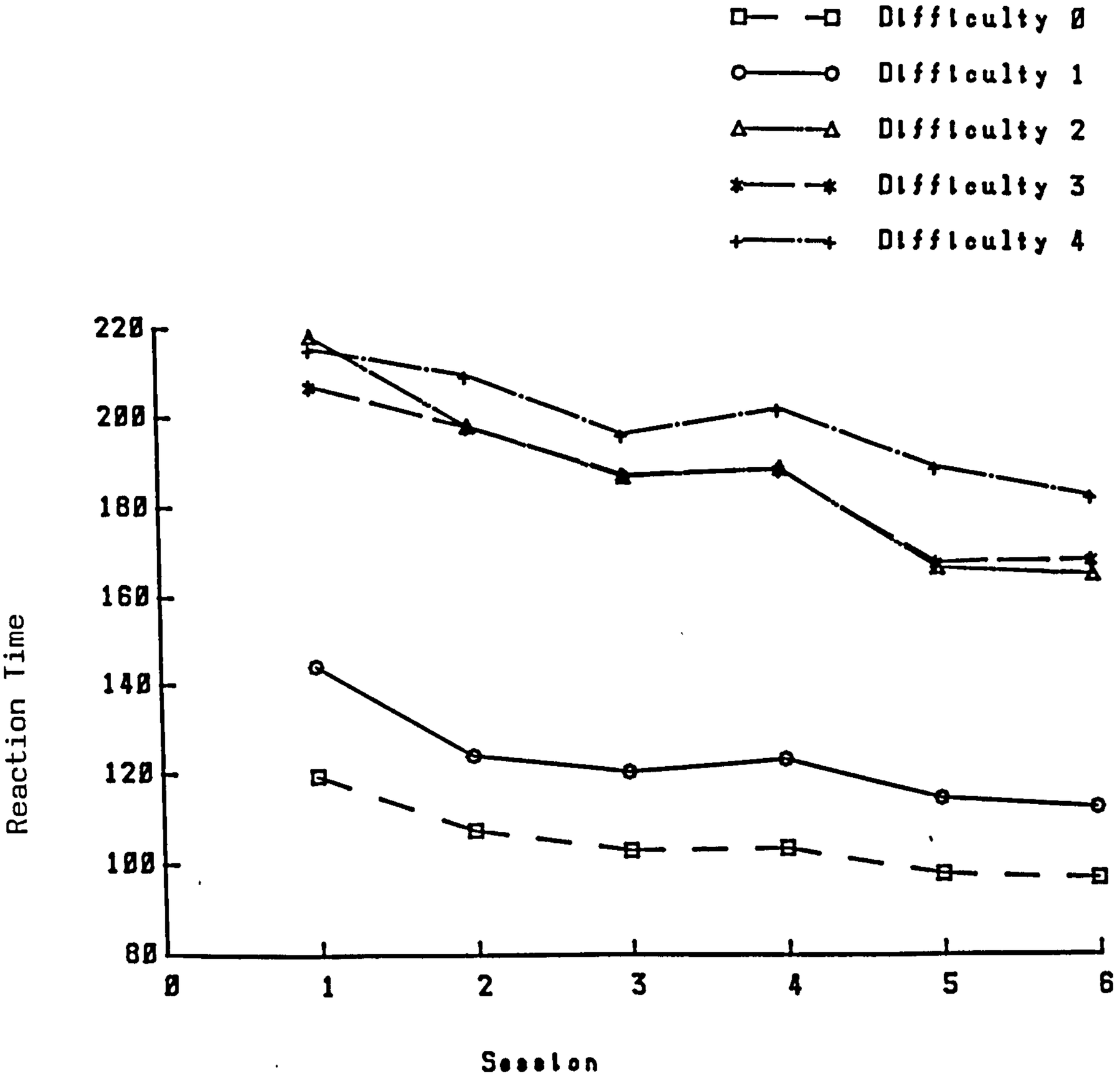


Figure 6.13 Mean RT of the current response as a function of the level of stimulus difficulty and level of practice

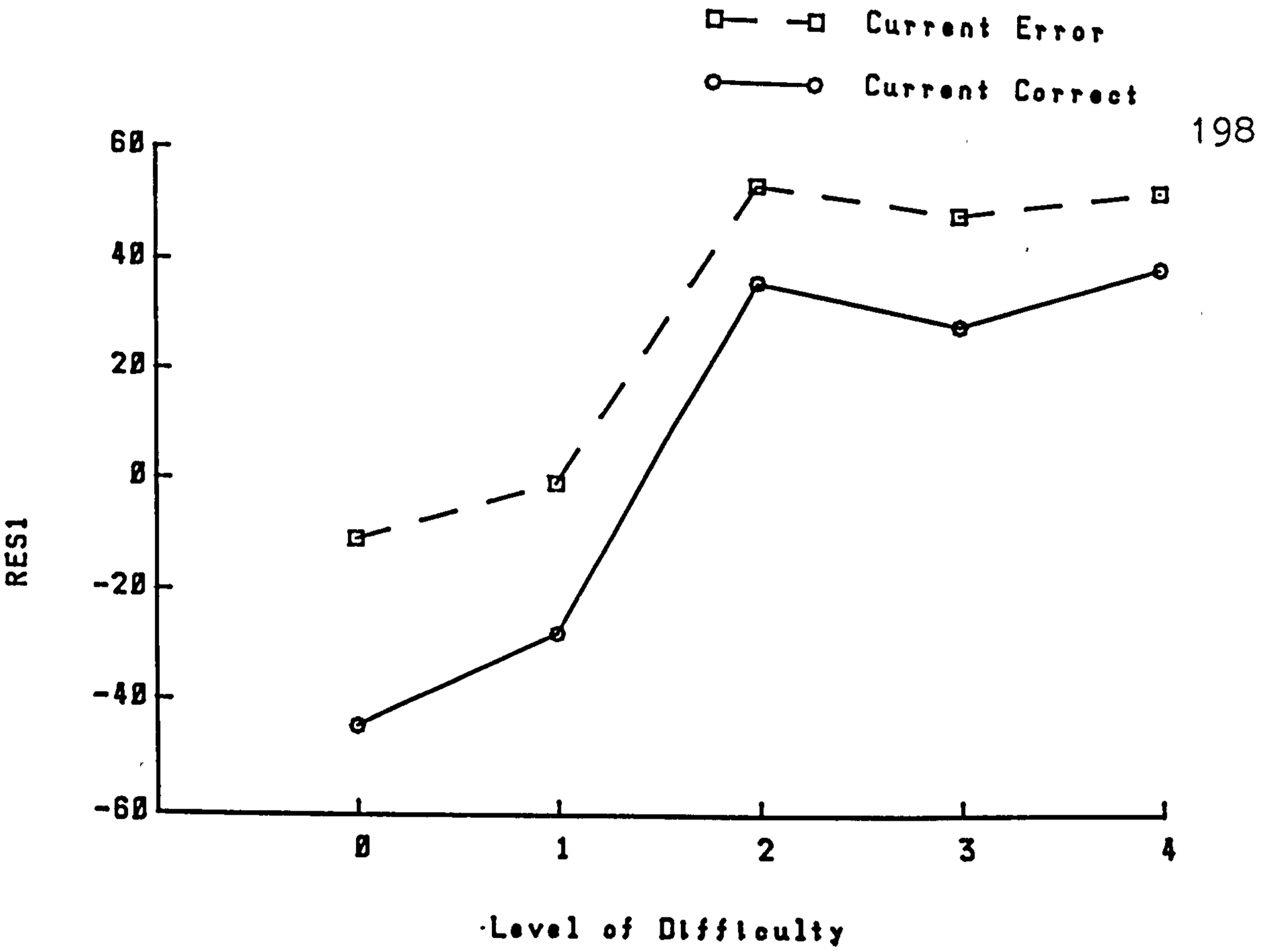


Figure 6.14 Mean residual RT of the current response as a function of the level of stimulus difficulty and accuracy of the current response

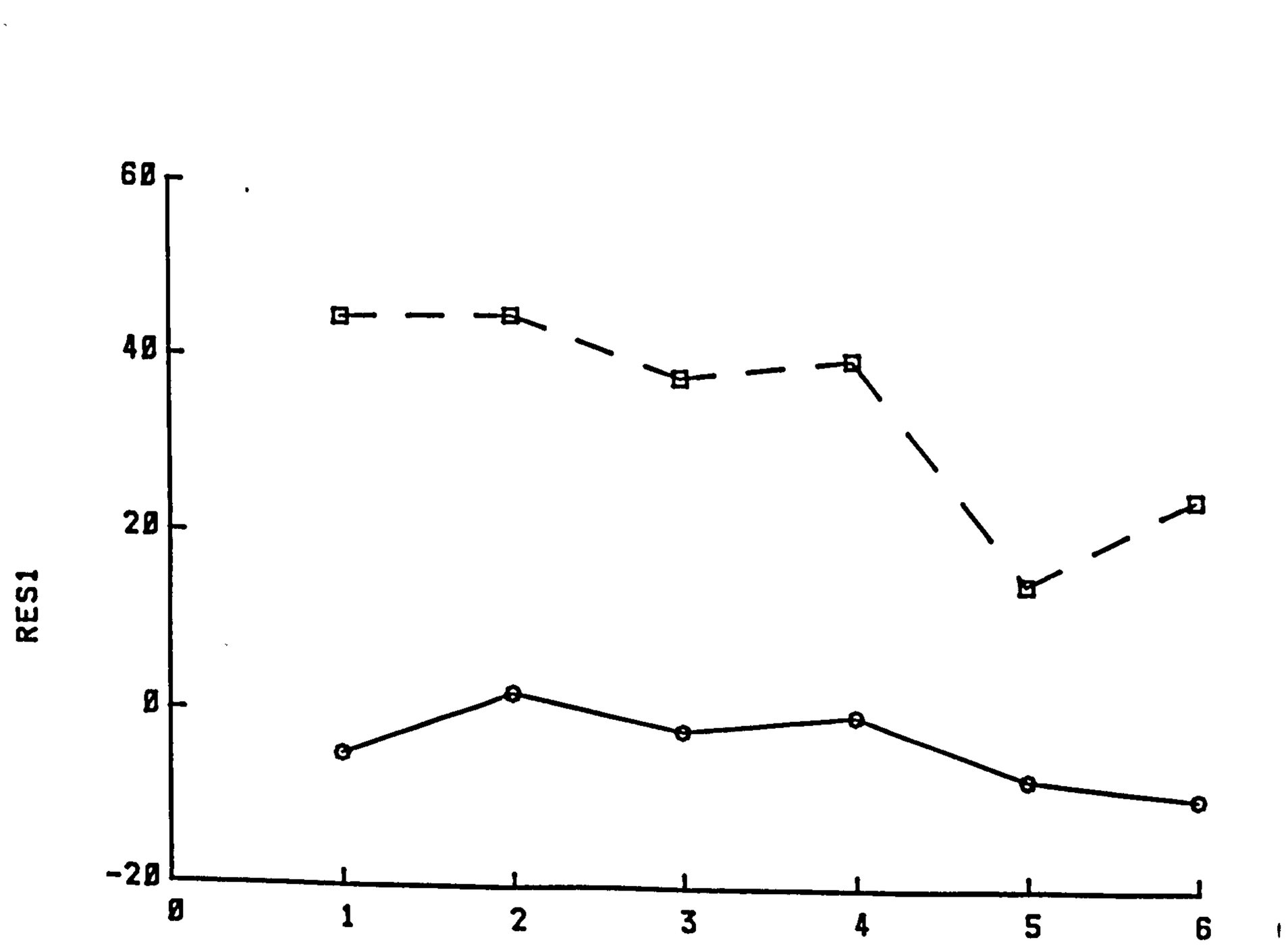


Figure 6.15 Mean residual RT of the current response as a function of practice and accuracy of the current response

Error responses were significantly slower than correct responses [F(1,19140)=264.20, p<.0001]. The difference between the residual RTs for error and correct responses varied significantly as a function of stimulus difficulty [F(4,19140)=5.36, p<.0001] (figure 6.14) but the effect did not appear systematic. Furthermore the difference between the residuals of error and correct responses was more pronounced earlier rather than later in practice (figure 6.15). This was revealed by the significant interaction between "accuracy of the current response" and practice "session" [F(5,19140)=4.73, p<.0001].

The Effect of Accuracy of the Preceding Response on the Speed of the Current Response

Similar to the data analysis of the previous two groups, further analysis was carried out to examine, for each trial, the effect of accuracy of the preceding response on the speed of the current response. Also, the combined effects of difficulty level and accuracy of both the preceding and the current responses, on speed, were examined by another analysis of variance. This analysis looked into the effects of "accuracy of the preceding response", "accuracy of the current response" and the "difficulty level of the current trial" on the residual RT values of the current response.

When the preceding response was an error, the speed of the current response was significantly reduced [F(1,18940)=33.41, p<.0001]. In other words the occurence of an error led to a decrease in the speed of the next response.

This effect was independent of the accuracy of the current response ($p>.5$) and also independent of the level of stimulus difficulty ($p>.8$). From another analysis of variance on the residual RT values and with the variables being "accuracy of the preceding response", "accuracy of the current response" and practice "session", it was found that the effect of accuracy of the preceding response on speed was also independent of the level of practice ($p>.5$). It can be concluded therefore that the occurence of an error slowed down the speed of the next response and this effect was independent of the accuracy of the current response, the level of stimulus difficulty or the level of practice.

Factors Predicting Speed

Similar to the analysis of the data for the FR and FA groups, a stepwise multiple correlation analysis was carried out to examine which factors best explained the variance in speed. The dependent and the predictor vcariables used in this analysis were the same as those used in the analyses of the other two groups. Namely, the dependent variable was the residual value of the reaction time of the current response and the predictor variables were: three measures of practice ["SESSION" (N=320), "BLOCK" (N=80) and "TRIAL"], difficulty level of the current trial (DIFF1), difficulty level of the preceding trial (DIFF0), speed of the preceding response (RES0) and accuracy of the preceding response (ACC0).

It was found that 33 per cent of the variance of speed could be explained by the difficulty level of the current

trial, the practice "session" and the speed of the preceding response. The obtained multiple correlation was R=.58 [F(3,18956)=3132.72, p<.0001] and the prediction equation was

Residual RT (csecs)

= 25.84(DIFF1)-6.69(SESSION)+0.06(RESO)-18.98

(Mean Residual RT = 0)

The rest of the predictor variables, except "trial", did enter the prediction equation at statistically significant levels but their contribution to the total variance accounted for was less than one per cent.

Accuracy Changes with Practice and Difficulty Level

A repeated measures analysis of variance was carried out on subjects' mean proportion of correct responses to examine how accuracy varied as a function of practice "sessions" and level of "stimulus difficulty". It was found that subjects' proportion of errors decreased significantly with practice [F(5,45)=10.65, p<.01] and increased with increased stimulus difficulty [F(4,36)=13.49, p<.025]. Finally, the significant interaction between practice "sessions" and "stimulus difficulty" [F(20,180)=1.86, p<.05] implied that the decrease in the proportion of errors because of practice was more pronounced for "difficult" rather than "easy" stimuli (figure 6.16) ("difficult" stimuli are considered to be those which require the depression of non-adjacent keys).

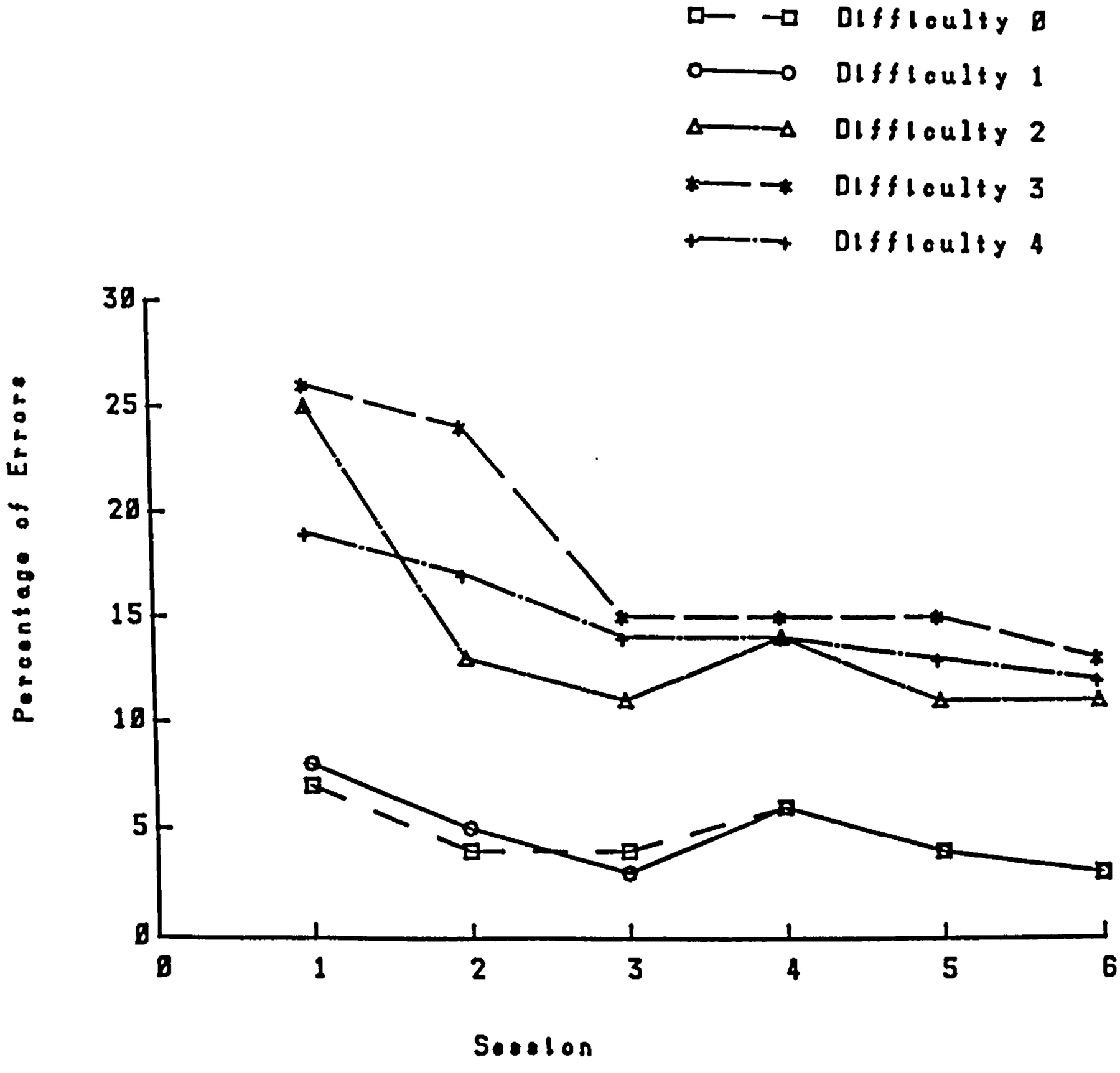


Figure 6.16 Mean percentage of errors as a function of practice and level of stimulus difficulty

The Effect of Accuracy of the Preceding Response on the Accuracy of the Current Response

The next question to be examined was whether the accuracy of the preceding response affected the accuracy of the current response. Similar to the analysis of the data for the other groups, all responses were crosstabulated in a two-by-two matrix, the variables being "accuracy of the preceding response" and "accuracy of the current response". This crosstabulation produced four categories of responses: errors preceded by errors, errors preceded by correct responses, correct responses preceded by errors and correct responses preceded by correct responses (table 6.7). (The values in each cell are: the actual number of responses, the expected number of responses and the percentage of total responses.)

	Previous Error	Previous Correct
Current Error	Obs. 203 Exp. 164.6 1.1%	1,565 1,603.4 8.3%
Current Correct	1,562 1,600.4 8.2%	15,630 15,592 82.04%

x^2 = 10.6217, df: 1, p<.0001
Phi = 0.0240

Table 6.7 Errors on current stimuli as a function of accuracy on the preceding response

The null hypothesis was that the two variables were independent, in other words that the accuracy of the preceding response did not affect the accuracy of the current response.

This hypothesis was tested using chi-square and the phi coefficient. The result of the chi-square test was statistically significant (x^2=10.62, p<.001) which implied that the two variables were not statistically independent. However, the phi coefficient value was very small (phi=0.0240). The fact that this low level of association was statistically significant was due to the large number of cases (N=18,960), as was the case for the FR group. This level of association is not substantial in real terms. In a total of 18,960 cases the difference between the actual and the expected number of responses in each category was only 39. Therefore, it can be concluded that there is a statistically significant but insubstantial relationship between accuracy of the preceding response and accuracy of the current response.

A second set of crosstabulations was made with the same two variables but controlling for level of stimulus difficulty. The same tests were carried out and, as expected, similar results were obtained. For each difficulty level, if there was a statistically significant relationship, it was due to the large number of cases and not to a substantial phi coefficient value (maximum value of phi was .0858, appendix4).

Finally, a set of crosstabulations was made with the same two variables but this time controlling for practice sessions. The same two statistical tests were carried out and similar results were obtained. For each session, there was either no significant relationship between the two variables or, if there was, it was insubstantial. The maximum phi coefficient

value obtained by this crosstabulation was 0.0664 (Appendix 4).

It can be quite safely concluded therefore that, although the null hypothesis that the accuracy of the preceding response and the accuracy of the current response were independent could not strictly be wholly rejected, the occurence of an error on a given trial did not substantially alter the probability of a correct response on the next trial.

Factors Predicting Accuracy

A multiple regression analysis was carried out to examine the factors that could account for the variance in the accuracy of the current response. The dependent variable was accuracy of the current response, and, similarly to all previous multiple regression analyses, the predictor variables were: "SESSION", "BLOCK", "TRIAL", the difficulty level of the preceding trial (DIFFO), the difficulty level of the current trial (DIFF1), the speed of the preceding response (RESO) and the accuracy of the preceding response (ACCO).

The results of this multiple regression (as for those for groups FR and FA) showed that only a poor prediction of the accuracy of the current response could be made. Only four variables entered the prediction equation (with $p<.05$): the difficulty level of the current trial, "session", the accuracy of the preceding response and the difficulty level of the preceding response.

Predicted ACC1 =

$$0.89-0.03(DIFF1)+0.0009(SESSION)+0.03(ACCO)+0.0004(DIFFO)$$

The multiple correlation coefficient was R=0.1357 [F(4,18955)=88.94, p<.0001] which means that only 1.8 per cent of the variance of the accuracy of the current response could be accounted for by the predictor variables.

Cohen's formula was used again to compute the maximum variance which could ideally be predicted. This (32%) was almost twenty times more than the variance actually accounted for. In conclusion, the variables "difficulty level of the current trial", "session", "accuracy of the preceding response" and "difficulty level of the preceding trial" explained 1.8% of the variance of the accuracy of the current response, while the larger part of the variance remained unaccounted for.

Summary of Speed and Accuracy Variations

In summary, the main speed and accuracy changes of the group who received training with the CR method were as follows.

(1) Both speed and accuracy increased with practice and decreased with increased difficulty.

(2) The speed changes because of practice varied as a function of stimulus difficulty.

(3) Error responses were slower than correct responses and this effect varied as a function of difficulty level as well as a function of practice.

(4) If the preceding response was an error, the speed of the current response slowed down and this effect was independent of the accuracy of the current response, the

level of stimulus difficulty and the level of practice.

(5) Accuracy on a given trial was generally independent from the accuracy of the preceding trial across all levels of stimulus difficulty and practice sessions.

(6) 33% of the variance in speed was explained by the difficulty level of the current trial, the practice "session" and the speed of the preceding response.

(7) Finally the variance in the accuracy of the current response was not explained in any satisfactory way.

6.5 Results of the Cumulative-Adaptive Mode of Training

Speed Changes with Practice, Stimulus Difficulty and Accuracy

All the statistical analyses used for the previous groups were also carried out for the last group of subjects who had received training with the CA method. First a repeated measures analysis of variance was carried out to examine the effects of "stimulus difficulty" and practice "sessions" on RT. It was found that RT decreased with practice [$F(5,45)=27.46$, $p<.001$] and increased with increased levels of "stimulus difficulty" [$F(4,36)=60.00$, $p<.001$]. The increase in RT because of "stimulus difficulty" varied as a function of practice "sessions" (figure 6.17). The latter was revealed by a significant interaction between "stimulus difficulty" and "sessions" [$F(20,180)=2.28$, $p<.001$].

The effects of practice ("session"), level of "stimulus difficulty" and "accuracy of the current response" on speed

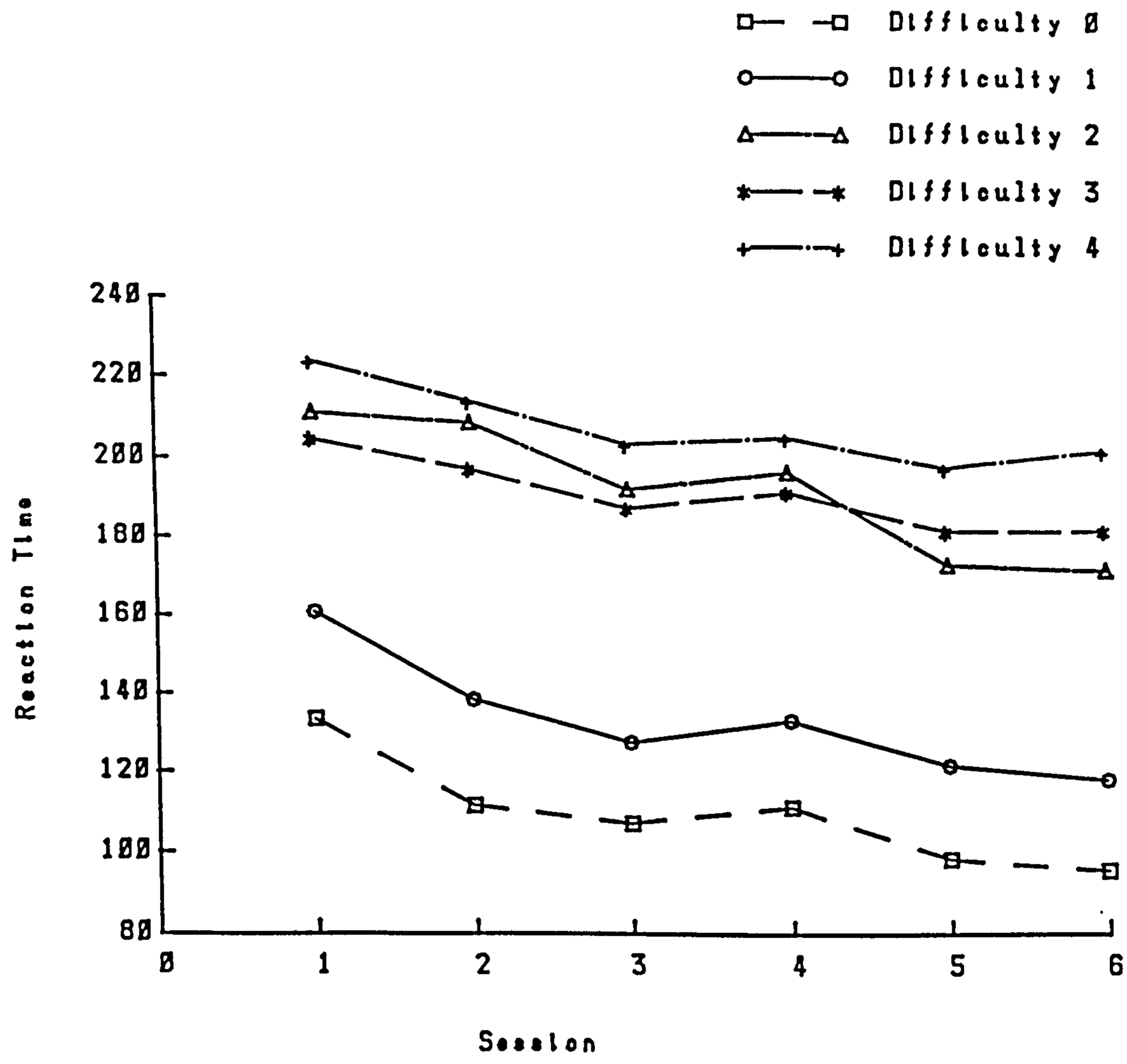


Figure 6.17 Mean RT of the current response as a function of practice and level of stimulus difficulty

were examined by an analysis of variance on the mean residual values of RT for each trial. (This of course meant that all between subjects variance had been removed.) Again, it was found that, in general, speed decreased with stimulus difficulty [$F(4,19140)=2253.80$, $p<.0001$] and increased with practice sessions [$F(5,19140)=160.71$, $p<.0001$]. The combined effect of "stimulus difficulty" and practice "sessions" was significant [$F(20,19140)=17.27$, $p<.0001$] but not systematic.

Error responses were again slower than correct responses [$F(1,19140)=59.42$, $p<.0001$]. This effect was independent of the level of practice ($p>.25$) but it varied as a function of "stimulus difficulty" [$F(4,19140)=6.92$, $p<.0001$]. The difference between the residual RTs of error and correct responses seemed more pronounced for the first two levels of stimulus difficulty (i.e. "easy" stimuli) (figure 6.18).

The Effect of Accuracy of the Preceding Response on the Speed of the Current Response

Similar to the data analysis of the other training groups, an analysis of variance was carried out to examine, for each trial, the effect of accuracy of the preceding response on the speed of the current response. This analysis of variance examined the effects of "accuracy of the preceding response", "accuracy of the current response" and "level of stimulus difficulty" on the residuals of RT.

When the preceding response was an error, the speed of the current response was significantly reduced [$F(1,18940)=176.47$, $p<.0001$]. This effect was independent of

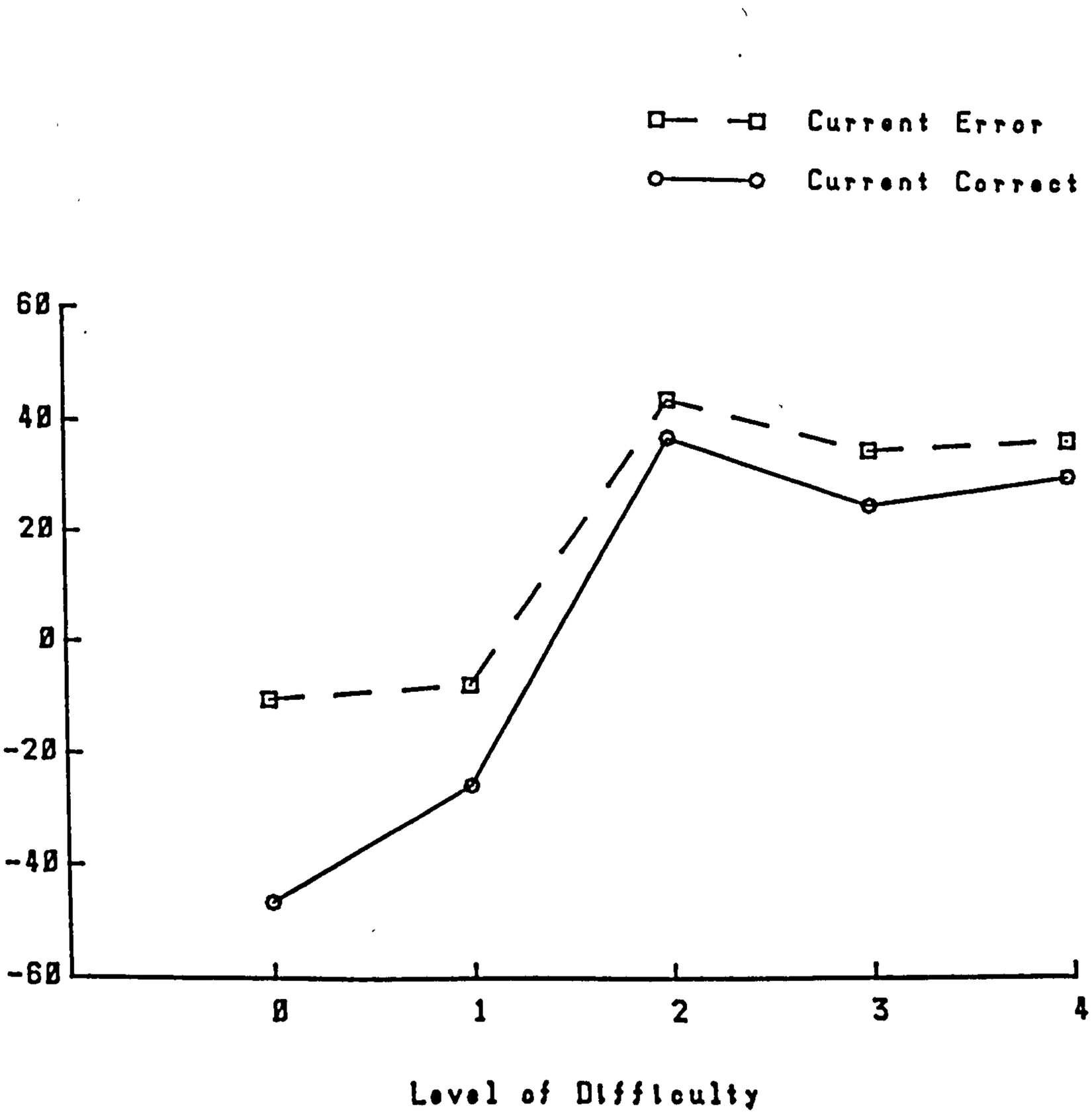


Figure 6.18 Mean residual RT of the current response as a function of the level of stimulus difficulty and accuracy of the current response

"stimulus difficulty" ($p>.1$) but varied according to the accuracy of the current response [$F(1,18940)=4.48$, $p<.05$]. (This latter interaction is going to be discussed further later on as a susequent analysis of variance appeared to give a contradictory result.) Finally the interaction of the variables "accuracy of the preceding response" and "accuracy of the current response" varied as a function of the level of stimulus difficulty as it was shown by the significant interaction among the three variables [$F(4,18940)=3.20$, $p<.01$]. The effect of the 3-way interaction however was not systematic (figure 6.19).

A third analysis of variance was carried out to examine the combined effects of "accuracy of the preceding response", "accuracy of the current response" and practice "session" on the residuals of RT. The results of this analysis showed that the effect of accuracy of the preceding response varied as a function of practice sessions [$F(5,18936)=3.19$, $p<.01$] but this effect was not systematic (figure 6.20).

This analysis of variance showed that the interaction of the variables "accuracy of the preceding response" and "accuracy of the current response" was not significant which is exactly the opposite of what was shown from the previous analysis of variance. This discrepancy can be explained by the fact that a least-squares method was used to carry out the analysis (because of the unequal cell frequencies - see above). In such analyses, discrepancies of this kind are not unusual as the inclusion (or exclusion) of one variable may mask the variability of the dependent variable. In this

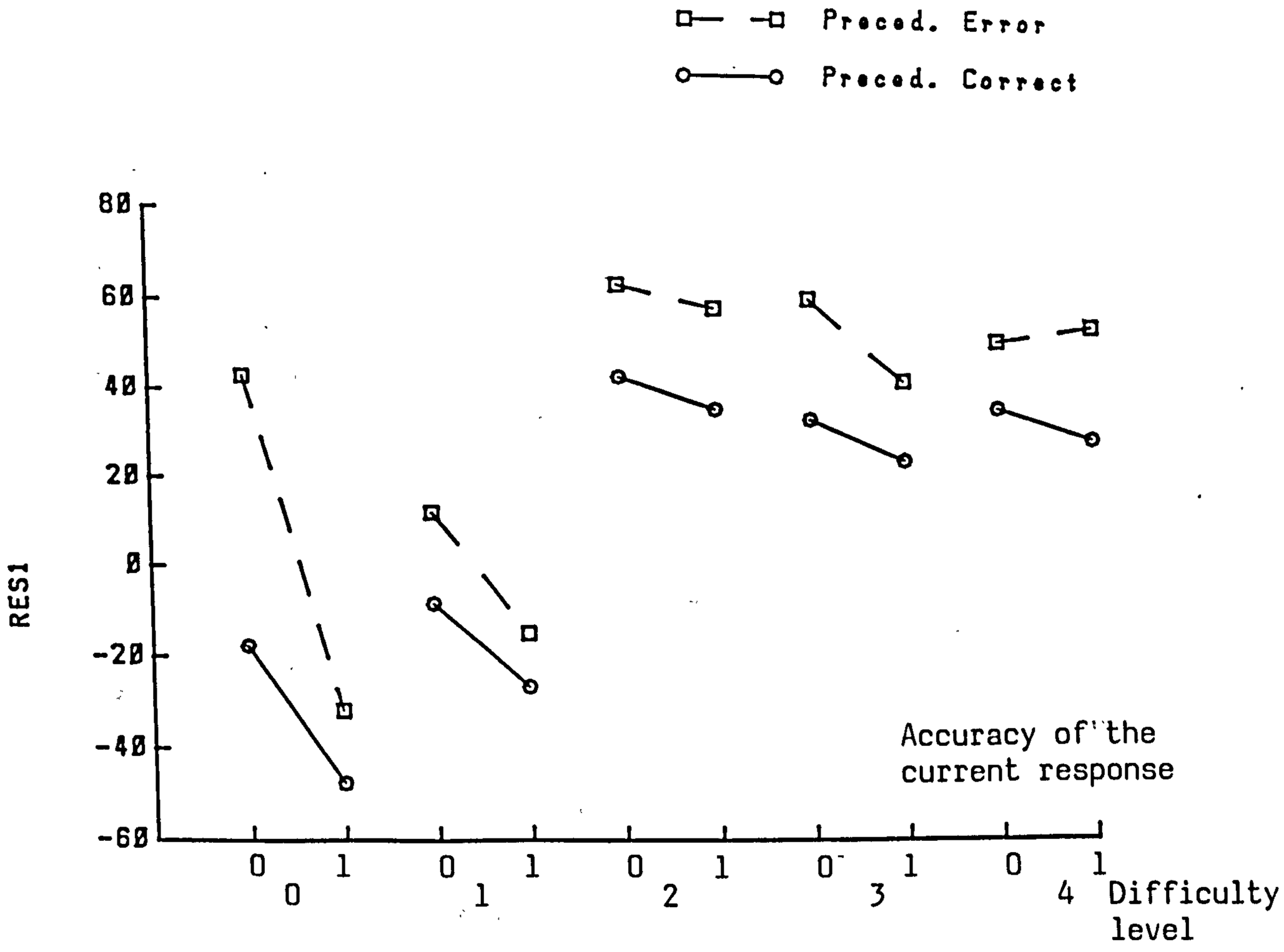


Figure 6.19 Mean residual RT of the current response as a function of stimulus difficulty and accuracy of the current and the preceding response

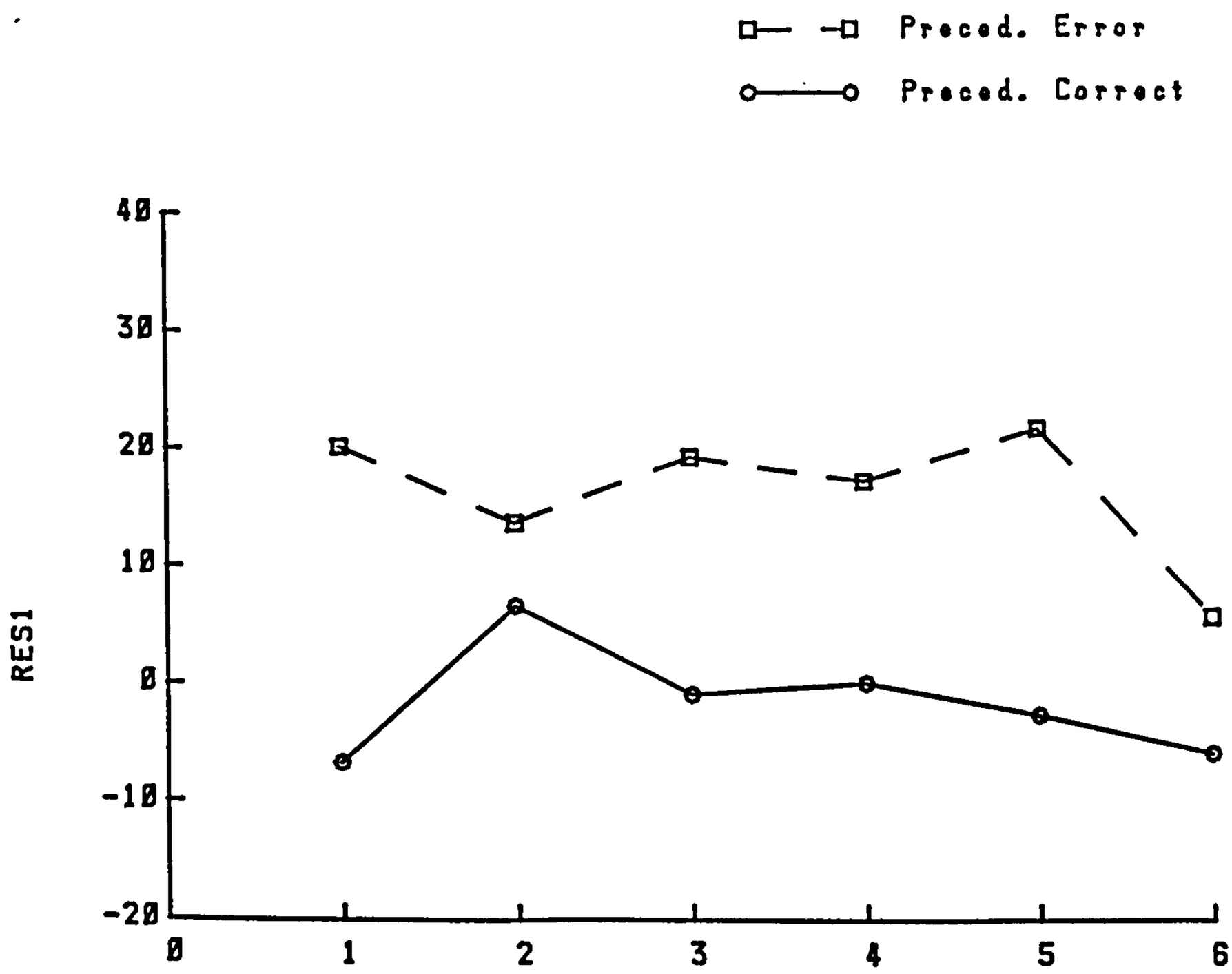


Figure 6.20 Mean residual RT of the current response as a function of practice and accuracy of the preceding response

particular case of the interaction between the variables "accuracy of the preceding response" and "accuracy of the current response", one would be more inclined to believe that the significance of the interaction is rather small and possibly insubstantial.

Factors Predicting Speed

As was done for all the other groups, a multiple regression was carried out to examine which factors best explained the variance in speed. The dependent variable was the residual of the current response. The predictor variables were the same as those used in all previous multiple regressions. Namely, they were: "SESSION", "BLOCK", "TRIAL", difficulty level of the current trial (DIFF1), difficulty level of the preceding trial (DIFF0), speed of the preceding response (RES0), and accuracy of the preceding response (ACC0).

It was found that 25 per cent of the variance of speed could be explained by the difficulty level of the current trial, the practice "session" and the accuracy of the preceding response. The obtained multiple correlation coefficient was $R=.50$ [$F(3,18956)=2060.19$, $p<.0001$] and the prediction equation was

Residual RT (csecs)

$$= 24.91(DIFF1)-5.97(SESSION)-17.20(ACC0)-4.43$$

(Mean Residual RT = 0)

All the other predictor variables entered the equation at statistically significant levels but their contribution to the

total variance was less than one per cent.

Accuracy Changes with Practice and Difficulty Level

A repeated measures analysis of variance was carried out on subjects' mean proportion of correct responses to examine how accuracy varied with changes in the level of practice and the level of stimulus difficulty. It was found that subjects' proportion of errors decreased with practice [$F(5,45)=8.60$, $p<.001$] and increased with increased stimulus difficulty [$F(4,36)=16.03$, $p<.001$]. These two effects were largely independent of each other (i.e. the interaction of the two variables was not statistically significant).

The Effect of Accuracy of the Preceding Response on the Accuracy of the Current Response

As for all the previous groups, the effect of the accuracy of preceding response on the accuracy of the current response was examined. All responses were crosstabulated in a two-by-two matrix, the variables being "accuracy of the preceding response" and "accuracy of the current response". As before, this crosstabulation produced four categories of responses: errors preceded by errors, errors preceded by correct responses, correct responses preceded by errors and correct responses preceded by correct responses (table 6.8).

The null hypothesis was that the two variables were independent and this hypothesis was tested by using the chi-square test and the phi coefficient. The result of the chi-square test was not significant ($p>.25$). This together with

the very small phi value (phi = 0.00865) meant that the two variables were independent.

	Previous Error	Previous Correct
Current Error	Obs. 105 Exp. 116.9 0.6%	1,383 1,371.1 7.3%
Current Correct	1,384 1,372.1 7.3%	16,088 16,100 84.9%

x^2 = 1.30, df: 1, not significant
Phi = 0.0087

Table 6.8 Errors on current stimuli as a function of accuracy on the preceding response

A second set of crosstabulations was made with the same two variables for each level of stimulus difficulty. The same two statistical tests were run and the same result was obtained. Generally, the two variables were independent irrespective of the level of difficulty (maximum phi value was 0.0361, appendix 4).

Finally a third set of crosstabulations of the same two variables was made but this time controlling for practice sessions. Again the same two statistical tests were carried out and the same results were obtained (appendix 4). The two variables were independent across all levels of practice.

Therefore, it can be quite safely concluded that the accuracy of the current response was not affected by the

accuracy of the preceding response. In other words, the occurence of an error did not alter the probability of a correct response in the next trial.

Factors Predicting Accuracy

A stepwise multiple regression was carried out to examine the factors which accounted for the variance in the accuracy of the current response. The dependent variable was accuracy of the current response, and as in all previous regression analyses, the predictor variables were "SESSION", "BLOCK", "TRIAL", difficulty level of the current trial (DIFF1), difficulty level of the preceding trial (DIFFO), speed of the preceding response (RESO) and accuracy of the preceding response (ACCO).

Although the maximum variance which could be explained was 30% (computed by using Cohen's formula: see earlier) only 1.1% was actually accounted for by the regression analysis. The variables which explained - poorly - this variance were the difficulty level of current trial, the practice session and the difficulty level of the preceding trial. The multiple correlation coefficient was $R=0.1072$ [$F(3,18956)=73.46$, $p<.0001$]. The prediction equation was

Predicted ACC1

$$= 0.92-0.02(DIFF1)+0.0008(SESSION)+0.0007(DIFFO)$$

The variables "block" and speed of the preceding response entered the prediction equation at statistically significant levels but their contribution to the total variance was less than 0.5%. The variables "trial" and accuracy of the preceding

response did not even enter the equation.

It can be concluded therefore that the variables "difficulty level of the current trial", practice "session" and "difficulty level of the preceding response" explained 1.1% of the variance in accuracy whereas the largest part of the variance remained unexplained.

Summary of Speed and Accuracy Variations

In summary, the main speed and accuracy changes of the group who received training with the CA method were the following.

(1) Both speed and accuracy increased with practice and decreased with increased difficulty.

(2) Speed changes, because of practice, varied as a function of stimulus difficulty.

(3) Error responses were slower than correct responses. This effect varied as a function of stimulus difficulty but it was independent of the level of practice.

(4) If the preceding response was an error, the speed of the current response slowed down. This effect was independent of the the accuracy of the current response and also idependent of the level of difficulty level of the current trial. However, it varied as a function of practice.

(5) Accuracy on a given trial was independent from the accuracy on the preceding trial across all levels of stimulus difficulty and practice.

(6) 25% of the variance in speed was explained by the variables difficulty level of the current trial, practice

session and accuracy of the preceding response.

(7) Finally the variance in the accuracy of the current response was very poorly explained.

6.6 SUMMARY OF SPEED-ACCURACY VARIATIONS ACROSS TRAINING MODES

Speed and accuracy changes for each one of the four training groups have already been examined in detail. The main findings across all four groups were the following.

The Effect of Practice and Stimulus Difficulty on Speed

Although the speed of performance increased significantly with practice, the rate of speed increase was not the same for all four groups. It varied as a function of the particular training method used by each group. The groups which were trained by cumulative methods (either adaptive, CA, or random, CR) were initially faster and showed less subsequent improvement in speed than the other two groups (figure 6.21). (These results have been discussed in Chapter 2.)

Speed decreased with an increase in the level of stimulus difficulty for all four groups. It seems that the rate of speed decrease with increases in stimulus difficulty was the same for all groups irrespective of the method by which they were trained (figure 6.22).

For all groups, there was a significant interaction between level of stimulus of difficulty and level of practice. In other words, for all groups, the decrease in speed because of stimulus difficulty varied as a function of the level of

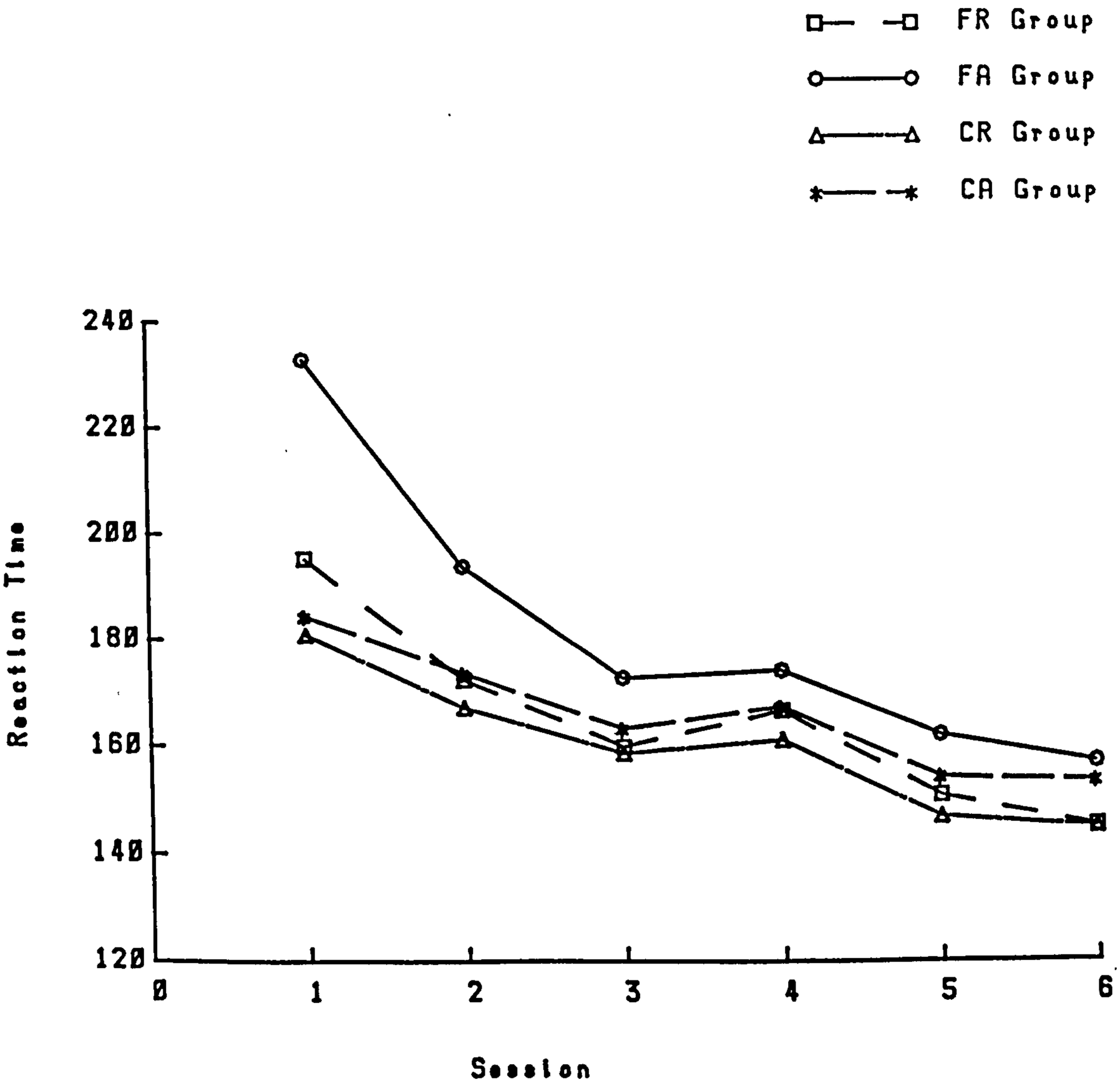


Figure 6.21 Mean RT of the current response of each training group as a function of practice session

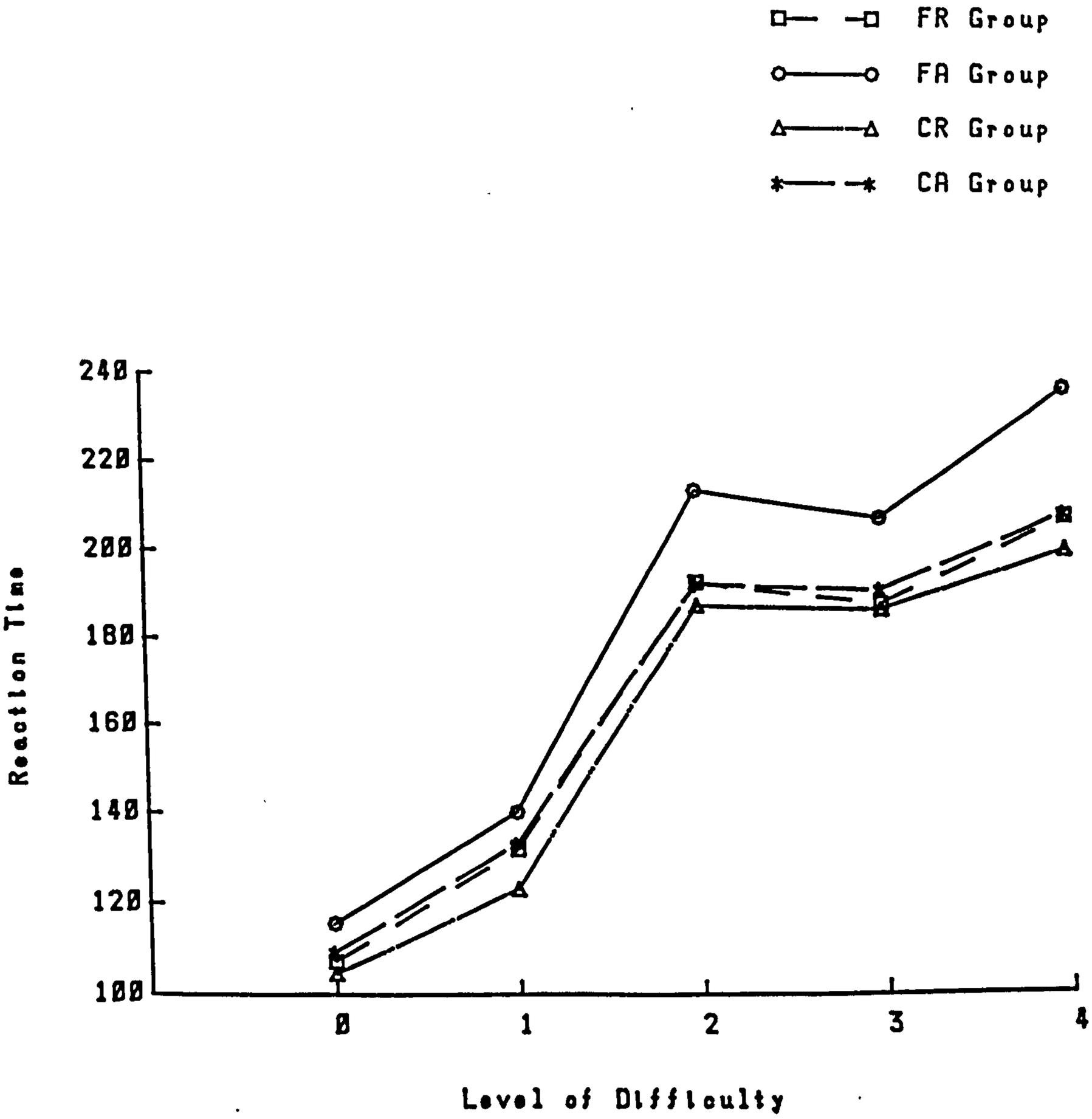


Figure 6.22 Mean RT of the current response of each training group as a function of stimulus difficulty

practice. These interactions (figures 6.1, 6.6, 6.13, 6.17) varied slightly across groups in that the curves for the cumulative modes were flatter than those for the random ones.

The Effect of Accuracy of the Current Response on Speed

Error responses were generally slower than correct responses across all four groups. The evidence for this finding was quite conclusive.

The difference in the speed of correct and error responses varied as a function of the level of stimulus difficulty. This significant effect was clear across all four groups but it did not appear very systematic, at least for the noncumulative modes of training. For all but the fixed-random (FR) mode, the difference between the residuals of RT for errors and correct responses was more pronounced for "easy" stimuli. However, this observation cannot be supported by conclusive statistical evidence due to the big standard errors of the means of the RT residuals for errors.

The effect of accuracy of the current response on speed varied as a function of the level of practice for the FA and CR modes of training. This effect however was not systematic. As mentioned above, these two-way interactions were very difficult to interpret given the large standard errors of the mean residual RTs for errors.

The Effect of Accuracy of the Preceding Response on Speed

When the preceding response was an error the speed of the current response was slower. This was consistent across all four groups.

The effect of "accuracy of the preceding response" varied for the noncumulative modes as a function of the level of difficulty. The difference in the residuals of RT for correct and error responses was greater for "difficult" stimuli than it was for "easy" stimuli.

The effect of "accuracy of the preceding response" also varied as a function of practice for the adaptive modes. This finding was expected, since the order of presentation of stimuli for the adaptive groups was partly a function of the accuracy of the preceding response.

Accuracy Changes with Practice and Stimulus Difficulty

As for speed, accuracy improved with practice for all four groups. The functions relating accuracy and practice did not differ markedly among the four groups. Accuracy decreased with increases in stimulus difficulty with the rate of increase being very similar for all four groups. The decrease in accuracy with practice varied significantly as a function of the level of difficulty for the non-adaptive groups. For the adaptive groups the change in accuracy with practice was independent of the level of stimulus difficulty. This result was not unexpected since "difficult" stimuli were presented more often than "easy" ones and therefore the effect of stimulus difficulty became non-significant as subjects received more practice.

The Effect of Accuracy of the Preceding Response on the Accuracy of the Current Response

While the accuracy of the preceding response affected the speed of the current response, the probability of a given response being correct was not greatly affected by the accuracy of the immediately preceding response. The evidence for this finding was quite conclusive for the adaptive groups, across all levels of stimulus difficulty and practice. The analyses of the data for the FR and CR training modes (non-adaptive modes) revealed an overall very low, but statistically significant, level of association between the accuracy of the preceding response and the accuracy of the current response. However, the values of the phi-coefficients were so small that this level of association could not be considered very substantial in real terms. The occurence of an error increased the probability of another error on the subsequent trial from 0.11 to 0.14.

It was concluded that, practically, accuracy of the current response was independent of the accuracy of the preceding response.

Prediction of Speed and Accuracy

For all but the CA group, the factors which explained around 34% of the variance in speed were the difficulty level of the current trial, the practice session and the speed of the preceding response. For the CA group 25% of the variance in speed was explained by the difficulty level of the current trial, the practice session and the accuracy of the preceding

response.

It was not possible to explain the variance in accuracy. The "best" prediction was made for the FR group but even this accounted for less than 3% of the total variance. Therefore it was concluded that none of the independent variables could account for any substantial amount of variance in accuracy. In short, speed of performance was predictable but accuracy was not.

Chapter 7

DISCUSSION OF MAIN FINDINGS

7.1 Speed of Correct and Error Responses

Error responses were consistently slower than correct responses across all modes of training. This result was very interesting in that it implied that subjects did not make errors because they traded accuracy for speed. In most CRT experiments it has been found that error responses are faster than correct responses. In other such experiments, however, it has been found that errors and correct responses have the same speed and in yet others that errors have slower response times than correct responses. The results of many such experiments have been published (e.g. Laming, 1968; Pickett, 1967; Pike, 1971; Emmerich, Gray, Watson & Tanis, 1972, etc.). Examples of such conflicting findings will be discusssed below and to make the matter more interesting all of them will be findings of the same experimenter and his associates.

Rabbitt (1966a) examined the latencies of responses preceding errors, of errors, of error-correction responses, and of responses following error-correction in a 4-choice and a 10-choice continuous performance choice-response task. The stimuli were lights set in a horizontal row. The responses involved the depression of one of the contact-grids mounted in a horizontal row below the signal lights on the same console. No one stimulus was presented on two consecutive trials. When subjects responded to a stimulus incorrectly, the stimulus did not change until the correct response to it was made.

Responses were classified as follows: Errors - when the correct contact-grid was not pressed; Error-Correcting Responses - the responses that followed an error when the same light remained switched on; Responses Following Error Correction - the first six responses following each Error Correcting Response (if any of these six responses was an error it was classified so); Correct Responses - all correct responses which were not Error-Correcting Responses or Responses Following Error Correction. Subjects always responded with the forefinger of their dominant hand and were instructed to respond as fast as possible. He found that on both the 10-choice task and the 4-choice task, Errors and Error-Correcting Responses were significantly faster than Correct Responses.

In another experiment, Rabbit and Vyas (1970) describe an experiment in which ten subjects had to respond to two repetitive sequences of letters. The first sequence ("2-cycle") required a repetitive sequence of taps on two adjacent grids, ABABABAB...etc. The second sequence ("4-cycle") required the repetitive sequence of four taps on three grids, ABCBABCBABCB...etc. They found that RTs for errors were significantly slower than RTs for correct responses for both sequences.

In yet another experiment, Rabbitt and Vyas (1970) asked a group of 10 elderly subjects and a group of 10 young subject to respond by saying "left" or "right" to identify the larger of two black circles which were briefly presented on either

side of a fixed point against a white background. They used ten different conditions of stimulus discriminability which produced a range of 1.3% to 48% errors. They found that when discriminations were easiest RTs for errors were slightly faster than RTs for correct responses for both groups. As discriminations became increasingly more difficult RTs for errors and RTs for correct responses showed no reliable difference for the young group whereas RTs for errors were slower than RTs for correct responses for the old group.

Such conflicting results can be found throughout the literature on speed-accuracy relationships. The validity of the results obtained by most of these experiments cannot and need not be argued. The main problem lies with the question of whether errors are faster than correct responses. No single yes or no answer can be given, as there is plenty of experimental evidence for either answer.

The present author has been unable to find any theoretical model which is able to account adequately for relative variations in the speed of errors and correct responses across experimental tasks. Most of the theoretical models predict either a "yes" or a "no" answer and then, in the best case, their authors try to explain "logically" a different result obtained from other experimental data or they claim that their model does not apply in that particular task. Two examples of such theoretical models are the following.

Rabbitt and Vyas (1970) discuss an early version of Rabbitt's (1981) "tracking model" which predicts that errors are caused because subjects respond too fast, which means that

errors are faster than correct responses. They attempt a taxonomy of errors, distinguishing between errors of perceptual analysis and errors of response selection and execution. In the experiment with repetitive sequences of letters, which has been described above, they found that errors are slower than correct responses. Their explanation was what they called "common-sense" (and what somebody else might call "speculative") and they admit it could not be "either substantiated or questioned by present evidence". They said that "subjects made errors only (or mainly) when they forgot the particular point in the sequence which they had reached. At such a crisis a S might pause and then respond at random. If he were accidentally correct he would score an unusually long RT, which would be collapsed with a large number of relatively fast correct RTs to derive the mean presented. If the S were accidentally wrong his error RT would be slow, and would be collapsed with a much smaller number of error RTs to give the means provided. Mean RT would thus tend to be slower than mean correct RT" (Rabbitt and Vyas, 1970, p. 65).

Laming (1968) discusses the Random Walk Model which predicts that errors are generally faster than correct responses. He refers to Pickett's (1967) findings - which are the opposite - by suggesting that subjects "appear to have striven for maximum accuracy without consideration of the time taken" (Laming, 1968, p. 29). He then added that such experiments "are not within the field of application of the

choice-reaction model" he presents. This argument however is not justified. The error rate reached as high as 20% and the lowest maximum and minimum latencies were recorded in the condition that produced the lowest error rate. Laming's argument has been further invalidated by Vickers (1970) who claims that enquiries to Pickett himself revealed that "it was made 'abundantly clear to subjects that response latencies were being recorded', the instructions being to respond 'as quickly and as accurately as possible'" (Vickers, 1970, p. 53).

The situation therefore is summarized as follows. The results of Experiment 4 were quite conclusive that errors were slower than correct responses. Examining the relevant literature, one can cite experiments which have found that error RTs are faster than correct RTs, or that error RTs are slower than correct RTs, or even that there is no reliable difference between error RT and correct RT. Theoretical models which consider speed-accuracy relationships are generally based on one kind of experimental task and predict one of the three alternative relationships between the speed of error and correct responses. However the issue remains open as long as there is not one single theoretical model which can account for all three possible relationships.

7.2 The Effect of Accuracy of the Preceding Response on the Speed and Accuracy of the Current Response

Another main finding of the data analysis was that when the preceding response was an error, the speed of the current

response was slower. [This varied as a function of stimulus difficulty for the non-cumulative groups. When the preceding response was an error and the current stimulus was difficult then the speed was slowest. It also varied as a function of practice for the adaptive modes of training but the way it varied was not systematic.] Moreover, it was found that for the non-adaptive modes of training, immediately after an error response there were significantly more errors than would have been expected by chance.

Relatively few researchers have examined subjects' performance immediately after the occurence of an error. Laming (1968, 1979) discusses changes in the probability of an error on trials following errors by refering to three of his experiments. The experiments employed two stimuli which were vertical white stripes and subjects had to respond by pressing one of two keys. He found that after making one error subjects were less likely than usual to make another error for the next five trials. Moreover, he found that RTs to signals following an error were generally increased. If the signal following an error was not a repetition of the signal which had triggered an error, RT increased quite substantially for the next three trials. If the new signal was a repetition the increase in RT was smaller and only appeared on just one trial following the error. Laming (1968, pp. 113-114) explains these results on the basis of the following four hypotheses he makes for a 2-choice experiment with inter-trial intervals of 1500 msecs or more:

(1) Errors are mainly due to the irrelevant information which

the subject inevitably samples, in a state of temporal uncertainty, before the signal is presented.

(2) After the occurence of an error the subject delays sampling of information for the next signal, frequently until after that signal has been presented. The possibility of another error is reduced because "the subject now samples much less irrelevant information, and frequently none at all" (Laming, 1968, p. 114). If the subject does not begin sampling until after the signal has been presented, it follows that there will also be a corresponding increase in reaction time.

(3) The boundary at which the erroneous decision was made is readjusted since the subject makes the accuracy criterion stricter. Laming says that although this reduces the probability of making the same response again, this reduction cannot be detected by itself because of (1) above. But the readjustment of the boundary - i.e. the stricter accuracy criterion - causes an increase in reaction time to the trial following an error if the alternative signal is presented.

(4) After a correct response, small contrary adjustments are made again to the limits of the boundaries to the time at which the sampling of information for the next decision begins. Laming (1968) claims that these parameters are thus determined by a stochastic process.

Laming's (1968, 1979) suggestion of a reduced probability of an error immediately after an error has been made does not

account for the results of the present experiment which suggest that the probability of an error either remains the same or is increased. However, Laming's hypotheses do account for the slowing of speed immediately after an error.

Unlike Laming, Rabbitt and Rodgers (1977) claim that "errors ... immediately follow one-another significantly more often than we would expect by chance" (p. 730). They ran an experiment with 2-choice, 4-choice and 8-choice reaction tasks. They found that responses following an error were slow and also inaccurate. They tried to explain these results by examining first, the nature of the errors and second, the speed of errors.

They found that a significant number of errors immediately following an error were "involuntary" Error Correction Responses (ECRs) to the first error. In other words, when subjects made an error they involuntarily tried to correct it - although they had been explicitly told not to do so - producing therefore a second error, if that signal was not a repetition of the signal which had triggered the first error. This finding was apparent in all three tasks and raised the question of whether the occurence of double errors could be entirely attributed to inappropriate ECRs or there was some other factor as well. Their analysis revealed that non-ECR double errors did not occur by chance. This finding was important as, according to Rabbitt and Rodgers (1977), it related to the theoretical issue of whether subjects can ever attend to the E+1 stimulus before responding to it. They analysed E+1 responses (responses immediately following an

error E) on the basis of whether E+1 stimuli were repetitions or alterations of the E stimuli. They found that the probability that a subject would make an involuntary ECR after an error was affected by the nature of the E+1 stimulus.

They concluded that after an error subjects were more likely to make an error than would be expected by chance. The accuracy of an E+1 response is dependent on the nature of E+1 stimulus. Responses to signal repetitions were accurate because "signal repetitions were quickly and accurately recognised on E+1 trials" (Rabbitt and Rodgers, 1977, p. 733).

Rabbitt and Rodgers (1977) also examined changes in the speed of responses immediately following an error. They classified E+1 responses in three categories.

(1) "Signal Repetitions" when the same signal of E trial was repeated on the E+1 trial.

(2) "Response Repetitions" when the response to the new signal was the same (incorrect) response which had been made on the E trial.

(3) "New Signals" when the E+1 signal was neither a repetition of the E signal nor when it required a repetition of the previous incorrect response (obviously the E+1 responses from the 2-choice task could not be classified in this third category, as there were only two stimuli).

They found that RTs of "Signal Repetitions" E+1 trials are faster than RTs of "Response Repetitions" E+1 trials which, in turn, are faster than RTs of "New Signals" E+1 trials. To account for these results Rabbitt and Rodgers

(1977) adopt the position of Rabbitt (1969) who had suggested that after subjects recognise an error they respond more slowly and more cautiously to the next signal which explains the slowing of "Response Repetitions" and "New Signals" E+1 RTs. Since it does not explain the difference in the size of the effect between the RTs of these responses and the E+1 RTs of "Signal Repetitions", they introduce a new possibility. They claim that when the E+1 signal is either a "Response Repetition" or a "New Signal" subjects have to suppress the tendency to make an ECR which produces an additional delay to the one due to the extra cautiousness after the occurence of an error.

Rabbitt and Rodgers (1977) therefore account for an increase in reaction time as well as in the probability of an error on trials immediately following an error. The suggestion of an increase in RT accounts for their data as well as for Laming's (1968) data and for the data of the present research. However, the suggestion of an increase in the probability of an error on trials following an error does not account for Laming's (1968) data and does not fully explain the present data either. To clarify this last point a little more, let us recapitulate on the explanations given by Laming (1968, 1979) and Rabbitt and Rodgers (1977).

Rabbitt and Rodgers (1977) claim that after an error subjects try to correct their error and as a consequence their ECR is an erroneous E+1 response. Therefore, the probability of a double error is increased. Although they did find that non-ECR double errors did not occur by chance and did

explicitly say that "not all second or third errors are ECRs, and we must look for other factors which may lead to E+1 response inaccuracy" (1977, p. 738), they did not explain the factor(s) which cause(s) an increased probability of error after the occurence of an error.

Laming (1968) claims that errors are caused because of the irrelevant information which sujects sample. After an error, he claims, subjects delay sampling information about the next signal "frequently until after [the] signal is presented" (Laming, 1968, p. 114). Therefore, he concludes, E+1 responses are slower but more accurate. Laming (1979) comments on Rabbitt's work specifically and says that Rabbitt and Rodgers (1977) have replicated HIS results at short Response-Stimulus intervals (RSI). Laming (1979), based on one of his own experiments (1968), claims that "the RSI is a critical parameter for all aspects of CR performance" (1979, p. 206). In that experiment he used five different RSIs (1, 8, 64, 512, and 4096 msecs) and found different patterns of performance on a 2-choice task. He found that when the RSI was 512 or 4096 msecs, the probability of an error decreased and RT increased on the trial immediately following an error. However, when the RSI was 1, 8 or 64 msecs, the increase in the RT of E+1 trial was much greater and the probability of an error increased. Laming therefore atributes the discrepancy between his findings and the findings of Rabbitt and Rodgers (1977) to the difference in RSI. However, this kind of explanation does not account for the results obtained by

Experiment 4 which had used RSIs of at least 1000 msecs. In fact it was even more than that after an error, because visual and auditory feedback were given to the subject.

This suggests that if the RSI is "a critical parameter", it is not the only one which determines the probability of an error immediately after an error. Furthermore, the fact that the probability of a double error was increased for two of the training modes whereas it was not altered for the other two (although the RSI was the same for all four modes) suggests that there must be more such factors.

The present results showed that for the groups who had received training with an adaptive mode, the occurence of an error on a given trial did not affect the probability of an error on the following trial. For these two groups, the training programme itself controlled for the difficulty of individual stimuli presented to subjects, so that, in general, subjects did not have to deal with signals more difficult than they were ready for. For the non-adaptive groups however, the results showed a small but nevertheless statistically significant increase in the probability of an error immediately after an error. For these two training groups there was no control over the individual signals which were presented to subjects. This meant that, at any time, subjects had to cope with stimuli of variable difficulty.

The fact that the difference between the two pairs of groups was only the sequence of stimulus presentation, and therefore the stimulus difficulty, may suggest that the probability of a double error might also be affected by the

level of stimulus difficulty. Unfortunately the present data cannot provide conclusive evidence about which factors affect the probability of an error after an error response.

7.3 Prediction of Speed and Accuracy

In the training sessions on the 10-key chord keyboard, the variables which account for most of the variance in speed are stimulus difficulty, level of practice, speed of the preceding response and accuracy of the preceding response. The total variance in speed (expressed by the residual value of RT) which could be accounted for by these variables ranged from 25% for the CA group to 36% for the FR group. This, therefore, provided a reasonably good prediction of speed. The variables which predicted this speed measure were expected to do so, especially the variables practice and stimulus difficulty. Practice is a major variable which accounts for the variance in RT in practically all CRT tasks (Rabbitt, 1981). The level of difficulty which was used as an independent variable in the multiple regression analysis was a complex variable in that there were a number of features which varied between the five levels of stimulus difficulty (e.g. number of hands required for the response, handedness, finger separation, etc.). The variations in these features had already been found (from Experiments 1 and 3) to account for the variance in RT.

Although is was possible to predict the speed of a response relatively well, it was not possible to predict whether a given response would be accurate or inaccurate. The

various statistical analyses which have been described above, have shown that there are various factors which affect accuracy, i.e. there was a positive relationship between the proportion of errors and practice and an inverse relationship between the proportion of errors and the level of stimulus difficulty. Furthermore, accuracy can often be predicted in such experiments and it had indeed been predicted reasonably well for Experiments 1 and 3. However, in those cases, the accuracy measure was the proportion of errors. This is a very different measure from the one used in the present analyses. In these analyses the measure of accuracy was whether a given response was correct or not. Therefore, it must be made clear that it is this specific measure of accuracy which is difficult to predict. In other words, while one can predict error rates associated with various conditions, one cannot predict whether a particular response will be right or wrong.

Chapter 8

SPEED-ACCURACY TRADEOFF THEORIES

As mentioned above, most CRT models have various serious limitations which make them rather inappropriate to be considered in relation to the present data.

8.1 Required Characteristics of a Speed-Accuracy Model

A sound theoretical model that could be used to account for the present experimental data should have the following basic characteristics.

(1) It should consider the subject as an active component of the person-task system. In other words, it should acknowledge that subjects have the ability to adjust their performance according to the demands and limitations of the task and therefore actively affect the system.

(2) Instead of focusing all attention on each transaction of signal identification and response production as an independent process, it should assert that subjects have the ability to exercise overall supervisory control and therefore to anticipate events and maximize sensitivity of signal detection which eventually would improve overall performance efficiency. It should also consider that subjects may already have relevent information in long-term memory which allows them to recognize stimuli easily and provides them with the appropriate motor programs to produce the correct response.

(3) It should actively consider accuracy of performance. The

existence of errors in subjects' performance should not be conceived as coincidental to overall performance. Instead accuracy should be incorporated in the model as a component of performance equal in importance to speed.

(4) It should consider the psychophysical features of stimuli (e.g. discriminability, complexity, etc). This is an important issue especially when the model is used to assess experimental results from experiments with different tasks.

(5) It should discuss variations of performance from trial to trial either between or within subjects and tasks. It should consider not only how average performance improves over time but also how and to what extent the system alters the maximum capabilities over time.

While most theoretical models do not comply with all the above characteristics they do deal with, at least, certain aspects of speed-accuracy relationships. The most important of these models are discussed below.

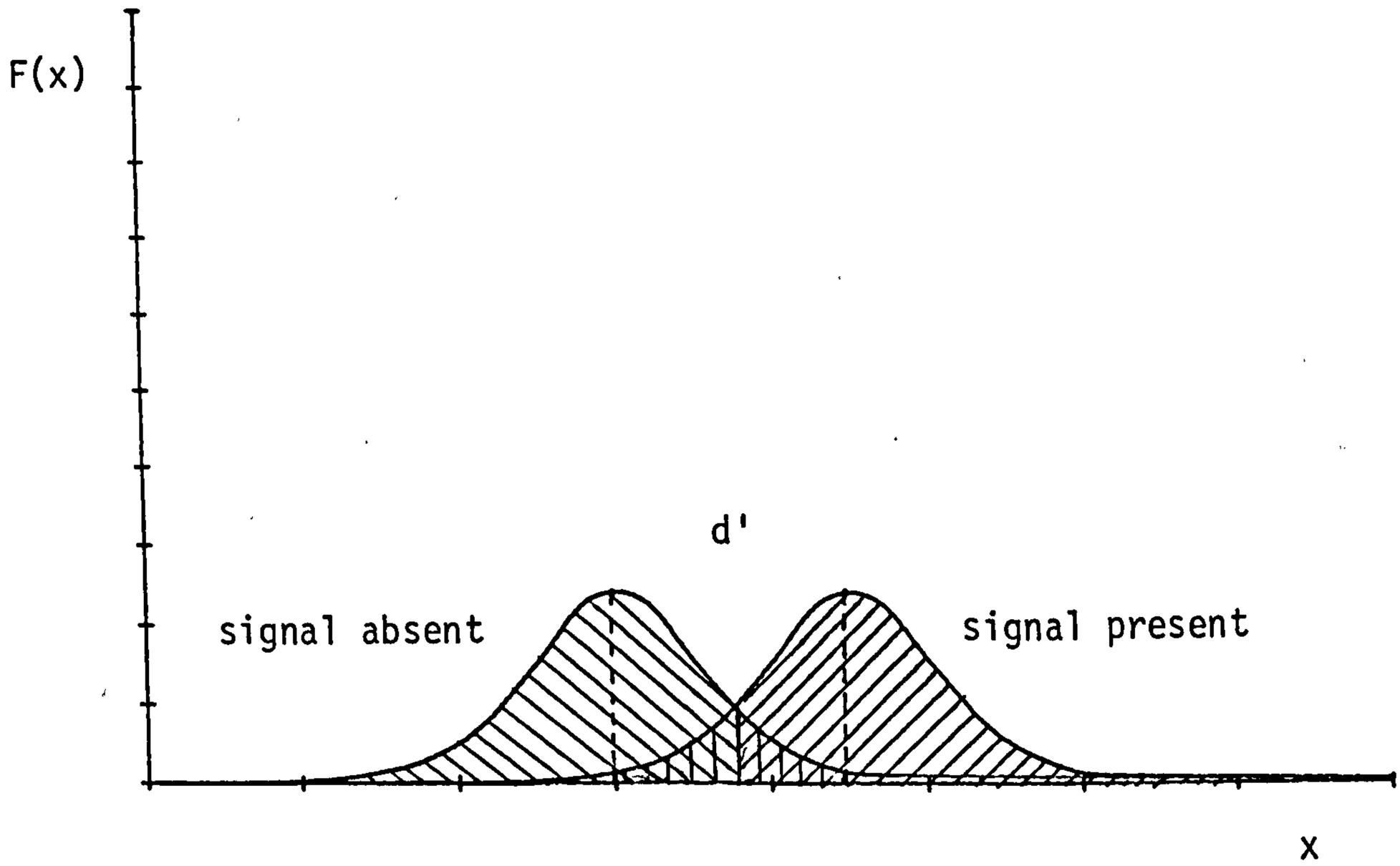
8.2 Signal Detection Theory

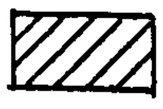
Signal Detection Theory (SDT) has been mainly established by the work of Tanner, Swets and their associates who worked on the detection of faint auditory and faint visual stimuli against a "noisy" background (Tanner and Swets, 1954; Swets, 1959, 1964; Swets, Tanner and Birdsall, 1961; Green and Swets, 1966). The theory incorporates into one model many diverse factors which affect the detection and recognition of signals embedded in noise. It emphasises the psychophysical features

of the task in the process of decision-making, in which the subjects' main job is to establish a criterion for responding appropriately in a continuously changing environment.

The basic situation which SDT considers is one in which there is a signal and background noise or there is background noise only. The subject has to decide whether a signal is present or absent. The response depends not only on the nature of the signal but also on subjects' decision process.

Swets et al. (1961, p. 303) use the term "observation" to refer to "the sensory datum on which the decision is based". They assume that this observation, x, varies continuously along a single dimension or axis. They further assume that any observation may arise, with specific probabilities, either from just noise or from signal plus noise. The fluctuations in noise and observations are represented in the form of probability distributions (figure 8.1). [For the purposes of this discussion it is assumed that the two probability distributions are normal and have equal variances. However, Green and Swets (1966) do discuss more complicated cases where the two distributions are exponential instead of normal or where they have unequal variances.] Given the inevitable existence of noise, the signal is not strong enough for the two probability distributions to be completely apart; there will always be some overlap. Moreover the mean of the distribution of signal plus noise will always be on the right of the mean of the noise distribution since the observations will tend to be of greater magnitude when a signal is




proportion of hits


proportion of false alarms


proportion of misses


proportion of correct negatives

Figure 8.1 Graphical Representation of Signal Detection Theory

present. The distance between the means of the two distributions is conventionally denoted as d' and is a measure of discriminability, or detectability, of the signal. The greater the strength of the signal and the smaller the level of noise, the larger the value of d'.

The theory assumes that, at some point, the observer decides that a response "signal" or "no signal" can be made. This cut-off point, or criterion, is denoted by beta. The position of beta is not arbitrary but can be defined mathematically as the likelihood ratio that "a central effect of the magnitude represented by x is due to signal-plus-noise as opposed to noise alone" (Welford, 1968, p. 32). Its optimal value depends on the relative frequency of occurence of "signals" and the payoffs which are associated with detecting or not detecting them when they are or are not present. Therefore, the parameter beta may be considered as a measure of caution exhibited by the observer, or as a measure of the observer's confidence. The observer plays the active role of making a judgement about the "best" position for locating the cutoff point.

The response pattern for a certain signal strength is reflected in a ROC (Receiver Operating Characteristic) curve (figure 8.2) which is "a function which summarises the possible performances of an observer faced with the task of detecting a signal in noise" (Egan, 1975, p. 3). The amount of "bow" in the curve is determined by d' and can be interpreted as a measure of the perceived signal strength, or possibly as a

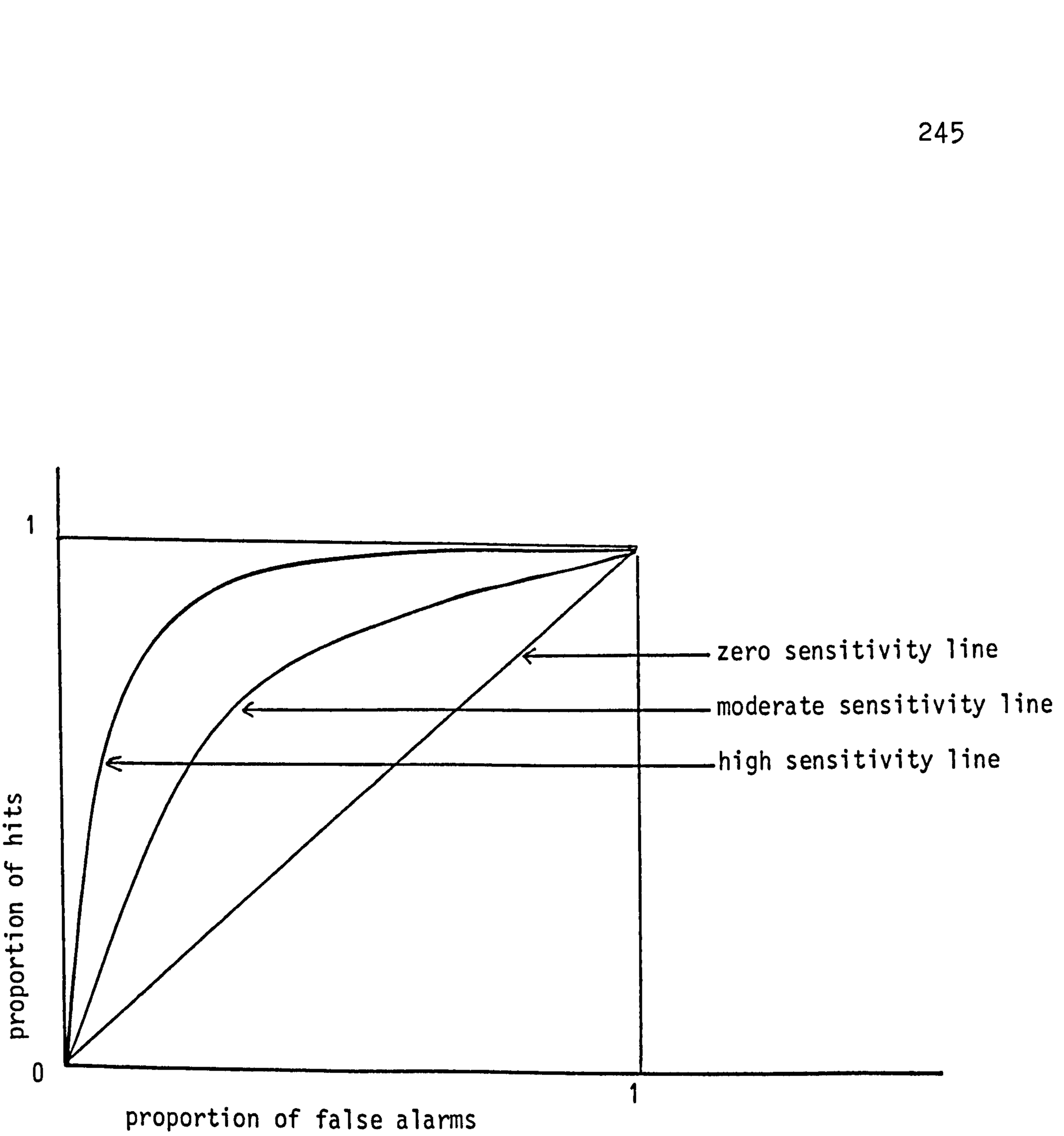


Figure 8.2 Receiver Operating Characteristic (ROC) Curves

measure of the observer's sensitivity to a particular signal. There is a positive relationship between d' and the sensitivity of the observer. In other words, a high d' indicates higher sensitivity whereas a low d' indicates a lower sensitivity.

The main features of Signal Detection Theory are the following. First, the subject is considered as an active participant who sets the criterion of performance for each signal in a continuously changing environment. Second, SDT "provides a means of evaluating the separate contributions to discriminative behaviour of an individual's sensitivity and response criterion" (D'Amato, 1970, p. 157). Finally the SDT treats accuracy as an integral part of performance.

Signal Detection Theory has been widely applied in psychological research, from animal psychophysics to subliminal perception (Green and Swets, 1966). However, the theory has certain limitations which make it inadequate to account for the present results. The first limitation is related to defining the "signal" and "no signal" situation. The stimuli which were used in the experiment required the simultaneous depression of a number of fingers.

(1) If each stimulus is treated as a group of sub-stimuli (i.e. a multidimensional stimulus) then SDT could be applied. However, it would treat the decision of whether to press or not to press each finger as independent from that for the other fingers. The results of Experiment 1 showed that the fingers could not be treated as operating independently from each other.

(2) If each stimulus is treated as one single signal the response to which might be either correct or incorrect, a lot of information about error responses is lost. A response may, in fact, be almost correct (e.g. the only error being the depression of an extra key). However, if SDT is applied all errors are treated as equivalent. Consequently, a lot of information about the nature of errors is lost, whereas it could be used to enhance the understanding of the way subjects learn such a task.

The second limitation of SDT is that it provides no means of discussing variations of performance from trial to trial either between or within subjects and/or tasks. Therefore, SDT cannot account for changes in performance because of practice, special training, etc.

Finally, although SDT involves the concept of the observer's "sensitivity", it does not explain adequately how the observer learns the task over time and it does not make any predictions about long term changes in performance.

Consequently, in its "classical" form, Signal Detection Theory cannot be adopted to discuss and explain the present results.

8.3 The Fast Guess Model

The Fast Guess model is one of the typical speed-accuracy tradeoff models. It was first conceived and presented by Ollman (1966) and is also presented in detail by Yellott (1971). According to the Fast Guess model, a choice reaction time task, subjects make two kinds of responses:

(1) Stimulus Controlled Responses (SCR) in which subjects take the time needed for an accurate response (although a small proportion of them might be incorrect) and

(2) Fast-Guess Responses which subjects initiate without processing the stimulus adequately.

The model assumes that on any given trial subjects have the option of either producing an SCR or a Fast-Guess response. The relative frequencies of these responses are controlled by the subjects and are expected to vary with instructions and payoffs. Subjects may increase the number of Fast-Guess responses to increase their overall speed or they may decrease this number to increase their overall accuracy. However, Yellott claims that "regardless of how the subject chooses to operate, it is possible to estimate the mean latency of the SCRs made over any set of trials" (1971, p. 160). With regard to the latency of SCRs and based on previous work (Ollman, 1966; Yellott, 1967), Yellott (1971, p. 161) claims that "under certain standard conditions, mean SCR latency does in fact remain approximately constant regardless of fluctuations in speed-accuracy". He further supports this claim by the results of three experiments he reports which involved 2-choice tasks.

The Fast Guess model does not claim that subjects can vary the speed-accuracy tradeoff only by varying the proportion of Fast-Guess responses. According to Yellott (1971, pp. 160-161), "...the Fast Guess model really does little more than formalize the notion that some (unspecified)

proportion of the subject's responses are guesses; nothing is assumed about possible effects of speed-accuracy bias on the latency of nonguess responses (i.e. SCRs). However the Fast Guess model does allow us to determine experimentally whether changes in SCR latency play any substantial role in the speed-accuracy tradeoff".

Regarding this latter point, Yellott (1971) has used a somewhat different way to plot the speed-accuracy relationship. On the ordinate, he plots the difference between average correct RT and average error RT each weighed by its respective proportion of trials (Pc*Mc-Pe*Me, where Pc and Pe are the proportions of correct and error responses and Mc and Me are the mean correct RTs and mean error RTs). On the abscissa, he plots the difference between the proportions of correct and error responses (Pc-Pe) (figure 8.3). In a 2-choice reaction time task, at one extreme when full attention is paid to accuracy, there are no errors, so Pc-Pe=Pc=1 and Pc*Mc-Pe*Me=Mc. At the other extreme, when speed is most important Pc=Pe and Mc=Me since all responses are Fast-Guess responses. Yellott (1971) claims that this function is linear. Finally the slope of the function is equal to the average duration of SCRs.

If Fast-Guess responses were the only reason for the occurence of errors, the formulations given by Yellott (1971) would provide "a powerful solution to the problem of variable error rates" (Pachella, 1974). However, the Fast-Guess model does not assume - and very rightly so - that Fast-Guesses are the only reason for errors. As it is though, the model does

nothing more than account for some errors. It makes no attempt to explain errors produced by SCRs or to compare the difference (if there is any) between correct and error SCR latencies.

The Fast Guess model cannot be used to explain the present data for a number of reasons.

(1) It cannot account for the fact that error responses were slower than correct responses.

(2) It has no way of dealing with stimuli of different levels of difficulty.

(3) It does not deal at all with trial to trial variations either between or within subjects (or tasks).

(4) It does not explain how subjects learn the task and how their performances changes with practice.

(5) Finally it does not discuss the role of the subjects as active participants in a system in which they can exercise any control (e.g. supervisory).

8.4 The Random Walk Model

The Random Walk model (Fitts, 1966; Laming, 1968) is a probabilistic model and shares characteristics with Signal Detection theory and some aspects of Bayesian decision making. The process considered by the model is more easily conceived, described and understood when there are only two alternatives, for example Sa and Sb. The model assumes that subjects have some initial opinion about the odds, D, which favours the occurence of one stimulus as opposed to the other (e.g. Sa is favoured over Sb). If it is assumed that the probabilities of

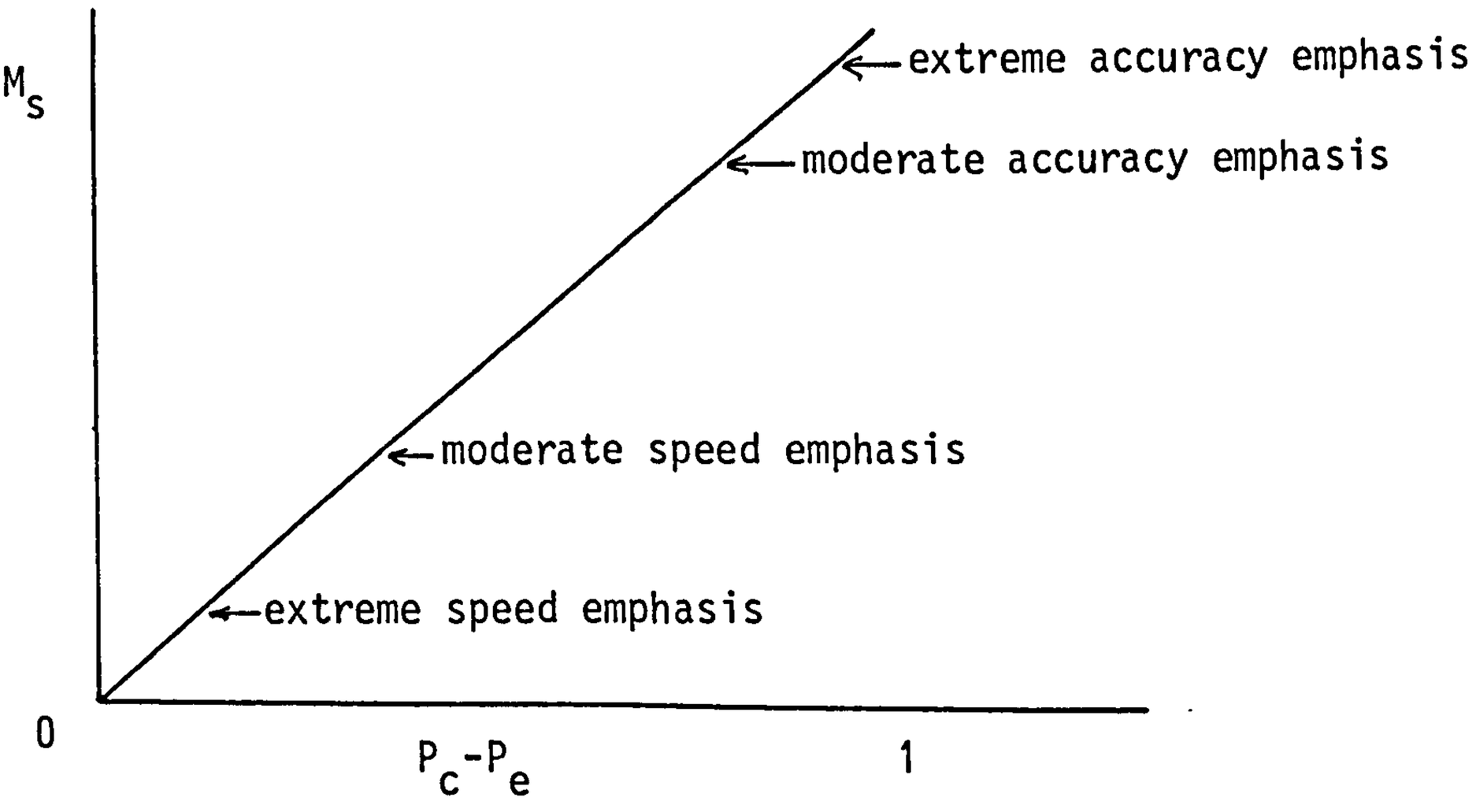


Figure 8. 3 The Form of Speed-Accuracy Operating Characteristic Assumed by the Fast-Guess Model

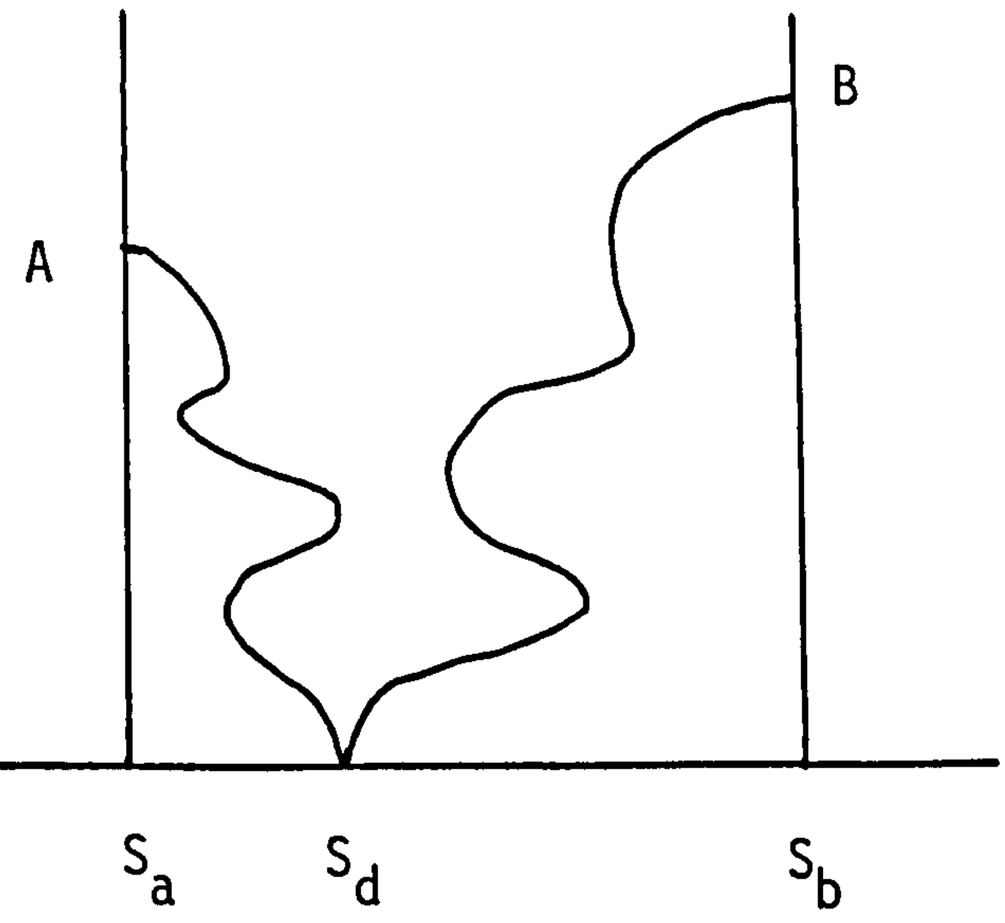


Figure 8.4 Graphical Representation of the Random Walk Model

the two stimuli are equal (i.e. P(Sa)=P(Sb)) then D(Sa) would be 1:1. When a stimulus is presented, subjects sample noisy information about it. Any increment of information favours one stimulus against the other. The process of sampling information takes the form of a random walk (figure 8.4) which "...is assumed to continue until it reaches some criterion boundary, which is defined in terms of the critical odds value" (Fitts, 1966, p. 850). The critical value is not necessarily the same for each of the two alternatives. If this value is high, in other words if the evidence which favours one alternative must exceed that of the other alternative by a large amount, the RTs will be generally long. In this situation, the decision will have been based on much evidence and, consequently, the probability of an error will be small. On the other hand, small critical values will lead to fast RT and high error rates. (It should be mentioned that the point from where the random walk begins is usually closer to one of the two boundaries - as opposed to being in the middle).

Laming (1968, pp. 36-38) defines the axioms which the Random Walk model uses:

(1) "Whilst a signal is present the subject extracts information from it continuously. The total information available to the subject at time t is a single valued random function of t.... "

(2) "The information stream shall contain all that is relevant to the discrimination to be made and nothing that is

irrelevant.... "

(3) "The signals to be discriminated are indeed distinguishable.... "

(4) "A response is made as soon as the probability of its being wrong reaches a certain small fixed value (which may be different for different responses).... "

(5) "In the absence of any experimental instructions or conditions to the contrary the subject will tend to optimize his decision process in the sense of achieving the fastest total performance compatible with a given proportion of errors...."

Fitts (1966, p. 850) claims that, in a 2-choice RT task, the Random Walk model has the following four qualitative properties.

(a) A symmetrical shift of both criterion boundaries toward more nearly equal odds (odds closer to 1:1) should result in faster mean RTs and increased errors for both responses (Ra and Rb).

(b) A unilateral shift of criterion boundary A toward more nearly equal odds should result in Ra being made more often, more rapidly, and with more errors, relative to Rb.

(c) An increase in the initial odds favouring Sa should result in faster RTs for Ra, and longer RTs for the alternative response, Rb, provided the criterion boundaries remain fixed (or alternatively should result in an asymmetric readjustment of the boundaries).

(d) An increase in the difficulty of discriminating between Sa and Sb, as would result from an increase in similarity or

a decrease in signal-to-noise ratio, should prolong the sampling process and result in slower RTs, providing again that the criterion boundaries remained fixed.

According to Laming (1968), the Random Walk model makes three predictions:

(1) In a 2-choice RT experiment, probable signals get fewer errors, and probable responses occur as errors. Let us assume that there are two stimuli Sa and Sb with corresponding responses Ra and Rb. If the probability of Sa is larger than the probability of Sb, then, the model predicts that, amongst the errors, there will be fewer Rb responses to Sa stimulus, than Ra responses to Sb stimulus. This has been found experimentally by Fitts, Peterson and Wolpe, 1963; Falmagne, 1965; Laming 1962; 1968.

(2) In a 2-choice experiment, a stimulus which elicits a faster response also has a smaller probability of error and vice versa. This prediction - which disagrees with the second qualitative property that has been Fitts (1966) presents - has been discussed by Laming (1968) and it implies that the increased speed of responses to probable stimuli is mainly due to a response bias rather than to better sensory input from the improbable stimulus (or as Laming expresses it (1968, p. 44) "if the subject is more ready to make one response rather than the other, there is a reduction in errors and reaction time conditional on the corresponding signal").

(3) For a given response, the RT distribution is the same irrespective of whether the response is correct or an error.

Laming (1968) made this prediction based on the work of Stone (1960) who proved it theoretically for a 2-choice task. Unfortunately, he failed to prove this prediction experimentally. Consequently, he modified the third prediction to "... In two-choice experiments errors are faster than the same response made correctly. The difference in mean between the distributions of error- and correct-reaction times is independent of most experimental conditions, but may be affected by excessive pressure on the subject to respond quickly, and will certainly decrease to zero with the intertrial interval" (1968, p. 82).

Alternatively, Fitts (1966) made the assumption that subjects set their criterion boundaries on the basis of the initial instructions they were given and then adjusted them accordingly on the basis of feedback they had from their own responses. Fitts argued that the changes in the criterion boundaries and the small number of errors made it difficult for the third assumption to be proved. However, he claimed that "similar RT distributions for correct and wrong responses should most likely be found when Os [Operators] are making a relatively large proportion of errors" (1966, p. 850). However, neither explanation is adequate to explain a large number of experiments which have found that RT distributions for correct and incorrect responses are not the the same.

Laming (1968, p. 44) stated very explicitly that the model "is concerned only with relations between reaction times and probabilities of error for a given set of signals". Even if it is accepted that the Random Walk model achieves this

goal and does explain this relation, the learning process is not explained and therefore the model is not useful to discuss the results of the present analyses. It cannot explain what happens from trial to trial and it cannot explain what happens with practice. However, the sampling of information about a stimulus, which the model introduces, is an interesting idea and can be further explored.

8.5 The Accumulator Model

The Accumulator model (Morton, 1969; Vickers, 1970) is another probabilistic model which shares features with Signal Detection Theory and some basic assumptions with the Random Walk model. It assumes that upon stimulus presentation, subjects accumulate noisy evidence about the stimulus and evaluate this evidence as to which of the possible alternatives is most likely. The reliability of this evidence increases as more information becomes available (and, of course, more time is taken). Because of the stress on speed, subjects have to set a response time criterion. A response is initiated either when sufficient information about a stimulus has been accumulated or when the response time criterion has been reached.

This model assumes that each alternative stimulus is associated with a separate representation in memory. Morton (1969) who worked on word recognition called such representations "logogens" ("logogen" meaning "word birth"). According to Morton (1969, p. 165) "the logogen is a device which accepts information from the sensory analysis mechanisms

concerning the properties of linguistic stimuli and from context-producing mechanisms." This representation, or logogen, is activated when there is sufficient evidence about the existence of the corresponding stimulus. For example, let us assume that there are two stimuli Sa and Sb which require responses Ra and Rb respectively. If Sa is presented, then both representations will become activated as a function of the degree to which they "see" evidence for their stimulus. It is more likely that the representation of Sa will accumulate sufficient evidence before the representation of Sb and thus the response Ra will be triggered. However, it is possible that because of common characteristics between Sa and Sb and because of random activity in the perceptual system (i.e. noise) the representation of Sb will reach the criterion first and the response Rb will be made incorrectly.

The main difference between the basic assumptions of the Random Walk and the Accumulator model is that whereas in the Random Walk model evidence is sampled so as to favour one stimulus against another, in the Accumulator model evidence about each alternative stimulus is accumulated independently from the others. However, this difference is not of great importance, at least at this point, as the only difference would be in the prediction of the RT distributions. The Accumulator model, however, seems to be more ingenious in the way it interprets speed and accuracy and more useful in the analysis of the learning process.

Morton (1969) not only discusses the way a response is

made but, in addition, he attempts to explain the learning process by discussing the "interaction of successive presentations". He also introduces the issue of "partially correct" responses which can be very useful in understanding the dynamics of the learning process. However, as Morton's work deals exclusively with word recognition, it is considered more relevant to refer to the general issues covered by the logogen model rather than the specific detailed procedures that it discusses.

Vickers' (1970) model is based on his work on psychophysical discrimination. In his experiments he presented subjects with two vertical lines and they had to choose which line was longer. In the 1970 presentation of his model, he discusses it in conjunction with other models (i.e. statistical decision models, the Recruitment model and the Runs model). In that presentation the scope of his model is rather limited as it does not discuss what happens with practice and how the parameters which affect speed and accuracy change in the long term. Furthermore, in that presentation he ignores subjects' capacity to adjust their performance by varying their response criteria (e.g. as they become more familiar with the task or as they become more confident).

Ten years later, Vickers (1980) extended the Accumulator model. He acknowledged subjects as active parts of the system. He pointed out that "... bias in responding may be exhibited by observers, who possess some capability for varying their relative readiness to make one response or the other" (1980,

p. 62). Vickers claimed that bias can be accounted for if observers are assumed to set unequal values to the criterion of each alternative. Moreover, he recognized that "... observers also possess some capacity for counteracting the variability in the sensory representation of stimuli, albeit at a cost of taking more time to respond" (1980, p. 62). He carries on to suggest that "... discrimination appears to be mediated by some kind of optional-stopping facility, whereby observers can tailor the number of observations to correspond with the discriminability of each stimulus pair" (1980, p. 62). In his experiments Vickers found quite systematically that RTs for errors were slower than RTs for correct responses. This led him to suggest that "the discrimination process is not memoryless, but probably involves some integration or accumulation of information over successive observations within a particular trial".

Vickers' (1980) revised Accumulator model provides some good explanations about the learning process in a choice task. Moreover, it provides an elegant approach to certain aspects of performance, like level of confidence or caution. However, the model is limited in its explanations of speed-accuracy relations. One problem is that it predicts incorrect responses should be slower than correct responses. While this is consistent with Vickers' data and the data of the present experiment, it only applies to tasks of some considerable difficulty (it is a model on "discrimination" after all), and cannot be used to explain a large amount of data (e.g. from

simple RT experiments, from relatively "easy" choice response tasks, etc.).

8.6 The Tracking Model

The Tracking model has been developed by Rabbitt (1981). He discusses three main issues which are essential for the Tracking model.

(1) He postulates that subjects have the capacity to exercise overall control and adjust their performance to deal efficiently with very complex tasks. This assumption may not seem necessary if one considers very simple tasks in which only one stimulus appears at a time and one discrete response is made before the next stimulus appears. However, this assumption becomes essential if one considers more difficult tasks. Rabbitt (1981) makes this assumption based on Shaffer's (1973) studies on typing where subjects learned to exercise adaptive control over very complex performance.

(2) Rabbitt (1981) discusses the issue of instructions to subjects. Namely, when the experimenter asks subjects to respond as fast and accurately as possible, he/she actually asks them to set their own SETOFs (Speed-Error TradeOff Functions). Rabbitt (1981) argues that when subjects themselves ask the experimenter (who instructs them to work as fast and accurately as possible) "Yes, but which matters more? Speed or errors?", they imply that they can "actively choose" whether to respond fast and make errors or to respond slowly and be accurate. As Rabbitt (1981, p. 161) points out "in formal terms they are saying that they can actively choose how

to structure their RT distributions in relation to their SETOFs". This implies that subjects either know or can find out their SETOFs for a particular task. Based on previous work (Pachella and Pew, 1968; Pew, 1969; Rabbitt and Vyas, 1970), Rabbitt (1981) claims that SETOFs vary from task to task. Subjects do not know their SETOF for a new task ab initio, they have to discover it and ajust their performance to the particular task demands. To achieve this, subjects must know: (1) when they make an error, otherwise a SETOF could not be recognised; and (2) what was the speed of their incorrect response. As Rabbitt (1981, p. 162) puts it, subjects "must be accurate in estimating, reproducing and adjusting their brief reaction times".

(3) Rabbitt (1981) considers that subjects are capable of detecting most of their errors (Long, 1976; Rabbitt, 1966a,b; 1968b; 1978b; Rabbitt and Vyas, 1970). They are even capable of detecting which occured two or more responses previously (Rabbitt, 1979a) and also detecting "... errors due to failures of perceptual analysis as well as errors due to failures of response selection and execution (Rabbitt, Cumming and Vyas, 1978)" (1981, p. 162).

Rabbitt (1981) claims that, from trial to trial, subjects respond faster and faster until an error occurs. Then, he says, they must recognise their error because their speed of responding changes, responses following errors are generally slow (Rabbitt, 1969). Following an error, slow responses continue for three or more trials. Then, subjects increase

their speed once more to approach the RT limit at which errors are likely to occur (Rabbitt and Vyas, 1970). Rabbitt (1981) argues that subjects adjust their RT within a certain "time band". The lower limit of the "time band" is determined by the SETOF; in other words, subjects learn to avoid these fast responses where the frequency of errors is increased. The upper limit of the "time band" is determined by an internal "clock" which warns subjects when they respond unnecessarily slowly. This notion of the "time band" is similar to the one mentioned earlier when referring to Wickelgren's (1977) methods of obtaining speed-accuracy tradeoff functions. However, in Rabbitt's (1981) case, the limits of the "time band" are defined by the subjects themselves and not by the experimenter.

According to Rabbitt (1981) the Tracking model makes predictions about the effects of prolonged practice on RT distributions as well as on arithmetic means of RT. It also predicts changing relationships between the distributions of RTs for correct and incorrect responses. It predicts:

(1) As subjects receive more practice, they learn to estimate more precisely (and therefore more safely) the lower limit of their "time band". Thus, from trial to trial, they can control their RTs to observe the limit so that they do not respond faster than it (and make errors). At the same time they try to keep their speed as close as possible to the limit.

(2) The model assumes that after subjects establish their SETOFs, they learn to avoid fast rates of responding when errors become too frequent. It follows therefore that subjects

would trangress SETOF limits early but not late in practice. Consequently, the model predicts that the difference between the means of RTs for correct and error responses decreases with practice. Rabbitt (1981) claims that this is the case. (3) Rabbitt (1981, p. 163) argues that subjects have precise feedback which enables them to understand when they respond too fast. However, their recognition that they are performing too slowly depends upon the accuracy of their internal 'clock' (i.e. on their estimation of time). Rabbitt (1981, p. 163) claims that there are two consequences from this statement: "First, RT distributions should be markedly skewed, with a sharp cut-off for fast responses and a straggling 'tail' of slower responses. This is the case. Second, if subjects are provided with external feedback (such as an elapsed-time signal) which informs them when they are slow, they should be able to shape the RT distributions accordingly and reduce arithmetical means of RT, without increasing errors, by shaping the distribution of RTs to avoid unnecessarily slow responses. Again this is just what happens (Rabbitt and Vyas, 1970)".

The Tracking model is a useful model to consider in order to explain the results obtained by the present statistical analyses. The assumptions of the model are quite realistic and have been experimentally substantiated. Moreover, the model considers subjects as active participants of the system. In addition, the model considers accuracy and its relationship to speed quite actively. Furthermore, it tries, and to a large

extent succeeds, in accounting for a number of issues (e.g. the effects of long practice on performance, variations of performance from trial to trial, the effect of errors on the following response, control of information stored in long-term memory, etc.).

However, the Tracking model has a rather important limitation. According to it, errors are faster than correct responses, since subjects "trade off" accuracy for speed, or vice versa. In some experiments (e.g Rabbitt, 1966) Rabbitt and his associates did experimentally confirm this relationship between speed and accuracy. In other experiments they failed to do so (e.g. Rabbitt and Vyas, 1970). Also the results of the present study have clearly established that error responses can be consistently slower than correct responses. However, the Tracking model clearly predicts that errors are faster than correct responses. This is probably the main limitation of an otherwise well-conceived model.

8.6 Conclusions

All of the above models have some contribution to make in explaining the results of the present experimental project. The Accumulator model and the Tracking model, in particular, account for a number of findings from the data. However, from what has been said so far it seems that none of the above models can provide a full account of the present results. It is felt that these models are not as incompatible with each other as their authors might claim. With this in mind, there is a need to present a "new" model, which represents an

integration of certain features of the existing ones. Before this can be done however, it will be necessary to look in some detail at the nature of subjects' response distributions and their errors.

Chapter 9

FURTHER EXAMINATION OF SUBJECT RESPONSES

The examination of speed-accuracy tradeoff functions carried out on the data of Experiment 4 revealed certain relationships (e.g. errors are slower than correct responses) which cannot be adequately explained by existing theories. Before an attempt is made to account for these results on a theoretical basis, some further analysis of the data is needed. First, subjects' RT distributions of correct and error responses are examined and compared and second, error responses are further analysed so as to establish the type of responses, instead of the correct ones, subjects produce.

9.1 RT Distributions Across Subjects

The relationship between speed and accuracy has been studied, thus far, using either the arithmetic mean RTs, or the mean residual RTs. However, the study of speed using only these two measures would be incomplete. As Broadbent (1971, p. 318) claims, this approach to the problem is "an arbitrary limitation". He argues that mean RT could remain constant even if there were systematic variations from trial to trial, or if there were differences between the minimum and maximum values of RT, or if there were any other changes in the process which affected features of subjects' performance. These considerations, as well as Rabbitt's (1981) suggestion that speed is plotted against trial number, were all taken into account. When the RT for each trial is plotted, there is a lot of noise. To avoid this noise mean RTs of every twenty trials

were calculated. This attempt to avoid noise was unsuccessful because even after that there was too much noise. Therefore another technique of studying subjects' RT distributions was used.

Distribution Histograms

Probability distribution histograms were produced for each one of the forty subjects (e.g. figure 9.1). The vertical axis represents the percentage of responses and the horizontal axis represents a range of RT class intervals of 20 csecs. Two columns were plotted for each RT interval: (a) the proportion of the total number of correct responses which were produced within a given RT interval (e.g. in the histogram of figure 9.1, 25% of correct responses were made within the interval of 80 and 100 csecs); and (b) the proportion of error responses produced within the same RT interval (e.g. in figure 9.1, almost 15% of error responses were produced within the RT interval of 80 and 100 csecs).

Two kinds of such histograms were made for each subject. First, one which showed the RT distributions of correct and error responses after having pooled all responses together; that is practice was not taken into account. Second, six other histograms for each subject, one histogram for every practice session. The histograms of the second kind were not as reliable as those of the first kind, because obviously they were produced from much less data, a remark which applies particularly to the RT distributions of errors.

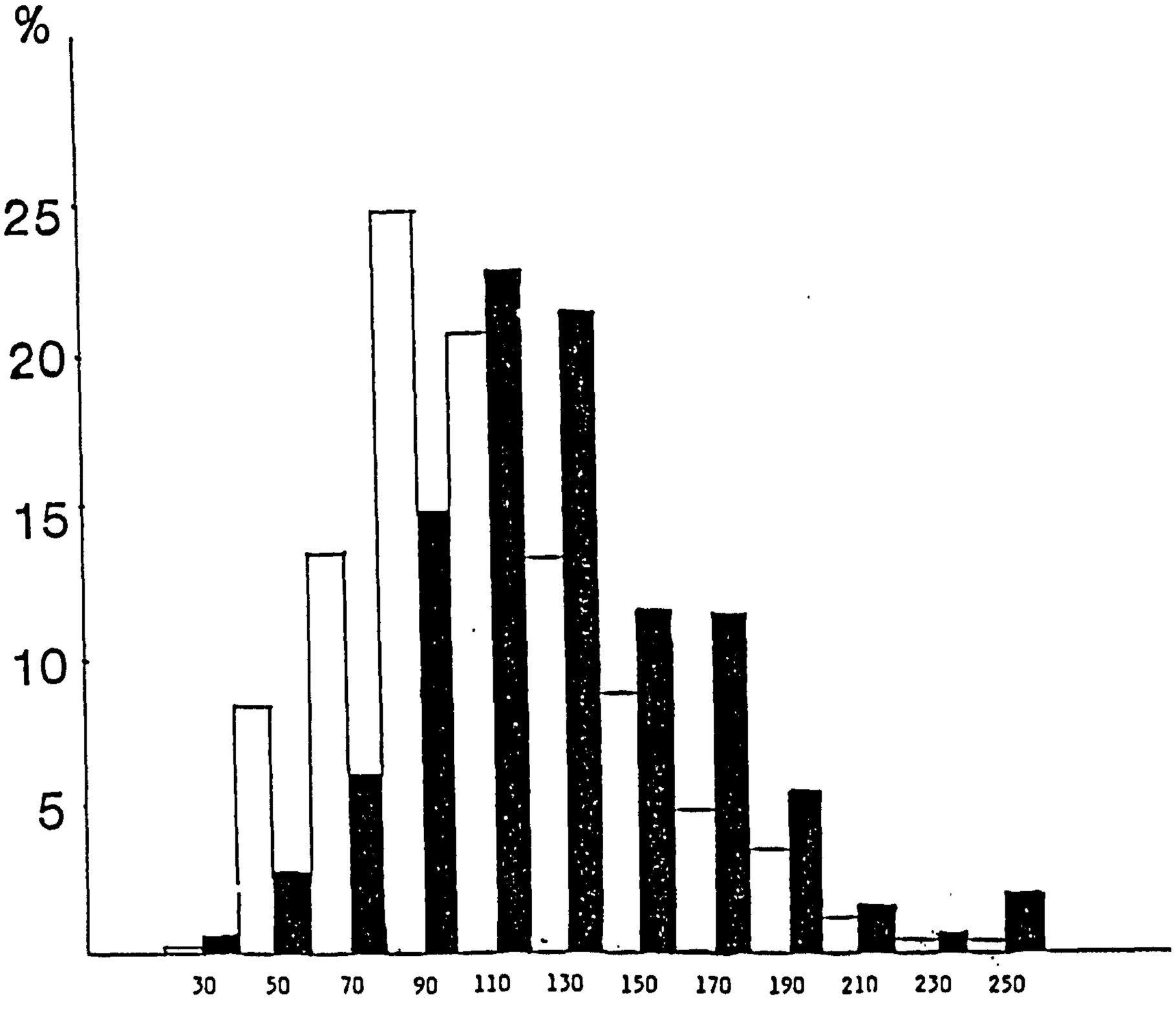


Figure 9.1 An example of the distribution of RTs of correct (white columns) and error (black columns) responses (data from Subject 37)

Overall RT Distributions

These histograms showed the same pattern across all subjects irrespective of the proficiency of each individual subject or the training mode by which subjects had been trained. The distributions of correct responses were skewed towards fast RT intervals. The error distributions were similar to those of correct responses but they were shifted towards slower RT intervals.

RT Distributions Across Practice Sessions

Similar results were also obtained from these histograms which were based on the data from each practice session.In general, RT distributions of correct responses were skewed to fast RT intervals, whereas the distributions of errors were similarly skewed but shifted towards slower RT intervals. Not surprisingly, the second kind of histograms were not as noise-free as the histograms of the overall RT distributions. This was the case during the early practice sessions especially. Later on in practice, the distributions became "smoother". The histograms from some subjects' performance show an interesting variation which is quite important to mention. Early in practice, the RTs of error responses showed a bimodal distribution. Although the majority of errors were skewed towards slow RT intervals, there was also another, smaller distribution involving a marked proportion of errors which occured within the fast RT intervals. This feature disappears when responses from all sessions are pooled together. However, it suggests that early in practice subjects, sometimes, seemed

to trade accuracy for speed whereas they did not do so later on in practice.

Discussion and Conclusions

These histograms show that the RT distributions of correct and error responses are similar and that the error RT distributions are always shifted to the right of correct RT distributions. This finding is rather similar to what other researchers have found before (e.g. Egeth and Smith, 1967; Fitts, 1966; Laming, 1968; Rabbitt, 1981).

Egeth and Smith (1967) used a word recognition task to study the relation between correct and incorrect response latencies by subjecting both types of latency to identical analyses. The task involved the presentation of a letter of the alphabet and subjects were expected to respond whether or not this letter was included in a predefined target set of letters by pressing either the "Yes" or "No" key. They found that the mean RTs of errors were a little faster than those of correct responses, but the RT distributions of both correct and error responses were very similar. They cite Smith's (1965) results to give more generality to their conclusion. Based on their results and those of Smith (1965 - that error RTs reflect the same relations as correct RTs with respect to number of targets, practice, and type of response) they suggested that "Such results seem to indicate that a conception of errors which makes them simply anticipations, a result of distraction, or some other kind of random response is untenable. The processes determining erroneous responses

must be essentially the same as those determining correct responses. Since errors are faster, it also seems likely that the difference between correct and incorrect responses is that these processes are not carried to the same degree of completion on those trials that eventuate in errors" (Egeth and Smith, 1967, p. 346). Unfortunately they do not explain how it follows from these results that error responses involve the same processes as correct responses but not to a sufficient degree of completion.

The finding about the RT distribution of correct responses is in agreement with what Rabbitt (1981) has claimed. Moreover, in both cases there is a big overlap between RT distributions of correct and error responses - as opposed to an extreme theoretical case where the two distributions would be completely apart. In relation to the RT distribution of errors, Rabbitt (1981) has suggested that the shape of the RT distribution of errors is the same as that for correct responses. The present error RT distributions are skewed similarly to the RT distributions of correct responses but they are shifted towards slow responses which is the opposite of what Rabbitt suggests. This finding implies that the majority of errors are slower than correct responses, a conclusion already well-established from the analyses on the data using the mean RTs of correct and error responses.

The examination of RT distributions across practice sessions revealed that in a number of cases, early in practice, subjects made a noticeable proportion of errors which were faster than their average mean RT for correct

responses. A possible explanation of this variation in the error RT distributions is that early in practice subjects have not yet adopted an effective speed-accuracy tradeoff criterion and therefore they make fast errors whereas later on in practice they learn to avoid such errors.

9.2 Classification of Error Responses

The above models concentrate on the relationship between speed and accuracy but do not adequately explain the reasons for the occurence of errors. According to Fitts (1966), errors are generally caused by misidentifications of stimuli, and their frequency is controlled by the criterion value. Laming (1968) considers that errors are caused by premature sampling. He claims that, under speed pressure, subjects start collecting information before a signal is presented. "The information so sampled is irrelevant to the discrimination between the signals" (Laming, 1968, p. 81). Therefore this irrelevant information makes discrimination and choice response more difficult. Both of these lines of explanation predict that the distributions of error RTs are the same as the distributions of correct RTs. In both cases, errors are considered as failures in stimulus identification and predict that error responses should be "similar" to correct ones and not random. Neither of the two models deals with the effects of practice on accuracy.

So far, the statistical analyses of the data have concentrated on examining variations in speed and accuracy in

relation to level of practice, stimulus difficulty, etc. However, there is more to be learnt about the causes of errors by a closer examination of the errors themselves. Are error responses random or are they systematically related to the stimuli? If the latter is the case, does the nature of this relationship help us understand the causes of errors?

Information on the systematic nature of errors is presented by Guest, Sime and Green (1972). In their experiment (which was not explicitly designed to examine the nature of errors), they used ten letters of the alphabet as stimuli. The required responses were two-key responses on a five-key chord keyboard (i.e. a keyboard similar to the one used in the present experiment but using only one hand). They found that errors were systematic and postulated two factors which related to the prediction of error responses: the presence or the absence of a common finger and the difference in finger separation (which they called Delta factor). Moreover, they found that mirror errors (i.e. responses which mirror the correct response about the middle finger) were quite frequent.

Method of Error Analysis

For each training group, a nine by thirty-two matrix was produced of all trials which involved an error made by the left hand and another such matrix of all trials which involved an error made by the right hand. Eight one-handed stimuli had been selected for the experimental task (these stimuli have been discussed in Chapter 4) whereas there were thirty-one possible responses to them as well as the thirty-second option

of not responding at all. This classification produced a confusion matrix in which all responses were plotted against the nine stimuli (eight stimuli plus the case where subjects did not have to depress any key). It showed, for each stimulus, the frequency with which each of the possible thirty-two responses were made. (The sum of the total error responses by the two matrices will be greater than the total number of error responses, as some errors involved mistakes by both hands.)

From these matrices, one way in which errors were examined related to whether errors had been made by the left or the right hand.

Second, errors were classified in the following three main categories, which represented three different "types of errors":

(1) "Omissions": when one or more keys were omitted;

(2) "Additions": when one or more keys were pressed in addition to the ones which should have been depressed; and

(3) "Combined Errors": when one or more keys were omitted and at the same time one or more keys were added.

Another way by which errors were classified was related to the "stimulus complexity". The "stimulus complexity" is a different measure from "stimulus difficulty" which has been used in the previous analyses, as in this section only one-handed stimuli are examined. The "stimulus complexity" was mainly a measure of the number of keys which were supposed to be depressed. There were five categories of "stimulus

complexity":

(1) 0-keys: In this category, no keys had to be depressed. This was the case for those trials, with a one-handed stimulus; when the error was made, the other hand, which normally should not have made any response, produced an error by pressing, incorrectly, one or more keys. Since most of the stimuli were two-handed, not many errors were expected in this category of "stimulus complexity". It should be mentioned that all of these errors would necessarily have to be "additions".

(2) 1-key: In this category, only one key had to be depressed.

(3) 2-keys-A: In this category, two adjacent keys had to be depressed.

(4) 2-keys-S: In this category two non-adjacent keys had to be depressed. The differentiation between the errors of this category and those of category (3) was made because it has been found that those stimuli which require the depression of non-adjacent keys are more difficult and produce a larger number of errors than the stimuli which require the depression of adjacent keys (Chapter 3).

(5) 3-keys: In this category of "stimulus complexity", three keys (non-adjacent) had to be depressed (obviously, there was only one possible pattern in this category of "stimulus complexity").

Finally, errors from each training group were classified separately, so that the effect of the mode of training on the "type of errors" could be examined.

All of these error classifications (i.e. handedness, "type of errors", "stimulus complexity", and mode of training) were studied separately and in combination.

Results

(1) Handedness

The total number of error responses was 7,219. However, the present error classification gave a total of 8,061 errors produced by both hands. This meant that, 842 errors (8,061-7,219=842) involved mistakes by both hands. From the 8,061 errors, 3,968 (49.22%) were made by the left hand and 4,093 (50.78%) were made by the right hand (table 9.1). This difference between the number of errors made by the left and the right hand can be considered negligible and it can be concluded that handedness had no effect on the accuracy of performance.

Hand	FR	FA	CR	CA	Overall
Left	1245 51.40%	987 48.48%	971 49.59%	765 46.50%	3968 49.22%
Right	1177 48.60%	1049 51.52%	987 50.41%	880 53.50%	4093 50.78%

Table 9.1 Number and proportion of errors as a function of handedness

(2) Type of Errors and Stimulus Complexity

Overall, when all errors from all four groups were classified, 51.42% of the errors were "omissions", 30.33% were "additions" and 18.25% were "combined" errors. This first

finding shows that errors were rather systematic in terms of "type of error". From all "omissions", 83.45% were omissions of one key while from all "additions" 97.91% were additions of one key (table 9.2). It was calculated that 72.61% of all errors were either an "omission" or an "addition" of one key. These errors can be considered as "almost correct" since the difference between them and what would have been considered "correct" was only one key.

	TRAINING MODE				
Type of Error	FR	FA	CR	CA	Overall
Omissions	1360 (56.15%)	970 (47.64%)	932 (47.60%)	883 (53.68%)	4145 (51.42%)
1 key	1076 (79.12%)	793 (81.75%)	821 (88.09%)	769 (87.09%)	3459 (83.45%)
2 keys	200 (14.71%)	132 (13.61%)	89 (9.55%)	97 (10.99%)	518 (12.50%)
3 keys	84 (6.17%)	45 (4.64%)	22 (2.36%)	17 (1.92%)	168 (4.05%)
Additions	676 (27.91%)	643 (31.58%)	699 (35.70%)	427 (25.96%)	2445 (30.33%)
1 key	662 (97.93%)	629 (97.82%)	682 (97.57%)	421 (98.59%)	2394 (97.91%)
2 keys	14 (2.07%)	13 (2.02%)	16 (2.29%)	6 (1.41%)	49 (2.00%)
3 keys	-	1 (0.16%)	1 (0.14%)	-	2 (0.09%)
Combined	386 (15.94)	423 (10.78%)	327 (16.70)	335 (20.36%)	1471 (18.25%)
TOTAL	2422 (30.05%)	2036 (25.26%)	1958 (24.29%)	1645 (20.40%)	8061

Table 9.2 Number and "type of errors" overall and for each training group

The nine one-handed stimuli which were combined in all possible ways to produce the stimuli used in the experiment included: one stimulus where no keys were involved, two stimuli involving one key, four stimuli involving two keys, and two stimuli involving three keys. Because of the different number of keys per stimulus and the different number of stimuli involving a given number of keys, the probabilities

of the occurence of certain "types of errors" (i.e. omissions, additions and combined errors) varied from stimulus to stimulus. For example a 1-key stimulus could produce thirty-two possible responses (as all one-handed stimuli could), one would be the correct response and thirty-one would be errors. If errors are produced by "random" responding (i.e. each key has a 0.50 probability of being pressed) then when an error is made each of the 32 possible patterns will be equally likely to occur. From these thirty-one possible error responses only one response could be an "omission" error, fifteen responses would be an "addition" error and fifteen responses would be "combined" errors. Given these different weightings as well as the actual number of errors produced by each kind of stimulus the expected values of the number of errors each "type" were calculated for each kind of stimulus (table 9.3, appendix 4).

Type of Error		0 keys	1 key		2 keys		3 keys		Overall	
Omissions	O	-	291	21.98%	1249	41.80%	2605	71.02%	4145	51.42%
	E	-	42.71	32.30%	289.16	9.68%	828.26	22.46%	1160.14	14.39%
Additions	O	81	759	57.33%	973	32.56%	632	17.23%	2445	30.33%
	E	81	640.65	48.39%	674.71	22.58%	354.97	9.68%	1751.33	21.73%
Combined	O	-	274	20.69%	766	25.64%	431	11.75%	1471	18.25%
	E	-	640.65	48.39%	2024.1	67.74%	2484.77	67.37%	5149.55	63.88%
TOTAL		81	1324		2988		3688		8061	

Table 9.3 Observed and expected number and proportion of "type of errors" overall and as a function of the number of keys involved in one-handed stimuli

It was found that, if errors were random responses and given that 8,061 errors were made, 1,160.14 (14.39%) of them were expected to be "omissions", 1,751.33 (21.73%) "additions"

and 5,149.55 (63.88%) "combined" errors. However, the observed frequencies were very different (table 9.3): there are far more "omissions" and far fewer combined errors than expected. Therefore, it has to be concluded that the "type of error" is not random.

Most errors (45.50%) were responses to stimuli which required the depression of three keys. 24.05% of all errors involved stimuli with two non-adjacent keys, 13.01% stimuli with two adjacent keys, 16.42% stimuli with one key, and, finally, only 1% fell into the category of "0-keys" (table 9.4). These results were generally expected since the "2-keys-S" and the "3-keys" categories were the most difficult and it had already been established that accuracy decreases with the level of stimulus difficulty.

Stimulus Complexity	FR	FA	CR	CA	Overall	
0 keys	21	12	28	20	81	1.00%
1 key	400	281	364	279	1324	16.42%
2 keys-a	301	220	256	272	1049	13.01%
2 keys-s	602	528	473	336	1939	24.05%
3 keys	1098	995	837	738	3668	45.50%
Total	2422	2036	1958	1645	8061	

Table 9.4 Number and proportion of errors as a function of "stimulus complexity", overall and for each training group

Apart from these main classifications, more questions were asked about the systematic distribution of errors. In the first instance, the data were examined to see whether stimuli with varying degrees of complexity produced different "types

of error". For this purpose, errors were classified in a five-by-three matrix, with the variables being "stimulus complexity" and "type of errors" (table 9.5). The expected values were calculated by taking into account the different weightings for each category of "stimulus complexity". The computation of the expected values was complicated by the fact that each category of stimulus complexity could produce different proportions of each "type of error".

Type of Error		0 key	1 key		2 keys-A		2 keys-S		3 keys	
Omissions	O	-	291	21.98%	511	48.71%	738	38.06%	2605	71.02%
	E	-	42.71	3.23%	101.52	9.68%	187.65	9.68%	828.16	22.46%
Additions	O	81	759	57.33%	303	28.88%	670	34.55%	632	17.23%
	E	81	640.65	48.39%	236.87	22.58%	437.84	22.58%	354.97	9.63%
Combined	O	-	274	20.69%	235	22.40%	531	27.39%	431	11.75%
	E	-	640.65	48.39%	710.61	67.74%	1313.52	67.74%	2484.77	67.37%
Total		81	1324		1049		1939		3688	

Table 9.5 Observed and expected number and proportion of errors as a function of "type of error" and "stimulus complexity"

There is no statistical test suitable for assessing variations in the difference between these observed and expected values across complexity. However the differences between the obtained and expected values are so marked that the two variables (i.e. "stimulus complexity" and "type of error") must be related in a systematic way. Across all categories of "stimulus complexity", there are more "omissions" (ranging from four to seven times more) and more

"additions" than expected. Moreover, the differences between the actual and the expected values vary across the different categories of "stimulus complexity". For example, in the category "2-keys-A" the expected proportion of "omissions" is 9.68% whereas the actual proportion is 48.71%; in the "2-keys-S" category, the expected proportion of errors is again 9.68% but the actual proportion is 38.06%.

A closer inspection of the "type of errors" (table 9.2) can reveal another interesting point. Of all the "additions", about 98% (plus or minus 0.50) are "additions" of one key, and this is the case across all training modes. Of all the "omissions", an average of 83.45% are "omissions" of one key (the proportions range from 79.12% to 88.01% across the four training modes). In other words, when there is an "addition" the error is most likely to be "very close" to the correct response. However, if the error is an "omission", the likelihood of it being "very close" to the correct response is somewhat less.

(3) Training Mode

The proportions of errors produced by each training group have already been discussed in detail (Chapters 4, 6). However, it should be briefly mentioned that 30.05% of all errors were made by the FR group, 25.26% by the FA group, 24.29% by the CR group and 20.41% by the CA group (table 9.2). The FR group produced the biggest number whereas the CA group produced the least number of errors.

It can be observed that the number of errors of each type

(i.e. "omissions", "additions", "combined" errors) varies among the four training modes (table 9.2). This raises the question of whether different training modes were associated with different distributions of "types of errors". The four training modes were the combinations of two different factors: (1) Fixed versus Cumulative sample size and (2) Random versus Adaptive stimulus sampling. A factorial chi-square was carried out to examine the effects of the possible interactions between "type of errors" and "Fixed vs Cumulative sample size", "type of errors" and "Random vs Adaptive stimulus sampling" as well as the possible 3-way interaction between all three variables.

All of these interactions were found to be significant (table 9.6).

Source	Chi-Square	df	p
T x (F vs C)	3.2615	2	.05
T x (R vs A)	25.4972	2	.01
T x (F vs C) x (R vs A)	46.1143	2	.01
Total	74.7830	6	.01

Table 9.6 Summary table of a factorial chi-square analysis of the factors "type of error" (T), "Fixed versus Cumulative" (F vs C) and "Random vs Adaptive" (R vs A)

(1) It was found that the Fixed sample groups (i.e. FR and FA) had significantly more "omissions" and fewer "additions" and "combined" errors than expected, while the Cumulative sample groups (i.e. CR and CA) had significantly fewer "omissions" and more "additions" and "combined" errors than expected.

Therefore there was a significant relationship between the two variables (i.e. "type of errors" and "Fixed vs Cumulative Sample size") (chi-square=3.26, df=2, $p<.05$).

(2) The Random stimulus sampling groups (i.e. FR and CR) had significantly more "omissions" and "additions" and fewer "combined" errors than expected. The Adaptive stimulus sampling groups (i.e. FA and CA) had fewer "omissions" and "additions" and more "combined" errors than expected. It can be concluded therefore that the two variables (i.e. "type of errors" and "Random vs Adaptive stimulus sampling") affect each other significantly (chi-square=25.50, df=2, $p<.01$).

(3) Finally, a significant 3-way interaction was revealed between the three variables (chi-square=46.14, df=2, $p<.01$). This means that the interaction of either factor (e.g. Fixed vs Cumulative) with the "type of errors" differed significantly as a function of the other factor (e.g. Random vs Adaptive).

Discussion

The error classification analyses revealed four main results. First, the "type of error" is generally systematic. Second, the majority of errors are "almost correct" responses. Third, there is a systematic relationship between "stimulus complexity" and "type of error". Finally, there are significant interactions between the factors involved in the modes of training and "type of error".

Error responses were systematic and "almost correct" responses. The implications of this finding become

particularly important if it is also recalled that the speed of errors was slower than the speed of correct responses (speed being expressed as mean RT, or mean residual value of RT, or even RT distribution). With these two points in mind, the following suggestion can be made with regard to the occurence of errors. Although subjects were trying to produce the correct response, for some reason they ran out of time and because of time pressure they produced a response, close to the correct one, but not quite. Most often, their response was either an "omission" or an "addition" of one key.

The occurence of more "omissions" than "additions" excludes the possibility that RT for errors was longer because subjects were depressing more keys. In fact, on average, they were depressing fewer keys than they should have done. Therefore the explanation of the increased RT for errors cannot be found by examining the "type of error" that occured.

It was found that the different factors involved in the training modes interacted significantly with the "type of errors" produced. However, it israther difficult to interpret these interactions in relation to the nature of errors.

To summarize the main findings and conclusion, errors are not random responses. Instead, they are systematically related to the stimulus. The fact that the majority of errors were "almost correct" responses possibly suggests that errors could have been avoided if subjects had taken a little longer before responding.

Chapter 10

PRESENTATION OF A NEW MODEL

10.1 Probability Functions

One way in which choice response tasks differ from each other is related to the emphasis placed on either speed or accuracy of performance. It can be assumed that this emphasis on either speed or accuracy varies along a continuum, with complete emphasis on speed at one end and complete emphasis on accuracy at the the other. Any speed-accuracy tradeoff can be represented as a point along this continuum. At one end of the continuum is the simple RT task in which all emphasis is placed on speed and at the other end is the "power" test in which only accuracy of performance matters. At any point along the continuum, tasks may vary as a function of a number of parameters (e.g. "difficulty", level of practice etc.). For example, a "power" test can be easy, in which case almost perfect accuracy can be achieved, or it can be so difficult that the probability of a correct response is very small. In both cases, however, there is the same emphasis on accuracy as opposed to speed.

Two probability functions are going to be considered: (a) the probability of a response being correct given that this response is made by time t, $P(C|R)$; and (b) the probability of making a response by time t, $P(R)$. Hence, as they are independent, the probability of making a correct response by time t, $P(C \cap R)$, will be the product of $P(C|R)$ and $P(R)$. I.e. $P(C \cap R)=P(C|R)*P(R)$

Simple RT Task

Let us consider first the case of a simple RT task in which subjects are expected to respond by pressing a single key every time a light is presented. In this kind of task, subjects do not have to discriminate between stimuli, they just have to decide whether a stimulus is present or not. If there is a stimulus they have to press the one and only key available.

The P(C|R) after the onset of the stimulus is 1, as only one response alternative exists (figure 10.1). However, it is possible for a response to be made before the onset of the stimulus and this is an error. Before the onset of the stimulus, the P(R) is relatively small and it increases sharply afterwards (figure 10.1). The P(R) function can be assumed to have an almost normal distribution about the mean RT. In a simple RT task, subjects will be a hundred per cent accurate if they wait until the stimulus is present before they make a response. If they respond before the stimulus is presented, this will be an error of anticipation. (If they respond very shortly after the stimulus is presented, though the response will be defined as correct, it will in fact be a response in anticipation of a stimulus.)

"Power" Test

Alternatively, let us consider the case of a multiple choice test in psychology which is given to psychology students under no time constraints whatsoever. In this case, speed is irrelevant as the main objective is to produce

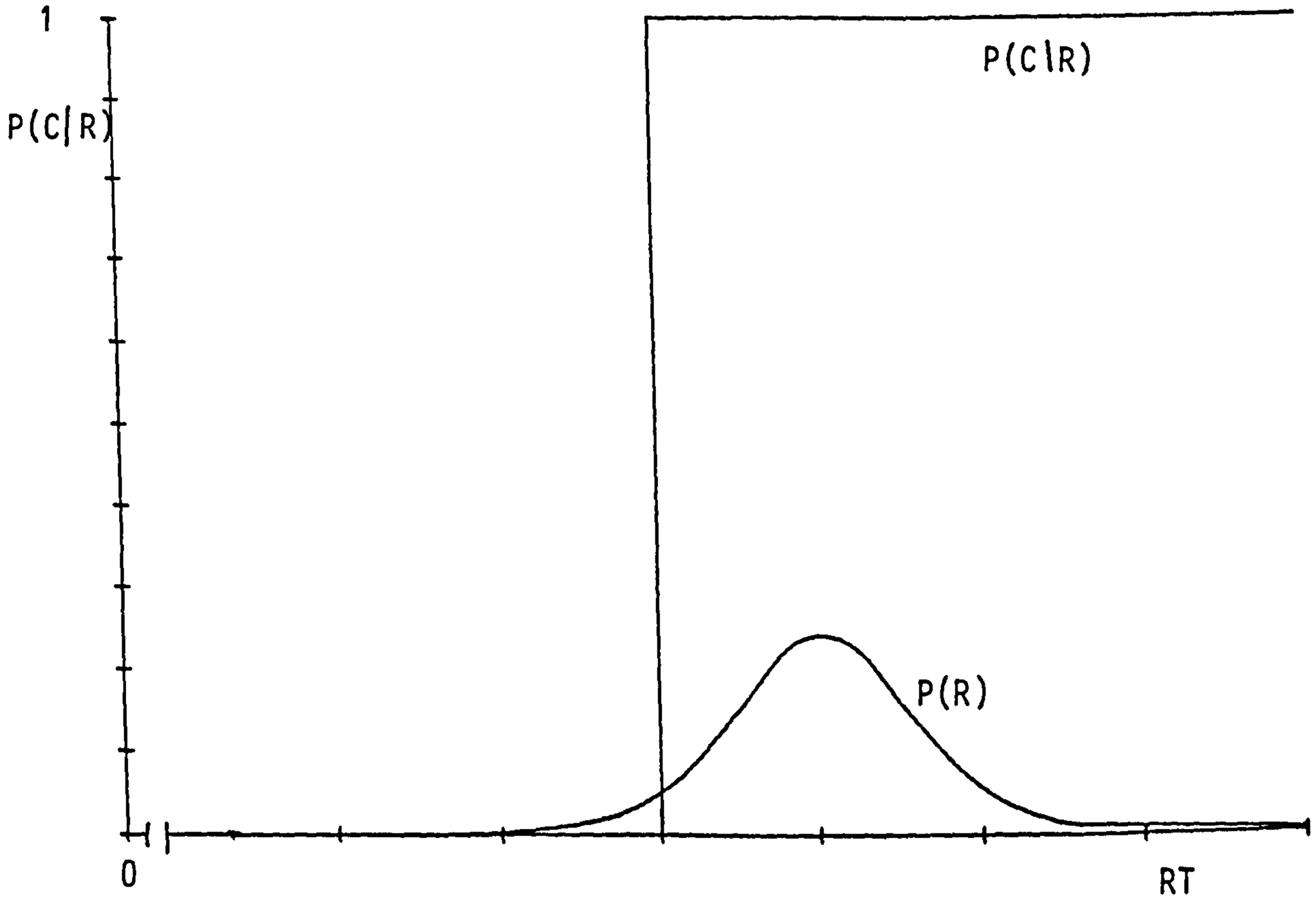


Figure 10.1 P(C|R) and P(R) Functions for a Simple RT Task

accurate responses. Subjects should respond only when they have accumulated sufficient information about the stimulus (the use of the expression "accumulation of sufficient information" is similar to that of the accumulator model and is discussed further later on).

The P(C|R) increases monotonically with time. At, and for some time after the onset of the stimulus, it is at chance level (figure 10.2) as information about the stimulus needs to be accumulated before the correct response is made. "Information about the stimulus being accumulated" does not refer only to the identification of the stimulus. It refers to the whole range of mental activities that take place from the moment a stimulus is presented until a response is made. This position is compatible with cases where the stimulus is difficult to identify but the required response is relatively easy, as well as cases where the stimulus is difficult but once it has been identified the response to it is easy. (The amount of information required is directly related to the "difficulty" of the stimulus. The stimulus "difficulty" depends on a number of factors which will be discussed later.)

As time elapses and more information is accumulated, the P(C|R) increases until it reaches a maximum level (its asymptote) of less than 1, as some of the multiple choice questions of the present example may exceed the students' knowledge and therefore some answers are expected to be wrong. In other words, errors of this kind arise from lack of knowledge and not from lack of time. These errors are

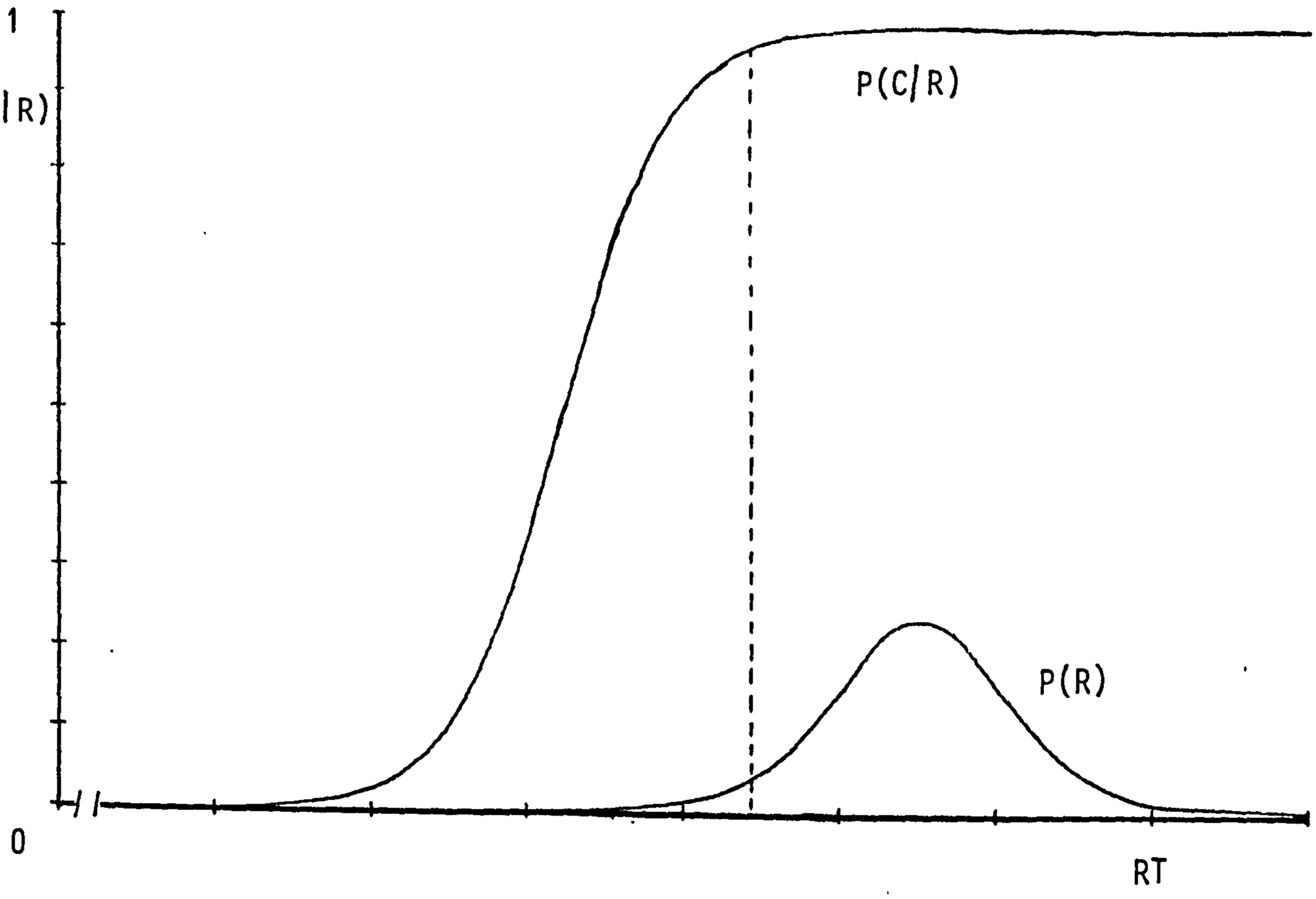


Figure 10.2 P(C/R) and P(R) FUnctions for a "Power" Test

different responses from the errors of anticipation mentioned above in connection to the simple RT task. A Logogen model type of explanation for these errors would be that subjects have not formed the appropriate memory representation (logogen). Therefore, despite the accumulation of information, the appropriate logogen cannot be activated and consequently the response is most likely incorrect.

For a task like the one in the present example, in which there is no speed pressure, subjects will wait until they have accumulated sufficient information about the stimulus before they make a response. In other words, most of the responses will be made around the point in time which corresponds to the maximum value of the P(C|R) function. The P(R) function is an approximately normal distribution, with most of the responses being made after the time t, which corresponds to the point at which P(C|R) reaches its maximum value (figure 10.2).

Typical Choice Response Tasks

For typical choice response tasks, speed-accuracy tradeoff functions can be varied to lie somewhere between the two extremes on the continuum from speed only to accuracy only. The present model attempts to explain the relationship between speed and accuracy for all these tasks.

In a typical choice response task where subjects are instructed to respond as fast and as accurately as possible, the P(C|R) increases monotonically with latency (figure 10.3). If a response is made before the onset of a stimulus, then this might be considered as an anticipation response.

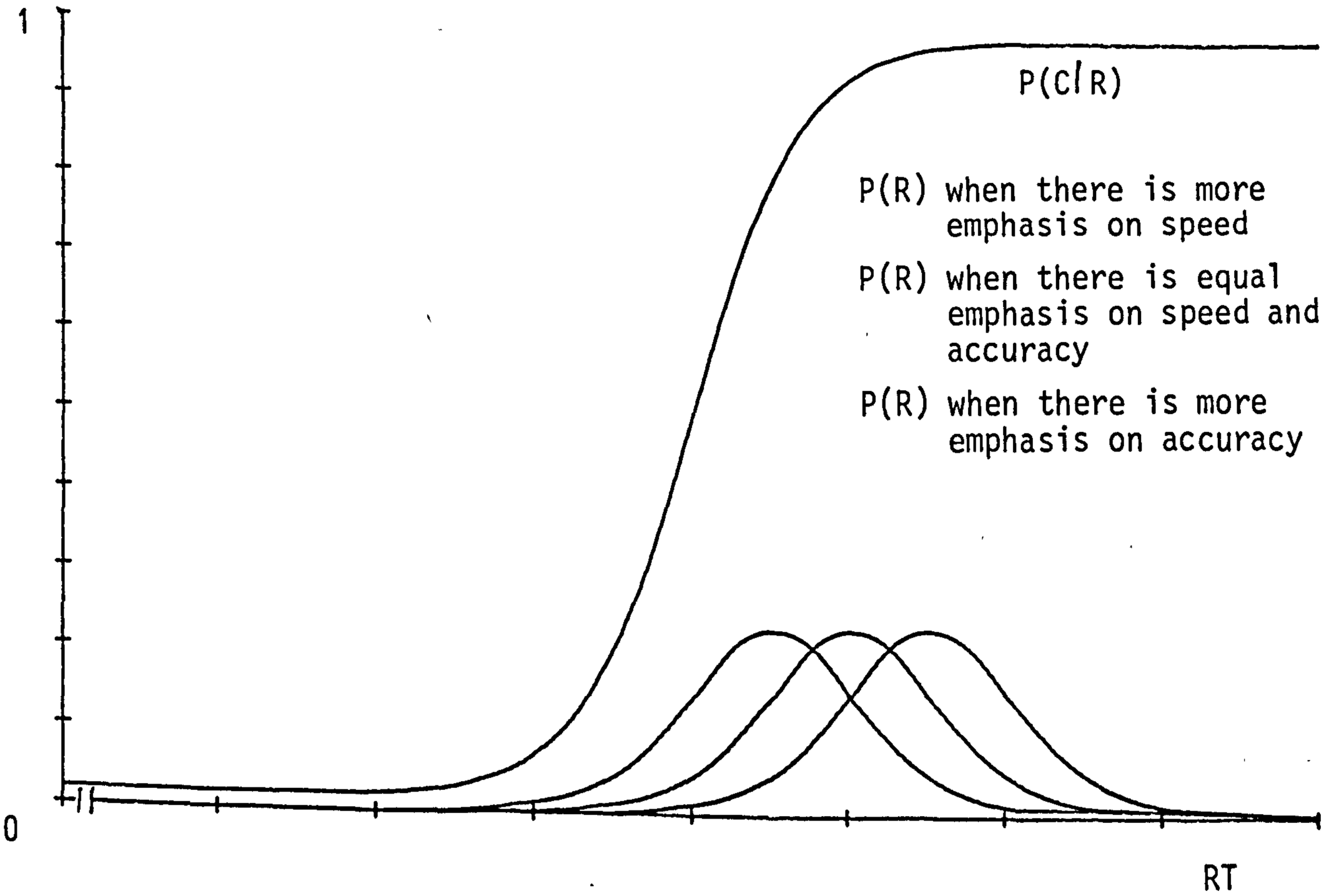


Figure 10.3 P(C|R) and P(R) Functions for a Typical Choice Response Task

Depending on the task, such a response might be registered as an error, correct (i.e. lucky guess) or not registered at all. Very soon after the onset of the stimulus and consequently before enough information has been accumulated, the probability of a response being correct is at chance level. If a response is made at this stage, it may be considered to be a fast-guess response (i.e. response initiated without adequate processing of the stimulus - as defined by the Fast-Guess model). As information is accumulated and more time is taken, the P(C|R) increases until it reaches a maximum level. The P(C|R) can be very high if enough time is taken before responding and if the task is "easy". However, because of time pressure a response may be made before the asymptote and, as a result, the probability of an error would be increased. The subjects' aim is to achieve an optimum balance between speed and accuracy. The way subjects actually perform in relation to the P(C|R) function determines their P(R) distribution (figure 10.3).

The P(R) distribution can be manipulated by the instructions given to subjects. Depending on the emphasis for speed and accuracy, as well as the relative payoffs a P(R) distribution may vary greatly from task to task. It is approximately normal if both speed and accuracy are equally important. If there is great emphasis on speed, the P(R) distribution may be skewed to the left as more responses will be made before subjects have accumulated sufficient information. On the other hand, if there is greater emphasis on accuracy, the P(R) may be shifted to the right. If there is

pressure for high accuracy and a time limit by which subjects must respond the P(R) distribution may be skewed to the right, as subjects will take longer, trying to accumulate as much information as possible before they respond (figure 10.3).

10.2 Probability of a Response Being Correct if it is Made

For a given stimulus, for a given subject, the P(C|R) is a function of the amount of information accumulated by time t. As there is variability in the rate of information accumulation, the P(C|R) for a given stimulus, for a given subject, on a particular trial varies as well. These variations will lie in an approximately normal distribution. The "general" P(C|R) function can be considered to represent a line passing through the means of these distributions for each point in time from t=0 to infinity (figure 10.4). It must be noted that the variance about the P(C|R) function varies with time. In the beginning there is zero variance; it increases with time until it reaches a maximum point and then it decreases again to almost zero as the P(C|R) function reaches its maximum level.

The variance about the P(C|R) function for any given time t is a function of stimulus "difficulty" (i.e. stimulus discriminability, stimulus complexity, etc.). As stimulus "difficulty" increases, the variability of the rate of information accumulation increases as well. In other words, for simple stimuli, the variance about the P(C|R) function is very small (figure 10.5a) whereas it is quite big for more complex stimuli (figure 10.5b). The effect of this variance

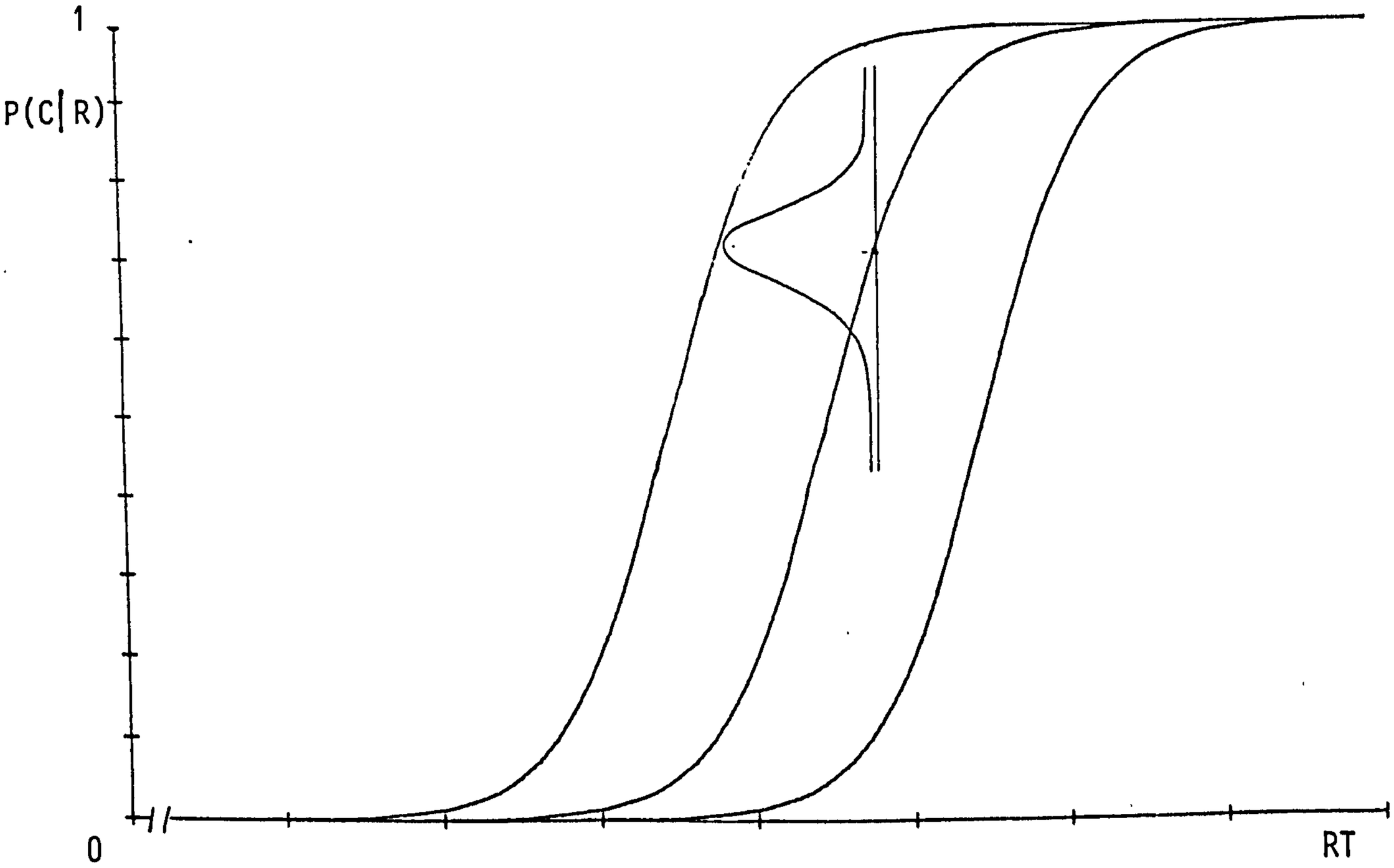


Figure 10.4 Distributions of the Amount of Information Accumulation the Means of Which Define the P(C/R) Function

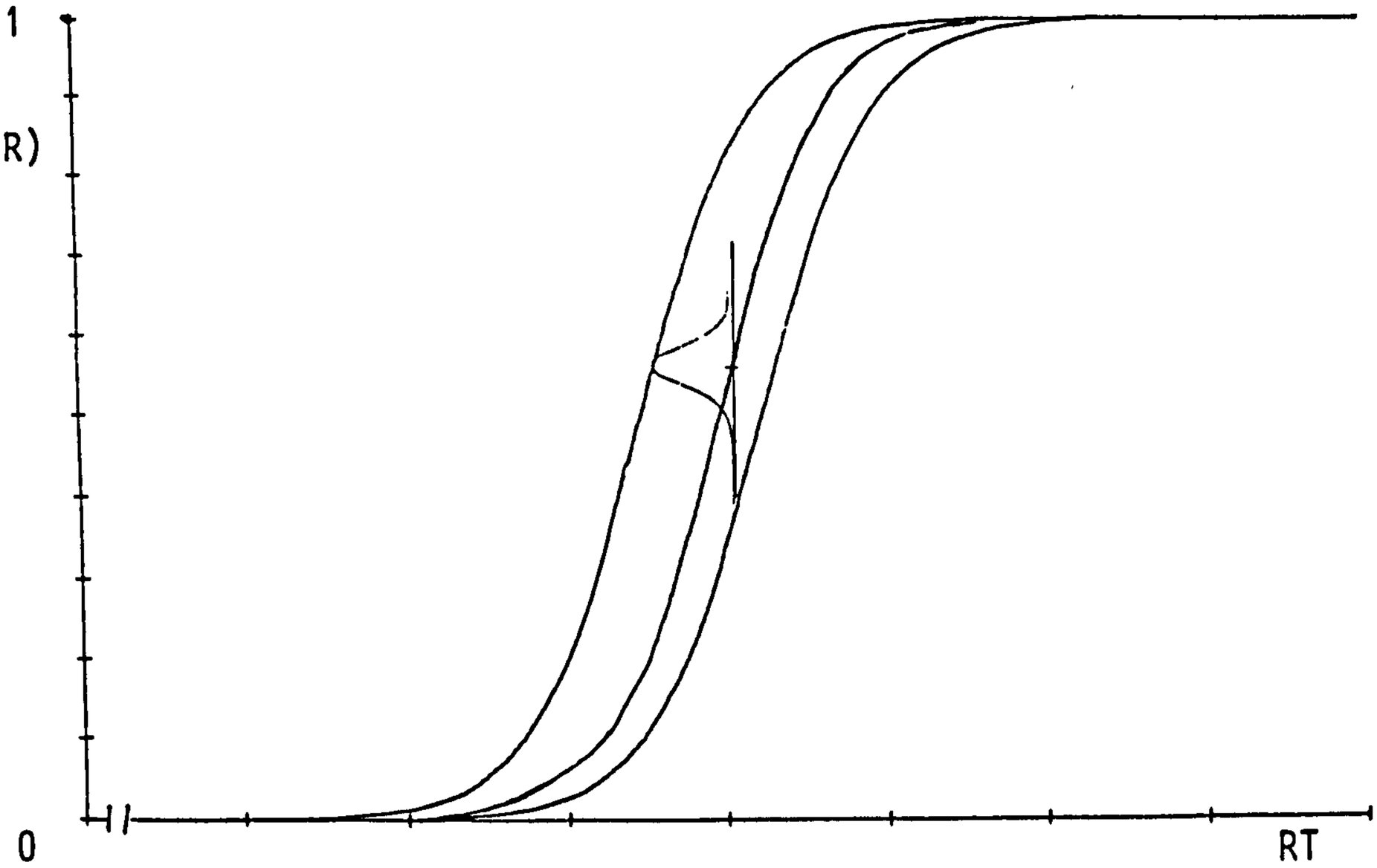


Figure 10.5a P(C|R) Functions for "Easy" Stimuli

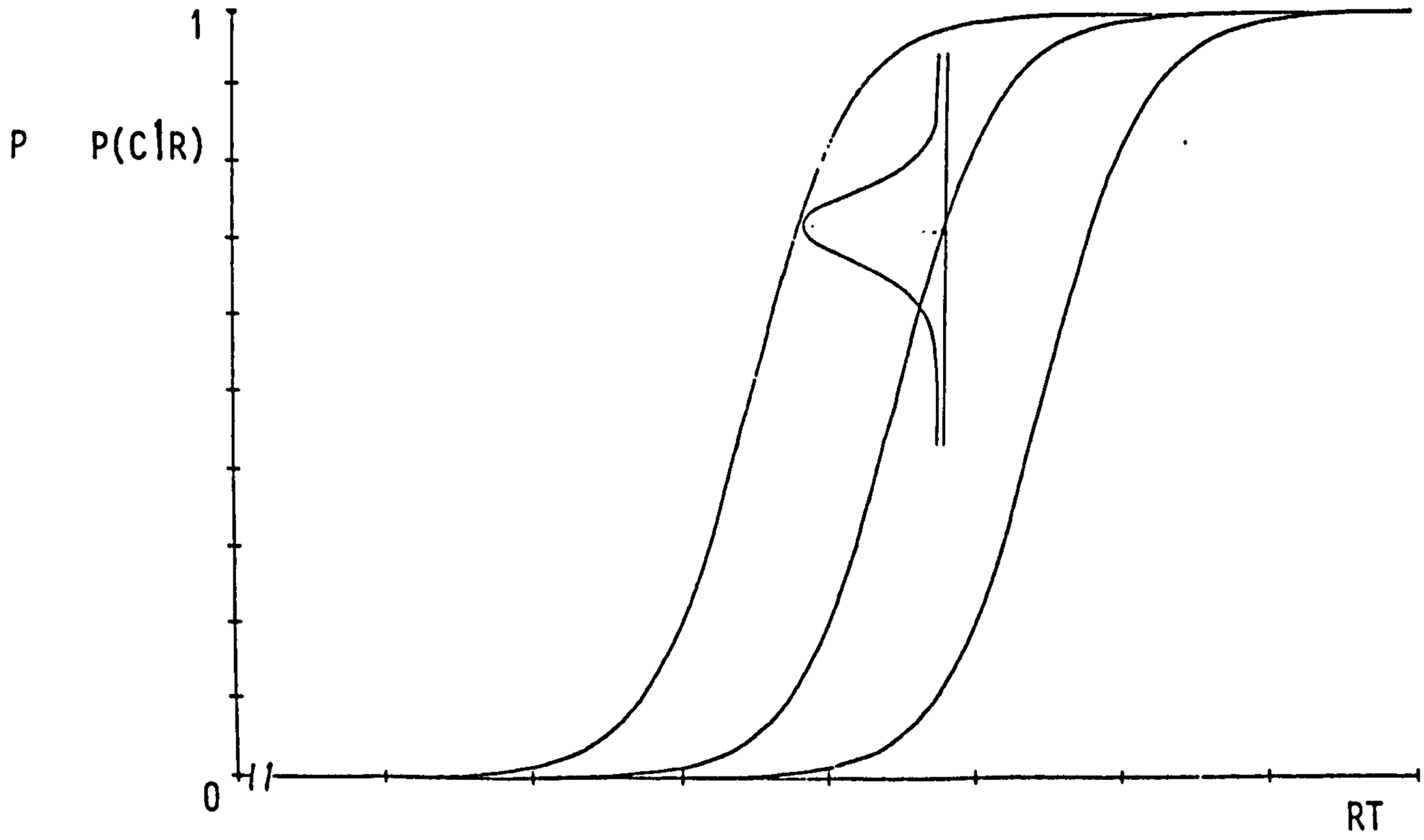


Figure 10.5b P(C|R) Functions for "Difficult" Stimuli

about the P(C|R) function on subjects' performance is going to be discussed further later on. For the time being, the "mean" P(C|R) function is discussed as if it were a single curve.

A. Fast-Guess Responses

The first part of the curve of the P(C|R) function represents fast-guess responses (figure 10.3). If a fast-guess response is made its probability of being correct depends on the number of response alternatives. If there is only one response alternative (i.e. in a simple RT task), the probability of a fast-guess response being correct is one. If there are two equiprobable alternative responses the probability of a fast-guess being correct (if such a response is made) decreases to 0.50. If there are four equiprobable response alternatives, it is 0.25. Generally, the probability of fast-guesses being correct (if these responses are made) is determined by the number of response alternatives and their relative frequencies. For equiprobable numbers of alternatives, if a fast-guess response is made, its probability of being correct follows the function of figure 10.6. It follows that in complicated tasks with many response alternatives, if a fast-guess is made, its probability of being correct is negligible.

B. Slope of the Curve of P(C|R)

The slope of the P(C|R) function depends on the "difficulty" of the stimulus. The term "difficulty" incorporates a number of variables which are known to affect speed and accuracy of performance on a choice response task

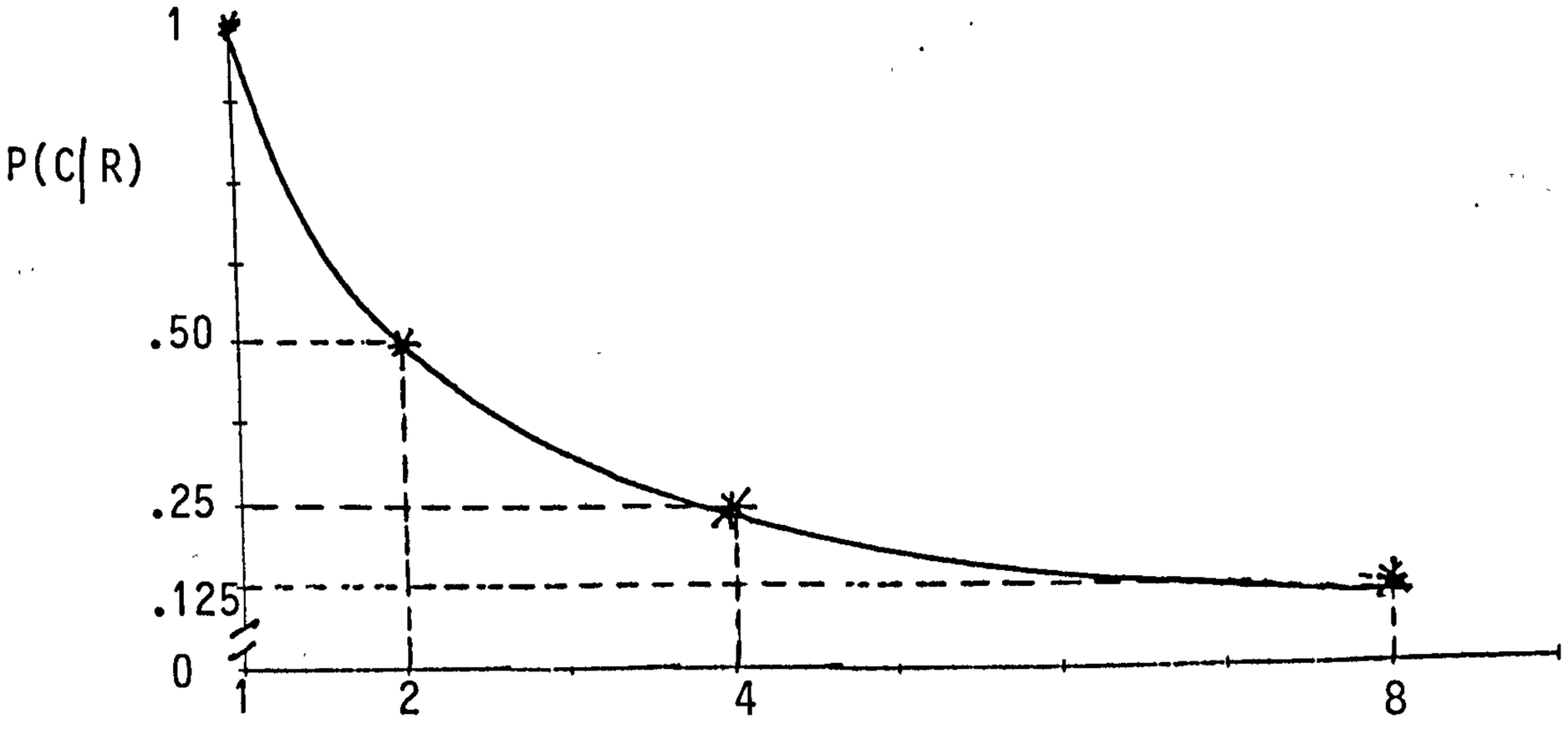


Figure 10.6 P(C/R) of Fast-guess Responses as a Function of the Number of Equiprobable Response Alternatives

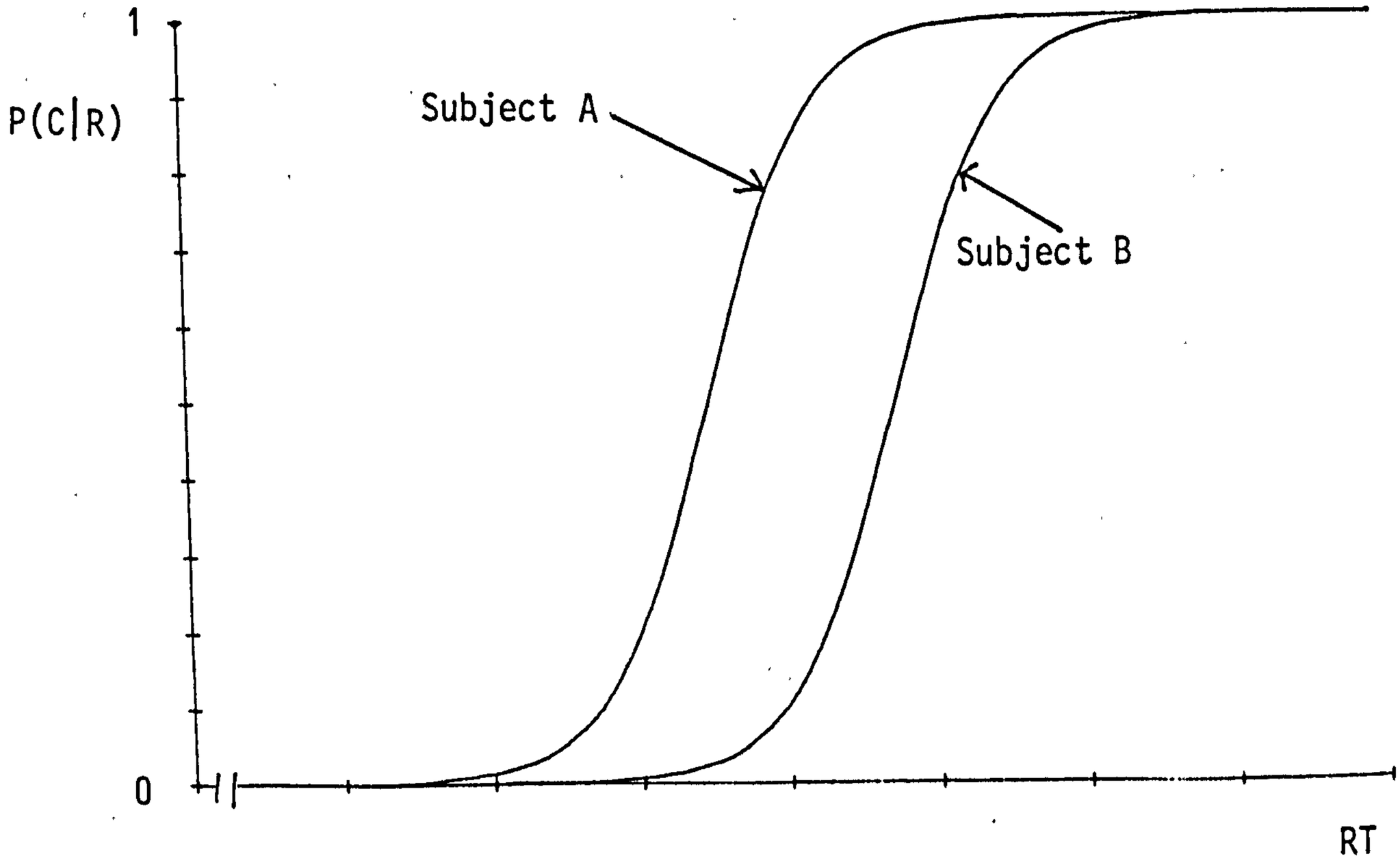


Figure 10.7 P(C/R) Functions for Subjects A and B, Where Subject A Has Received More Practice than Subject B

(e.g. practice, number of stimulus-response alternatives, probabilities, stimulus-response compatibility, stimulus discriminability, etc). The way each one of these variables affects the slope of the P(C|R) function can only be operationally defined for each particular task. However, in theory and assuming that everything else is equal, each one of the above variables affects the slope of the probability function in the following way.

(1) Practice

More practised subjects can accumulate information about a stimulus more efficiently (i.e. at a faster rate) than less practised subjects. Generally, increased levels of practice lead to a faster increase (i.e. steeper slope) in the P(C|R). For example, if subject A has had more practice than subject B, the P(C|R) for A increases more steeply and reaches its maximum point before that for subject B (figure 10.7).

(2) Number of Stimulus-Response Alternatives

(a) Equiprobable Alternatives: The bigger the number of stimulus-response alternatives the lower the probability that fast-guess responses are going to be correct if they are made, and the longer it takes for the P(C|R) function to increase to maximum level. Let us consider stimuli S1, S2, S3, ..., Sn which are selected from sets with 1, 2, 3,..., n number of equiprobable stimulus-response alternatives. When S1 is presented (i.e. simple RT task) the P(C|R) is one. For stimulus S2 which is selected from a set with only two equiprobable stimulus-response alternatives, the P(C|R) upon the onset of the stimulus is 0.50. For a stimulus S4 which is

selected from a set with four equiprobable stimulus-response alternatives the P(C|R) upon the onset of the stimulus is 0.25. These probability functions increase with time until they reach their maximum value (figure 10.8). For a given value of P(C|R) which is equal to Px, the latencies associated with these probability functions will be t1, t2, t3, ..., tn (where, t1<t2<t3<...<tn). The bigger the number of equiprobable stimulus-response alternatives, the lower the starting point of the P(C|R) functions.

(b) Alternatives with Unequal Probabilities of Occurence: If a given stimulus, Sa, has a higher probability of being presented than another stimulus Sb, it follows that stimulus Sa receives more practice than stimulus Sb. As has been argued above, increased levels of practice lead to steeper increases of the P(C|R) functions. Therefore, the P(C|R) for Sa is higher than the P(C|R) for Sb. Generally, the higher the probability of a stimulus being presented the faster the increase of the P(C|R) function and the higher the guessing rate for this stimulus.

(3) Stimulus-Response Compatibility

The more compatible the stimulus-response situation, the faster the rate of accumulating information about the stimulus. It follows therefore that an increase in S-R compatibility leads to a steeper increase in the P(C|R) function. To illustrate the relationship let us consider two stimuli Sa and Sb, for which the S-R compatibility for Sa is higher than that for Sb. The curve of P(C|R) for Sa starts at

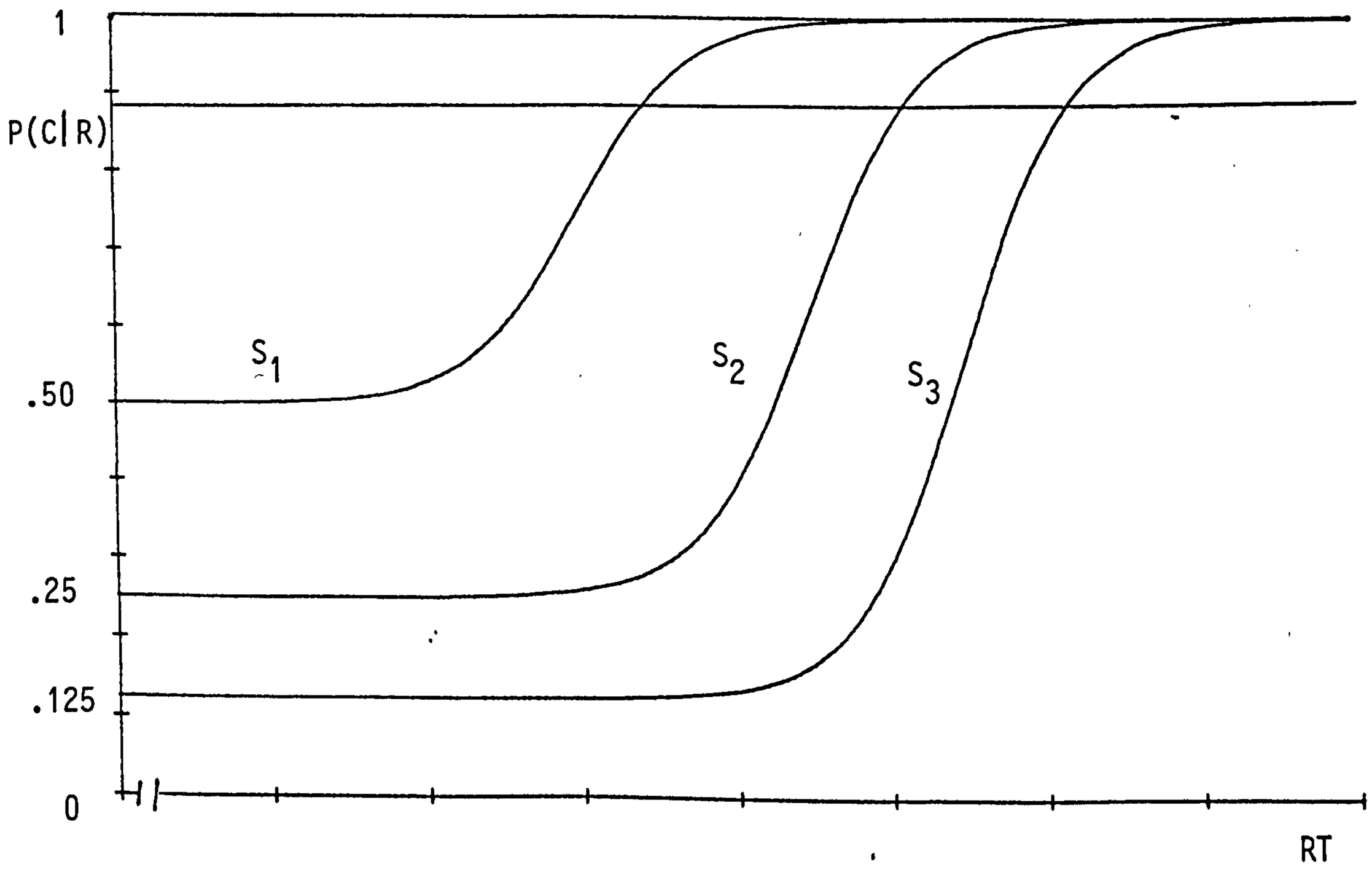


Figure 10.8 P(C|R) for Stimuli S_1, S_2, S_3,..., S_n Which Are Selected from Sets with 1,2,3,..., n^n Equiprobable Stimulus Alternatives

the same point but it reaches the maximum level faster than the P(C|R) for Sb. It can be concluded that the relationship between S-R compatibility and P(C|R) is analogous to that of practice and P(C|R), in other words the higher the S-R compatibility the steeper the increase of the function.

(4) Stimulus Discriminability

When there is a high signal-noise ratio, then it can be said that there is high stimulus discriminability. Generally, the higher the stimulus discriminability, the higher the rate of accumulating information about the stimulus and therefore the steeper the P(C|R). For two stimuli Sa and Sb, where Sa is easier to discriminate than Sb, the function of P(C|R) for Sa should be steeper than the function P(C|R) for Sb. Generally, the better the stimulus discriminability the faster the increase of the P(C|R) function.

(5) Stimulus Confusability

Stimulus confusability refers to the degree to which stimuli are confused with each other because of similar characteristics. The higher the confusability between stimuli, the slower the rate of accumulating information and therefore the shallower the function of P(C|R). Let us consider two stimuli Sa and Sb, where Sa is easily confused with other stimuli from its set whereas Sb is very distinct. The P(C|R) function for Sa is expected to be shallower than the P(C|R) for Sb. In summary, stimulus confusability affects the increase of the P(C|R) function.

The variables which have just been discussed are some of

the most common to affect the P(C|R) function. Depending on the specific experimental task it is possible that other variables affect the difficulty of a stimulus (and therefore the slope of the P(C|R) function). For example in the task of Experiment 4, the number of keys involved in each stimulus does have a considerable effect on the "difficulty" of the stimulus and, consequently, it would have affected the slope of P(C|R).

Some of the variables which affect stimulus "difficulty" are related to the stimulus itself (e.g. stimulus discriminability); others are related to the processing which takes place before responding (e.g. S-R compatibility); still others are related to the response production (e.g. response complexity). The effects of all these variables on the P(C|R) function may be additive and/or interactive. (For example, according to Stenberg (1969) - who assumes the existence of processing stages - these effects are additive.) This implies that, in this case, the slope of P(C|R) and the way it varies will be much more difficult to estimate. It will require an examination of the effect of each variable as well as an examination of the effect of the exact interactions between these variables on the P(C|R) function.

C. Maximum Level of Accuracy

After a certain point in time the increase in the P(C|R) function is very small when considered in relation to the additional time it has taken. In other words, the P(C|R) function follows what is sometimes referred to as the "law of

diminishing returns". The increase in the function becomes smaller until the function eventually reaches its asymptote. This maximum value of accuracy is the third parameter which determines the curve of P(C|R). It is directly related to the threshold value for the given task and can be defined as the maximum accuracy which can possibly be achieved on a given task if there is no time constraint on subjects to perform. For those tasks on which there are speed constraints maximum accuracy will hardly ever be achieved.

10.3 Probability of a Response Being Made

As already mentioned, in a typical choice response task, the subjects' aim is to achieve an optimum balance between speed and accuracy. The balance they adopt between speed and accuracy is reflected in the P(R) function, which describes how the probability of a response being made varies as a function of time. The P(R) is affected by two main sets of variables: (1) variables related to the way subjects should perform (e.g. instructions about the emphasis on speed and accuracy, payoffs); and (2) variables related to the level of subjects' "confidence" at the time of their response.

(1) Speed-Accuracy Tradeoff Functions

Two major variables affecting the P(R) distribution are related to the emphasis placed on speed as opposed to accuracy by the instructions and the consequences of the speed-accuracy tradeoff which is adopted by the subject. If there is greater emphasis on speed, most responses will be fast. Therefore, the P(R) distribution should be either skewed or shifted to the

left, depending on the penalties which subjects have to pay for slow or inaccurate performance. By analogy, if there is greater emphasis on accuracy, subjects will take longer to achieve higher accuracy. So, the P(R) distribution is expected to be either skewed or shifted to the right. When speed and accuracy are equally important, the P(R) distribution should be approximately normal. Irrespective of the skewness of the distribution, its mode represents the point in time at which subjects place the cutoff point. Up to a certain point during a trial, subjects wait before they make a response because the rate of accumulating information at that level is so fast that there will be a big gain in accuracy in a short time. From a certain point onwards the gain in accuracy is so small that it does not justify the extra time taken.

The optimal cutoff point represents the ideal point in time when subjects must respond. If this point is equal to A, the curve of P(R) should be like the one shown by Optimal P(R) (figure 10.9). Subjects' aim should be to approximate this ideal function (i.e. zero variance with a mean at the asymptote). Although this ideal function would be impossible to achieve, the actual P(R) function should be as close as possible to it (figure 10.9).

(2) Subjects' Level of "Confidence"

The probability that subjects will make a response by a given time t also depends on the P(C|R) function and its variability as perceived by subjects themselves. In other words, it depends on the extent to which subjects feel "confident" that they have assessed their "odds" correctly.

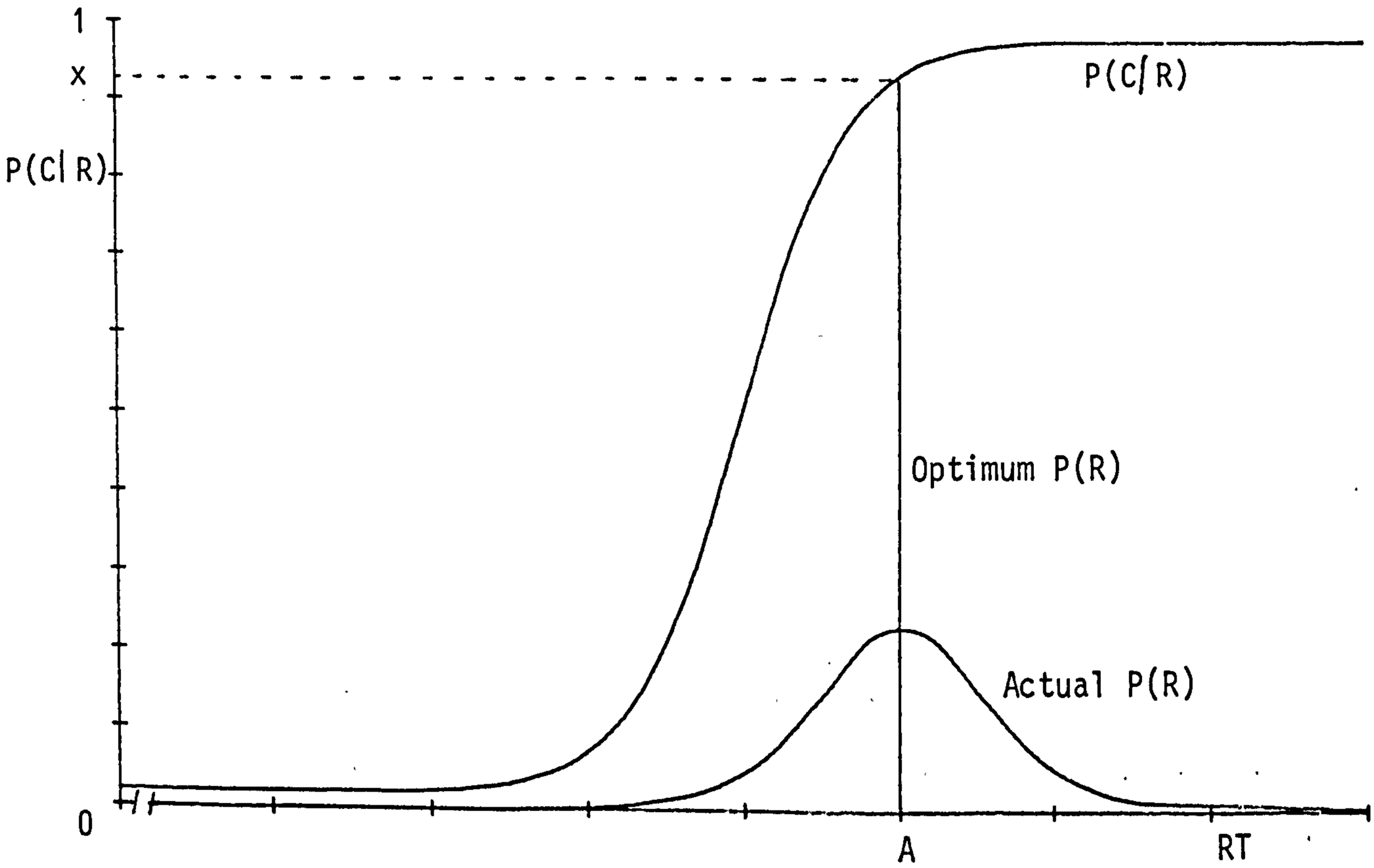


Figure 10.9 Optimum and Actual Functions of P(R) when the Ideal Cutoff Point Is x

The term "confidence" is used to refer to the extent to which subjects "feel" that they would be correct if they responded. After a stimulus has been presented, the P(R) is relatively small as subjects' "confidence" that they can make the correct response is very low. As time elapses and more information is accumulated, subjects become increasingly "confident" about the way they should respond. Therefore, the probability that they do respond increases.

It is reasonable to suppose that subjects' rate of increase in "confidence" is closely related to the moment to moment changes in P(C|R). When the P(C|R) increases quite slowly, subjects' "confidence" increases slowly and more time is needed to accumulate sufficient information before they feel confident enough to respond. On the other hand, when the rate of information accumulation is quite fast and the increase in the P(C|R) quite steep, subjects' "confidence" will increase rapidly. In other words, subjects' "confidence" functions can be treated as their internal representation of P(C|R).

Subjects' levels of "confidence" are not constant from trial to trial. They are affected by a number of variables and, by implication, so are the P(R) distributions. On one hand, "confidence" may vary as a result of variables such as the level of practice or stimulus "difficulty". Such variables affect the slope of the "confidence" function. For example, on a trial where the stimulus is relatively unpractised and/or "difficult" subjects' levels of "confidence" are expected to

be low as opposed to those on a trial where the stimulus is relatively practised and/or "easy". On the other hand, "confidence" may vary as a result of subjects' increased "cautiousness" on a trial, as a result of variables such as their performance on the preceding trial. For example, if an error was made on the preceding trial, it would be expected that subjects become more cautious on the current trial. Therefore, on the current trial, subjects would need to accumulate more information, and by implication take longer, before they feel "confident" enough to respond. This would explain why responses following errors are generally slower even though the P(C|R) function on each trial should be independent.

In summary, a subject's P(R) is affected by: (1) variables which define the speed-accuracy tradeoff function, as they determine the subject's choice of the cutoff point; and (2) the subject's level of "confidence", which is related to the way the subject, him/herself, assesses the P(C|R).

10.4 Choice of the Cutoff Point

It has been argued that subjects learn to adjust their P(R) so as to achieve the optimum balance between speed and accuracy. This is achieved by discovering, for each trial, the optimal cutoff point after which gains in accuracy do not justify the extra time required, but before which the level of error rate is unacceptable.

Although the optimal cutoff point should ideally be just a point, this is not the case in reality. Because the choice

of the cutoff point cannot be perfect it must be considered as an "approximation" to the optimal one. If it is postulated that for a particular subject, for a particular stimulus, at a particular level of practice, the optimal cutoff point is point A (figure 10.9), the subject should, generally, make a response as close as possible to the point in time that corresponds to A. However, because of "imperfect" performance, their response is expected to lie on a distribution around this point in time. Since there is time pressure subjects usually respond earlier rather than later than this point. Therefore, the response distribution will become skewed to the left (figure 10.9). The variance of this distribution should decrease as subjects learn to respond more efficiently (i.e. with practice).

The choice of the cutoff point is directly related to subjects' "confidence". As already discussed, subjects' "confidence" refers to the way they "feel" about the P(C|R) and is affected by a number of variables (e.g. long-term practice, stimulus "difficulty", performance on the preceding trial, etc.). This suggests that the effect of these variables on "confidence" affects the choice of the cutoff point on a given trial.

The Long-Term Effects of Practice on Choice of the Cutoff Point

The P(C|R) function for any stimulus varies as a function of practice as already discussed. This means that subjects have to assess continuously the P(C|R) and adjust the P(R) according to the new P(C|R). Another way of describing this

process was actually noted by one of the subjects. He said that when he saw the stimulus he "knew" if it was an "easy" or a "difficult" one and whether he could produce a quick and correct response or needed to take longer. In other words, the subject was claiming that at the start of the trial, he was making a judgement about the "difficulty" of the stimulus (i.e. assessing the P(C|R)) and proceeded with the actual response when he felt sufficiently "confident" about his response. However, it cannot be expected that such judgements can be made reliably early in practice.

Early in practice, subjects are not yet in a position to assess the P(C|R) function efficiently and variations in their judgement and "confidence" may be relatively uncorrelated with variations in difficulty. This suggests that the choice of the cutoff point will be inefficient at a low level of practice. In addition, it would be expected that, because of lack of experience, subjects may be cautious, therefore needing to accumulate more information before they are "confident" enough to respond.

As subjects receive more practice, through trial and error, they learn to assess the P(C|R) function more efficiently. Subsequently, their ability to judge difficulty and set realistic levels of "confidence" increases and the choice of the optimal cutoff point becomes more precise. In other words, the variance of the response distribution around the optimal cutoff point decreases.

In summary, it can be argued that practice gives rise to

a mechanism (judgement of difficulty and realistic level of "confidence") which allows subjects to assess, increasingly more efficiently, the probability of making a correct response if they make a response at all, within a certain time. This mechanism makes it possible for subjects to respond so as to achieve the optimal combination of speed and accuracy.

The Effect of Stimulus "Difficulty" on the Choice of the Cutoff Point

The P(C|R) function of a "difficult" stimulus and the P(C|R) function of an "easy" stimulus differ from each other in the steepness of the slope. Irrespective of the source of stimulus "difficulty", when the stimulus is "difficult" the P(C|R) function is more shallow than when the stimulus is "easy". Subjects' general level of "confidence" also varies with stimulus "difficulty". It is higher for "easy" stimuli and lower for "difficult" stimuli.

When a stimulus is "easy", subjects will rapidly reach high levels of "confidence" and will make their response relatively fast and accurately. On the other hand, when the stimulus is "difficult", subjects will be less "confident" about their ability to make a fast and accurate response. They take longer and at the same time they anticipate a lower level of accuracy. Let us discuss an example to illustrate and clarify this point. Let us assume that, in a choice response task, a subject has established an average RT of 1.2 secs and an average error rate of 5%. This average 5% error rate will not be the same across all stimuli, irrespective of "difficulty". For very "easy" stimuli the error rate could be

as low as, say, 1% and for very "difficult" ones it could be as high as 9% (table 10.1). However, the point is not only to maintain an average level of accuracy but also to maintain control over variance. Subjects are in a position to adjust their performance so that they have a 5% error rate for "easy" as well as for "difficult" stimuli, in which case there is no error variance across stimuli. The former tradeoff function would result in faster mean RTs than the latter. However, it would be unacceptable to have such high error rates for "easy" stimuli and such slow latencies for "difficult" ones. Therefore, a "compromise" tradeoff function is achieved with something like 3% error rate for "easy" stimuli and 7% for "difficult" ones.

Level of Difficulty	"Easy"		"Difficult"		Overall	
	% Err	RT	% Err	RT	% Err	RT
P(C|R) (without attending to error variance)	99%	1.2	91%	1.2	95%	1.2
P(C|R) (when error variance is zero)	95%	1.0	95%	2.6	95%	1.9
P(C|R) (while compromising speed of performance and error variance)	97%	1.1	93%	1.7	95%	1.4

Table 10.1 Speed-accuracy tradeoff functions for two different levels of difficulty and a "compromise" between speed of performance and error variance

Performance Variations from Trial to Trial

As can be inferred from the above discussion, the cutoff point varies from trial to trial as the P(C|R) function and the level of subjects' "confidence" vary from trial to trial. It must be mentioned that in tasks where the stimulus "difficulty" varies from trial to trial and/or individual stimuli receive different amounts of training, there is an interaction between the effects just discussed and the choice of the cutoff point. In this case, learning to choose the cutoff point is more complicated and requires more extensive practice.

To summarise, the production of a response is determined by a cutoff point which is manipulated by the subject so as to maintain an optimum balance between speed or accuracy. The cutoff point is chosen by subjects for each individual stimulus on each trial. The main factors which affect this choice are: (1) the emphasis on speed and accuracy; (2) the payoffs in terms of speed and accuracy which are considered by continuously assessing the function of P(C|R); (3) the level of practice which subjects have had; (4) the relative "difficulty" of each stimulus and therefore the optimum speed and accuracy expected (by subjects) for that particular stimulus; (5) the general level of "confidence" the subject has about his/her performance on the task.

The Occurrence of an Error

Questions related to the nature and sources of errors

have been asked for a long time and various explanations have been presented throughout the literature. The present model attempts, and hopefully succeeds, to provide some answers. An a priori assumption is that the stimuli presented to subjects are within the range of subjects' capacities. For example, if the task involves the translation of words from Chinese to English, no matter how long English-speaking subjects take to respond, they cannot give an accurate answer if they do not speak Chinese.

Provided that the above assumption is true in a particular situation, three different kinds of errors may occur:

(1) Unintentional Movements: Error responses may occur as a result of unintentional movements. For example, as a subject has his/her fingers over the keyboard (as was the case in Experiment 4) one of the hands may slip and a key may be depressed accidentally. This accidental movement is recorded as an error response. This kind of error might occur at any time until the subject executes the "chosen" response. In a typical choice response task, a small proportion of such errors would be expected.

(2) Fast Error Responses: A response may be made before the stimulus is adequately processed and an error might occur as a result (i.e. the type of fast-guess errors discussed by the Fast-Guess model). These errors would be generally expected in those choice response tasks which involve easily distinguishable stimuli (e.g. a simple 2-choice response task) where there is greater emphasis on speed. In this case

subjects' "confidence" does not vary greatly from trial to trial as stimulus "difficulty" is fairly constant and consequently the P(C|R) function is almost the same from trial to trial. In other words, it can be assumed that the two curves which represent the limits of P(C|R) for high and low stimulus difficulty respectively have come so close to each other that they practically ovelap (figure 10.10). It follows that, in this case, when subjects respond faster they are more likely to make an error, whereas the longer they take, the more likely they are to make the correct response. In other words, such errors are faster than correct responses.

(3) Slow Error Responses: When stimulus "difficulty" varies from trial to trial (as in Experiment 4), so do the P(C|R) function and the related "confidence" function. Let us assume that for an "easy" stimulus, A, the probability of a correct response being produced is P (figure 10.11). For a more "difficult" stimulus, B, there are three options a subject may choose.

(1) The subject may respond by time t1 with a low probability, P1, of being correct. This option produces unacceptable levels of accuracy (too many errors).

(2) The subject may respond by time t2 with a probability of P (the same as for the "easy" stimulus A) of being correct. This option produces unacceptable levels of speed (too slow).

(3) The subject may respond by time tc (where $t1<tc<t2$) with a probability of Pc of making a correct response (where

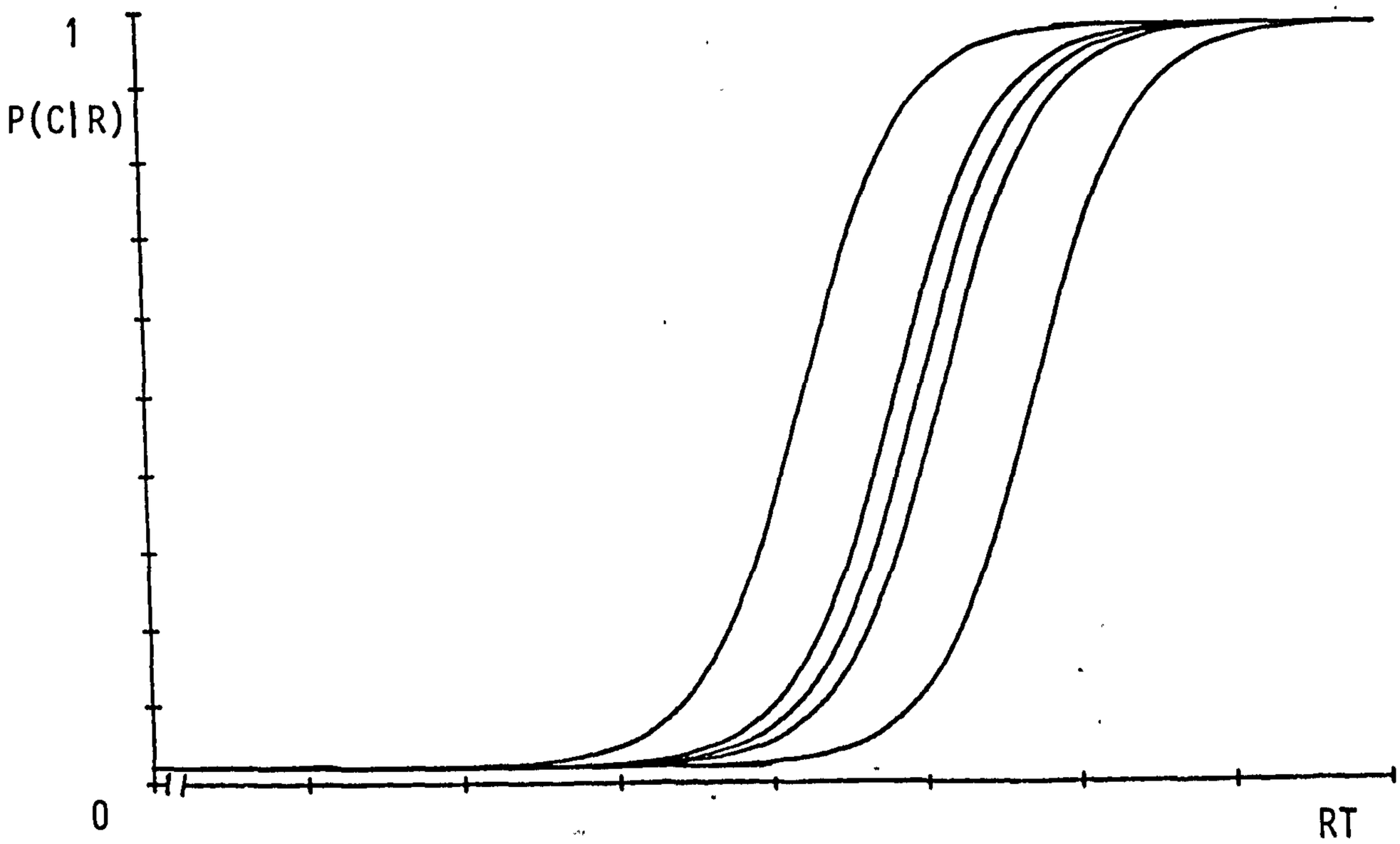


Figure 10.10 P(C|R) Functions as Affected by Subjects' "Confidence". (When "confidence" is constant from trial to trial, the P(C|R) is a single curve)

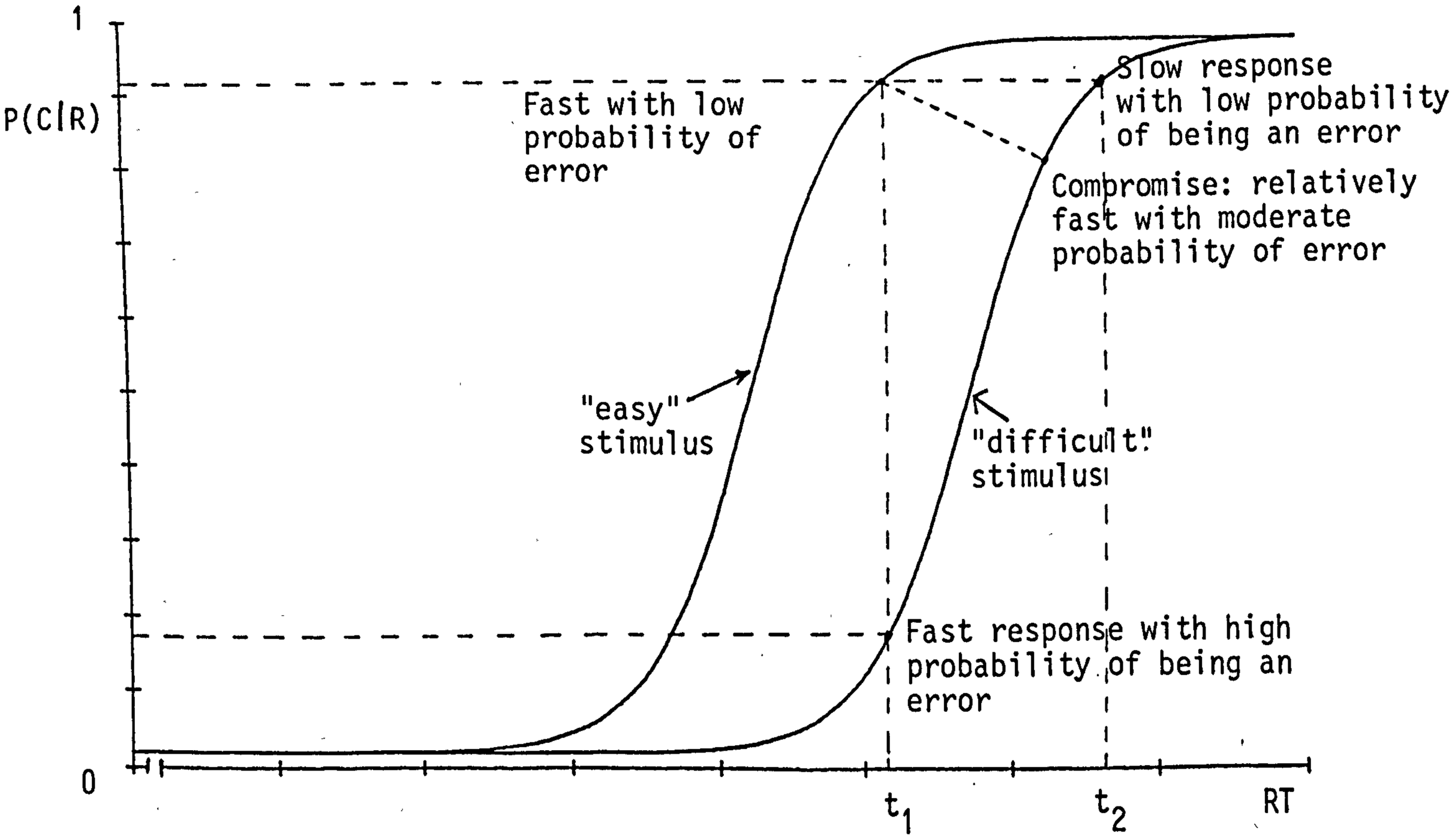


Figure 10.11 P(C/R) Functions and Cutoff Points for "Easy" and "difficult" Stimuli and Various Speed-Accuracy Tradeoffs

P1<Pc<P). This option represents the "compromise" which has been discussed above and it has been argued that this is the option subjects tend to choose.

Errors made when such a "compromise" has been chosen are slower than correct responses such as that to stimulus A, when the stimulus is "easy" and the subject can accumulate information about it fast and accurately. Subjects have to resort to such "compromise" solutions during relatively difficult tasks (as the task of Experiment 4) when their "confidence" varies from trial to trial.

Although task difficulty is a major factor that affects the relationship between the speed of correct and error responses, it is not the only one. Even when the stimulus difficulty is the same on two different trials, in one case the response might be correct and in the other it might be an error. It is argued that if the task from which the stimulus is selected is difficult, an error will be generally slower than a correct response to the same stimulus (or a stimulus of equal difficulty). The explanation for this lies with the concept of "confidence". On one occasion, the subject may feel "confident" about a stimulus S and therefore respond quickly and accurately. On another occasion, however, the subject may feel much less "confident" about the same stimulus S. There may be a number of possible reasons for variations in "confidence": the subject has made a fairly high number of errors during the last trials, or in the previous trial when the subject felt quite "confident" but the response was an error, or the subject's attention was momentarily distracted

and therefore the response had to be produced under a greater time pressure, etc.

In cases such as the above, when the subject does not feel "confident" about a stimulus on a given trial, he/she will need longer to accumulate more information about the stimulus before feeling sufficiently "confident" to respond. However, after a certain time length, the time pressure will be so high that the subject will produce a response though he/she has not reached a sufficient level of confidence. In this case the response will be slow and will have a higher probability of being an error.

Chapter 11

ASSESSMENT OF THE NEW MODEL

The above model was developed following the failure of already known models to explain the results of Experiment 4. Having presented this model, its merits and limitations will be discussed. The model should be assessed against the requirements which, it has been argued, a good model should satisfy. Moreover this model must be used to account for the main results of Experiment 4 while also accounting for relationships between speed and accuracy established throughout literature.

11.1 Basic Characteristics of the New Model

The present model meets all the basic requirements for a good model of speed-accuracy tradeoff functions as outlined and discussed earlier (Chapter 8).

(1) It considers the subject as an active component of the person-task system by promoting the idea that subjects make judgements about stimuli and decisions about the speed and nature of their response.

(2) To assert that subjects can make judgements about stimuli and decisions about their responses, the model assumes that subjects have the ability to exercise overall supervisory control during a trial. By implication, subjects are also considered as using relevant information from long-term memory.

(3) It actively considers accuracy of performance. The model

argues that the production of a given response depends on the extent to which subjects "feel" that they would be correct if they responded. Phrased differently, subjects' assessment of the probability of a response being correct if it is made, P(C|R), is a major consideration of the present model.

(4) It emphasises the importance of stimulus "difficulty" and discusses how various task features which alter the level of "dificulty" also affect the speed-accuracy tradeoff function and, consequently, subjects' performance. It argues that the relationship between speed and accuracy is largely dependent on the "difficulty" of the task. Any variations in task "difficulty" will affect the P(C|R) functions, subjects' perception of the P(C|R) functions (i.e. their level of "confidence") and, consequently, speed-accuracy relationships. Therefore, it seems futile for any model to try to predict such relationships (e.g. the speed of errors versus the speed of correct responses) unless the "difficulty" of the task is taken into account. So far, this point has been largely neglected by most experimenters. It is believed that once this point is taken into account, there will be more progress in the study of human performance on choice response tasks.

(5) Finally the model discusses variations of performance between trials and between subjects. The use of probability functions makes it possible to study trial-to-trial variations and differences between subjects. Consequently, it can be argued that the model can explain how and to what extent the system alters its maximum capabilities over time.

11.2 How It Accounts for the Main Speed-Accuracy Relationships

(1) The Effect of Stimulus "Difficulty" on Speed and Accuracy

The term stimulus "difficulty" has been used in this model to refer to a number of variables which are known to affect speed and accuracy of performance (e.g. practice, stimulus discriminability, stimulus complexity, S-R compatibility, etc.). In the present research, practice and stimulus complexity were of greater interest as they were two of the main variables in Experiment 4 (where stimulus complexity is defined by the number of lights - corresponding to keys - involved in each stimulus) .

According to the present model, practice gives rise to a mechanism which allows subjects to make increasingly reliable judgements about the "difficulty" of stimuli and develop a level of "confidence" to produce a response with realistic chances of being correct. It follows that, with increased practice, subjects can respond faster and more accurately. Furthermore, with the rise of this mechanism, subjects' variance of performance, in terms of both speed and accuracy, decreases over time. These suggestions are fully consistent with the results of Experiment 4 which showed that speed and accuracy systematically improved with practice and that variance of both speed and accuracy decreased.

This effect of practice on speed and accuracy is generally accepted by practically every researcher who studies speed-accuracy relationships. The way it is explained, however, varies from theorist to theorist. Most provide no

means of explaining why speed and accuracy improve with practice, while only a few exceptions offer any explanations which, though ingenious, are yet not sufficient (e.g. Crossman, 1959, Rabbitt, 1981).

When the level of stimulus "difficulty" (e.g. number of lights per stimulus) is increased, subjects need longer to accumulate information so as to reach the level of "confidence" needed before they produce a response. Therefore, their speed decreases. This is exactly what has been revealed by the results of Experiment 4.

Lack of practice and increased stimulus "difficulty" (from whatever source) affect the P(C|R) functions and, consequently, subjects' level of "confidence" and the rate at which they accumulate information. If the rate of information accumulation is slow, they either take longer to accumulate sufficient information or respond within a certain time (i.e. before they have accumulated enough information) at the increased risk of making errors. When the instructions stress accuracy (at least as much as speed - as in Experiment 4), subjects tend to choose the first option of taking longer to accumulate more information, and increase their level of "confidence", before they produce a response. When the instructions stress speed more than accuracy, subjects tend to choose the second option of responding even if they have not accumulated sufficient information and are not very "confident" about the response.

(2) Speed of Correct and Error Responses

According to the present model, the key to the relationship between the speed of correct and error responses lies with the concept of "confidence". When subjects are presented with a certain stimulus they start accumulating information until they are "confident" enough to respond. The longer they take, the more information they accumulate and, consequently, the better their chances of being correct. It follows from this argument that, as a general rule, it can be suggested that the faster a response the greater its probability of being an error. This explains a large number of experimental findings which show that errors are faster than correct responses (e.g. Laming, 1968; Rabbitt, 1966a).

The above rule, though basically true, cannot be used to predict the relationship between the speed of correct and error responses unless other variables about the task are defined. Error responses are faster than correct responses only under certain conditions: when the stimulus-response situation is relatively "easy" (i.e. well-practised stimulus, with low complexity, high discriminability, easy to identify stimulus-response mapping, etc.) and there is substantial pressure on subjects for speed. In this case, the variance of $P(C|R)$ functions is relatively small, as stimuli are "easy" to identify and process. As a result, subjects do not need very long to accumulate sufficient information before they are "confident" enough to respond. However, subjects may respond earlier, because of pressure for speed, at a higher risk of making an error. In other words, when subjects "judge" that

the stimulus is "easy", they usually respond incorrectly, if they rush their response. In this case errors are faster than correct responses.

The above prediction however cannot be generalised. When the stimulus-response situation is "difficult" (e.g. not well-practised stimulus, with high complexity, low discriminability, difficult to identify stimulus-response mapping, etc.), the variance of the $P(C|R)$ functions is quite substantial. This means that on one trial a subject's level of "confidence" by time t could be quite high, whereas, on another trial, his/her level of "confidence" by the same time could be quite low. Thus the time taken to accumulate sufficient information to reach the same level of "confidence" will be highly variable. However, if there is substantial pressure for speed, subjects cannot spend as much time as they need to accumulate information. They make a "compromise" and respond after a longer period of time (in comparison to the other trial) but before they have actually reached an equivalent level of "confidence". Therefore, their response is slow, as they have exhausted all the time they could afford, and it has a relatively high probability of being an error, as they are not very "confident" about their decision when they respond.

To phrase it differently, when subjects judge that a given stimulus is "difficult" they do not respond until they either reach the desired level of "confidence" or "feel" they have spent too long and must respond even though their

"confidence" is low. If the latter is the case, their response is not only slow but, also, based on insufficient information. It follows that under these circumstances, errors may, on average, be slower than correct responses.

The second prediction accounts for the results of Experiment 4 as well as the results of other experiments in the literature (e.g. Pickett, 1967; Rabbitt and Vyas, 1970) in which the experimental task was more "difficult" than the 2-, 4-, 8-choice response tasks which were used by researchers like Laming (1968) or Rabbitt (1966) whose results support the first prediction.

(3) Speed Reduction after an Error

According to the present theory, after subjects make an error, they become more "cautious"; in other words they need to reach a somewhat higher level of "confidence" before they respond. To reach this higher level of "confidence", they have to accumulate more information and, consequently, take longer. The implication of this is that speed is reduced after the occurence of an error.

Indeed, this is what happens in reality. It was revealed by the results of Experiment 4. Furthermore it is supported by the results of other researchers (e.g. Laming, 1968, 1979; Rabbitt, 1968; Rabbitt and Rodgers, 1977).

(4) Errors Are Systematically Related to Stimuli

The present model argues that errors are not random but arise from insufficient accumulation of information. Subjects make their decision about the response on the basis of all the

information they have accumulated up to the moment they select their response. Though this information is not sufficient for the correct response to be chosen, their response is not very different from the correct one. This implies that had they waited a little longer to reach a higher level of "confidence", their response could have been correct.

The above argument is well supported by the results of Experiment 4. The error classification analysis (Chapter 9) revealed that errors were systematically related to the stimulus and were "almost correct". The majority of errors were found to be either an omission or an addition of one key. There is more support for the above argument by the results of Guest, Sime and Green (1972) who also found that errors were systematically related to stimuli.

11.3 Limitations of the New Model

The present model was developed as a result of the need to account for relationships which were discovered while studying speed-accuracy tradeoff functions. It provides a new theoretical framework within which these functions can be studied. However, it needs further development before it can be claimed that it makes any great contribution to the study of speed-accuracy problems.

First, the present model leaves certain questions regarding the speed and accuracy functions unanswered. For example, it cannot predict the effect of the occurence of an error on the accuracy of the following response. Further research should be carried out to investigate this specific

question.

In addition, the model, in its present form, is too general to make exact predictions about the effects of various variables on speed and accuracy. Ideally, it should be developed to become a mathematical model the use of which should specify the extent to which speed and accuracy vary as a result of variations in the level of practice, stimulus complexity, S-R compatibility, etc. In its mathematical form, the model would define "easy" and "difficult" tasks and make predictions about speed and accuracy.

The model should be extended to account for variations in performance due to changes brought about by other factors, such as stress, fatigue, use of drugs and age. In its present form, the model makes allowances for such variations by using the concept of "confidence" functions and its relationship to $P(C|R)$ which are assumed to change as result of these factors.

Chapter 12

GENERAL SUMMARY AND CONCLUSIONS

The research conducted and discussed in the foregoing chapters has been in two parts. The research in Part A was carried out to provide information needed to select a new keyboard design for foreign destination letter coding desks at the Post Office as well as information needed for the development of the first stage of a computer-based training programme for Post Office letter sorters. This work was supported by contracts from the Post Office and therefore was applied research bearing certain constraints and allowing fairly little diversion from the specific objectives. The research in Part B was more theoretical, free from contract constraints and examined the issue of speed-accuracy tradeoff relationships in choice response tasks.

The results of the research have already been discussed in detail. The objective of this last chapter is to recapitulate on the major findings, draw general conclusions and discuss implications for applied research.

12.1 Introduction of the 10-Key Chord Keyboard

Four keyboard designs were evaluated: the 2x6 chord keyboard, the 4x4 chord keyboard, the 10-key sequential keyboard and the 10-key chord keyboard. Following two pilot experiments investigating certain aspects of the 10-key chord and the 10-key sequential keyboards (Experiments 1 and 2) the main evaluation study was carried out (Experiment 3). The analysis of the results revealed clear differences between the

four keyboards. After six weeks of about one hour of practice a day (a total of 26,000 trials per subject) the 10-key chord keyboard produced the fastest and most accurate performance and the 10-key sequential the slowest and least accurate. The 4x4 keyboard though more efficient than the currently used 2x6 was less efficient than the 10-key chord one. It was also shown that at the end of the six weeks performance continued to improve with the group using the 10-key chord keyboard showing the greatest rate of improvement. Consequently, the 10-key chord keyboard was recommended to the Post Office, a recommendation which has already been adopted together with some suggestions for a more ergonomically "sound" 10-key chord keyboard.

One observation from Experiment 3 was that whereas the subjects using the 10-key chord keyboard showed clear improvement in both speed and accuracy throughout the six weeks of training (with some fluctuations of course), this was not the case for the subjects using the other three keyboards. All of them showed very systematic improvement in speed. However, in term of accuracy, the subjects using the 4x4 keyboard maintained a fairly stable level of accuracy and the subjects using the 2x6 keyboard showed a slight decrease in accuracy. The subjects using the sequential keyboard showed slight improvement in the first two weeks but, later, their accuracy deteriorated to the extent that they were 2-3% less accurate in the end than they were in the beginning of their training.

The implications of the above observation though not directly related to the selection of a keyboard for the Post Office, provided one of the very interesting points which emerged from the applied research. A finding like this can suggest nothing less than that subjects were consciously trading accuracy for speed although they had been instructed quite explicitly that speed and accuracy were equally important. The experimenter's opinion of the reasons of this speed-accuracy tradeoff is the following. Subjects were highly motivated to do well (financial incentives make a lot of impact on students!) and they knew that the amount of money they would be given in the end would depend on their final performance and overall rate of improvement. Therefore they were interested not only in their own performance but also the performance of the other subjects. The experimental set-up was such that each subject met one subject from each one of the other groups between sessions and at the end of each day's practice. There was no way they could have been prevented from exchanging information about the feedback they received at the end of each session. This must have been a confounding variable in that it probably affected subjects' speed-accuracy tradeoff functions.

Although there was awareness on the part of the experimenter of this confounding variable, the final result was not actually affected as the comparison study would be based on measures of both speed and accuracy. In fact, the existence of these variable tradeoff functions adopted by the four experimental groups emphasised even more the superiority

of the 10-key chord keyboard. Examining the speed-accuracy tradeoff functions of these groups could have been very interesting but, because of shortage of disk space, the appropriate data (i.e. the stimulus and the exact response for each trial) were not recorded. This was to be borne in mind for the following experiment.

12.2 Operator Training Procedures

The computer-based training system for operators of the foreign destination letter coding desks consists of three stages, Keyboard Familiarization, Area-Name Coding and Address Encoding. The choice of these stages was beyond the scope of the present research. The present research concentrated on the first stage of training, i.e. on Keyboard Familiriazation, during which operators develop the skill of keying various chord patterns.

The first three experiments had revealed big individual differences among subjects and this was an important factor to be taken into account in the design of a training system. Given that operators would have to receive individual training, it would be best if the training system could "adapt" to individual operators' needs by pacing itself and varying the training content according to individual performance.

Two types of adaptation were investigated: the first type of adaptation (Cumulative versus Fixed) exercised control over the size of the population from which stimuli were selected; the second (Adaptive versus Random) exercised control over the

individual items presented to subjects on each trial. Four training modes (all possible combinations of the cumulative versus fixed and adaptive versus random factors) were investigated:

(1) Fixed-Random in which items were randomly selected from the whole population;
(2) Fixed-Adaptive in which there was individual item selection from the whole population;
(3) Cumulative-Random in which items were randomly selected from the current item set; and
(4) Cumulative-Adaptive in which there was individual item selection from the current item set.

The comparison and evaluation study of these four training procedures revealed that the CR mode produced the fastest performance within the shortest training time whereas the Adaptive modes produced the most accurate performance. It was concluded that the CR mode was most appropriate for the Keyboard Familiarization stage of training whereas the Adaptive modes should be introduced in the later stages of training when operators have already reached a good level of the basic motor skills required for the task and have to increase their accuracy level and/or they have to be trained on the more complex tasks of stages two and three which require more than just motor skills.

12.3 The Study of Speed-Accuracy Relationships

Following the completion of Experiment 4, the applied part of the research was complete and attention was directed

to the issue of speed-accuracy tradeoff relationships in choice response tasks. What affects speed and accuracy of performance? What is the relationship between speed and accuracy during various stages of practice? Are speed and accuracy relationships affected by the nature of the task? The investigation of these questions was interesting for various reasons.

First, it is important to understand the theoretical background of speed-accuracy relationships when one develops a whole system for training operators on speed and accuracy. Second, answers to such questions can provide answers to specific problems about training needs (e.g. if speed and accuracy are affected differently at different stages of practice, it might be advisable to stress speed/accuracy at some stages and accuracy/speed at others). Third, the examination of speed and accuracy curves obtained from the data of Experiment 3 suggested that the four groups adopted differential speed-accuracy tradeoff functions. This sparked a theoretical interest in the matter especially when a review of the literature revealed a gap in the theoretical knowledge of RT and the way this knowledge can be used to account for experimental findings on speed-accuracy relationships in choice response tasks. Furthermore, this type of information can be used to increase the efficiency of training of perceptual-motor skills. Therefore such research would be useful not only for the specific Post Office projects but also for any projects involving training on any choice response

task, from typists, to word processor operators, to computer operators, etc.

The data obtained from Experiment 4 were analysed again to investigate, specifically, speed-accuracy relationships. Following detailed analyses of the data of all four groups, the main findings were the following.

(1) Both speed and accuracy improved with practice and decreased with an increase in stimulus difficulty.

(2) Error responses were generally slower than correct responses.

(3) The speed of a response was reduced as a result of the occurence of an error on the preceding trial.

(4) Speed can be predicted by the level of stimulus difficulty, the practice session and the accuracy of the preceding response. The accuracy of a given response cannot be predicted by variables such as stimulus difficulty, practice session, accuracy of the preceding response, etc.

The finding that errors were slower than correct responses was of particular interest. In Experiment 1, a high positive correlation between RT and error rate had been obtained and similar results were obtained from the data of the 10-key chord keyboard group in Experiment 3. These findings suggested that subjects did not trade speed for accuracy (or vice versa); either they knew how to respond and they did so fast and accurately or they did not and they were slow and inaccurate. The same sort of relationship between the speed of correct and error responses was found in Experiment

4. On the other hand, as mentioned above, in Experiment 3, it had been found that the 4x4 and the 2x6 keyboard groups did try to increase speed and made more errors and that the 10-key sequential group showed a clear speed-accuracy tradeoff with speed increasing at a great expense of accuracy.

Following a comprehensive literature review in an area which has been studied since at least 1868 when Donders published his work, it was concluded that there is no adequate theoretical model which can account for the present data as well as the data of other experimental studies while also fulfilling certain major theoretical requirements outlined and discussed in Chapter 8. The majority of theories and models account for some results but not others. Moreover, it seems that most theories and models are not really as incompatible or different as they have been presented to be. At the moment there are hundreds of articles based on various experimental designs, dealing with many different tasks, controlling for certain variables but not others, etc., and all this knowledge and many possibly valuable findings may remain unnoticed. It is the author's opinion that a taxonomy of all relevant studies and experiments should be developed so that all knowledge which has been accumulated by so many reasearchers over so many years of work can be organised and studied in a systematic way. This would by no means be a simple task. On the contrary, it would be a major undertaking which would require the work of many researchers over a long time.

It was impossible to develop such a taxonomy within the

limited time of a PhD and no model could adequately account for some speed-accuracy tradeoff relationships revealed by the analyses on the data of Experiment 4. It was necessary therefore to develop a theoretical framework within which the obtained results could be explained. This was achieved by developing a "new" model. This model is considered "new" in the sense that although it was basically an integration of existing models with relatively few new elements, it provides a relatively new way of looking at such relationships.

12.4 The Development of a New Theoretical Model

The model considered two probability functions: (1) the conditional probability P(C|R) of a response being correct given that this response is made by a certain time t; and (2) the probability P(R) of a response being made by a time t. The model argued that subjects would respond if they "felt" reasonably "confident" that they would be correct. In other words, on each trial, the P(R) depended on subjects' assessment of the P(C|R).

The model argues that the P(C|R) increases monotonically with time and varies as a function of a stimulus "difficulty" (e.g. stimulus discriminability, stimulus complexity, stimulus-response compatibility, etc.). As the stimulus "difficulty" increases so does the rate of information accumulation. Clearly, this type of information can be very useful for designers of display-control systems. If it is borne in mind that the P(C|R) must be optimized, those variables which increase the efficient operation of the system

can be incorporated in the design.

The P(R) function is affected by two sets of variables: (a) variables related to the way subjects should respond as determined by instructions about the emphasis on speed and accuracy, payoffs, etc.; and (b) variables related to the subject's individual assessment of P(C|R), i.e. the subject's level of "confidence" at the time of their response. Obviously, it would be of great help to designers of training programmes and to trainers if these variables were defined and specified.

This model fulfills the requirements discussed in Chapter 8. In addition, it explains the findings of Experiment 4. It discusses the effect of practice and stimulus difficulty on speed and accuracy and accounts for the present findings. It explains why it is possible for errors to be faster, slower or of equal speed to correct responses. It discusses the reasons why speed decreases after the occurence of an error and why errors are systematically related to stimuli. If used appropriately, the model may explain other experimental findings as well.

One and possibly the major limitation of this model is that it is too general to make any specific predictions about speed or accuracy. It was suggested in Chapter 11 that it should be developed to become a mathematical model. It was also suggested that the model should be extended to account for changes in performance as a result of stress, fatigue, use of drugs and age. There is no easy way for these to be achieved. The development of a taxonomy of studies and

experiments would be the ideal source of information and data. Until then, nobody will be able to make predictions about speed and accuracy of performance with any confidence.

REFERENCES

Alden, D.G., Daniels, R.W. and Kanarick, A.F. (1972) Keyboard Design and Operation: A Review of the Major Issues. Human Factors, 14, 275-293.

Audley, R.J. (1960) A Stochastic Model for Individual Choice Behaviour. Psychological Review, 67, 1-15.

Audley, R.J. (1973) Some Observations on Theories of Choice Reaction Time: A Tutorial Review. In S. Kornblum (Ed.) Attention and Performance IV. New York: Academic Press.

Audley, R.J., Coudrey, D.J., Howell, P. and Powell, D.J. (1975) Reaction Time Exchange Functions in Choice Tasks. In P.M.A. Rabbitt and S. Dornic (Eds.) Attention and Performance V. London: Academic Press.

Bartram, D. The Design and Evaluation of Computer-Based Training System for Foreign Destination Coding Desk Operators. In J. Patrick and K. Duncan (Eds.) Human Decision Making and Control. Elsevier Science Publishers (in press).

Bartram, D. and Feggou, O. (1985a) A Pilot Investigation of the Ten-Key Sequential Keyboard. Technical Report ERG/Z6545/4a, University of Hull.

Bartram D. and Feggou, O. (1985b) An Evaluation of Four Different Keyboard Designs for Foreign Destination Coding Desks. Technical Report ERG/Z6545/4c, University of Hull.

Bartram, D. and Feggou, O. (1986) An Evaluation of Four Different Keyboard Designs for Foreign Destination Coding Desks. In D.J. Oborne (Ed.) Contemporary Ergonomics 1986. London: Taylor and Francis.

Bartram, D. and Lee, M.P.L. (1985) A Pilot Investigation of the 10-Key Chord Keyboard. Technical Report ERG/Z6545/4b, University of Hull.

Bartram, D., Feggou, O. and Hasenauer, J.A. The Relative Efficacy of Adaptive and Cumulative Training (in preparation).

Bertelson, P. (1961) Sequential Redundancy and Speed in a Serial Two-Choice Responding Test. Quarterly Journal of Experimental Psychology, 13, 90-102.

Bertelson, P. (1963) S-R Relationships and Reaction Time to New Versus Repeated Signals in a Serial Task. Journal of Experimental Psychology, 65, 478-484.

Bertelson, P. (1965) Serial Choice Reaction Time as a Function of Response versus Signal-and-Response Repetition. Nature, 206, 217-218.

Bilodeau, E.A. (1955) Variations in Knowledge of Component Performance and Its Effects upon the Part-Part and Part-Whole Relations. Journal of Experimental Psychology, 50, 215-224.

Bilodeau, E.A. (1957) Patterns of Internal Consistency in Multipart Skilled Performances.American Journal of Psychology, 70, 550-559.

Bilodeau, E.A. and Bilodeau, I.McD. (1954) The Contribution of Component Activities to the Total Psychomotor Task. Journal of Experimental Psychology, 47, 37-46.

Bowen, H.M. and Guinness, G.V. (1965) Preliminary Experiments on Keyboard Design for Semiautomatic Mail Sorting. Journal of Applied Psychology, 49, 194-198.

Bower, G.H. (1959) Choice-Point Behavior. In R.R. Bush and W.K.

Estes (Eds). Studies in Mathematical Learning Theory. Stanford: Stanford University Press.

Briggs, G.E. and Naylor, J.C. (1962) The Relative Efficiency of Several Training Methods as A Function of Transfer Task Complexity. Journal of Experimental Psychology, 64, 505-512.

Broadbent, D.E. (1971) Decision and Stress. London: Academic Press.

Campbell, J. (1971) Personnel Training and Development. Annual Review of Psychology, 22, 565-602.

Christie, L.S. and Luce, R.D. (1956) Decision Structure and Time Relations in Simple Choice Behaviour. Bulletin of Mathematical Biophysics, 18, 89-111.

Cohen, J. (1983) The Cost of Dichotomization. Applied Psychological Measurement, 7, 249-253.

Conrad, R. (1960) Experimental Psychology in the Field of Telecommunications. Ergonomics, 3, 289-295.

Conrad, R. (1960) Letter Sorting Machines - Paced, "Lagged" or Unpaced? Ergonomics, 3, 149-157.

Conrad, R. (1962) Practice, Familiarity and Reading Rate for Words and Nonsense Syllables. Quarterly Journal of Experimental Psychology, 14, 71-76.

Conrad, R. and Hull, A.J. (1968) The Preferred Layout for Numerical Data-Entry Keysets. Ergonomics, 11, 165-173.

Conrad, R. and Longman, D.J.A. (1965) Standard Typewriter versus Chord Keyboard - An Experimental Comparison. Ergonomics, 8, 77-88.

Cornog, J.A., Hockman, U.F. and Craig, J.C. (1963) Address

Encoding - A Study of the Double Binary Keyboard as a Link in the Machine-Sorting of Mail. No. 63. WA 338, New York. ASME (1-8).

Crossman, E.R.F.W. (1955) The Measurement of Discriminability. Quarterly Journal of Experimental Psychology, 7, 176-195.

Crossman, E.R.F.W. (1956) The Measurement of Perceptual Load in Manual Operations. Unpublished Ph.D. Thesis, University of Birmingham.

Crossman, E.R.F.W. (1959) A Theory of the Acquisition of Speed Skill. Ergonomics, 2, 153-166.

D'Amato, M.R. (1970) Experimental Psychology: Methodology, Psychophysics and Learning. New York: McGraw-Hill Book Company.

Dean, C. and Whitlock, Q. (1983) A Handbook of Computer-Based Training. London: Kogan Press.

Deininger, R.L. (1960) Human Factors Engineering Studies of the Design and Use of Pushbutton Telephone Sets. The Bell System Technical Journal, 39, 995-1012.

Donders, F.C. Over de snelheid van psychische processen. (On the speed of mental processes.) Onderzoekingen begaan in het physiologisch Laboratorium der Ugtrechtsche Hoogeschod, 1868-69, Tweede reeks, 11, 92-130. Translated by W.G. Koster. Acta Psychologica, 1969, 30, 412-431.

Dvorak, A., Merrick, N.F., Dealey, W.L. and Ford, G.C. (1936) Typewriting Behavior: Psychology Applied to Teaching and Learning Typewriting. New York: American Book Co.

Eckstrand, G.A. (1964) Current Status of Technology and Training. Wright-Patterson A.F.B. Aerospace Medical Labs, Report AMRL-

TDR-64-86.

Egan, J.P. (1975) Signal Detection Theory and ROC Analysis. New York: Academic Press.

Egeth, H. and Smith, E.E. (1967) On the Nature of Errors in a Choice Reaction Task. Psychonomic Science, 8, 345-346.

Emmerich, D.S., Gray, J.L., Watson, C.S. and Tanis, D.C. (1972) Response Latency, Confidence, and ROCs in Auditory Signal Detection. Perception and Psychophysics, 11, 65-72.

Endfield, C. (1978) A Typewriter in Your Pocket. Computer Weekly, 11 May, p.4.

Falmagne, J.C. (1965) Stochastic Models for Choice Reaction Time with Applications to Experimental Results. Journal of Mathematical Psychology, 2, 77-124.

Favreau, O. (1964) Proactive Decremental Effects on Response Speed in a Continuous DRT Task. Psychonomic Science, 1, 319-320.

Fitts, P.M. (1966) Cognitive Aspects of Information Processing III: Set of Speed vs Accuracy. Journal of Experimental Psychology, 71, 849-857.

Fitts, P.M. and Deininger, R.L. (1954) S-R Compatibility: Correspondence among Paired Elements within Stimulus and Response Codes, Journal of Experimental Psychology, 48, 483-492.

Fitts, P.M. and Seeger, C.M. (1953) S-R Compatibility: Spatial Characteristics of Stimulus and Response Codes. Journal of Experimental Psychology, 46, 199-210.

Fitts, P.M. and Switzer, G. (1962) Cognitive Aspects of

Information Processing: I. The Familiarity of S-R Sets and Subsets. Journal of Experimental Psychology, 63, 321-329.

Fitts, P.M., Peterson, J.R. and Wolpe, G. (1963) Cognitive Aspects of Information Processing: II. Adjustments to Stimulus Redundancy. Journal of Experimental Psychology, 65, 423-432.

Fleishman, E.A. and Quaintance, M.K. (1984) Taxonomies of Human Performance. London: Academic Press.

Gatewood, E.L. (1920) Individual Differences in Finger Reactions. The Psychological Monographs, 18, No.4.

Goldstein, I. (1980) Training in Work Organizations. Annual Review of Psychology, 31, 229-272.

Green, D.M. and Swets, J.A. (1966) Signal Detection Theory and Psychophysics. New York: Wiley.

Guest, D.J., Sime, M.E. and Green, T.R.G. (1972) Error Patterns in a Chord Keyboard Response Task. Quarterly Journal of Experimental Psychology, 24, 139-148.

Haaland, J. (1962) Anatomic and Physiological Consideration of the Use of the Thumb for Manual Control in Space Flight. Honeywell Memorandum, July 11.

Haaland, J., Wingert J. and Olson, B.A. (1963) Force Required to Actuate Switches, Maximum Fingerpushing Force, and Coefficient of Friction of Mercury Gloves. Honeywell Memorandum, Feb. 23.

Hale, D.J. (1969) Speed-Error Tradeoff in a Three-Choice Serial Reaction Task. Journal of Experimental Psychology, 81, 428-435.

Harm, O.J. and Lappin, J.S. (1973) Probability, Compatibility, Speed and Accuracy. Journal Of Experimental Psychology, 100,

416-418.

Hayes, V. and Halpin, G. (1978) Reaction Time of the Fingers with Responses Measured on a Typewriter Keyboard. Perceptual and Motor Skills, 47, 863-867.

Hayes, V., Wilson, G.D. and Schafer, R.L. (1977) Typewriting Rate as a Function of Reaction Time. Perceptual and Motor Skills, 45, 1179-1184.

Hick, W.E. (1952) On the Rate of Gain of Information. Quarterly Journal of Experimental Psychology, 4, 11-27.

Hillix, W.A. and Coburn, R. (1961) Human Factors in Keyset Design, Navy Electronics Laboratory/Report 1023, San Diego, California, 22 March.

Hinrichs, J. (1976) Personnel Training. In M.D. Dunnette (Ed) Handbook of Industrial and Organizational Psychology. New York: Rand McNally.

Hobday, S. (1981) "The Quick Brown Fox..." or Keyboard Alternative. The Inventor, 21, 16-18.

Howell, W.C. and Kreidler, D.L. (1963) Information Processing under Contradictory Instructional Sets. Journal of Experimental Psychology, 65, 39-46.

Howell, W.C. and Kreidler, D.L. (1964) Instructional Sets and Subjective Criterion Levels in a Complex Information-Processing Task. Journal of Experimental Psychology, 68, 612-614.

Hyman, R. (1953) Stimulus Information as a Determinant of Reaction Time. Journal of Experimental Psychology, 45, 188-196.

Kamouri, A.L. (1984) Computer-Based Training: A Cognitive Framework of Evaluating Systems' Designs. Journal of Educational Technology Systems, 12, 287-309.

Kantowitz, B.H. (1978) On Accuracy of Speed-Accuracy Tradeoff. Acta Psychologica, 42, 79-80.

Kelley, C.R. (1969) What Is Adaptive Training? Human Factors, 11, 547-566.

Kirby, N.H. (1976) Sequential Effects in Two-Choice Reaction Time: Automatic Facilitation or Subjective Expectancy? Journal of Experimental Psychology: Human Perception and Performance, 2, 567-577.

Klemmer, E.T. (1958) A Ten-Key Typewriter. IBM Research Memorandum, RL-65, IBM Research Center, Yorktown Heights, N.Y.

Klemmer, E.T. (1971) Keyboard Entry. Applied Ergonomics, 2.1, 2-6.

Kornblum, S. (1969) Sequential Dependencies as a Determinant of Choice Reaction Time: A Summary. In W.G. Koster (Ed.) Attention and Performance II. Amsterdam: North-Holland Publishing Company.

Kornblum, S. (1969) Sequential Determinants of Information Processing In Serial and Discrete Choice Reaction Time. Psychological Review, 76, 113-131.

Kornblum, S. (1973) Sequential Effects in Choice Reaction Time: A Tutorial Review. In S. Kornblum (Ed.) Attention and Performance IV. New York: Academic Press.

Kroemer, K.H. (1972) Human Engineering the Keyboard. Human Factors, 14, 51-63.

LaBerge, D. (1962) A Recruitment Theory of Simple Behaviour.

Psychometrika, 27, 375-396.

LaBerge, D. & Tweedy, J.R. (1964) Presentation Probability and Choice Time. Journal of Experimental Psychology, 68, 477-481.

Lamb, J. and Kaufman, H. (1965) Information Transmission with Unequally Likely Alternatives. Perceptual and Motor Skills, 21, 255-259.

Laming, D.R.J. (1962) A Statistical Test of a Prediction from Information Theory in a Card-Sorting Situation. Quarterly Journal of Experimental Psychology, 14, 38-48.

Laming, D.R.J. (1968) Information Theory of Choice Reaction Times. London: Academic Press.

Laming, D.R.J. (1979) Autocorrelation of Choice Reaction Times. Acta Psychologica, 43, 381-412.

Landauer, T.K. (1964) Recency and recall Time. Psychonomic Science, 21, 255-259.

Lappin, J.S. and Disch, K. (1972a) Latency Operating Characteristic: I. Effects of Stimulus Probability on Choice Reaction Time. Journal of Experimental Psychology, 92, 419-427.

Lappin, J.S. and Disch, K. (1972b) The Latency Operating Characteristic: II. Effects of Visual Stimulus Intensity on Choice Reaction Time. Journal of Experimental Psychology, 93, 367-372.

Lappin, J.S. and Disch, K. (1973) Latency Operating Characteristic: III. Temporal Uncertainty Effects. Journal of Experimental Psychology, 98, 279-285.

Lee, M.P.L. (1984) The Practicability of a 10-Key Chord Keyboard. Unpublished MSc Thesis, University of Hull.

Leonard, J.A. (1959) Tactual Choice Reactions: I. Quarterly Journal of Experimental Psychology, 11, 76-83.

Leonard, J.A. and Newman, R.C. (1965) The Acquisition and Maintenance of High Speed and High Accuracy in a Keyboard Task. Ergonomics, 8, 281-304.

Levine, J.M. and Teichner, W.H. (1972) Plans for the Development of a System Language. In E.A. Fleishman, W.H. Teichner and R.W. Stephenson Development of a Taxonomy of Human Performance: A Review of the Second Year's Progress. JSAS Catalog of Selected Documents in Psychology, 2, 39-40. (Ms. No.112)

Levine, J.M. and Teichner, W.H. (1973) Development of a Taxonomy of human Performance: An Information-Theoretic Approach. JSAS Catalog of Selected Documents in Psychology, 3, 28. (Ms. No. 325).

Levy, M. (1955) The Electronics Aspects of the Canadian Sorting of Mail Systems. Proceedings of the National Electronic Conference, 10, 1-14.

Lewis, B.N. and Pask, G. (1965) The Theory and Practice of Adaptive Teaching Systems. In R. Glaser (Ed.) Teaching Machines and Programmed Learning. Washington: National Education Association.

Links, S.W. (1971) Applying RT Deadlines to Discrimination Reaction Time. Psychonomic Science, 25, 355-358.

Long, J. (1976) Effects of Delayed Irregular Feedback on Unskilled Keying Performance. Ergonomics, 19, 183-202.

McCormick, E.J. and Ilgen, D.R. (1980) Industrial Psychology, 7th Ed. London: George Allen and Unwin.

McGill, W.J. (1963) Stochastic Latency Mechanisms. In R.D. Luce, R.R. Bush and E. Galanter (Eds.) Handbook of Mathematical Psychology, 1. New York: Wiley, 309-360.

Martin, J.M. (1980) A Literature Review of Chord Keyboard Research. HUSAT Memo No. 202 Report 0015, Loughborough University of Technology.

Michaels, S.E. (1971) QWERTY versus Alphabetic Keyboard for Operators of Various Skills. Human Factors, 13, 419-426.

Minor, F.J. and Revesman, S.L. (1962) Evaluation of Input Devices for a Data Setting Task. Journal of Applied Psychology, 46, 332-336.

Morin, R.E. and Forin, B. (1962) Mixing Two Types of S-R Associations in a Choice Reaction Time Task. Journal of Experimental Psychology, 64, 137-141.

Morin, R.E. and Forin, B. (1963) Response Equivocation and Reaction Time. Journal of Experimental Psychology, 66, 30-36.

Morton, J. (1969) Interaction of Information in Word Recognition. Psychological Review, 76, 165-178.

Mowbray, G.H. and Rhoades, M.V. (1959) On the reduction of choice reaction times with practice. Quarterly Journal of Experimental Psychology, 11, 16-23.

Nash, A.N., Muczyk, J.P. and Vettori, F.L. (1971) The Relative Practical Effectiveness of Programmed Instruction. Personnel Psychology, 24, 397-418.

Naylor, J.C. (1962) Parameters Affecting the Relative Efficiency of Part and Whole Practice Methods: A Review of the Literature. Technical Report 950-1, U.S. Naval Training Devices Center, February.

Naylor, J.C. and Briggs, G.E. (1963) The Effect of Task Complexity and Task Organization on the Relative Efficiency of Part and Whole Training Methods. Journal of Experimental Psychology, 65, 217-224.

Nickerson R.S. and Fechrer, C.E. (1964) Stimulus Categorization and Response Time. Perceptual and Motor Skills, 18, 785-793.

Ollman, R.T. (1966) Fast Guesses in Choice Reaction Time. Psychonomic Science, 6, 155-156.

Overall, J.E. and Spiegel, D.K. (1969) Concerning Least Squares Analysis of Experimental Data. Psychological Review, 72, 311-322.

Pachella, R.G. (1974) The Interpretation of Reaction Time in Information-Processing Research. In B.H. Kantowitz (Ed.) Human Information Processing: Tutorials in Performance and Cognition. New Jersey: Lawrence Erlbaum Associates, Publishers.

Pachella, R.G. and Fisher, D.F.(1969) Effect of Stimulus Degradation and Similarity on the Tradeoff between Speed and Accuracy in Absolute Judgements. Journal of Experimental Psychology, 81, 7-9.

Pachella, R.G. and Fisher, D.F. (1972) Hick's Law and the Speed-Accuracy Tradeoff in Absolute Judgements. Journal of Experimental Psychology, 92, 378-384.

Pachella, R.G. and Pew, R.W. (1968) Speed-Accuracy Tradeoff in

Reaction Time: Effect of Discrete Criterion Times. Journal of Experimental Psychology, 76, 19-24.

Pachella, R.G. Fisher, D.F. and Karsh, R. (1968) Absolute Judgements in Speeded Tasks: Quantification of the Tradeoff between Speed and Accuracy. Psychonomic Science, 12, 225-226.

Pask, G. (1960) Adaptive Teaching with Adaptive Machines. In A.A. Lumsdaine and R. Glaser (Eds.) Teaching Machines and Programmed Learning: A Source Book. Washington: National Education Association of the United States.

Pask, G. (1964) The Adaptive Teaching System. In K. Austwick (Ed.) Teaching Machines and Programming. Oxford: Pergamon Press.

Pew, R.W. (1969) The Speed-Accuracy Operating Characteristic. In W.G. Koster (Ed.) Attention and Performance II. Amsterdam:North-Holland Publishing Company.

Pickett, R.M. (1967) Response Latency in a Pattern Perception Situation. Acta Psychologica, 27, 160-169.

Pike, A.R. (1971) The Latencies of Correct and Incorrect Responses in Discrimination and Detection Tasks: Their Interpretation in Terms of a Model Based on Simple Counting. Perception and Psychophysics, 9, 455-460.

Posner, M.I. and Mitchell, R.F. (1967) Chronometric Analysis of Classification. Psychological Review, 74, 392-409.

Rabbitt, P.M.A. (1966a) Errors and Error Correction in Choice-Response Tasks. Journal of Experimental Psychology, 71, 264-272.

Rabbitt, P.M.A. (1966b) Error Correction Time Without External Error Signals. Nature, 212, 438.

Rabbitt, P.M.A. (1968a) Repetition Effects and Signal Classification Strategies in Serial Choice Response Tasks. Quarterly Journal of Experimental Psychology, 20, 232-240.

Rabbitt, P.M.A. (1968b) Three Kinds of Error-Signalling Response in a Serial Choice Task. Quarterly Journal of Experimental Psychology, 20, 179-188.

Rabbitt, P.M.A. (1969) Psychological Refractory Delay and Response-Stimulus Interval Duration in Serial Choice-Response Tasks. In W.G. Koster (Ed.) Attention and Performance II. Amsterdam: North-Holland Publishing Company.

Rabbitt, P.M.A. (1978) Detection of Errors by Skilled Typists. Ergonomics, 21, 945-958.

Rabbitt, P.M.A. (1981) Sequential Reactions. In D.H. Holding (Ed.) Human Skills. Chichester: John Wiley and Sons.

Rabbitt, P.M.A. and Rodgers, B. (1977) What Does a Man Do After He Makes an Error? An Analysis of Response Programming. Quarterly Journal of Experimental Psychology, 29, 727-743.

Rabbitt, P.M.A. and Vyas, S.M. (1970) An Elementary Preliminary Taxonomy for Some Errors in Laboratory Choice RT Tasks. Acta Psychologica, 33, 56-76.

Rabbitt, P.M.A., Cumming, G. and Vyas, S. (1978) Some Errors of Perceptual Analysis in Visual Search Can Be Detected and Corrected. Quarterly Journal of Experimental Psychology, 30, 319-332.

Ratz, H.C. & Ritchie, D.K. (1961) Operator Performance on a Chord Keyboard. Journal of Applied Psychology, 45, 303-308.

Rochester, N., Bequaert, F.C. and Sharp, E.M. (1978) The Chord Keyboard. Computer, 11, 57-63.

Rohmert, W. and Luczak, H. (1978) Ergonomics in the Design and Evaluation of a System for "Postal Video Letter Coding". Applied Ergonomics, 9, 85-95.

Schmitt, J.C. and Scheirer, C.J. (1977) Empirical Approaches to Information Processing: Speed-Accuracy Tradeoff Functions on Reaction Time. Acta Psychologica, 41, 321-325.

Schouten, J.F. and Bekker, J.A.M. (1967) Reaction Time and Accuracy. Acta Psychologica, 27, 143-153.

Seibel, R. (1962) Performance on a Five-Finger Chord Keyboard. Journal of Applied Psychology, 46, 165-169.

Seibel, R. (1963) Discrimination Reaction Time for a 1,023-Alternative Task. Journal of Experimental Psychology, 66, 215-226.

Seibel, R. (1964) Learning Curves, and System Values for Human Performance on Complex Tasks. Human Factors, 6, 293-298.

Seibel, R. (1970) Data Entry Devices and Procedures. In H. Van Cott and R. Kinkade, (Eds.) Human Engineering Guide to Equipment Design. Wasington, D.C.: American Institute for Research.

Seibel, R. (1972) Data Entry Devices and Procedures. In Human Engineering Guide to Equipment Design, U.S. Printing Office, Washington. Cited in H.L. Snyder (1984) Keyboard Design. In Ergolab AB Sweden (Eds.) Ergonomic Principles in Office Automation 2nd Ed.. Sweden: Ericson Information Systems AB.

Shaffer, L.H. (1973) Latency Mechanisms in Transcription. In S.

Kornblum (Ed.) Attention and Performance IV. New York: Academic Press.

Smith, E.E. (1968). Choice Reaction Time: An analysis of the Major Theoretical Positions. Psychological Bulletin, 69, 77-110.

Snodgrass, J.G., Luce, R.D. and Galanter, E. (1967) Some Experiments on Simple and Choice Reaction Time. Journal of Experimental Psychology, 75, 1-17.

Snyder, H.L. (1984) Keyboard Design. In Ergolab AB Sweden (Eds.) Ergonomic Principles in Office Automation, 2nd Ed. Sweden: Ericson Information Systems.

Stanovich, K. and Pachella, R. (1977) Encoding, Stimulus-Response Compatibility and Stages of Processing. Journal of Experimental Psychology: Human Perception and Performance, 3, 411-421.

Sternberg, S. (1964) Two Operations in Character Recognition: Some Evidence from Raction-Time Measurements. Paper presented at the AFCRL Symposium, Models for the Perception of speech and visual form, Boston, November.

Sternberg, S. (1966) High-Speed Scanning in Human Memory. Science, 153, 652-654.

Sternberg, S. (1969) The Discovery of Processing Stages: Extensions of Donders' Method. Acta Psychologica, 30, 276-315.

Sternberg, S. (1975) Memory Scanning: New Findings and Current Controversies. Quarterly Journal of Experimental Psychology, 27, 1-32.

Sternberg, S., Knoll, R.L. and Wright, C.E. (1978) Experiments on

Temporal Aspects of Keyboard Entry. In J.P. Duncanson (Ed.) Getting It Together: Research and Applications in Human Factors. Santa Monica, CA: Human Factors.

Stone, M. (1960) Models for Choice Reaction Time. Psychometrika, 25, 251-260.

Strong, E.P. (1956) A Comparative Experiment in Simplified Keyboard Supplementary Training. Washington, D.C.: General Services Administration. Cited in D.G. Alden, R.W. Daniels and A.F. Kanarick (1972) Keyboard Design and Operation: A Review of the Major Issues. Human Factors, 14, 275-293.

Swenson, R.G. (1972) The Elusive Tradeoff: Speed vs Accuracy in Visual Discrimination Tasks. Perception and Psychophysics, 12, 16-32.

Swenson, R.G. and Edwards, W. (1971) Response Strategies in a Two-Choice Reaction Time Task with a Continuous Cost for Time.Journal of Experimental Psychology, 88, 67-81.

Swets, J.A. (1959) Indices of Signal Detectability Obtained with Various Psychophysical Procedures. J. Acoust. Soc. Amer., 31, 511-513.

Swets, J.A. (Ed.) (1964) Signal Detection and Recognition by Human Observers. New York: Wiley.

Swets, J.A., Tanner, W.P. and Birdsall, T.G. (1961) Decision Processes in Perception. Psychological Review, 68, 301-340.

Tanner, W.P. and Swets, J.A. (1954) A Decision-Making Theory of Visual Detection. Pychological Review, 61, 401-409.

Taylor, D.A. (1976) Stage Analysis of Reaction Time.

Psychological Bulletin, 83, 161-191.

Teichner, W.H. (1969) An Information-Theoretic Approach to Task Classification. Paper presented at the meeting of the American Psychological Association, Washington D.C., September.

Thurmond, J.B. and Alluisi, E.A. (1963) Choice Time as a Function of Stimulus Dissimilarity and Discriminability. Canadian Journal of Psychology, 17, 326-337.

Townsend, J.T. (1972) Some Results Concerning the Identifiability of Parallel and Serial Processes. British Jou rnal of Mathematical and Statistical Psychology, 25, 168-197.

Townsend, J.T. (1974) Issues and Models Concerning the Processing of a Finite Number of Inputs. In B.H. Kantowitz (Ed.) Human Information Processing: Tutorials in Performance and Cognition. Hillsdale, N.J.: Lawrence Erlbaum.

Townsend, J.T. (1976) Serial and Within-Stage Independent Parallel Model Equivalence on the Minimum Completion Time. Journal of Mathematical Psychology, 14, 219-238.

Vickers, D. (1970) Evidence for an Accumulator Model of Psychophysical Discrimination. Ergonomics, 13, 37-58.

Vickers, D. (1980) Discrimination. In A.T. Welford (Ed.) Reaction Times. London: Academic Press.

Welford, A.T. (1960) The Measurement of Sensory-Motor Performance: Survey and Reappraisal of Twelve Years' Progress. Ergonomics, 3, 231-244.

Welford, A.T. (1968) Fundamentals of Skill. London: Methuen and Co Ltd.

Welford, A.T. (1976) Skilled Performance. Glenview, Illinois:

Scott, Foresmen and Company.

Wickelgren, W.A. (1977) Speed-Accuracy Tradeoff and Information Processing Dynamics. Acta Psychologica, 41, 67-85.

Yellot, J.T. (1967) Correction for Fast-Guessing in Choice Reaction Time. Psychonomic Science, 28 349-351.

Yellott, J.I. (1971) Correction for Fast-Guessing and the Speed-Accuracy Tradeoff in Choice Reaction Time. Journal of Mathematical Psychology, 8, 159-199.

Zipp, P., Haider, E. Halpern, N. and Romhert, W. (1983) Keyboard Design through Physiological Strain Measurements. Applied Ergonomics, 14, 117-122.

ADDITIONAL BIBLIOGRAPHY

Anderson, J.R. (1981) Cognitive Skills and their Acquisition. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Annett, J. (1971) Acquisition of Skill. British Medical Bulletin, 27, 266-271.

Annett, J. and Duncan, K.D. (1967) Task Analysis and Training Design. Occupational Psychology, 41, 211-221.

Baddeley, A.D. and Longman, D.J.A. (1978) The Influence of Length and Frequency of Training Session on the Rate of Learning to Type. Ergonomics, 21, 627-635.

Bass, B.M. and Vaughan, J.A. (1966) Training in Industry: The Management of Learning. London: Tavistock Publications.

Broadbent, D.E. (1964) Perception and Communication. Oxford: Pergamon Press.

Broadbent, D.E. and Gregory, M. (1965) On the Interaction of S-R Compatibility with Other Variables Affecting Reaction Time. British Journal of Psychology, 56, 61-67.

Caro Jr., P.W. (1969) Adaptive Training - An Application to Flight Simulation. Human Factors, 11, 569-576.

Duncan, J. (1978) Response Selection in Spatial Choice Reaction: Further Evidence Against Associative Models. Quarterly Journal of Experimental Psychology, 30, 429-440.

Edwards, W. (1965) Optimal Strategies for Seeking Information: Models for Statistics, Choice Reaction Times and Human Information Processing. Journal of Mathematical Psychology, 2, 312-329.

Evans, S.H. and Anastasio, E.J. (1968) Misuse of Analysis of Covariance when Treatment Effect and Covariate Are Confounded. Psychological Bulletin, 69, 225-234.

Fitts, P.M. and Posner, M.I (1967) Human Performance. Belmont: Brooks/Cole Publishing Co.

Fitts, P.M. and Switzer, G. (1962) Aspects of Information Processing: I. The Familiarity of S-R Sets and Subsets. Journal of Experimental Psychology, 63, 321-329.

Glaser, R. (Ed.) (1962) Training Researchand Education. New York: John Wiley and Sons Inc.

Glaser, R. (Ed.) (1965) Teaching Machines and Programmed Learning II. Washington, D.C.: National Education Association of the United States.

Hasenauer, J.A. (1985) Adaptive Training in Chord Keyboard Operation: A Comparative Study. Unpublished MSc thesis, University of Hull.

Kantowitz, B.H. (Ed.) (1974) Human Information Processing: Tutorials in Performance and Cognition. New Jersey: Lawrence Erlbaum Associates, Publishers.

Krinchik, E. (1974) Probability Effects in Choice Reaction Time Tasks. Perception and Psychophysics, 15, 131-144.

Laming, D. (1979) Choice Reaction Performance Following an Error. Acta Psychologica, 43, 119-224.

Laszlo, J.I., Bairstow, P.J. and Baker, J.E. (1979) Dependence on Feedback Following Practice. Perceptual and Motor Skills, 48, 1225-1226.

Leonard, J.A. and Carpenter, A. (1964) On the Correlation between

a Serial Choice Reaction Task and Subsequent Achievement at Typewriting. Ergonomics, 7, 197-204.

Luce, R.D. (1986) Response Times. New York: Oxford University Press.

Lumsdaine, A.A. and Glaser, R.(Eds.) (1960) Teaching Machines and Programmed Learning. Washington, D.C.: National Education Association of the United States.

Lundervold, A. (1958) Electromyographic Investigations During Typewriting. Ergonomics, 1, 226-233.

Noyes, J. (1983) Chord Keyboards. Applied Ergonomics, 14, 55-59.

Pask, G. (1975) The Cybernetics of Human Learning and Performance. London: Hutchinson.

Pask, G. and Scott, C.E. (1972) Learning Strategies and Individual Competence. International Journal of Man-Machine Studies, 4, 217-253.

Poulton, E.C. (1966) Engineering Psychology. Annual Review of Psychology, 17, 177-200.

Prinz, W. and Sanders, A.F. (Eds.) (1984) Cognition and Motor Processes. Berlin: Springer-Verlag.

Rabbitt, P.M.A. (1967) Time to Detect Errors as a Function of Factors Affecting Choice Response Time. Acta Psychologica, 27, 131-142.

Rabbitt, P.M.A. (1971) Times for the Analysis of Stimuli and for the Selection of Responses. British Medical Bulletin, 27, 259-265.

Rabbitt P.M.A. and Rodgers, B. (1979) "The Last of the Old Lag?" A Reply to Welford's Comment on Rabbitt and Rodgers. Quarterly Journal of Experimental Psychology, 31, 543-548.

Reigeluth, C.M. (1983) Instructional-Design Theories and Models: An Overview of their Current Status. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Remington, R.J. and Rogers, M. (1969) Keyboard Literature Survey. Phase I. Bibliography. Technical Report TR29.0042, IBM Systems Development Division. Research Triangle Park, North Carolina, January.

Rumelhart, D.E. and Norman, D.A. (1982) Simulating a Skilled Typist: A Study of Skilled Cogniitve-Motor Performance. Cognitive Science, 6, 1-36.

Schmidt, R.A. (1982) Motor Control and Learning. Champaign, Ill.: Human Kinetics Publishers.

Stelmach, C.E. (Ed.) (1976) Motor Control: Issues and Trends. New York: Academic Press.

Stelmach, C.E. (Ed.) (1978) Information Processing in Motor Control Learning. New York: Academic Press.

Stelmach, C.E. and Requin, J. (Eds.) (1980) Tutorials in Motor Behavior. Amsterdam: North-Holland Publishing Company.

Shaffer, L.H. (1978) Timing in the Motor Programming of Typing. Quarterly Journal of Experimental Psychology, 30, 333-345.

Welford, A.T. (1971) What Is the Basis of Choice Reaction-Time? Ergonomics, 14, 679-693.

Welford, A.T. (1979) Comment on the Paper "What Does a Man Do After He Makes an Error? by Patrick Rabbitt and Bryan Rodgers. Quarterly Journal of Experimental Psychology, 31, 539-542.

Welford, A.T. (Ed.) (1980) Reaction Times. London: Academic

Press.

Wickelgren, W.A. and Becker, G.M. (1965) Decisions Based on Conflicting and Inaccurate Observations. Journal of Mathematical Psychology, 2, 180-189.

Wickens, C.D. (1984) Engineering Psychology and Human Performance. Columbus: Charles E. Merrill Publishing Company.

Wickens, C.D., Sandry, D.L. and Vidulich, M. (1983) Compatibility and Resource Competition between Modalities of Input, Central Processing, and Output. Human Factors, 25, 227-248.

APPENDIX 1

Experiment 1 / Chapter 1

Analysis of Variance on RT Data

Factor A: Left versus Right Hand - Levels: 2
Factor B: One-handed versus Two-handed Stimuli - Levels: 2

Analysis of Variance Summary Table

Source	Sums of Squares	df	Variance	F ratio	p
BETWEEN Ss	307228.38	14	21944.9		
WITHIN Ss	268138.79	45	5958.64		
Factor A	91.07	1	91.07	0.65	NS
Factor B	225304.07	1	225304.07	80.50	0.001
AxB	69.65	1	69.65	0.63	NS
SxA	1948.15	14	139.53		
SxB	39184.34	14	2798.88		
SxAxB	1541.51	14	110.11		
TOTAL	575367.18	59			

Analysis of Variance on ERR Data

Factor A: Left versus Right Hand - Levels: 2
Factor B: One-handed versus Two-handed Stimuli - Levels: 2

Analysis of Variance Summary Table

Source	Sums of Squares	df	Variance	F ratio	p
BETWEEN Ss	0.10	14	0.01		
WITHIN Ss	0.12	45	0.00		
Factor A	0.00	1	0.00	0.06	NS
Factor B	0.05	1	0.05	20.43	.01
AxB	0.00	1	0.00	0.05	NS
SxA	0.02	14	0.00		
SxB	0.04	14	0.00		
SxAxB	0.01	14	0.00		
TOTAL	.22	59			

APPENDIX 2

Experiment 3 / Chapter 3

Summary Tables of analyses of variance showing no difference reaction time (RT) and percentage of errors (ERR) on the Pretest among the four Keyboard Groups

Analysis of Variance on RT Data

Source	Sums of Squares	df	Variance	F ratio	p
BETWEEN Ss	8286.10	15	542.41		
Groups	211.21	3	70.40	.10	NS
Ss within Groups	8074.88	12	672.91		
TOTAL	8286.10	15			

Analysis of Variance on ERR Data

Source	Sums of Squares	df	Variance	F ratio	p
BETWEEN Ss	0.09	15	0.01		
Groups	0.01	3	0.00	.80	NS
Ss within Groups	0.07	12	0.01		
TOTAL	0.09	15			

Experiment 3 / Chapter 3

Analyses of variance on reaction time (RT), percentage of errors (ERR) and adjusted scores (ASC) data comparing the performance of the four groups during the last two weeks of practice

Analysis of Variance on RT Data

Factor A: Weeks (Week 5 and Week 6) - Levels: 2
Factor B: Practice (10 sessions per week) - Levels: 10
Groups Factor: Keyboard (2x6, 4x4, 10-key sequential, 10-key chord) - Levels: 4

Mean RTs (csecs)

Session	Week 5	Week 6	Mean
1	107.81	98.13	102.97
2	106.88	96.63	101.75
3	106.25	97.13	101.69
4	105.00	95.94	100.47
5	102.19	95.31	98.75
6	101.13	94.88	98.00
7	101.06	93.25	97.16
8	99.88	93.69	96.78
9	100.00	92.81	96.41
10	98.50	92.44	95.47
Mean	102.87	95.02	98.94

Group	Week 5	Week 6	Mean
2x6	106.45	97.50	101.98
4x4	95.48	89.63	92.55
Seq.	115.50	107.80	111.65
Chord	94.05	85.15	89.60
Mean	102.87	95.02	98.94

Session	2x6	4x4	Seq.	Chord
1	107.63	95.63	114.25	94.38
2	105.63	94.88	114.00	92.50
3	105.00	96.38	112.50	92.88
4	103.00	93.13	113.88	91.88
5	100.38	92.88	111.38	90.38
6	101.38	91.25	112.00	87.38
7	99.63	90.75	110.63	87.63
8	99.25	90.38	111.25	86.25
9	99.25	90.38	108.25	87.75
10	98.63	89.88	108.38	85.00

Experiment 3 / Chapter 3

Mean	101.98		92.55		111.65		89.60	
	W E E K 5				W E E K 6			
Session	2x6	4x4	Seq.	Chord	2x6	4x4	Seq.	Chord
1	112.25	99.50	119.75	99.75	103.00	91.75	108.75	89.00
2	109.75	99.25	120.50	98.00	101.50	90.50	107.50	87.00
3	110.00	100.25	117.00	97.75	100.00	92.50	108.00	88.00
4	108.25	97.00	119.00	95.75	97.75	89.25	108.75	88.00
5	105.00	95.50	114.25	94.00	95.75	90.25	108.50	86.75
6	105.50	92.72	114.75	91.50	97.25	89.75	109.25	83.25
7	105.00	93.00	113.75	92.50	94.25	88.50	107.50	82.75
8	103.25	93.00	113.50	89.75	95.25	87.75	109.00	82.75
9	103.25	92.00	111.50	93.25	95.25	88.75	105.00	82.25
10	102.25	92.50	111.00	88.25	95.00	87.25	105.75	81.75
Mean	106.45	93.48	115.50	94.05	97.50	89.63	107.80	85.15

Analysis of Variance Summary Table

Source	Sums of Squares	df	Variance	F ratio	p
BETWEEN Ss	52378.39	15	3491.89		
Groups	23905.84	3	7968.61	3.36	.1
Ss within Groups	28472.55	12	2372.71		
WITHIN Ss	9660.60	304	31.78		
Factor A	4929.80	1	4929.80	103.04	.001
Factor B	1959.80	9	217.76	28.20	.001
AxB	177.06	9	19.67	3.09	NS
Groups x A	126.70	3	42.23	0.88	NS
Groups x B	243.85	27	9.03	1.17	NS
Groups x A x B	128.44	27	4.76	0.75	NS
Ss within GroupsxA	574.10	12	47.84		
Ss within GroupsxB	833.95	108	7.72		
Ss within Groups x A x B	686.90	108	6.36		
TOTAL	62038.99	319			

Experiment 3 / Chapter 3

Analysis of Variance on ERR Data

Factor A: Weeks (Week 5 and Week 6) - Levels: 2
Factor B: Practice (10 sessions per week) - Levels: 10
Groups Factor: Keyboards (2x6, 4x4, 10-key Sequential, 10key Chord) - Levels: 4

Mean ERR (Percentage of Errors)

Session	Week 5	Week 6	Mean
1	3.29	4.17	3.73
2	3.35	4.40	3.87
3	3.94	4.51	4.23
4	4.26	4.76	4.51
5	3.62	4.61	4.11
6	4.52	5.08	4.80
7	3.98	5.37	4.67
8	4.36	5.33	4.85
9	4.26	5.16	4.71
10	5.35	5.50	5.42
Mean	4.09	4.89	4.49

Group	Week 5	Week 6	Mean
2x6	3.48	4.40	3.94
4x4	3.90	3.68	3.79
Seq.	6.11	8.75	7.43
Chord	2.88	2.73	2.80
Mean	4.09	4.89	4.49

Session	2x6	4x4	Seq.	Chord
1	2.98	3.19	6.62	2.12
2	3.38	3.60	6.67	1.85
3	3.88	3.55	7.62	1.86
4	3.93	3.74	7.76	2.62
5	3.93	3.74	6.95	1.83
6	3.86	3.98	7.02	4.36
7	4.17	3.83	8.24	2.45
8	4.67	3.38	7.52	3.81
9	3.93	4.41	7.86	2.64
10	4.67	4.48	8.05	4.50
Mean	4.40	3.60	3.68	2.73

Experiment 3 / Chapter 3

	WEEK 5				WEEK 6			
Session	2x6	4x4	Seq.	Chord	2x6	4x4	Seq.	Chord
1	3.14	2.43	5.34	2.24	2.81	3.95	7.90	2.00
2	3.24	3.00	5.00	2.14	3.53	4.19	8.33	1.57
3	3.53	3.29	7.10	1.86	4.24	3.81	8.14	1.86
4	3.76	4.14	6.00	3.14	4.10	3.33	9.53	2.09
5	3.34	3.33	5.95	1.86	4.53	4.14	7.95	1.81
6	3.05	4.38	5.29	5.38	4.67	3.57	8.76	3.33
7	3.10	4.48	6.24	2.10	5.24	3.19	10.24	2.81
8	4.29	3.81	6.29	3.05	5.05	2.95	8.76	4.57
9	3.00	4.96	6.86	2.24	4.86	3.86	8.86	3.05
10	4.33	5.19	7.05	4.81	5.00	3.76	9.05	4.19
Mean	3.48	3.90	6.11	2.88	4.40	3.68	8.75	2.73

Analysis of Variance Summary Table

Source	Sums of Squares	df	Variance	F ratio	p
BETWEEN Ss	2688.72	15	179.25		
Groups	983.18	3	327.73	2.31	NS
Ss within Groups	1705.54	12	142.13		
WITHIN Ss	1029.64	304	3.39		
Factor A	50.90	1	50.90	4.39	.05
Factor B	75.16	9	8.35	3.13	.01
AxB	8.93	9	0.99	0.44	NS
Groups x A	107.27	3	35.76	3.09	.05
Groups x B	57.45	27	2.13	0.80	NS
Groups x A x B	57.95	27	2.15	0.95	NS
Ss within GroupsxA	139.04	12	11.59		
Ss within GroupsxB	288.57	108	2.67		
Ss within Groups x A x B	244.37	108	2.26		
TOTAL	3718.36	319			

Experiment 3 / Chapter 3

Analysis of Variance on ASC Data

Factor A: Weeks (Week 5 and Week 6) - Levels: 2
Factor B: Practice (10 sessions per week) - Levels: 10
Groups Factor: Keyboards (2x6, 4x4, 10-key Sequential, 10key Chord) - Levels: 4

Mean ASC

Session	Week 5	Week 6	Mean
1	111.63	102.50	107.06
2	110.38	101.31	105.84
3	110.31	101.94	106.13
4	109.06	100.94	105.00
5	106.06	100.06	103.06
6	105.81	99.69	102.75
7	105.13	98.75	101.94
8	104.63	99.31	101.97
9	104.56	98.00	101.28
10	104.06	98.13	101.09
Mean	107.16	100.06	103.61

Group	Week 5	Week 6	Mean
2x6	110.25	101.83	106.04
4x4	98.98	92.85	95.91
Seq.	122.93	118.05	120.49
Chord	96.50	87.53	92.01
Mean	107.16	100.06	103.61

Session	2x6	4x4	Seq.	Chord
1	111.00	98.63	122.38	96.25
2	109.13	98.00	122.00	94.25
3	109.13	99.00	121.75	94.63
4	107.38	96.25	123.38	93.00
5	104.50	96.25	119.50	92.00
6	104.25	94.75	120.50	91.50
7	103.75	94.00	120.25	89.75
8	104.25	93.38	120.38	89.88
9	103.38	94.75	117.25	89.75
10	103.63	94.13	117.50	89.13
Mean	106.04	95.91	120.49	92.01

Experiment 3 / Chapter 3

	WEEK 5				WEEK 6			
Session	2x6	4x4	Seq.	Chord	2x6	4x4	Seq.	Chord
1	115.75	102.00	126.75	102.00	106.25	92.25	118.00	90.50
2	113.50	101.50	126.50	100.00	104.75	94.50	117.50	88.50
3	113.75	102.25	125.75	99.50	104.50	95.75	117.75	89.75
4	112.75	100.50	126.50	96.50	102.00	92.00	120.25	89.50
5	108.75	98.50	121.25	95.75	100.25	94.00	117.75	88.25
6	108.25	97.00	121.25	96.75	100.25	92.50	119.75	86.25
7	108.25	97.00	121.25	94.00	99.25	91.00	119.25	85.50
8	108.00	96.50	121.25	92.75	100.50	90.25	119.50	87.00
9	106.50	97.00	119.75	95.00	100.25	92.50	114.75	84.50
10	107.00	97.50	119.00	92.75	100.25	90.75	116.00	85.50
Mean	110.25	98.98	122.93	96.50	101.83	95.91	118.05	87.53

Analysis of Variance Summry Table

Source	Sums of Squares	df	Variance	F ratio	p
BETWEEN Ss	65137.25	15	4342.48		
Groups	38759.70	3	12919.90	5.88	.01
Ss within Groups	26377.55	12	2198.13		
WITHIN Ss	8590.70	304	28.26		
Factor A	4032.80	1	4032.80	104.56	.001
Factor B	1390.45	9	154.49	16.41	.001
AxB	145.08	9	16.12	1.82	NS
Groups x A	223.45	3	74.48	1.93	NS
Groups x B	211.55	27	7.84	0.84	NS
Groups x A x B	150.42	27	5.57	0.63	NS
Ss within GroupsxA	462.85	12	38.57		
Ss within GroupsxB	1016.70	108	9.41		
Ss within Groups x A x B	957.40	108	8.87		
TOTAL	73727.95	319			

Experiment 3 / Chapter 3

Chord patterns rank-orderd by mean RT (csecs) for the final 20 sessions of training (NF = Number of fingers used; Pcr = proportion correct).

Rank	Left-hand	Right-hand	NF	Pcr	RT
1:	0 0 0 1 0	1 0 0 0 0	2	0.99	71.3
2:	1 0 0 0 0	0 0 0 0 1	2	0.98	71.9
3:	0 0 0 0 1	1 0 0 0 0	2	1.00	72.5
4:	0 0 0 0 1	0 0 0 0 1	2	0.99	72.8
5:	1 0 0 0 0	1 0 0 0 0	2	0.99	73.1
6:	0 1 0 0 0	0 0 0 1 0	2	1.00	73.1
7:	0 0 0 1 0	0 1 0 0 0	2	1.00	73.5
8:	0 1 0 0 0	0 0 0 0 1	2	1.00	73.9
9:	0 1 0 0 0	0 1 0 0 0	2	0.98	74.1
10:	0 1 0 0 0	1 0 0 0 0	2	0.97	74.2
11:	0 0 0 1 0	0 0 0 0 1	2	1.00	74.8
12:	1 0 0 0 0	1 1 0 0 0	3	0.98	75.5
13:	1 0 0 0 0	0 1 0 0 0	2	0.98	75.8
14:	0 0 0 1 0	0 0 0 1 0	2	0.99	75.8
15:	0 0 0 0 1	0 1 0 0 0	2	0.95	76.5
16:	1 0 0 0 0	0 0 0 1 0	2	0.99	76.6
17:	0 0 1 0 0	1 0 0 0 0	2	0.99	77.3
18:	0 0 1 1 0	1 0 0 0 0	3	0.98	78.2
19:	0 0 0 0 1	0 0 0 1 0	2	1.00	78.2
20:	1 0 0 0 1	1 0 0 0 0	3	0.98	78.7
21:	0 1 0 0 0	1 1 0 0 0	3	0.97	78.8
22:	0 1 1 1 0	0 1 1 1 0	6	1.00	78.9
23:	0 0 0 1 0	1 1 0 0 0	3	0.98	78.9
24:	1 0 0 0 0	0 1 1 1 0	4	1.00	79.6
25:	0 0 0 1 1	1 1 0 0 0	4	1.00	79.6
26:	1 0 0 0 0	0 1 1 0 0	3	0.97	79.7
27:	0 1 1 1 0	1 0 0 0 0	4	0.98	79.7
28:	0 0 1 0 0	0 0 0 0 1	2	0.97	79.7
29:	0 0 1 1 0	0 1 1 0 0	4	0.99	80.5
30:	0 0 0 0 1	1 1 0 0 0	3	1.00	80.5
31:	1 0 0 0 1	1 0 0 0 1	4	0.99	80.6
32:	0 0 1 1 0	0 0 0 0 1	3	0.99	80.6
33:	0 1 1 1 0	0 0 0 0 1	4	1.00	80.6
34:	0 0 0 0 1	0 1 1 1 0	4	0.97	80.7

Rank	Left-hand	Right-hand	NF	Pcr	RT
35:	0 1 0 0 0	1 0 0 0 1	3	0.98	81.2
36:	0 0 0 1 0	0 1 1 1 0	4	0.96	81.2
37:	1 0 0 0 0	0 0 1 0 0	2	0.98	81.3
38:	1 0 0 0 0	1 0 0 0 1	3	0.96	81.4
39:	0 0 0 0 1	1 0 0 0 1	3	0.99	81.4
40:	0 0 1 0 0	0 0 1 0 0	2	0.99	81.9
41:	0 1 1 1 0	0 1 0 0 0	4	0.99	82.0
42:	0 1 0 0 0	0 1 1 0 0	3	0.99	82.3
43:	0 0 1 1 0	0 0 0 1 0	3	0.99	82.3
44:	0 1 1 0 0	1 0 0 0 0	3	1.00	82.8
45:	0 0 1 0 0	1 1 0 0 0	3	0.99	82.9
46:	0 1 0 0 0	0 0 1 0 0	2	0.92	83.0
47:	0 0 0 1 1	0 0 0 0 1	3	1.00	83.0
48:	0 0 1 1 0	0 1 0 0 0	3	0.98	83.1
49:	0 0 1 1 0	0 0 1 1 0	4	0.97	83.2
50:	0 1 0 0 0	0 1 1 1 0	4	0.95	83.2
51:	0 0 0 1 1	0 1 0 0 0	3	0.99	83.3
52:	0 0 0 0 1	0 1 1 0 0	3	0.99	83.3
53:	1 0 0 0 1	0 0 0 1 0	3	0.99	83.3
54:	1 1 1 1 0	0 1 0 0 0	5	0.99	83.5
55:	0 1 0 0 0	0 0 0 1 1	3	0.97	83.5
56:	1 0 0 0 0	0 0 1 1 0	3	0.98	83.7
57:	0 0 0 1 0	1 0 0 0 1	3	1.00	83.7
58:	1 0 0 0 0	0 0 0 1 1	3	0.99	83.7
59:	0 0 0 1 1	1 0 0 0 0	3	0.99	83.8
60:	0 1 1 1 0	0 0 0 1 0	4	0.99	83.9
61:	1 1 1 1 0	0 1 1 1 0	7	1.00	84.1
62:	0 0 0 1 0	0 1 1 0 0	3	0.96	84.3
63:	0 0 1 0 0	0 0 0 1 0	2	0.98	84.5
64:	0 0 0 0 1	0 0 1 0 0	2	0.95	84.6
65:	1 1 0 0 0	1 0 0 0 0	3	0.97	84.6
66:	0 0 0 0 1	0 0 1 1 0	3	0.98	84.7
67:	0 1 1 1 0	0 1 1 0 0	5	0.97	84.7
68:	0 0 0 0 1	0 0 0 1 1	3	0.98	84.8
69:	1 0 0 0 1	0 0 0 0 1	3	0.98	85.0
70:	0 0 0 1 1	0 0 0 1 0	3	0.99	85.2
71:	1 0 0 0 1	1 1 0 0 0	4	0.98	85.2
72:	1 1 0 0 0	1 1 0 0 0	4	0.97	85.2
73:	0 0 1 1 0	0 1 1 1 0	5	0.98	85.2
74:	0 1 1 0 0	0 0 0 0 1	3	0.96	85.3
75:	0 0 0 0 1	1 1 1 1 0	5	0.97	85.4
76:	1 1 1 1 0	1 0 0 0 0	5	0.99	85.4
77:	1 0 0 0 0	0 1 1 1 1	5	0.98	85.5
78:	0 1 1 0 0	1 1 0 0 0	4	0.96	85.5
79:	0 1 0 0 0	0 0 1 1 0	3	0.97	85.6
80:	0 1 1 1 0	1 0 0 0 1	5	0.98	85.6
81:	1 1 0 0 0	0 0 0 0 1	3	0.98	85.7
82:	1 0 0 0 0	1 1 1 1 0	5	0.98	85.8
83:	0 0 1 0 0	1 0 0 0 1	3	0.99	85.9

Rank	Left-hand	Right-hand	NF	Pcr	RT
84:	0 1 1 0 0	0 0 0 1 0	3	0.99	86.3
85:	1 1 1 1 0	0 0 0 0 1	5	0.98	86.3
86:	0 1 0 0 0	0 1 1 1 1	5	0.98	86.8
87:	0 1 1 0 0	0 0 1 1 0	4	0.99	87.0
88:	0 0 1 0 0	0 0 1 1 0	3	0.99	87.0
89:	1 0 0 0 1	0 1 1 0 0	4	0.97	87.0
90:	0 1 1 1 0	1 1 0 0 0	5	0.93	87.0
91:	0 1 1 0 0	1 0 0 0 1	4	0.99	87.2
92:	1 1 0 0 0	0 0 0 1 1	4	1.00	87.2
93:	1 0 0 0 1	0 1 0 0 0	3	0.98	87.3
94:	0 1 1 0 0	0 1 1 0 0	4	0.99	87.3
95:	0 0 1 0 0	0 1 1 1 0	4	0.98	87.4
96:	1 0 0 0 0	1 1 1 0 0	4	0.96	87.5
97:	0 0 1 1 0	1 0 0 0 1	4	0.98	87.5
98:	0 1 1 0 0	0 1 0 0 0	3	0.98	87.6
99:	0 0 1 0 0	0 1 0 0 0	2	0.89	87.9
100:	1 0 0 0 1	0 0 1 0 0	3	0.97	87.9
101:	0 0 0 1 0	0 0 1 0 0	2	0.94	88.0
102:	0 0 1 0 0	0 0 0 1 1	3	0.99	88.0
103:	0 0 1 0 0	0 1 1 0 0	3	1.00	88.0
104:	0 1 0 0 0	1 1 1 0 0	4	0.97	88.1
105:	1 1 0 0 0	1 0 0 0 1	4	0.98	88.1
106:	1 1 1 1 0	0 1 1 0 0	6	0.97	88.1
107:	1 1 1 1 0	1 1 0 0 0	6	0.98	88.1
108:	1 0 0 0 1	0 0 0 1 1	4	0.99	88.2
109:	0 0 0 0 1	0 1 1 1 1	5	0.98	88.2
110:	1 1 0 0 0	0 0 0 1 0	3	0.99	88.2
111:	0 0 0 1 0	0 0 0 1 1	3	0.98	88.3
112:	0 0 0 0 1	1 1 1 0 0	4	0.96	88.4
113:	0 1 1 1 0	0 0 1 0 0	4	0.99	88.4
114:	0 0 1 1 0	0 0 1 0 0	3	0.98	88.4
115:	0 0 0 1 1	0 0 1 0 0	3	0.93	88.5
116:	1 1 1 1 0	0 0 0 1 0	5	0.98	88.6
117:	0 0 0 1 1	0 0 0 1 1	4	1.00	88.7
118:	0 0 0 1 0	0 0 1 1 0	3	0.97	88.8
119:	0 1 1 1 1	1 0 0 0 0	5	0.97	88.8
120:	1 1 1 1 0	1 0 0 0 1	6	0.96	89.1
121:	0 0 0 1 1	1 0 0 0 1	4	0.96	89.1
122:	0 0 0 1 1	0 1 1 1 0	5	0.95	89.6
123:	0 0 1 1 0	1 1 0 0 0	4	0.96	89.9
124:	1 0 0 0 1	0 1 1 1 0	5	0.99	89.9
125:	0 0 0 1 0	0 1 1 1 1	5	0.99	90.0
126:	0 0 0 1 0	1 1 1 1 0	5	0.97	90.4
127:	0 0 1 1 0	1 1 1 0 0	5	0.98	90.5
128:	0 1 1 1 0	0 0 1 1 0	5	0.97	90.5
129:	0 0 1 0 0	1 1 1 0 0	4	0.95	90.7
130:	1 1 0 0 0	0 1 0 0 0	3	0.91	90.7
131:	1 0 0 0 1	0 1 1 1 1	6	0.98	90.7
132:	0 0 1 1 0	0 0 0 1 1	4	0.99	90.7

Rank	Left-hand	Right-hand	NF	Pcr	RT
133:	0 0 1 0 0	1 1 1 1 0	5	0.98	91.1
134:	0 1 1 1 1	0 1 0 0 0	5	0.98	91.1
135:	1 1 1 1 1	1 0 0 0 1	7	0.97	91.1
136:	0 0 1 1 1	0 0 0 0 1	4	0.96	91.1
137:	0 0 1 1 0	0 1 1 1 1	6	0.95	91.3
138:	1 0 0 0 0	1 1 1 1 1	6	0.96	91.5
139:	0 1 1 1 0	1 1 1 1 0	7	0.98	91.6
140:	0 1 1 0 0	0 0 1 0 0	3	0.97	91.6
141:	0 1 0 0 0	1 1 1 1 1	6	0.93	91.9
142:	0 1 1 1 1	0 1 1 1 0	7	0.99	92.0
143:	0 1 0 0 0	1 1 1 1 0	5	0.98	92.1
144:	1 1 1 1 1	0 0 0 1 0	6	0.97	92.2
145:	0 0 1 1 0	1 1 1 1 0	6	0.94	92.3
146:	0 1 1 1 1	1 1 1 1 0	8	0.95	92.3
147:	0 0 0 1 0	1 1 1 1 1	6	0.96	92.4
148:	1 1 1 1 1	1 0 0 0 0	6	0.99	92.4
149:	0 1 1 1 1	0 0 0 0 1	5	0.98	92.4
150:	0 0 0 1 1	0 1 1 1 1	6	0.94	92.7
151:	0 1 1 1 1	0 0 0 1 0	5	0.98	92.7
152:	1 1 1 1 1	0 0 0 0 1	6	0.98	92.7
153:	0 1 1 0 0	0 1 1 1 0	5	0.97	93.0
154:	0 1 1 1 0	0 1 1 1 1	7	0.96	93.0
155:	1 1 1 0 0	0 0 0 0 1	4	0.93	93.0
156:	1 1 1 1 0	0 1 1 1 1	8	0.98	93.1
157:	1 1 0 0 0	0 1 1 0 0	4	0.94	93.2
158:	0 0 0 1 1	0 0 1 1 0	4	0.93	93.4
159:	1 1 1 1 1	1 1 0 0 0	7	0.94	93.5
160:	1 0 0 0 1	1 1 1 0 0	5	0.96	93.6
161:	1 0 0 0 1	1 1 1 1 0	6	0.97	93.6
162:	1 0 0 0 1	0 0 1 1 0	4	0.99	93.6
163:	0 0 0 1 0	1 1 1 0 0	4	0.97	93.8
164:	0 0 0 0 1	1 1 1 1 1	6	0.98	93.8
165:	1 1 1 1 0	0 0 1 1 0	6	0.96	93.8
166:	0 0 0 1 1	0 1 1 0 0	4	0.98	93.8
167:	0 1 1 1 1	1 0 0 0 1	6	1.00	94.2
168:	0 1 1 1 0	0 0 0 1 1	5	0.97	94.2
169:	0 0 1 1 1	0 1 1 0 0	5	0.99	94.2
170:	0 0 1 0 0	0 1 1 1 1	5	0.97	94.2
171:	0 0 1 1 1	0 0 0 1 0	4	0.99	94.2
172:	0 0 1 1 1	0 1 0 0 0	4	0.97	94.4
173:	1 0 0 0 1	1 1 1 1 1	7	0.98	94.4
174:	0 1 1 1 1	1 1 0 0 0	6	0.96	94.6
175:	1 1 1 1 1	0 1 1 1 0	8	0.98	94.7
176:	1 0 0 0 0	0 0 1 1 1	4	0.96	94.8
177:	1 1 1 1 0	1 1 1 1 0	8	0.93	94.8
178:	0 1 1 0 0	1 1 1 1 0	6	0.98	94.8
179:	0 1 1 1 1	0 0 1 0 0	5	0.95	94.8
180:	1 1 1 1 0	0 0 1 0 0	5	0.97	94.9
181:	0 0 1 1 1	1 0 0 0 0	4	0.96	94.9

Rank	Left-hand	Right-hand	NF	Pcr	RT
182:	1 1 1 0 0	1 0 0 0 0	4	0.91	95.0
183:	0 1 1 1 1	0 1 1 0 0	6	0.96	95.1
184:	0 1 1 1 0	1 1 1 0 0	6	0.94	95.4
185:	0 0 0 0 1	0 0 1 1 1	4	0.98	95.6
186:	0 0 1 1 1	1 0 0 0 1	5	0.96	95.7
187:	0 1 1 0 0	0 0 0 1 1	4	0.97	95.7
188:	0 0 1 0 0	1 1 1 1 1	6	0.95	95.9
189:	1 1 0 0 0	0 0 1 0 0	3	0.92	95.9
190:	0 0 0 1 0	0 0 1 1 1	4	0.95	96.0
191:	1 1 1 1 1	0 1 0 0 0	6	0.97	96.0
192:	1 1 1 1 1	1 1 1 1 1	10	0.97	96.1
193:	1 1 1 1 1	0 0 1 0 0	6	0.95	96.1
194:	1 1 1 1 1	0 1 1 0 0	7	0.94	96.1
195:	0 0 0 1 1	1 1 1 1 1	7	0.95	96.2
196:	1 1 1 1 0	1 1 1 1 1	9	0.98	96.2
197:	0 0 1 1 0	1 1 1 1 1	7	0.91	96.2
198:	0 1 1 0 0	1 1 1 0 0	5	0.92	96.5
199:	0 1 1 1 1	0 0 1 1 0	6	0.95	96.5
200:	1 1 0 0 0	0 1 1 1 0	5	0.97	96.5
201:	1 1 0 0 0	1 1 1 0 0	5	0.95	96.6
202:	0 0 0 1 1	1 1 1 0 0	5	0.98	96.6
203:	0 0 0 1 1	1 1 1 1 0	6	0.94	96.8
204:	0 0 1 1 1	0 0 1 0 0	4	0.98	96.9
205:	1 1 0 0 0	1 1 1 1 0	6	0.96	97.0
206:	0 0 1 0 0	0 0 1 1 1	4	0.94	97.3
207:	0 1 1 1 0	1 1 1 1 1	8	0.97	97.6
208:	0 1 0 0 0	0 0 1 1 1	4	0.92	97.7
209:	1 1 1 0 0	1 0 0 0 1	5	0.96	97.7
210:	0 1 1 1 1	1 1 1 0 0	7	0.99	97.9
211:	1 1 1 1 0	0 0 0 1 1	6	0.97	98.0
212:	1 1 1 1 1	0 0 0 1 1	7	0.95	98.1
213:	0 0 1 1 1	0 1 1 1 0	6	0.99	98.2
214:	1 1 0 0 0	0 0 1 1 0	4	0.93	98.4
215:	0 1 1 0 0	0 1 1 1 1	6	0.98	98.7
216:	0 1 1 0 0	0 0 1 1 1	5	0.98	98.9
217:	1 1 1 0 0	0 0 0 1 0	4	0.90	98.9
218:	1 1 1 1 1	1 1 1 1 0	9	0.98	99.0
219:	1 0 0 0 1	0 0 1 1 1	5	0.94	99.1
220:	1 1 1 1 0	1 1 1 0 0	7	0.92	99.4
221:	1 1 1 1 1	0 0 1 1 0	7	0.99	99.4
222:	0 1 1 1 1	1 1 1 1 1	9	0.95	99.5
223:	0 0 1 1 0	0 0 1 1 1	5	0.95	99.6
224:	0 1 1 1 1	0 1 1 1 1	8	0.99	99.6
225:	0 0 1 1 1	1 1 1 0 0	6	0.95	99.7

APPENDIX 3

Experiment 4 / Chapter 4

Mean Reaction Time (csecs) of Subjects in the Fixed-Random Group

	SESSION					
Subject	1	2	3	4	5	Test
4	117.60	111.37	96.53	102.16	102.98	92.02
8	136.11	120.42	119.90	121.60	113.02	110.73
12	185.38	173.36	157.60	162.83	148.76	145.78
16	147.95	127.34	123.09	129.83	127.10	110.12
20	167.78	146.55	139.07	159.80	137.35	143.32
24	184.38	168.39	152.41	166.91	170.03	160.54
28	241.29	213.03	200.10	189.82	174.28	174.91
32	261.52	245.64	221.22	213.41	191.76	177.55
36	206.73	182.76	178.01	183.10	164.76	155.59
40	208.13	194.30	182.35	188.11	168.94	165.67
Mean	185.69	168.32	157.03	161.76	149.90	143.62
SD	43.16	40.72	36.87	32.90	27.52	28.14

Mean Reaction Time (csecs) of Subjects in the Fixed-Adaptive Group

	SESSION					
Subject	1	2	3	4	5	Test
1	210.16	166.56	155.34	167.31	152.31	135.04
5	167.29	143.10	123.20	124.66	109.34	111.95
9	176.45	158.82	154.12	145.48	135.48	131.13
13	196.21	177.13	161.06	152.07	138.36	134.30
17	216.29	177.92	148.33	163.11	142.23	125.47
21	235.76	164.63	137.68	138.36	129.20	121.09
25	269.20	235.21	221.63	225.57	193.84	175.71
29	332.94	305.85	243.05	220.70	197.28	196.70
33	238.25	199.12	177.08	181.17	161.16	146.16
37	289.83	255.74	238.84	245.16	252.98	246.36
Mean	233.24	198.41	176.03	176.36	161.22	152.39
SD	49.23	48.83	40.84	38.81	40.15	39.74

Experiment 4 / Chapter 4

Mean Reaction Time (cssecs) of Subjects in the Cumulative-Random Group

	SESSION					
Subject	1	2	3	4	5	Test
2	131.93	132.83	117.94	121.62	121.43	113.78
6	151.60	162.58	158.80	158.47	145.02	139.90
10	144.85	157.79	157.16	155.18	145.07	142.51
14	108.52	109.34	100.67	99.11	94.47	90.66
18	149.00	157.14	165.57	163.19	137.16	133.00
22	158.80	167.82	166.44	170.94	172.04	181.18
26	149.15	157.65	133.16	151.18	130.25	121.33
30	148.56	148.36	154.72	155.00	151.05	134.40
34	150.85	163.90	160.02	159.74	157.21	168.35
38	156.90	159.84	163.78	161.93	172.56	180.57
Mean	145.02	151.65	147.83	149.64	142.63	140.57
SD	13.99	16.85	21.57	20.90	22.38	27.75

Mean Reaction Time (csecs) of Subjects in the Cumulative-Adaptive Group

	SESSION					
Subject	1	2	3	4	5	Test
3	135.02	145.58	130.76	127.07	113.89	118.31
7	122.54	119.03	110.03	104.49	100.79	90.99
11	164.42	180.31	175.41	181.56	165.36	192.90
15	163.27	169.19	175.11	167.42	182.04	152.03
19	160.81	165.70	164.16	168.85	172.42	157.23
23	157.70	171.82	171.70	162.30	165.61	175.53
27	152.95	164.67	167.16	162.90	162.32	138.21
31	143.92	135.49	122.04	125.28	107.41	97.89
35	161.37	198.23	175.88	180.34	204.28	205.01
39	160.14	180.58	193.21	199.91	187.15	190.49
Mean	152.21	163.06	158.55	158.01	156.13	151.86
SD	13.29	22.29	26.07	28.39	34.13	38.11

Experiment 4 / Chapter 4

Mean Error Rate of Subjects in the Fixed-Random Group

	SESSION					
Subject	1	2	3	4	5	Test
4	0.63	2.81	3.44	2.50	0.94	4.38
8	24.38	15.00	11.88	19.06	11.86	8.13
12	12.81	2.81	3.44	4.06	2.50	3.13
16	6.56	4.38	5.00	5.63	1.88	3.44
20	20.94	13.75	10.94	13.13	9.38	5.94
24	11.88	8.75	10.00	13.44	1.88	2.50
28	21.25	16.56	16.88	19.38	14.38	10.00
32	24.06	14.38	11.88	17.19	11.88	13.13
36	16.88	15.31	15.94	13.75	7.19	8.13
40	31.88	19.38	15.31	17.19	13.75	9.69
Mean	17.13	11.31	10.47	12.53	7.56	6.85
SD	8.83	5.80	4.77	5.96	5.09	3.35

Mean Error Rate of Subjects in the Fixed-Adaptive Group

	SESSION					
Subject	1	2	3	4	5	Test
1	16.25	6.88	5.31	9.38	6.56	5.63
5	10.94	4.69	4.38	10.63	7.19	9.07
9	11.25	9.38	2.81	9.69	6.88	5.94
13	12.50	8.44	5.63	3.75	6.25	4.06
17	8.44	5.63	4.38	6.88	4.06	4.38
21	19.38	18.13	24.69	20.00	14.06	9.69
25	24.69	8.13	10.00	7.81	10.94	4.38
29	8.75	3.44	4.38	6.56	6.25	7.50
33	12.50	8.13	8.75	11.25	10.00	9.06
37	28.44	15.63	13.75	12.19	6.88	4.69
Mean	15.31	8.85	8.41	9.81	7.91	6.44
SD	6.48	4.41	6.27	4.15	2.76	2.09

Experiment 4 / Chapter 4

Mean Error Rate of Subjects in the Cumulative-Random Group

	SESSION					
Subject	1	2	3	4	5	Test
2	9.06	5.31	4.06	5.00	4.69	3.75
6	5.31	5.00	3.44	1.56	3.44	0.94
10	10.31	9.38	4.69	9.38	8.44	6.25
14	12.81	7.81	7.50	9.69	6.25	7.81
18	14.69	10.00	8.13	8.44	10.94	7.81
22	7.50	5.00	4.38	9.69	6.88	7.19
26	7.81	3.13	6.88	7.81	7.19	8.75
30	15.63	24.69	25.00	22.81	18.44	18.13
34	17.50	14.69	8.13	12.19	9.69	5.00
38	10.94	12.81	11.25	12.81	11.56	15.00
Mean	11.16	9.78	8.35	9.94	8.75	8.06
SD	3.73	6.08	6.00	5.30	4.05	4.83

Mean Error Rates of Subjects in the Cumulative-Adaptive Group

	SESSION					
Subject	1	2	3	4	5	Test
3	6.56	6.88	5.63	5.94	7.50	4.38
7	10.63	8.75	5.63	5.00	3.44	4.06
11	17.19	5.31	7.19	11.88	4.38	7.81
15	16.88	11.56	10.94	21.25	16.88	10.00
19	5.63	3.44	3.13	6.88	6.25	3.75
23	7.19	10.00	2.50	7.19	10.63	7.50
27	10.63	9.38	9.38	8.75	6.56	6.88
31	7.81	7.81	5.94	7.19	5.94	3.75
35	4.38	5.94	3.13	4.38	4.38	3.75
39	9.38	13.13	12.50	10.31	9.06	10.00
Mean	9.63	8.22	6.60	8.88	7.50	6.19
SD	4.17	2.80	3.24	4.67	3.76	2.44

Experiment 4 / Chapter 4

Mean Adjusted Scores of Subjects in the the Fixed-Randon Group

	SESSION					
Subject	1	2	3	4	5	Test
4	118.68	114.97	101.97	105.16	103.79	96.39
8	188.50	147.33	137.74	157.26	131.08	122.09
12	218.52	179.36	163.79	169.61	153.94	151.49
16	159.65	134.34	130.78	139.50	129.81	114.77
20	216.68	174.53	159.32	184.68	154.11	153.68
24	211.37	188.25	171.46	199.50	173.84	165.01
28	319.38	263.72	245.21	249.84	208.26	196.06
32	360.26	292.95	251.16	261.81	223.87	211.94
36	249.09	217.13	217.00	213.78	176.76	169.62
40	321.58	247.41	219.86	230.65	196.46	185.48
Mean	236.37	196.00	179.76	191.18	165.19	156.65
SD	72.82	55.43	48.31	47.22	35.94	35.06

Mean Adjusted Scores of Subjects in the Fixed-Adaptive Group

	SESSION					
Subject	1	2	3	4	5	Test
1	258.68	181.14	164.80	188.52	165.18	145.01
5	189.74	150.61	129.57	141.60	118.74	125.04
9	202.43	178.83	160.50	164.18	146.84	140.45
13	227.65	194.44	171.81	159.46	150.91	142.16
17	239.51	189.36	155.16	175.03	149.60	132.72
21	292.62	201.73	187.53	178.73	153.85	136.60
25	367.86	256.79	249.85	246.67	222.18	185.67
29	365.88	315.77	254.73	234.79	213.88	217.75
33	274.20	219.18	195.41	207.74	183.79	164.21
37	422.37	307.66	278.13	281.34	275.03	258.82
Mean	284.09	219.55	194.75	197.81	178.00	164.84
SD	73.75	53.06	46.96	41.91	44.11	41.21

Experiment 4 / Chapter 4

Mean Adjusted Scores of Subjects in the Cumulative-Random Group

	SESSION					
Subject	1	2	3	4	5	Test
2	146.28	141.62	123.24	128.56	128.18	118.05
6	163.58	173.67	164.71	161.23	150.53	141.50
10	168.80	178.39	167.11	174.71	161.31	152.94
14	127.09	119.57	110.09	110.87	101.37	100.38
18	183.49	179.58	184.77	179.57	155.76	146.66
22	178.10	181.32	178.47	196.87	188.55	199.93
26	163.79	164.12	145.18	166.03	141.91	134.77
30	190.27	215.53	224.77	218.23	190.23	174.11
34	193.72	200.07	175.49	185.65	176.68	177.78
38	182.34	188.65	188.87	195.43	196.79	221.11
Mean	169.75	174.25	166.27	171.72	159.13	156.72
SD	19.64	26.14	31.57	30.56	28.78	34.99

Mean Adjusted Scores of Subjects in the Cumulative-Adaptive Group

	SESSION					
Subject	1	2	3	4	5	Test
3	145.56	157.39	139.20	135.24	124.14	124.09
7	139.40	131.85	117.63	110.97	104.68	95.19
11	204.80	192.98	193.25	213.50	174.16	212.47
15	203.79	196.83	198.70	221.36	222.61	171.80
19	272.20	172.79	169.28	183.83	185.30	163.06
23	172.27	194.83	176.77	177.00	190.57	193.03
27	175.72	186.37	187.34	179.37	174.33	150.86
31	157.15	146.15	129.63	136.65	114.76	102.36
35	170.18	214.00	181.91	188.06	214.28	214.85
39	179.22	213.92	221.68	227.84	211.04	215.79
Mean	172.03	180.71	171.54	177.38	171.59	164.35
SD	20.34	26.54	31.34	37.13	40.62	43.38

Experiment 4 / Chapter 4

Size of set (number of items per set from which stimuli were sampled) for each session, for each subject of the two cumulative groups

Cumulative-Random Group

Subject	Session	Set Size
2	1	80
	2	80
	3	80
	4	80
	5	80
	6	80
6	1	64
	2	80
	3	80
	4	80
	5	80
	6	80
10	1	64
	2	80
	3	80
	4	80
	5	80
	6	80
14	1	80
	2	80
	3	80
	4	80
	5	80
	6	80
18	1	52
	2	64
	3	68
	4	76
	5	80
	6	80
22	1	40
	2	48
	3	56
	4	64
	5	68
	6	80

Cumulative-Adaptive Group

Subject	Session	Set Size
3	1	80
	2	80
	3	80
	4	80
	5	80
	6	80
7	1	80
	2	80
	3	80
	4	80
	5	80
	6	80
11	1	24
	2	36
	3	40
	4	40
	5	48
	6	80
15	1	40
	2	52
	3	64
	4	64
	5	76
	6	80
19	1	28
	2	40
	3	44
	4	48
	5	60
	6	80
23	1	36
	2	40
	3	48
	4	48
	5	60
	6	80

Experiment 4 / Chapter 4

Cumulative-Random Group			Cumulative-Adaptive Group		
Subject	Session	Set Size	Subject	Session	Set Size
26	1	76	27	1	40
	2	80		2	52
	3	80		3	52
	4	80		4	60
	5	80		5	72
	6	80		6	80
30	1	52	31	1	76
	2	72		2	80
	3	80		3	80
	4	80		4	80
	5	80		5	80
	6	80		6	80
34	1	40	35	1	32
	2	52		2	40
	3	60		3	40
	4	76		4	40
	5	80		5	48
	6	80		6	80
38	1	40	39	1	36
	2	48		2	36
	3	48		3	52
	4	56		4	52
	5	56		5	60
	6	80		6	80

Experiment 4 / Chapter 4

Summary table of analysis of co-variance on RT data over the first three sessions adjusted for covariate pretest adjusted score (ASC)

Source	Sums of Squares	df	Variance	F ratio	p
BETWEEN Ss	180937.07	39	4639.41		
A vs R	13164.07	1	13164.07	5.49	.05
C vs F	33369.20	1	33369.20	13.92	.01
(A vs R) x (C vs F)	3745.04	1	3745.04	1.56	NS
Covariates	46755.96	1	46755.96	19.50	.01
Ss within Groups	83902.73	35	2397.22		
WITHIN Ss	27479.15	80	343.49		
Practice	7379.96	2	3689.98	45.23	.01
Practice x (A vs R)	783.66	2	391.83	4.80	.05
Practice x (C vs F)	12102.74	2	6051.37	74.17	.01
Practice x (A vs R) x (C vs R)	1338.33	2	669.17	8.20	.01
Ss within Groups x Practice	5874.45	72	81.59		
TOTAL	208416.15	119			

Summary table of analysis of co-variance on ERR (percentage of errors) data over the first three sessions adjusted for covariate pretest adjusted score (ASC)

Source	Sums of Squares	df	Variance	F ratio	p
BETWEEN Ss	0.3278	39	0.0084		
A vs R	0.0109	1	0.0109	1.83	NS
C vs F	0.0251	1	0.0251	4.21	.05
(A vs R) x (C vs F)	0.0003	1	0.0003	0.05	NS
Covariates	0.0828	1	0.0828	13.88	.01
Ss within Groups	0.2088	35	0.0060		
WITHIN Ss	0.1330	80	0.0017		
Practice	0.0523	2	0.0262	27.51	.01
Practice x (A vs R)	0.0000	2	0.0000	0.02	NS
Practice x (C vs F)	0.0122	2	0.0061	6.40	.01
Practice x (A vs R) x (C vs F)	0.0000	2	0.0000	0.02	NS
Ss within Groups x Practice	0.0685	72	0.0010		
TOTAL	0.4609	119			

Experiment 4 / Chapter 4

Summary table of analysis of co-variance on ASC (Adjusted Score) data over the first three training sessions adjusted for covariate pretest adjusted score ASC

Source	Sums of Squares	df	Variance	F ratio	p
BETWEEN Ss	313583.8	39	8040.6		
A vs R	8307.3	1	8307.3	2.24	NS
C vs R	63287.4	1	63267.4	17.09	.01
(A vs R) x (C vs F)	4298.6	1	4298.6	1.16	NS
Covariates	108158.9	1	108158.9	29.22	.01
Ss within Groups	129551.8	35	3701.5		
WITHIIN Ss	76010.1	80	950.1		
Practice	28565.6	2	14282.8	65.85	.01
Practice x (A vs R)	1149.6	2	574.8	2.65	NS
Practice x (C vs F)	28901.2	2	14450.6	66.63	.01
Practice x (A vs R) x (C vs F)	1777.9	2	889.0	4.10	.05
Ss within Groups x Practice	15615.7	72	216.9		
TOTAL	389593.1	119			

A vs R : Adaptive versus Random
C vs F : Cumulative versus Fixed

Experiment 4 / Chapter 4

Mean reaction time (csecs) on the test session, averaged across subjects, for each set of chord difficulty

Set	FR	FA	CR	CA
1	98.91	100.78	96.40	95.16
2	116.23	119.61	112.08	117.46
3	164.48	175.94	164.42	170.46
4	162.82	174.03	165.99	179.30
5	181.00	196.37	180.78	199.63

Summary table of analysis of co-variance on RT data over the test session, adjusted for covariate pretest adjusted score (ASC)

Source	Sums of Squares	df	Variance	F ratio	p
BETWEEN Ss	236186.8	39	6056.1		
A vs R	3795.4	1	3795.4	0.97	NS
C vs F	49.5	1	49.5	0.01	NS
(A vs R) x (C vs R)	1.3	1	1.3	0.00	NS
Covariates	95890.8	1	95890.8	24.60	
Ss within Groups	136449.9	32	3898.6		
WITHIN Ss	299315.7	160	1870.7		
Chord	246318.2	4	61579.5	175.22	.001
Chord x (A vs R)	1733.0	4	433.2	1.23	NS
Chord x (C vs F)	522.7	4	138.2	0.39	NS
Chord x (A vs R) x (C vs F)	104.5	4	26.1	0.07	NS
Ss within Groups x Practice	50607.3	144	351.4		
TOTAL	535502.5	199			

Experiment 4 / Chapter 4

Mean percentage of errors on the test session, averaged across subjects, for each set of chord pattern difficulty

Set	FR	FA	CR	CA
1	3.4	1.8	3.2	2.5
2	2.3	2.1	3.1	1.9
3	9.3	10.6	10.7	8.5
4	8.7	8.0	12.8	11.0
5	10.3	9.2	11.3	7.7

Summary table of analysis of co-variance on ERR (percentage of error) data over the test session, adjusted for covariate pretest adjusted score (ASC)

Source	Sums of Squares	df	Variance	F ratio	p
BETWEEN Ss	0.25	39	0.01		
A vs R	0.01	1	0.01	1.22	NS
C vs F	0.00	1	0.00	0.43	NS
(A vs R) x (C vs F)	0.00	1	0.00	0.45	NS
Covariates	0.04	1	0.04	6.43	
Ss within Groups	0.20	35	0.01		
WITHIN Ss	0.57	160	0.00		
Chord	0.26	4	0.06	30.84	.01
Chord x (A vs R)	0.00	4	0.00	0.26	NS
Chord x (C vs F)	0.01	4	0.00	1.26	NS
Chord x (A vs R) (C vs F)	0.00	4	0.00	0.34	NS
Ss within Groups x Chord	0.30	144	0.00		
TOTAL	0.82	199			

Experiment 4 / Chapter 4

Mean adajusted scores on the test session, averaged across subjects, for each set of chord pattern difficulty

Set	FR	FA	CR	CA
1	102.99	103.22	100.01	97.07
2	119.33	122.50	116.05	120.35
3	182.82	197.00	184.59	187.22
4	180.99	191.60	199.44	205.53
5	205.68	216.54	209.04	217.97

Summary table of analysis of co-variance on ASC (adjusted scores) data over the test session, adjusted for covariate pretest adjusted score (ASC)

Source	Sums of Squares	df	Variance	F ratio	p
BETWEEN Ss	321716.2	39	8249.1		
A vs R	1685.7	1	1685.7	0.34	NS
C vs F	126.1	1	126.1	0.03	NS
(A vs R) x (C vs F)	175.9	1	175.9	0.04	NS
Covariates	146881.5	1	146881.5	29.74	
Ss within Groups	172847.1	35	4938.5		
WITHIN Ss	492910.4	160	3080.7		
Chord	390805.3	4	97701.3	143.25	.01
Chord x (A vs R)	810.2	4	202.5	0.30	NS
Chord x (C vs F)	2977.1	4	744.3	1.09	NS
Chord x (A vs R) x (C vs F)	225.1	4	56.3	0.08	NS
Ss within Groups x Chord	98092.7	144	681.2		
TOTAL	814626.6	199			

Experiment 4 / Chapter 4

Total training time (mins) over five sessions for each subject

Fixed-Random Group	
Subject	Total Time
4	28.42
8	33.71
12	44.58
16	35.25
20	40.74
24	45.63
28	56.27
32	61.86
36	49.16
40	51.48
Mean	44.71
SD	9.93

Fixed-Adaptive Group	
Subject	Total Time
1	46.22
5	35.90
9	41.75
13	44.50
17	45.49
21	43.60
25	62.03
29	69.45
33	51.70
37	69.59
Mean	51.02
SD	11.28

Cumulative-Random Group	
Subject	Total Time
2	33.58
6	41.76
10	41.51
14	27.63
18	42.19
22	45.92
26	38.93
30	43.47
34	43.43
38	44.74
Mean	40.32
SD	5.35

Cumulative-Adaptive Group	
Subject	Total Time
3	34.99
7	30.00
11	47.27
15	46.92
19	44.71
23	44.92
27	43.86
31	33.94
35	49.33
39	50.05
Mean	42.60
SD	6.67

APPENDIX 4

Experiment 4 / Chapter 6

Analysis of variance on RT data of the Fixed-Random group

Factor A: Sessions - Levels: 6
Factor B: Difficulty - Levels: 5
Subjects: 10

Mean RT

	DIFFICULTY					
Session	0	1	2	3	4	Mean
1	124.64	157.54	235.70	221.49	236.78	195.23
2	108.22	136.46	202.01	194.89	220.00	172.32
3	104.99	126.46	183.46	185.15	198.90	159.77
4	106.81	132.91	194.50	189.57	208.47	166.45
5	100.16	121.82	172.96	166.78	192.10	150.76
6	99.19	116.29	164.46	163.63	181.44	145.00
Mean	107.33	131.89	192.18	186.92	206.28	164.92

Summary Table

Source	Sums of Squares	df	Variance	F ratio	p
BETWEEN Ss	405013.01	9	45001.4		
WITHIN Ss	628326.64	290	2166.64		
Factor A	79978.64	5	15995.7	25.33	.01
Factor B	440690.71	4	110173	67.46	.01
A x B	9045.82	20	452.29	7.14	.01
S x A	28412.81	45	631.40		
S x B	58790.35	36	1633.07		
S x A x B	11408.32	180	63.38		
TOTAL	1033339.65	299			

Experiment 4 / Chapter 6

Analysis of variance on Percentage Correct (Pcr) data of the Fixed-Random group

Factor A: Sessions - Levels: 6
Factor B: Difficulty - Levels: 5
Subjects: 10

Mean Pcr

	DIFFICULTY					
Session	0	1	2	3	4	Mean
1	0.94	0.90	0.74	0.76	0.78	0.82
2	0.97	0.95	0.82	0.83	0.86	0.89
3	0.95	0.94	0.84	0.87	0.87	0.89
4	0.96	0.93	0.83	0.82	0.83	0.87
5	0.96	0.96	0.90	0.90	0.90	0.92
6	0.96	0.98	0.91	0.91	0.90	0.93
Mean	0.96	0.94	0.84	0.85	0.85	0.89

Summary Table

Source	Sums of Squares	df	Variance	F ratio	p
BETWEEN Ss	0.85	9	0.09		
WITHIN Ss	2.26	290	0.01		
Factor A	0.37	5	0.07	16.49	.01
Factor B	0.79	4	0.20	16.55	.01
A x B	0.11	20	0.01	2.96	.01
S x A	0.20	45	0.00		
S x B	0.43	36	0.01		
S x A x B	0.35	180	0.02		
TOTAL	3.11	299			

Experiment 4 / Chapter 6

Crosstabulation of "accuracy of current response" by "accuracy of preceding response", controlling for difficulty level of the current trial and overall

	Pre Error	Pre Correct	Total	
Level 0				
Cur Error	28	125	153	x^2 = 6.89,
Cur Correct	378	3036	3567	df=1, p<.01
Total	406	3161	3567	Phi = 0.05
Level 1				
Cur Error	39	199	238	x^2 = 5.94,
Cur Correct	440	3556	3996	df=1, p<.01
Total	479	3755	4234	Phi = 0.04
Level 2				
Cur Error	93	489	582	x^2 = 10.65,
Cur Correct	341	2730	3071	df=1, p<.001
Total	434	3219	3653	Phi = 0.06
Level 3				
Cur Error	59	487	546	x^2 = 0.92
Cur Correct	283	2732	3015	df=1, NS
Total	342	3219	3561	Phi = 0.02
Level 4				
Cur Error	82	483	565	x^2 = 10.05
Cur Correct	337	3043	3380	df=1, P<.005
Total	419	3526	3945	Phi = 0.05
OVERALL				
Cur Error	301	1783	2084	x^2 = 28.52
Cur Correct	1779	15097	16876	df=1, p<.001
Total	2080	16880	18960	Phi = 0.39

Experiment 4 / Chapter 6

Crosstabulation of "accuracy of current response" by "accuracy of preceding response", controlling for practice sessions

	Prev Error	Prev Correct	Total	
Session 1				
Cur Error	121	421	542	$x^2 = 11.39$
Cur Correct	424	2196	2618	df=1, p<.001
Total	545	2615	3160	Phi = 0.06
Session 2				
Cur Error	51	308	359	$x^2 = 3.10$
Cur Correct	306	2495	2801	df=1, NS
Total	357	2803	3160	Phi = 0.03
Session 3				
Cur Error	50	282	332	$x^2 = 7.91$
Cur Correct	280	2548	2828	df=1, p<.005
Total	330	2830	3160	Phi = 0.05
Session 4				
Cur Error	41	356	397	$x^2 = 1.64$
Cur Correct	352	2411	2763	df=1, NS
Total	393	2767	3160	Phi = 0.02
Session 5				
Cur Error	20	219	239	$x^2 = 0.12$
Cur Correct	220	2701	2921	df=1, NS
Total	240	2920	3160	Phi = 0.01
Session 6				
Cur Error	18	197	215	$x^2 = 0.65$
Cur Correct	197	2748	2945	df=1, NS
Total	215	2945	3160	Phi = 0.02

Experiment 4 / Chapter 6

Analysis of variance on RT data of the Fixed-Adaptive group

Factor A: Sessions - Levels: 6
Factor B: Difficulty - Levels: 5

Mean RT

	DIFFICULTY					
Session	0	1	2	3	4	Mean
1	144.71	175.20	273.01	263.38	307.18	232.69
2	117.26	150.80	238.09	216.93	246.14	193.84
3	113.08	135.40	202.03	194.85	218.08	172.69
4	113.60	134.48	200.57	196.95	224.55	174.03
5	103.87	124.86	183.46	187.93	209.42	161.91
6	101.23	120.01	179.70	178.07	204.39	156.68
Mean	115.63	140.13	212.81	206.35	234.96	181.97

Summary Table

Source	Sums of Squares	df	Variance	F ratio	p
BETWEEN Ss	656297.17	9	72921.91		
WITHIN Ss	1089722.16	290	3757.66		
Factor A	195261.60	5	39052.32	25.68	.001
Factor B	630355.81	4	157589	37.90	.001
A x B	22454.38	20	1122.72	8.58	.001
S x A	68421.42	45	1520.48		
S x B	149683.77	36	4157.88		
S x A x B	23545.18	180	130.81		
TOTAL	1746019.33	299			

Experiment 4 / Chapter 6

Analysis of variance on Percentage Correct (Pcr) data of the Fixed-Adaptive group

Factor A: Sessions - Levels: 6
Factor B: Difficulty - Levels: 5
Subjects: 10

Mean Pcr

	DIFFICULTY					
Session	0	1	2	3	4	Mean
1	0.95	0.94	0.77	0.82	0.83	0.86
2	0.96	0.97	0.88	0.89	0.91	0.92
3	0.96	0.97	0.89	0.90	0.89	0.92
4	0.97	0.97	0.87	0.87	0.87	0.91
5	0.97	0.97	0.88	0.92	0.90	0.93
6	0.98	0.98	0.89	0.92	0.91	0.94
Mean	0.97	0.97	0.86	0.89	0.88	0.91

Summary Table

Source	Sums of Squares	df	Variance	F ratio	p
BETWEEN Ss	0.32	9	0.04		
WITHIN Ss	1.62	290	0.01		
Factor A	0.18	5	0.04	7.21	.001
Factor B	0.57	4	0.14	20.71	.001
A x B	0.06	20	0.00	1.43	NS
S x A	0.22	45	0.00		
S x B	0.25	36	0.01		
S x A x B	0.35	180	0.00		
TOTAL	1.94	299			

Experiment 4 / Chapter 6

Crosstabulation of "accuracy of current response" by "accuracy of preceding response", controlling for level of difficulty of the current trial and overall

	Prev Error	Prev Correct	Total	
Level 0				
Cur Error	7	91	98	$x^2 = 0.40$
Cur Correct	266	2508	2774	df=1, NS
Total	273	2599	2872	Phi = 0.02
Level 1				
Cur Error	11	103	114	$x^2 = 0.00$
Cur Correct	327	2962	3289	df=1, NS
Total	338	3065	3403	Phi = 0.00
Level 2				
Cur Error	51	527	578	$x^2 = 0.11$
Cur Correct	339	3281	3620	df=1, NS
Total	390	3808	4198	Phi = 0.01
Level 3				
Cur Error	54	407	461	$x^2 = 2.15$
Cur Correct	330	3166	3496	df=1, NS
Total	384	3573	3957	Phi = 0.02
Level 4				
Cur Error	57	487	544	$x^2 = 1.39$
Cur Correct	352	3634	3986	df=1, NS
Total	409	4121	4530	Phi = 0.02
OVERALL				
Cur Error	180	1615	1795	$x^2 = 0.67$
Cur Correct	1614	15551	17165	df=1, NS
	1794	17166	18960	Phi = 0.01

Experiment 4 / Chapter 6

Crosstabulation of "accuracy of current response" by "accuracy of preceding response", controlling for practice session

	Prev Error	Prev Correct	Total	
Session 1				
Cur Error	76	410	486	$x^2 = 0.03$
Cur Correct	407	2267	2674	df=1, NS
Total	483	2677	3160	Phi = 0.00
Session 2				
Cur Error	26	252	278	$x^2 = 0.02$
Cur Correct	256	2626	2882	df=1, NS
Total	282	2878	3160	Phi = 0.00
Session 3				
Cur Error	29	237	266	$x^2 = 1.93$
Cur Correct	238	2656	2894	df=1, NS
Total	267	2893	3160	Phi = 0.03
Session 4				
Cur Error	21	290	311	$x^2 = 3.09$
Cur Correct	286	2563	2849	df=1, NS
Total	307	2853	3160	Phi = 0.03
Session 5				
Cur Error	16	234	250	$x^2 = 0.67$
Cur Correct	235	2675	2910	df=1, NS
Total	251	2909	3160	Phi = 0.02
Session 6				
Cur Error	12	192	204	$x^2 = 0.04$
Cur Correct	192	2674	2956	df=1, NS
Total	204	2956	3160	Phi = 0.01

Experiment 4 / Chapter 6

Analysis of variance on the RT data of the Cumulative-Random group

Factor A: Sessions - Levels: 6
Factor B: Difficulty - Levels: 5
Subjects: 10

Mean RT

	DIFFICULTY					
Session	0	1	2	3	4	Mean
1	119.64	144.25	218.09	206.89	215.31	180.84
2	107.31	124.04	197.61	197.45	209.01	167.09
3	102.78	120.47	186.47	186.89	195.73	158.46
4	103.36	123.38	188.19	187.98	201.43	160.87
5	97.70	114.49	166.05	167.16	188.53	146.79
6	96.71	112.37	164.53	167.84	182.09	144.71
Mean	104.58	123.16	186.82	185.70	198.68	159.79

Summary Table

Source	Sum of Squares	df	Variance	F ratio	p
BETWEEN Ss	219011.56	9	24334.62		
WITHIN Ss	560821.53	290	1933.87		
Factor A	44782.85	5	8956.56	27.90	.001
Factor B	438247.55	4	109561.9	91.61	.001
A x B	5733.77	20	286.69	3.54	.01
S x A	14445.77	45	321.02		
S x B	43054.19	36	1195.95		
S x A x B	14557.27	180	80.87		
TOTAL	779833.09	299			

Experiment 4 / Chapter 6

Analysis of variance on Percentage Correct (Pcr) data of the Cumulative-Random group

Factor A: Sessions - Levels: 6
Factor B: Difficulty - Levels: 5
Subjects: 10

Mean Pcr

	DIFFICULTY					
Session	0	1	2	3	4	Mean
1	0.93	0.92	0.75	0.74	0.81	0.83
2	0.96	0.95	0.87	0.76	0.83	0.87
3	0.96	0.97	0.89	0.85	0.86	0.91
4	0.94	0.94	0.86	0.85	0.86	0.89
5	0.96	0.96	0.89	0.85	0.87	0.91
6	0.97	0.97	0.89	0.87	0.88	0.92
Mean	0.95	0.95	0.86	0.82	0.85	0.89

Summary Table

Source	Sum of Squares	df	Variance	F ratio	p
BETWEEN Ss	1.17	9	0.13		
WITHIN Ss	2.58	290	0.01		
Factor A	0.26	5	0.05	10.65	.01
Factor B	0.91	4	0.23	13.49	.025
A x B	0.10	20	0.00	1.86	.05
S x A	0.22	45	0.00		
S x B	0.61	36	0.02		
S x A x B	0.48	180	0.00		
TOTAL	3.75	299			

Experiment 4 / Chapter 6

Crosstabulation of "accuracy of current response" by "accuracy of preceding response", controlling for level of difficulty of the current trial and overall

	Prev Error	Prev Correct	Total	
Level 0				
Cur Error	24	240	264	x^2 = 0.00
Cur Correct	433	4313	4746	df=1, NS
Total	457	4553	5010	Phi = 0.00
Level 1				
Cur Error	33	201	234	x^2 = 4.32
Cur Correct	436	4055	4491	df=1, p<.05
Total	469	4256	4725	Phi = 0.03
Level 2				
Cur Error	56	501	557	x^2 = 0.18
Cur Correct	318	3072	3390	df=1, NS
Total	374	3573	3947	Phi = 0.01
Level 3				
Cur Error	62	363	425	x^2 = 19.97
Cur Correct	187	2215	2402	df=1, p<.0001
Total	249	2578	2827	Phi = 0.09
Level 4				
Cur Error	28	260	288	x^2 = 0.22
Cur Correct	188	1975	2163	df=1, NS
Total	216	2235	2451	Phi = 0.01
OVERALL				
Cur Error	105	1383	1488	x^2 = 1.30
Cur Correct	1384	16088	17472	df=1, NS
Total	1489	17471	18960	Phi = 0.01

Experiment 4 /Chapter 6

Crosstabulation of "accuracy of current response" by "accuracy of preceding response", controlling for practice session

	Prev Error	Prev Correct	Total	
Session 1				
Cur Error	42	311	353	x^2 = 0.17
Cur Correct	309	2498	2807	df=1, NS
Total	351	2809	3160	Phi = 0.01
Session 2				
Cur Error	41	264	305	x^2 = 4.69
Cur Correct	268	2587	2855	df=1, p<.05
Total	309	2851	3160	Phi = 0.04
Session 3				
Cur Error	38	226	264	x^2 =13.06
Cur Correct	225	2671	2896	df=1, p<.001
Total	263	2897	3160	Phi = 0.07
Session 4				
Cur Error	36	277	313	x^2 = 0.84
Cur Correct	276	2571	2847	df=1, NS
Total	312	2848	3160	Phi = 0.02
Session 5				
Cur Error	26	252	278	x^2 = 0.08
Cur Correct	249	2633	2882	df=1, NS
Total	275	2885	3160	Phi = 0.01
Session 6				
Cur Error	20	235	255	x^2 = 0.00
Cur Correct	235	2670	2905	df=1, NS
Total	255	2905	3160	Phi = 0.00

Experiment 4 / Chapter 6

Analysis of variance on RT data of the Cumulative-Adaptive group

Factor A: Sessions - Levels: 6
Factor B: Difficulty - Levels: 5
Subjects: 10

Mean RTs

	DIFFICULTY					
Session	0	1	2	3	4	Mean
1	133.42	160.58	210.69	204.11	223.35	180.43
2	111.33	138.29	207.97	196.62	213.21	173.48
3	106.97	127.54	191.68	186.85	202.67	163.14
4	111.14	133.18	195.93	191.03	204.55	167.16
5	98.25	121.64	172.51	181.00	196.93	154.07
6	95.30	117.93	171.10	181.00	200.77	153.22
Mean	109.40	133.19	191.65	190.10	206.91	166.25

Summary Table

Source	Sums of Squares	df	Variance	F ratio	p
BETWEEN Ss	371688.77	9	41298.75		
WITHIN Ss	571720.51	290	1971.45		
Factor A	39414.85	5	7882.97	57.46	.001
Factor B	431527.37	4	107881.9	60.00	.001
A x B	4677.96	20	233.90	2.28	.001
S x A	12918.37	45	287.08		
S x B	64730.92	36	1798.08		
S x A x B	18451.04	180	102.51		
TOTAL	943409.27	299			

Experiment 4 / Chapter 6

Analysis of variance on Percentage Correct (Pcr) data of the Cumulative-Adaptive group

Factor A: Sessions - Levels: 6
Factor B: Difficulty - Levels: 5
Subjects: 10

Mean Pcr

	DIFFICULTY					
Session	0	1	2	3	4	Mean
1	0.93	0.91	0.81	0.83	0.89	0.87
2	0.97	0.95	0.88	0.88	0.91	0.92
3	0.98	0.98	0.91	0.89	0.93	0.94
4	0.96	0.95	0.87	0.88	0.90	0.91
5	0.97	0.98	0.92	0.89	0.93	0.94
6	0.98	0.98	0.92	0.89	0.92	0.94
Mean	0.96	0.96	0.88	0.88	0.92	0.92

Summary Table

Source	Sums of Squares	df	Variance	F ratio	p
BETWEEN Ss	0.31	9	0.00		
WITHIN Ss	1.16	290	0.00		
Factor A	0.16	5	0.00	8.60	.001
Factor B	0.39	4	0.01	16.06	.001
A x B	0.00	20	0.00	1.43	NS
S x A	0.16	45	0.00		
S x B	0.22	36	0.01		
S x A x B	0.20	180	0.00		
TOTAL	1.47	299			

Crosstabulation of "accuracy of current response" by "accuracy of preceding response", controlling for level of difficulty of the current trial and overall

	Prev Error	Prev Correct	Total	
Level 0				
Cur Error	27	195	222	$x^2 = 5.02$
Cur Correct	332	3952	4284	df=1, p<.05
Total	359	4147	4506	Phi = 0.04
Level 1				
Cur Error	7	202	209	$x^2 = 5.44$
Cur Correct	359	4109	4468	df=1, p<.05
Total	366	4311	4677	Phi = 0.04
Level 2				
Cur Error	34	501	535	$x^2 = 1.13$
Cur Correct	335	3987	4322	df=1, NS
Total	369	4488	4857	Phi = 0.02
Level 3				
Cur Error	24	331	355	$x^2 = 1.06$
Cur Correct	215	2305	2520	df=1, NS
Total	239	2636	2875	Phi = 0.02
Level 4				
Cur Error	13	154	167	$x^2 = 0.00$
Cur Correct	143	1735	1878	df=1, NS
Total	156	1889	2045	Phi = 0.00
OVERALL				
Cur Error	203	1565	1768	$x^2 = 10.62$
Cur Correct	1562	15592	17192	df=1, p<.001
Total	1765	17195	18960	Phi = 0.02

Crosstabulation of "accuracy of current response" by "accuracy of preceding response", controlling for practice session

	Prev Error	Prev Correct	Total	
Session 1				
Cur Error	33	272	305	x^2 = 0.36
Cur Correct	273	2582	2855	df=1, NS
Total	306	2854	3160	Phi = 0.01
Session 2				
Cur Error	17	243	260	x^2 = 0.81
Cur Correct	242	2658	2900	df=1, NS
Total	259	2901	3160	Phi = 0.02
Session 3				
Cur Error	14	193	207	x^2 = 0.00
Cur Correct	196	2757	2953	df=1, NS
Total	210	2950	2953	Phi = 0.00
Session 4				
Cur Error	20	262	282	x^2 = 0.97
Cur Correct	260	1618	2878	df=1, NS
Total	280	2880	3160	Phi = 0.02
Session 5				
Cur Error	12	224	236	x^2 = 1.78
Cur Correct	225	2699	2924	df=1, NS
Total	237	2923	3160	Phi = 0.03
Session 6				
Cur Error	9	189	198	x^2 = 0.74
Cur Correct	188	2774	2962	df=1, NS
Total	197	2963	3160	Phi = 0.02

Fixed-Random Group: Analysis of Variance on the effects of "accuracy of the current response", "difficulty level of the current response" and "practice session" on the residual RT of the current repsonse

RES1 Current Residual RT
BY ACC1 Current Response
DIFF1 Current Difficulty
SESSION

SESSION

ACC1	1	2	3	4	5	6	Mean
0	57.02 (548)	35.42 (362)	13.98 (335)	24.23 (401)	.20 (242)	-3.88 (219)	27.37 (2107)
1	22.93 (2652)	4.26 (2838)	-7.27 (2865)	-2.32 (2799)	-14.20 (2958)	-20.54 (2981)	-3.37 (17093)
Mean	28.77 (3200)	7.78 (3200)	-5.05 (3200)	1.01 (3200)	-13.11 (3200)	-19.40 (3200)	0.00 (19200)

SESSION

DIFF1	1	2	3	4	5	6	Mean
0	-39.92 (636)	-56.41 (553)	-59.51 (613)	-57.62 (629)	-64.39 (547)	-64.66 (615)	-56.86 (3593)
1	-6.84 (727)	-28.36 (748)	-38.48 (701)	-31.54 (706)	-42.64 (734)	-48.28 (695)	-32.54 (4311)
2	70.74 (597)	36.71 (628)	18.56 (612)	29.67 (631)	8.12 (611)	.07 (630)	27.06 (3709)
3	56.74 (595)	30.35 (646)	20.06 (564)	24.77 (586)	1.85 (637)	-.67 (563)	22.28 (3591)
4	71.99 (645)	55.45 (625)	34.69 (710)	43.98 (648)	27.45 (671)	16.61 (697)	41.09 (3996)
Mean	28.77 (3200)	7.78 (3200)	-5.05 (3200)	1.01 (3200)	-13.11 (3200)	-19.40 (3200)	0.00 (19200)

* * * A N A L Y S I S O F V A R I A N C E * * *

Source of Variation	Sum of Squares	DF	Mean Square	F	p
Main Effects	32524483.27	10	3252448.33	1265.22	0.0
ACC1	76762.39	1	76762.39	29.86	.000
DIFF1	26610636.69	4	6652659.17	2587.91	0.0
SESSION	4826281.50	5	965256.30	375.49	0.0
2-way Interactions	595139.00	29	20522.03	7.98	0.0
ACC1 DIFF1	33355.85	4	8338.96	3.24	.012
ACC1 SESSION	17415.72	5	3483.14	1.36	.238
DIFF1 SESSION	515645.00	20	25782.25	10.03	0.0
3-way Interactions	91915.20	20	4595.76	1.79	.017
ACC1 DIFF1 SESSION	91915.20	20	4595.76	1.79	.017
Explained	33211537.48	59	562907.42	218.97	0.0
Residual	49202528.98	19140	2570.67		
Total	82414066.46	19199	4292.62		

Fixed-Random Group: Analysis of Variance on the effects of "accuracy of the previous response", "accuracy of the current response" and "difficulty level of the current response" on the residual RT of the current response

```
    RES1      Current Residual RT
BY  ACCO      Previous Response
    ACC1      Current Response
    DIFF1     Current Difficulty
```

ACC1

ACC1	0	1
0	36.52 (301)	9.61 (1779)
1	25.11 (1783)	-5.19 (15097)

DIFF1

ACCO	0	1	2	3	4	Mean
0	-51.01 (406)	-20.02 (479)	43.69 (434)	43.30 (342)	58.75 (419)	13.50 (2080)
1	-57.63 (3161)	-34.60 (3755)	24.43 (3219)	19.68 (3219)	38.72 (3526)	-1.99 (16880)
Mean	-56.88 (3567)	-32.95 (4234)	26.72 (3653)	21.95 (3561)	40.85 (3945)	-0.29 (18960)

DIFF1

ACC1	0	1	2	3	4	Mean
0	-43.87 (153)	-18.69 (238)	34.67 (582)	36.38 (546)	47.59 (565)	26.76 (2084)
1	-57.46 (3414)	-33.80 (3996)	25.21 (3071)	19.33 (3015)	39.72 (3380)	-3.63 (16876)
Mean	-56.88 (3567)	-32.95 (4234)	26.72 (3653)	21.95 (3561)	40.85 (3945)	-0.29 (18960)

* * * A N A L Y S I S O F V A R I A N C E * * *

Source of Variation	Sum of Squares	DF	Mean Square	F	Signif of F
Main Effects	27783850.68	6	4630641.78	1648.76	0.0
ACCO	482164.55	1	482164.55	171.68	.000
ACC1	231952.45	1	231952.45	82.59	.000
DIFF1	25691044.89	4	6422761.22	2286.86	0.0
2-way Interactions	92146.16	9	10238.46	3.65	.000
ACCO ACC1	4589.47	1	4589.47	1.63	.201
ACCO DIFF1	63155.40	4	15788.85	5.62	.000
ACC1 DIFF1	28215.05	4	7053.76	2.51	.040
3-way Interactions	49514.78	4	12378.70	4.41	.002
ACCO ACC1 DIFF1	49514.78	4	12378.70	4.41	.002
Explained	27925511.63	19	1469763.77	523.32	0.0
Residual	53194024.53	18940	2808.56		
Total	81119536.16	18959	4278.68		

Fixed-Adaptive Group: Analysis of Variance on the effects of "accuracy of the current response", "difficulty lrvrl of the current response" and "practice session" on the residual RT of the current response

RES1 Current Residual RT
BY ACC1 Current Response
DIFF1 Current Difficulty
SESSION

SESSION

ACC1	1	2	3	4	5	6	Mean
0	87.04 (490)	24.98 (283)	1.18 (269)	7.09 (314)	7.24 (253)	-1.53 (206)	29.63 (1815)
1	56.66 (2710)	14.62 (2917)	-11.35 (2931)	-11.81 (2886)	-26.24 (2947)	-35.17 (2994)	-3.09 (17385)
Mean	61.31 (3200)	15.53 (3200)	-10.30 (3200)	-9.95 (3200)	-23.59 (3200)	-33.01 (3200)	-0.00 (19200)

SESSION

DIFF1	1	2	3	4	5	6	Mean
0	-45.16 (437)	-73.60 (403)	-78.62 (448)	-77.17 (502)	-89.19 (511)	-88.97 (611)	-76.68 (2912)
1	-14.91 (455)	-40.37 (547)	-54.36 (613)	-54.46 (591)	-63.04 (596)	-69.76 (659)	-51.41 (3461)
2	82.43 (663)	49.94 (778)	13.51 (742)	13.49 (666)	-5.24 (694)	-11.53 (705)	23.72 (4248)
3	74.33 (655)	26.34 (642)	7.30 (665)	6.90 (694)	-1.41 (776)	-12.40 (570)	16.76 (4002)
4	120.57 (990)	55.05 (830)	28.31 (732)	33.87 (747)	19.88 (623)	15.13 (655)	50.99 (4577)
Mean	61.31 (3200)	15.53 (3200)	-10.30 (3200)	-9.95 (3200)	-23.59 (3200)	-33.01 (3200)	-0.00 (19200)

		DIFF1				
	ACC1	0	1	2	3	4
SESSION 1						
	0	33.33 (23)	-14.08 (30)	90.43 (153)	73.99 (122)	120.02 (162)
	1	-49.52 (414)	-14.97 (425)	80.03 (510)	74.41 (533)	120.68 (828)
SESSION 2						
	0	-68.35 (15)	-29.43 (16)	34.80 (98)	22.25 (74)	43.86 (80)
	1	-73.80 (388)	-40.69 (531)	52.12 (680)	26.88 (568)	56.24 (750)
SESSION 3						
	0	-86.36 (17)	-54.13 (16)	-.94 (84)	9.15 (70)	25.49 (82)
	1	-78.32 (431)	-54.37 (597)	15.36 (658)	7.08 (595)	28.66 (650)
SESSION 4						
	0	-66.57 (17)	-36.37 (22)	5.20 (89)	4.20 (92)	35.21 (94)
	1	-77.54 (485)	-55.16 (569)	14.77 (577)	7.32 (602)	33.68 (653)
SESSION 5						
	0	-54.86 (16)	-49.29 (19)	5.89 (83)	21.38 (66)	25.29 (69)
	1	-90.30 (495)	-63.49 (577)	-6.75 (611)	-3.53 (710)	19.21 (554)
SESSION 6						
	0	-78.00 (11)	-39.22 (13)	-10.52 (76)	-.89 (46)	31.55 (60)
	1	-89.17 (600)	-70.38 (646)	-11.65 (629)	-13.42 (524)	13.47 (595)

* * * A N A L Y S I S O F V A R I A N C E * * *

Source of Variation	Sum of Squares	DF	Mean Square	F	p
Main Effects	56162378.42	10	5616237.84	1177.33	0.0
ACC1	16873.51	1	16873.51	3.54	.060
DIFF1	36552304.00	4	9138076.00	1915.61	0.0
SESSION	14285576.46	5	2857115.29	598.94	0.0
2-way Interactions	702438.29	29	58704.77	12.31	0.0
ACC1 DIFF1	81241.33	4	20310.33	4.26	.002
ACC1 SESSION	120898.89	5	24179.78	5.07	.000
DIFF1 SESSION	1521286.18	20	76064.31	15.95	0.0
3-way Interactions	139795.88	20	6989.79	1.47	.082
ACC1 DIFF1 SESSION	139795.88	20	6989.79	1.47	.082
Explained	58004612.58	59	983129.03	206.09	0.0
Residual	91303780.53	19140	4770.31		
Total	149308393.11	19199	7776.88		

Fixed-Adaptive Group: Analysis of Variance on the effects of "accuracy of the previous response", "accuracy of the current response" and "difficulty level of the current response" on the residual RT of the current response

RES1 Current Residual RT
BY ACCO Previous Response
ACC1 Current Response
DIFF1 Current Difficulty

ACC1

ACCO	0	1
0	68.36 (180)	16.98 (1614)
1	24.88 (1615)	-5.52 (15551)

DIFF1

ACCO	0	1	2	3	4	Mean
0	-59.83 (273)	-44.07 (338)	45.33 (390)	44.87 (384)	88.08 (409)	22.13 (1794)
1	-78.71 (2599)	-52.82 (3065)	20.94 (3808)	13.34 (3573)	46.93 (4121)	-2.66 (17166)
Mean	-76.92 (2872)	-51.95 (3403)	23.21 (4198)	16.40 (3957)	50.64 (4530)	-0.31 (18960)

DIFF1

ACC1	0	1	2	3	4	Mean
0	-46.54 (98)	-34.73 (114)	28.98 (578)	27.39 (461)	58.13 (544)	29.24 (1795)
1	-77.99 (2774)	-52.55 (3289)	22.29 (3620)	14.95 (3496)	49.62 (3986)	-3.40 (17165)
Mean	-76.92 (2872)	-51.95 (3403)	23.21 (4198)	16.40 (3957)	50.64 (4530)	-0.31 (18960)

* * * A N A L Y S I S O F V A R I A N C E * * *

Source of Variation	Sum of Squares	DF	Mean Square	F	Signif of F
Main Effects	42397017.39	6	7066169.57	1292.07	0.0
ACCO	1089981.10	1	1089981.10	199.31	.000
ACC1	184190.85	1	184190.85	33.68	.000
DIFF1	39683579.64	4	9920894.91	1814.07	0.0
2-way Interactions	279244.14	9	31027.13	5.67	.000
ACCO ACC	23236.84	1	23236.84	4.25	.039
ACCO DIFF1	174364.25	4	43591.06	7.97	.000
ACC1 DIFF1	61004.80	4	15251.20	2.79	.025
3-way Interactions	145974.20	4	36493.55	6.67	.000
ACCO ACC1 DIFF1	145974.20	4	36493.55	6.67	.000
Explained	42822235.73	19	2253801.88	412.16	0.0
Residual	103580264.83	18940	5468.86		
Total	146402500.56	18959	7722.06		

Cumulative-Random Group: Analysis of Variance on the effects of "accuracy of the current response", "difficulty level of the current response" and "practice session" on the residual RT of the current response

RES1 Current Residual RT
BY ACC1 Current Response
DIFF1 Current Difficulty
SESSION

SESSION

ACC1	1	2	3	4	5	6	Mean
0	44.47 (357)	44.85 (313)	37.76 (267)	39.72 (318)	14.60 (280)	24.01 (258)	35.09 (1793)
1	-4.78 (2843)	2.22 (2887)	-2.02 (2933)	-.16 (2882)	-7.39 (2920)	-9.43 (2942)	-3.61 (17407)
Mean	.71 (3200)	6.39 (3200)	1.30 (3200)	3.80 (3200)	-5.47 (3200)	-6.73 (3200)	0.00 (19200)

SESSION

DIFF1	1	2	3	4	5	6	Mean
0	-29.03 (1503)	-43.33 (761)	-47.73 (741)	-46.02 (795)	-51.35 (589)	-53.45 (671)	-42.42 (5060)
1	-6.05 (900)	-25.98 (906)	-29.36 (811)	-26.69 (737)	-35.46 (758)	-37.72 (677)	-26.08 (4789)
2	70.05 (517)	50.43 (780)	40.07 (765)	39.27 (681)	17.56 (652)	14.84 (619)	38.26 (4014)
3	52.67 (203)	45.73 (439)	35.45 (475)	37.88 (535)	16.64 (654)	17.20 (552)	30.88 (2858)
4	57.66 (77)	55.84 (314)	38.85 (408)	47.36 (452)	31.63 (547)	31.10 (681)	39.42 (2479)
Mean	.71 (3200)	6.39 (3200)	1.30 (3200)	3.80 (3200)	-5.47 (3200)	-6.73 (3200)	0.00 (19200)

		DIFF1				
	ACC1	0	1	2	3	4
SESSION 1						
	0	8.75 (115)	27.39 (70)	77.17 (124)	79.79 (39)	30.21 (9)
	1	-32.16 (1388)	-8.87 (830)	67.81 (393)	46.22 (164)	61.29 (68)
SESSION 2						
	0	.33 (30)	-13.71 (47)	59.02 (113)	66.12 (87)	62.58 (36)
	1	-45.12 (731)	-26.65 (859)	48.98 (667)	40.69 (352)	54.97 (278)
SESSION 3						
	0	-24.36 (29)	-15.29 (26)	53.35 (90)	46.11 (67)	59.89 (55)
	1	-48.68 (712)	-29.82 (785)	38.30 (675)	33.70 (408)	35.58 (353)
SESSION 4						
	0	-28.91 (45)	4.56 (42)	60.50 (103)	51.03 (76)	69.83 (52)
	1	-47.05 (750)	-28.58 (695)	35.49 (578)	35.70 (459)	44.44 (400)
SESSION 5						
	0	-36.52 (24)	-23.16 (29)	17.48 (74)	24.95 (93)	33.69 (60)
	1	-51.98 (565)	-35.95 (729)	17.57 (578)	15.26 (561)	31.38 (487)
SESSION 6						
	0	-44.38 (23)	-25.52 (21)	28.15 (65)	34.10 (70)	44.72 (79)
	1	-53.78 (648)	-38.11 (656)	13.28 (554)	14.75 (482)	29.31 (602)

* * * A N A L Y S I S O F V A R I A N C E * * *

Source of Variation	Sum of Squares	DF	Mean Square	F	p
Main Effects	27881760.79	10	2788176.08	1322.71	0.0
ACC1	556923.17	1	556923.17	264.20	.000
DIFF1	25061300.58	4	6265325.15	2972.26	0.0
SESSION	2348481.92	5	469696.38	222.82	0.0
2-way Interactions	485528.03	29	16742.35	7.94	0.0
ACC1 DIFF1	45218.76	4	11304.69	5.36	.000
ACC1 SESSION	49865.21	5	9973.04	4.73	.000
DIFF1 SESSION	348696.85	20	17434.84	8.27	0.0
3-way Interactions	99184.86	20	4959.24	2.35	.001
ACC1 DIFF1 SESSION	99184.86	20	4959.24	2.35	.001
Explained	28466473.67	59	482482.61	228.89	0.0
Residual	40345848.30	19140	2107.93		
Total	68812321.97	19199	3584.16		

Cumulative-Random Group: Analysis of Variance on the effects of "accuracy of the previous response", "accuracy of the current response" and "difficulty level of the current response" on the residual RT of the current response

	RES1	Current Residual RT
BY	ACCO	Previous Response
	ACC1	Current Response
	DIFF1	Current Difficulty

ACC1

ACCO	0	1
0	42.64 (203)	0.99 (1562)
1	33.64 (1565)	-4.39 (15630)

DIFF1

ACCO	0	1	2	3	4	Mean
0	-37.21 (457)	-20.14 (469)	46.44 (374)	39.21 (249)	44.08 (216)	5.78 (1765)
1	-43.19 (4553)	-27.18 (4256)	37.14 (3573)	29.83 (2578)	38.81 (2235)	-0.93 (17195)
Mean	-42.65 (5010)	-26.48 (4725)	38.02 (3947)	30.66 (2827)	39.28 (2451)	-0.30 (18960)

DIFF1

ACC1	0	1	2	3	4	Mean
0	-10.84 (264)	-.72 (234)	52.88 (557)	47.37 (425)	51.21 (288)	34.67 (1768)
1	-44.42 (4746)	-27.82 (4491)	35.58 (3390)	27.70 (2402)	37.69 (2163)	-3.90 (17192)
Mean	-42.65 (5010)	-26.48 (4725)	38.02 (3947)	30.66 (2827)	39.28 (2451)	-0.30 (18960)

* * * A N A L Y S I S O F V A R I A N C E * * *

Source of Variation	Sum of Squares	DF	Mean Square	F	Signif of F
Main Effects	25347461.44	6	4224576.91	1893.52	0.0
ACCO	74537.48	1	74537.47	33.41	.000
ACC1	693460.64	1	693460.64	310.82	.000
DIFF1	22908301.76	4	5727075.44	2566.96	0.0
2-way Interactions	72860.89	9	8095.65	3.63	.000
ACCO ACC1	669.89	1	669.89	.30	.584
ACCO DIFF1	2550.92	4	637.73	.29	.887
ACC1 DIFF1	69398.36	4	17349.59	7.78	.000
3-way Interactions	12682.96	4	3170.74	1.42	.224
ACCO ACC1 DIFF1	12682.96	4	3170.74	1.42	.224
Explained	25433005.29	19	1338579.23	599.97	0.0
Residual	42256597.98	18940	2231.08		
Total	67689603.27	18959	3570.32		

Cumulative-Adaptive Group: Analysis of Variance on the effects of "accuracy of the current response", "difficulty level of the current response" and "practice session" on the residual RT of the current response

RES1 Current Residual RT
BY ACC1 Current Response
DIFF1 Current Difficulty
SESSION

	SESSION						
ACC1	1	2	3	4	5	6	Mean
0	21.18 (308)	35.22 (263)	18.29 (211)	28.07 (284)	23.21 (240)	22.96 (198)	25.09 (1504)
1	-6.71 (2892)	4.82 (2937)	-.56 (2989)	-.74 (2916)	-2.77 (2960)	-6.82 (3002)	-2.13 (17696)
Mean	-4.02 (3200)	7.32 (3200)	.68 (3200)	1.82 (3200)	-.82 (3200)	-4.98 (3200)	-0.00 (19200)

	SESSION						
DIFF1	1	2	3	4	5	6	Mean
0	-24.56 (1555)	-49.54 (638)	-51.91 (619)	-48.45 (683)	-62.62 (497)	-63.20 (580)	-44.36 (4572)
1	-4.32 (1008)	-20.16 (957)	-31.74 (715)	-27.09 (764)	-37.63 (572)	-40.76 (713)	-24.87 (4729)
2	51.60 (345)	54.59 (1047)	37.34 (1080)	42.96 (1021)	22.67 (759)	11.76 (663)	37.47 (4951)
3	40.24 (200)	26.86 (241)	22.90 (512)	25.64 (507)	25.29 (893)	23.79 (564)	25.80 (2917)
4	41.54 (92)	33.69 (317)	18.09 (274)	12.21 (225)	21.37 (479)	42.01 (680)	29.52 (2067)
Mean	-4.02 (3200)	7.32 (3200)	.68 (3200)	1.82 (3200)	-.82 (3200)	-4.98 (3200)	-0.00 (19200)

		DIFF1				
	ACC1	0	1	2	3	4
SESSION 1						
	0	3.58 (129)	6.91 (82)	64.03 (65)	40.50 (29)	52.71 (3)
	1	-27.11 (1426)	-5.32 (926)	48.71 (280)	40.20 (171)	41.17 (89)
SESSION 2						
	0	-19.51 (23)	-1.74 (45)	54.86 (141)	42.00 (27)	34.15 (27)
	1	-50.66 (615)	-21.07 (912)	54.55 (906)	24.95 (214)	33.65 (290)
SESSION 3						
	0	-8.50 (15)	-40.23 (19)	30.27 (89)	26.56 (68)	12.51 (20)
	1	-52.98 (604)	-31.51 (696)	37.98 (991)	22.34 (444)	18.53 (254)
SESSION 4						
	0	-19.21 (30)	-12.64 (39)	48.55 (138)	34.25 (60)	16.76 (17)
	1	-49.79 (653)	-27.86 (725)	42.09 (883)	24.48 (447)	11.84 (208)
SESSION 5						
	0	-51.92 (15)	-40.25 (11)	20.52 (52)	35.86 (111)	34.19 (51)
	1	-62.96 (482)	-37.58 (561)	22.83 (707)	23.79 (782)	19.84 (428)
SESSION 6						
	0	-61.37 (14)	-20.41 (14)	16.64 (55)	34.12 (63)	50.49 (52)
	1	-63.25 (566)	-41.17 (699)	11.32 (608)	22.50 (501)	41.30 (628)

* * * A N A L Y S I S O F V A R I A N C E * * *

Source of Variation	Sum of Squares	DF	Mean Square	F	p
Main Effects	24915882.47	10	2491588.25	952.63	0.0
ACC1	155400.77	1	155400.77	59.42	.000
DIFF1	23579004.16	4	5894751.04	2253.80	0.0
SESSION	2101658.31	5	420331.66	160.71	0.0
2-way Interactions	1027553.14	29	35432.87	13.56	0.0
ACC1 DIFF1	72414.20	4	18103.55	6.92	.000
ACC1 SESSION	17066.78	5	3413.36	1.31	.259
DIFF1 SESSION	903594.21	20	45179.71	17.27	0.0
3-way Interactions	49807.91	20	2490.40	.95	.519
ACC1 DIFF1 SESSION	49807.91	20	2490.40	.95	.519
Explained	25993243.52	59	440563.45	168.45	0.0
Residual	50060160.02	19140	2615.47		
Total	76053403.53	19199	3961.32		

Cumulative-Adaptive Group: Analysis of Variance on the Effects of "accuracy of the previous response", "accuracy of the current response" and "difficulty level of the current response" on the residual RT of the current response

	RES1	Current Residual RT
BY	ACC0	Previous Response
	ACC1	Current Response
	DIFF1	Current Difficulty

ACC1

ACC0	0	1
0	51.54 (105)	14.03 (1384)
1	23.05 (1383)	-3.73 (16088)

DIFF1

ACC0	0	1	2	3	4	Mean
0	-26.18 (359)	-14.55 (366)	57.71 (369)	42.54 (239)	51.87 (156)	16.68 (1489)
1	-46.21 (4147)	-25.93 (4311)	35.62 (4488)	23.95 (2636)	27.66 (1889)	-1.61 (17471)
Mean	-44.61 (4506)	-25.04 (4677)	37.30 (4857)	25.49 (2875)	29.51 (2045)	-0.17 (18960)

DIFF1

ACC1	0	1	2	3	4	Mean
0	-10.32 (222)	-7.76 (209)	43.33 (535)	34.11 (355)	35.43 (167)	25.06 (1488)
1	-46.39 (4284)	-25.85 (4468)	36.55 (4322)	24.28 (2520)	28.98 (1878)	-2.32 (17472)
Mean	-44.61 (4506)	-25.04 (4677)	37.30 (4857)	25.49 (2875)	29.51 (2045)	-0.17 (18960)

* * * A N A L Y S I S O F V A R I A N C E * * *

Source of Variation	Sum of Squares	DF	Mean Square	F	Signif of F
Main Effects	23041155.97	6	3840192.66	1406.36	0.0
ACCO	481861.69	1	481861.69	176.47	.000
ACC1	259316.82	1	259316.82	94.97	.000
DIFF1	21541935.49	4	5385483.87	1972.28	0.0
2-way Interactions	173024.04	9	19224.89	7.04	.000
ACCO ACC1	12220.71	1	12220.71	4.48	.034
ACCO DIFF1	20535.64	4	5133.91	1.88	.111
ACC1 DIFF1	130585.38	4	32646.34	11.96	0.0
3-way Interactions	34994.31	4	8748.58	3.20	.012
ACCO ACC1 DIFF1	34994.31	4	8748.58	3.20	.012
Explained	23249174.33	19	1223640.75	448.12	0.0
Residual	51717291.36	18940	2730.59		
Total	74966465.69	18959	3954.14		