AUDITORY ALARMS IN THE INTENSIVE CARE UNIT: EXPERIMENTAL AND OBSERVATIONAL STUDIES

MEREDITH, CHRISTINA

http://hdl.handle.net/10026.1/2264

http://dx.doi.org/10.24382/4419
University of Plymouth

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.
AUDITORY ALARMS IN THE INTENSIVE CARE UNIT: EXPERIMENTAL AND OBSERVATIONAL STUDIES.

by

CHRISTINA MEREDITH

A thesis submitted to the University of Plymouth in partial fulfilment of the degree of

DOCTOR OF PHILOSOPHY

Department of Psychology,
Faculty of Human Sciences

In collaboration with the Royal Aircraft Establishment,
Farnborough.

June 1995
ABSTRACT

Auditory Alarms In The Intensive Care Unit: Experimental And Observational Studies.

By

Christina Meredith

There are many problems associated with the number of auditory warnings in hospital environments such as the Intensive Care Unit. As the amount of medical technology used to monitor a patient's condition increases, there is a concomitant increase in the number of auditory warnings. Each piece of equipment has its own alarm and often the sounds used are inappropriate. For example, the sounds are often too loud, too insistent and are irritating to staff, distracting them from other tasks. A further feature of sounds used for auditory alarms is that there is, at present, no agreement between manufacturers on the types of sounds used. This means that the same item of equipment can have different alarms if produced by different manufacturers. Subsequently there is the potential for confusion between alarms to occur if sounds are similar.

The research presented in this thesis aims to investigate the psychological dimension of confusion between alarm sounds and the correct identification of a set of twelve auditory warnings currently in use in the I.C.U. Derriford Hospital, Plymouth. Hence, the first set of experiments examines the learning and retention of the set of auditory warnings in a laboratory setting. However, the many problems regarding auditory warnings should not be considered in isolation and in order to determine the types of activities undertaken by staff in the I.C.U. environment when alarms are activated, two observational studies were undertaken. The first study used a video camera and the second study involved direct observation using two observers.

A series of tasks were developed that used the multiple resources literature as a framework and also represented tasks undertaken in the environment of the I.C.U. In the second experiment, participants were again required to learn and retain the set of auditory warnings. The tasks were introduced during the return stage of the experiment in order to examine first, whether there was an effect on the primary task of correctly identifying the sounds and whether the confusions between sounds increased or changed, and second to examine performance on the secondary tasks.

The results showed that for all experiments in general participants required few trials to learn the sounds and the information was retained for a period of over one week. The results also showed that features of some sounds were easier to learn than other sounds and that certain sounds were consistently confused during each experiment. When the tasks were introduced performance on the primary task remained fairly constant, with no overall change or increase in the number of confusions between sounds. However, there was a decrement in the performance of the secondary tasks, as predicted by the dual-task literature.

In conclusion, the results suggest that identification of sounds may depend on a global, overall label for a sound, such as a 'melodic' sound or a 'continuous' sound, with the more intricate details undetected by participants. The results also suggest that participants in the laboratory may alter their strategies to maintain performance on the primary task, by either responding more rapidly to task demands or by consciously deciding not to respond to one of the secondary tasks.
Chapter 1 What is the Problem With Hospital Alarms?

1.1 Introduction ........................................................................................................ 1

1.1.2 Background ....................................................................................................... 2

1.1.3 Specific Auditory Warning Problems
1.1.3.1 False Alarms ................................................................................................. 3
1.1.3.2 Theory of Signal Detection .......................................................................... 3
1.1.3.3 Noise .............................................................................................................. 8
1.1.3.4 Loudness of Auditory Warnings ................................................................... 10

1.1.4 Attempts to Remedy the Problems of Auditory Warnings ....................... 12
1.1.4.1 Design of new Auditory Warnings ............................................................. 12
1.1.4.2 How Many Auditory Alarms ........................................................................ 12
1.1.4.3 The Nature of Warning Sounds ................................................................. 14
1.1.4.4 Perceived Urgency of Auditory Warnings .................................................. 15

1.1.5 Recommendations for Medical Equipment Alarms .................................... 17

1.1.6 Clinical Trials of Proposed Auditory Warnings ........................................... 20

1.1.7 How Much Monitoring? .................................................................................... 21

1.1.8 Sustained Attention .......................................................................................... 23

1.1.9 Do Monitors Aid Vigilance ................................................................................ 25

1.2 Information Processing ....................................................................................... 27

1.2.1 Divided Attention ............................................................................................. 27

1.2.2 Time Sharing ..................................................................................................... 29

1.2.3 Postponement (Bottleneck) Theories ............................................................... 30

1.2.4 Single Resources Model ..................................................................................... 33

1.2.5 Multiple Resources Model ................................................................................ 34
1.2.5.1 Processing Stages .......................................................................................... 40
1.2.5.2 Perceptual Modalities .................................................................................. 40
1.2.5.3 Processing Codes ............................................................................................ 40

1.2.6 Task Difficulty and Practice ............................................................................. 41

1.2.7 Automated and Controlled Processing ......................................................... 42

1.2.8 Workload ........................................................................................................... 43

1.2.9 Comparison With Physical Workload ............................................................. 45

1.2.10 Mental Workload ............................................................................................. 46
1.2.10.1 Task Analytic Measures ............................................................................ 47
1.2.10.2 Primary Task ............................................................................................... 48
1.2.10.3 Secondary Task Measures .................................................. 50
1.2.10.4 Subjective Measures .......................................................... 54
1.2.10.5 Physiological Measures ...................................................... 57
1.2.11 Comparison and Dissociation Between Measures of Workload .... 58
1.2.12 Workload in the Hospital Environment .................................... 63
1.2.13 Rationale for Current Research ............................................... 64

Chapter 2: A Series of Experiments Examining Auditory Confusion Between Hospital Alarms.

2.1 Introduction ........................................................................... 67
2.2 Method ................................................................................... 72
2.2.1 Participants ......................................................................... 72
2.2.2 Apparatus/Stimuli ............................................................... 72
2.2.3 Procedure .......................................................................... 78
2.2.4 Results ............................................................................... 79
2.2.4.1 Introduction ..................................................................... 79
2.2.4.2 Number of Trials ............................................................. 79
2.2.4.3 Correct Responses ............................................................ 83
2.2.4.4 Confusions ....................................................................... 101
2.3 Discussion ............................................................................. 110

Chapter 3 An Exploratory Video Study to Examine Activities and Auditory Warnings on The Intensive Care Unit

3.1 Introduction ........................................................................... 116
3.1.1 Observation Studies ............................................................. 117
3.1.2 Sampling Strategy ............................................................... 119
3.1.3 Video Studies ..................................................................... 120
3.1.4 Reactivity ........................................................................... 121
3.2 Current Study ........................................................................ 124
3.2.1 Introduction ........................................................................ 124
3.2.2 Method ............................................................................... 124
3.2.2.1 Procedure ...................................................................... 124
3.2.3 Results ............................................................................... 127
3.2.3.1 Introduction ..................................................................... 127
3.2.3.2 Absolute Frequency of Alarms and Miscellaneous Sounds .... 129
3.2.3.3 Frequency of Alarms and Miscellaneous Sounds Occurring During 15 Minute Modified Frequency Scores ......................... 133
3.2.3.4 Relative Frequency (Probability) of Sounds Occurring............ 140
3.2.3.5 Absolute Duration of Alarms and Miscellaneous Sounds........... 143
3.2.3.6 Relative Duration (Percent Total Duration) of All Sounds........ 146
3.2.3.7 Relative Duration (Percent Total Duration) of Total
  Session Time.................................................................................. 148
3.2.3.8 Comparison Between Frequency and Duration Data............... 150

3.3 Discussion........................................................................................... 152

Chapter 4 The Development and Implementation of a Direct Observation Procedure for Use in the I.C.U.

4.1 Introduction............................................................................................. 159
4.2 Which Observational Technique to Use: Participant or Direct
  Observation?.......................................................................................... 160
  4.2.1 Summary....................................................................................... 161
4.3 Deciding to Use Passive Observation: What to Observe?................. 161
  4.3.1 Informal and Systematic Observation........................................... 162
4.4 Classification of Data............................................................................ 162
  4.4.1 Molecular and Molar Categories.................................................. 164
  4.4.2 Focal Sampling............................................................................. 165
  4.4.3 Summary....................................................................................... 166
4.5 Deciding on a Sampling Strategy.......................................................... 166
  4.5.1 Real Time Sampling..................................................................... 167
  4.5.2 Time Sampling.............................................................................. 168
  4.5.3 Event Sampling............................................................................. 168
  4.5.4 Interval Recording....................................................................... 169
  4.5.5 Summary....................................................................................... 170
4.6 Some Disadvantages of Using Direct Observational Techniques...... 171
  4.6.1 Reactivity...................................................................................... 171
  4.6.2 Observer Bias............................................................................... 172
  4.6.3 Observer Drift............................................................................... 172
  4.6.4 Inter-observer Reliability.............................................................. 173
4.7 Current Study.......................................................................................... 175
  4.7.1 Introduction................................................................................... 175
  4.7.2 Method........................................................................................ 176
    4.7.2.1 Development of the Coding Strategy.................................... 176
    4.7.2.2 The Informal Observation Phase....................................... 177
    4.7.2.3 The Development of the Systematic Observation Phase....... 184
    4.7.2.4 Implementation of the Systematic Phase............................ 187
Chapter 5  Experiment 4 : An Examination of the Effect of the Introduction of Tasks on the Identification of Auditory Warnings.

5.1  Introduction............................................................................................. 229
5.1.2 Background......................................................................................... 229
5.1.3 Explanation of the Decrement............................................................ 234
5.1.4 Operator Workload................................................................................ 237
5.1.5 Summary............................................................................................. 239
5.1.6 Multiple Resources Model.................................................................... 240
  5.1.6.1 Processing Stages............................................................................ 241
  5.1.6.2 Perceptual Modalities.................................................................... 242
  5.1.6.3 Processing Codes........................................................................... 243
5.2  Current Study.......................................................................................... 246
5.2.1 Introduction........................................................................................... 246
5.2.2 Experimental Overview........................................................................ 247
5.2.3 The Tasks.............................................................................................. 248
  5.2.3.1 Mental Arithmetic Task.............................................................. 248
  5.2.3.2 Observation Task (F/E Task)...................................................... 251
  5.2.3.3 Talking Task................................................................................. 255
  5.2.3.4 Auditory Task................................................................................. 257
  5.2.3.5 Presentation of the Tasks in Each Phase of Experiment 4............. 260
5.2.4 Phase 1 Learning.................................................................................. 263
  5.2.4.1 Introduction................................................................................... 263
  5.2.4.2 Method........................................................................................ 264
  5.2.4.3 Participants................................................................................. 264
  5.2.4.4 Stimuli........................................................................................ 264
  5.2.4.5 Procedure................................................................................... 264
5.2.5 The Return Phase................................................................................ 265
5.2.6 Phase 3 and Phase 4............................................................................ 268
5.3  Results..................................................................................................... 271
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.3.1 Introduction</td>
<td>271</td>
</tr>
<tr>
<td>5.3.2 Correct Responses</td>
<td>271</td>
</tr>
<tr>
<td>5.3.2.1 Analysis of the Interaction</td>
<td>274</td>
</tr>
<tr>
<td>5.3.2.2 Main Effects</td>
<td>283</td>
</tr>
<tr>
<td>5.3.3 Latency</td>
<td>285</td>
</tr>
<tr>
<td>5.3.3.1 Introduction</td>
<td>285</td>
</tr>
<tr>
<td>5.3.4 Correct Responses for Secondary Tasks</td>
<td>289</td>
</tr>
<tr>
<td>5.3.5 Comparison of Primary and Secondary Task Performance</td>
<td>292</td>
</tr>
<tr>
<td>5.3.6 Subjective Ratings</td>
<td>310</td>
</tr>
<tr>
<td>5.3.7 Significant Confusions Between Sounds for Each Phase</td>
<td>311</td>
</tr>
<tr>
<td>5.3.8 Comparison of Overall Performance, All Phases</td>
<td>317</td>
</tr>
<tr>
<td>5.3.9 Comparison Between Experiment 1 and Experiment 4</td>
<td>320</td>
</tr>
<tr>
<td>(Learning Phase)</td>
<td></td>
</tr>
<tr>
<td>5.4 Discussion</td>
<td>323</td>
</tr>
<tr>
<td>5.4.1 Primary and Secondary Task Performance</td>
<td>323</td>
</tr>
<tr>
<td>5.4.2 Correct Identification of Sounds</td>
<td>330</td>
</tr>
<tr>
<td>5.4.3 Significant Confusions</td>
<td>331</td>
</tr>
<tr>
<td>5.4.4 Subjective Measures</td>
<td>334</td>
</tr>
<tr>
<td>5.4.5 Future Research</td>
<td>336</td>
</tr>
<tr>
<td>Chapter 6 Discussion</td>
<td></td>
</tr>
<tr>
<td>6.1 Overview</td>
<td>339</td>
</tr>
<tr>
<td>6.2 General Discussion of Results</td>
<td>340</td>
</tr>
<tr>
<td>6.2.1 Latency</td>
<td>344</td>
</tr>
<tr>
<td>6.2.2 Performance on Secondary Tasks</td>
<td>345</td>
</tr>
<tr>
<td>6.3 Strategies and Workload</td>
<td>348</td>
</tr>
<tr>
<td>6.4 Factors Relevant to the I.C.U. and Hospital Environment</td>
<td>351</td>
</tr>
<tr>
<td>6.5 Subjective Ratings</td>
<td>353</td>
</tr>
<tr>
<td>6.6 Future Research</td>
<td>354</td>
</tr>
<tr>
<td>6.7 Conclusion</td>
<td>357</td>
</tr>
<tr>
<td>References</td>
<td>359</td>
</tr>
</tbody>
</table>
List of Tables

Chapter 2

2.1 Description of Equipment.............................................................................. 75
2.2 Stimuli........................................................................................................... 76
2.3 Description of Stimuli................................................................................... 77
2.4 Number of Trials for Experiment (Learning and Return Phases)............... 82
2.5 Mean Number of Trials for Phase (All Experiments).................................. 82
2.6 Mean Number of Trials for Experiment and Phase....................................... 82
2.7 Summary Two-Way Anova Table for Number of Trials.............................. 82
2.8 Mean Number of Correct Responses for the Learning Phase
   Experiments 1, 2, and 3................................................................................ 84
2.9 Mean Number of Correct Responses for the Return Phase
   Experiments 1, 2, and 3................................................................................ 85
2.10 Summary Table of Three-Way Anova Correct Responses
   Experiments 1/2........................................................................................... 86
2.11 Adjustment of Degrees of Freedom.......................................................... 86
2.12 Mean Number of Correct Responses for Sounds in Each Phase
   (Experiments 1/2)....................................................................................... 89
2.13 Mean Number of Correct Responses for Main Effect of Sounds
   (Experiments 1/2)....................................................................................... 91
2.14 Mean Number of Correct Responses for Phase (Experiments 1/2)............... 91
2.15 Mean Number of Correct Responses for Sounds (Experiments 1/2)........... 91
2.16 Summary Table of Three-Way Anova Correct Responses
   Experiments 2/3........................................................................................... 92
2.17 Adjustment of Degrees of Freedom.......................................................... 92
2.18 Mean Number of Correct Responses for Sounds in Each Experiment
   (Experiment 2/3)....................................................................................... 95
2.19 Mean Number of Correct Responses to Sounds in Each Phase
   Experiments 2/3........................................................................................... 98
2.20 Mean Number of Correct Responses for Phase Experiments 2/3.............. 99
2.21 Mean Number of Correct Responses to Sounds............................................ 100
2.22 Significant Confusions Between Warnings................................................ 102
2.23 Significant Non-Confusions Between Warnings.......................................... 103
2.24 Experiment 1 Learning Phase Significant Confusions and Confusions
   Approaching Significance.......................................................................... 104
2.25 Experiment 2 Learning Phase Significant Confusions and Confusions
   Approaching Significance.......................................................................... 105
Chapter 3

3.1 Total Frequency of Sounds Recorded During the Observation Study........ 129
3.2 Absolute Frequency (Number) of Individual Alarms......................... 132
3.3 Absolute Frequency (Number) of Miscellaneous Sounds...................... 132
3.4 Relative Frequency of Individual Alarms Occurring........................ 142
3.5 Relative Frequency of Miscellaneous Sounds Occurring..................... 142
3.6 Total Duration of Alarms Recorded During the Observation Study........ 144
3.7 Total Duration of Miscellaneous Sounds Recorded During the Observation Study...................................................... 144
3.8 Absolute Duration (seconds) for Alarms During Each Observation Session............................................................................. 145
3.9 Absolute Duration (seconds) for Miscellaneous Sounds During Each Observation Session.............................................................. 145
3.10 Absolute Duration (seconds) for Alarms and Miscellaneous Sounds...... 146
3.11 Relative Duration (seconds) for Alarms............................................ 147
3.12 Relative Duration (seconds) for Miscellaneous Sounds..................... 147
3.13 Relative Duration (%) for Alarms...................................................... 149
3.14 Relative Duration (%) for Miscellaneous Sounds............................. 149
3.15 Comparison Between Frequency and Duration of Alarms ...................... 151
3.16 Comparison Between Frequency and Duration of Miscellaneous Sounds..... 151

Chapter 4

4.1 An Example of the Contingency Table Used in the Calculation for the Degree of Agreement Between the Observers................................. 190
4.2 Inter-Observer Agreement for Activities............................................ 192
4.3 Inter-Observer Agreement for Alarms................................................ 194
4.4 Proportion of Each Activity as a Proportion of the Total Activities......... 197
4.5 Proportion of Each Alarm as a Proportion of the Total Alarms.............. 198
4.6 The Total Number of Activities Occurring When Each Alarm was Activated .......................................................... 199
4.7 One Activity Occurring with One Alarm During Each Two-Minute Interval .................................................. 215
4.8 Two Activities Occurring with One Alarm During Each Two-Minute Interval .................................................. 216
4.9 Three Activities Occurring with One Alarm During Each Two-Minute Interval .................................................. 217
4.10 Four Activities Occurring with One Alarm During Each Two-Minute Interval .............................................. 218
4.11 Activities Occurring With Two Alarms Within the Same Two-Minute Interval ............................................... 219

Chapter 5
5.1 Combination of Resources Involved in the Tasks for Each Phase of Experiment 4 .................................................. 262
5.2 Summary Table for Two-Way Anova (Phase by Correct Response to Sounds) .................................................. 273
5.3 Adjustment Of Degrees of Freedom ............................................................................................................... 273
5.4 Mean Number of Correct Responses for Sounds for Each Phase .............................................................. 277
5.5 Simple Main Effects of Phase ...................................................................................................................... 278
5.6 Significant Differences for Correct Responses to Sounds Across Each Phase .......................................................... 280
5.7 Simple Main Effects of Sounds For Each Phase ............................................................................................... 281
5.8 Significant Differences for Correct Responses to Sounds Within Each Phase of Experiment 4 ................................................. 282
5.9 Mean Number of Correct Responses for Phase ............................................................................................ 283
5.10 Overall Mean number of Correct Responses to Sounds (all phases) ............................................................ 284
5.11 Summary Table for Two Way Anova (Phase by Sound) ....................................................................................... 286
5.12 Adjustment of H-F Degrees of Freedom ........................................................................................................... 286
5.13 Mean Number of Response Latency for Each Phase of the Experiment .......................................................... 287
5.14 Ranks of Correct Responses to Sounds .......................................................................................................... 305
5.15 Ranks for the number of Correct Responses for the Mental Arithmetic Calculations .................................................. 306
5.16 Summary Table for Comparisons of Correct Responses for Mental Arithmetic Task .............................................. 307
5.17 Ranks for the Correct Responses for the Observation Task ............................................................................... 308
5.18 Summary Table for Comparisons of Correct Responses for Observation Task ........................................................... 309
5.19 Participants Rating for Perceived Difficulty of Each Phase......................... 310
5.21 Significant Confusions Between Warnings.................................................. 312
5.22 Experiment 4 Phase 2 (Mental Arithmetic) Significant Confusions and
    Confusions Approaching Significance......................................................... 313
5.23 Experiment 4 Phase 2 (Observation) Significant Confusions and
    Confusions Approaching Significance......................................................... 314

5.24 Experiment 4 Phase 3 Significant Confusions and Confusions Approaching
    Significance.................................................................................................. 315
5.25 Experiment 4 Phase 4 Significant Confusions and Confusions Approaching
    Significance.................................................................................................. 316
5.26 Summary Table of Comparison Between Correct Responses to Tasks,
    Latency, Subjective Ratings and Confusions Between Sounds....................... 319
5.27 Number of Correct Responses for the Learning Phases Experiment 1 and
    Experiment 4............................................................................................... 320
5.28 Comparison Between Significant Confusions  Experiment 1 and
    Experiment 4............................................................................................... 321
5.29 Experiment 4 Learning Phase Significant Confusions and Confusions
    Approaching Significance............................................................................. 322
List of Figures

Chapter 1
1.1 The Proposed Structure of Processing Resources ......................................... 39

Chapter 2
2.1 Interaction of Phase and Correct Responses to Sounds (Experiment 1/2)......... 88
2.2 Interaction Between Experiment and Sound (Experiment 2/3) ....................... 94
2.3 Interaction Between Correct Responses and Phase (Experiment 2/3) .......... 97

Chapter 3
3.1 Diagram of the Four Bed Unit ........................................................................ 125
3.2 Absolute Frequency for All Sounds Recorded During the Observation Study ................................................................. 130
3.3 Comparisons Of the Absolute Frequency of Alarms and Miscellaneous Sounds Occurring During the Observation Study ........................................... 131
3.4 Frequency of Sounds During the Time Period 12 midday to 20.00hours (15 minute intervals) .............................................................................................. 134
3.5 Frequency of Sounds During the Time Period 16.00hours to midnight (15 minute intervals) ................................................................. 135
3.6 Frequency of Sounds During the Time Period midnight to 08.00hours (15 minute intervals) .............................................................................................. 136
3.7 Frequency of Sounds During the Time Period 08.00hours to midnight (15 minute intervals) .............................................................................................. 137
3.8 Frequency and Duration of Sounds During a 1 Hour Period (15.00 to 16.00hours) .............................................................................................. 139

Chapter 4
4.1 Physical Workload Checklist ........................................................................ 185
4.2 Auditory Sounds Checklist ............................................................................ 186
4.3 Total Number of Activities Recorded by the Two Observers ......................... 193
4.4 Total Number of Alarms Recorded by the Two Observers ............................. 193
4.5 Alarms and Activities Day 1 ........................................................................ 202
4.6 Alarms and Activities Day 2(i) .................................................................... 203
4.7 Alarms and Activities Day 2(ii) .................................................................... 204
4.8 Alarms and Activities Day 3 ........................................................................ 205
4.9 Alarms and Activities Day 4 ........................................................................ 206
4.10 Alarms and Activities Day 5 ....................................................................... 207
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.11</td>
<td>Alarms and Activities Day 6</td>
</tr>
<tr>
<td>4.12</td>
<td>Alarms and Activities Day 7(i)</td>
</tr>
<tr>
<td>4.13</td>
<td>Alarms and Activities Day 7(ii)</td>
</tr>
<tr>
<td>4.14</td>
<td>Alarms and Activities Day 8(i)</td>
</tr>
<tr>
<td>4.15</td>
<td>Alarms and Activities Day 8(ii)</td>
</tr>
<tr>
<td>5.1</td>
<td>Mental Arithmetic Task</td>
</tr>
<tr>
<td>5.2</td>
<td>Resources Used in the Mental Arithmetic Task</td>
</tr>
<tr>
<td>5.3</td>
<td>Observation Task</td>
</tr>
<tr>
<td>5.4</td>
<td>Resources Used in the Observation Task</td>
</tr>
<tr>
<td>5.5</td>
<td>Talking Task</td>
</tr>
<tr>
<td>5.6</td>
<td>Resources Used in the Talking Task</td>
</tr>
<tr>
<td>5.7</td>
<td>Representation of the Names Presented on the Archimedes Screen</td>
</tr>
<tr>
<td>5.8</td>
<td>Auditory Task</td>
</tr>
<tr>
<td>5.9</td>
<td>Resources Used in the Auditory Task</td>
</tr>
<tr>
<td>5.10</td>
<td>Layout of the Equipment Used in the Return Phase of Experiment 4</td>
</tr>
<tr>
<td>5.11</td>
<td>Interaction of Phase And Correct Responses to Sounds</td>
</tr>
<tr>
<td>5.12</td>
<td>Latency of Response to the Sounds in Each Phase</td>
</tr>
<tr>
<td>5.13</td>
<td>Observation Scores</td>
</tr>
<tr>
<td>5.14</td>
<td>Correct Scores for Mental Arithmetic Task for Each Phase</td>
</tr>
<tr>
<td>5.15</td>
<td>Correct Responses Subject 1 (All Phases All Tasks)</td>
</tr>
<tr>
<td>5.16</td>
<td>Correct Responses Subject 2 (All Phases All Tasks)</td>
</tr>
<tr>
<td>5.17</td>
<td>Correct Responses Subject 3 (All Phases All Tasks)</td>
</tr>
<tr>
<td>5.18</td>
<td>Correct Responses Subject 4 (All Phases All Tasks)</td>
</tr>
<tr>
<td>5.19</td>
<td>Correct Responses Subject 5 (All Phases All Tasks)</td>
</tr>
<tr>
<td>5.20</td>
<td>Correct Responses Subject 6 (All Phases All Tasks)</td>
</tr>
<tr>
<td>5.21</td>
<td>Correct Responses Subject 7 (All Phases All Tasks)</td>
</tr>
<tr>
<td>5.22</td>
<td>Correct Responses Subject 8 (All Phases All Tasks)</td>
</tr>
<tr>
<td>5.23</td>
<td>Correct Responses Subject 9 (All Phases All Tasks)</td>
</tr>
<tr>
<td>5.24</td>
<td>Correct Responses Subject 10 (All Phases All Tasks)</td>
</tr>
<tr>
<td>5.25</td>
<td>Correct Responses Subject 11 (All Phases All Tasks)</td>
</tr>
</tbody>
</table>
Acknowledgements

I would like to thank colleagues in the Department of Psychology, University of Plymouth, and in the Faculty of Health, Buckinghamshire College, for their support and advice during the time taken to complete this thesis.

In particular I would like to thank my Director of Studies, Dr. Judy Edworthy for her guidance throughout the period of study undertaken for this thesis, Dr David Rose for his help in undertaking the observation study, Professor Jonathan Evans for writing the computer program used in the final set of experiments, and Dr Ian Dennis and Dr Graham Crocker for their advice regarding statistical procedures used in this thesis. I would also like to thank Dr Pam Sutton for her support and help, particularly during the final stages of submitting this thesis.

I would like to thank all the staff on the I.C.U. Derriford Hospital for their enthusiastic support and interest during the observation studies. Finally, I would like to thank the Science and Engineering Research Council for supporting me with a studentship throughout the three years of research undertaken for this thesis.
Author's Declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other university award.

This study was financed with the aid of a case studentship from the Science and Engineering Research Council (Award no 90590754) and carried out in collaboration with the Royal Aircraft Establishment, Farnborough.

A programme of advanced study was undertaken during the period of study. This involved a programme of directed reading; a SPSS-PC4 course held at the Computer Centre, University of Plymouth. Other courses attended were programming the Archimedes, and lectures on covariance structure, exploratory factor analysis and principle component analysis held at the Department of Psychology, University of Plymouth. Study was also undertaken to learn new statistical packages in order to carry out statistical analysis.

Relevant scientific seminars and conferences were regularly attended and papers presented.

Conference papers presented are as follows;

Auditory Warnings and Confusion in the Intensive Care Unit presented to "The Human Factors of Alarm Design", Aston University, October 1992

Auditory Confusion in the Intensive Care Unit presented to "Information Management Systems in Intensive Care" British Medical Informatics Society, Watford November 1992

Auditory Warnings and Workload presented at the Annual Conference of the Ergonomics Society Edinburgh, April 1993
Publications include the following;


Signature: Meredith

Date: 30.6.95
CHAPTER 1 What is the Problem With Hospital Alarms?

1.1 Introduction

There are many problems with existing auditory warnings in hospitals. For example, there are too many auditory warnings and there is an urgent need for rationalisation. The sounds are often irritating and annoying, distracting the staff from more urgent problems, particularly when many of the warnings sounded are false alarms, and may result in the alarms being turned off. Another problem is that there is often no relationship between the sound of the alarm and the urgency of the situation it is signalling. There is also very little agreement between manufacturers on the sounds used for medical equipment, which can create confusion if two sounds are very similar.

Auditory warnings in the hospital environment and in particular areas such as the Intensive Care Unit (I.C.U.) and the Operating Room (O.R.) are necessary for two important reasons. First because the equipment used to monitor patients is constantly increasing in terms of complexity and sophistication. It could become increasingly difficult for staff to efficiently observe all the information displayed by the machine and to perform other tasks, without a device to alert them to an adverse situation. The second problem is the fairly recent increase in litigation, which can, should a serious incident occur, cost thousands of pounds paid in damages. Alarms are therefore installed on virtually every piece of equipment to alert the operator when a variable changes from a pre-set limit.

Kerr (1985) stated that because of the number of problems associated with auditory alarms in the hospital environment they should not be considered in isolation and a rational framework should be developed in order to provide the most efficient protection for the patient. However, often the limited amount of literature regarding hospital auditory warnings seems to be expressed as subjective opinions and anecdotal evidence, with few objective, empirical studies. There is considerable literature in which opinions are
expressed, many of which highlight the problems of existing auditory warnings.

This chapter examines the problems of auditory warnings and discusses the literature which proposes specific solutions to some of the problems as the result of empirical studies. The second section of the chapter attempts to relate these problems to the underlying psychological issues at a more theoretical level. Vigilance, attention, and workload in multi-task environments are discussed and related to the increasing cognitive demands made upon staff working in the environment of the I.C.U. The I.C.U. environment was the specific focus in this thesis because of the constantly increasing amount of equipment and subsequent numerous alarms. At present research concerning auditory warnings does not take into account factors such as changes in workload, particularly in a multi-task environment. The final section of the introduction looks at the rationale for the studies and experiments undertaken in this project.

1.1.2 Background

One of the main reasons why problems with hospital alarms have occurred during the past decade is because of the vast increase in the amount of technology in the I.C.U. and the O.R. designed to aid the monitoring of vital signs of patients. As Samuel (1986) comments, monitoring in the past consisted of a finger on a pulse or using a stethoscope, while today there are many different controls, displays and perhaps as many as twenty or thirty visual and auditory alarms for each patient. As more sophisticated equipment has been developed the basic ventilator or anaesthetic machine used on the O.R. has had additional pieces of new technology added to it, each piece with its own alarm. There is little, if any, standardisation amongst manufacturers regarding the sound of the auditory warnings, with the concomitant problem that the nurse or anaesthetist may be unable to detect quickly enough which piece of equipment is alarming. Samuels, in fact, complains that it is often impossible to tell which monitor is alarming and this can cause considerable anxiety to the operator.
The term 'alarm' can have many definitions (Stanton and Booth 1990), for example a means of signalling state changes, a means of attracting attention or arousing someone and the response of an operator to state changes. Patterson (1985) states that the purpose of an auditory warning is to cut through noise and command peoples' attention to signal danger or potential danger. In the hospital situation, Hyman and Drinker (1983) add that an alarm alerts the operator to adverse changes in the device or patient variables which are outside those of a pre-selected range. With the increasing numbers of monitors and alarms that have been introduced into both the I.C.U. and the O.R., supposedly to aid staff in the observation of the patient, there has been an accompanying increase of concern from the very staff the monitors have been designed to help. Firstly concerning the problem of the monitors themselves and secondly, the associated alarms. Instead of producing an aid to monitoring patient care, a Pandora's Box of new problems has been opened.

1.1.3 Specific Auditory Warning Problems Identified

1.1.3.1 False Alarms

Auditory warnings are often silenced because of the frequent number of false alarms, sometimes the operator forgetting to turn the alarm back on again, or not even being aware that the alarm is off. Reports of death, injury and near mishaps occur when critical alarms, (i.e. those alarms that indicate that an essential variable is beyond a safe or pre-selected limit), are improperly used or not used at all, as reported by The Emergency Care Research Institute (ECRI 1987). False alarms caused by artefact, such as patient movement, respiratory tract suction or the loosening of electrodes, are design problems and subsequently the realm of the engineers.

1.1.3.2 Theory of Signal Detection

The Theory of Signal Detection (Green and Swets 1966) can be related to the high incidence of false alarms sounding in the I.C.U. In fact it can be used to look both at the relationship between the situation which is being signalled and the presence or absence of an
alarm, and the relationship between an alarm sounding and the recipient's detection of that alarm. The first of these is considered in this instance, as the second is an issue which does not directly address the problem of false alarms, although the issue of the detection is itself an important factor in auditory warnings research, and any problems here would be compounded by system reliability problems.

Signal Detection Theory is applicable in any situation in which there are two conditions, one condition being the presence of a signal indicating that there is a problem, the second condition being when a signal occurs but there is not a corresponding problem. A two by two matrix can be produced, generating four classes of events which are hits, misses, false alarms and correct rejections. A hit occurs when a situation warranting an alarm is present, and is actually signalled by an alarm. A miss occurs when a situation warrants signalling, but an alarm does not sound. A false alarm occurs when an alarm is presented but the situation does not warrant the attention of an operator (e.g. there is not a genuine emergency). The final category, correct rejection, occurs when no situation requires signalling and no alarm sounds.

It has been shown recently (Bliss et al. in press) that people are very good at matching their response rates to the reliability of alarm systems. For example, if an alarm system is only 25% reliable (so that 75% of the alarms which are heard are not actually signalling events which warrant alarms) then participants will modify their response rates to alarms so that they respond roughly a quarter of the time that the alarm sounds; if the system is almost wholly accurate, then participants will respond nearly all the time to that particular alarm. Of course, this problem is further complicated by the fact that specific responses of the participant do not necessarily coincide with the correct soundings of the alarm, so systems which are unreliable will tend to become more so, as responses to such systems tend to degenerate in this way.

If there are repeated false alarms or misses, the I.C.U. staff may be inclined to switch off the alarm (e.g. McIntyre 1985) thereby risking
few false alarms or misses. However, there is the serious risk that should a genuine emergency or malfunction occur this might not produce an appropriate response. Conversely, the staff could leave the alarm systems on thereby detecting most of the genuine signals but risk many false alarms. In the I.C.U. environment it would appear from the literature discussed below that the latter situation occurs. Whilst genuine problems are detected a large number of false alarms also cause problems. As Westman and Walters (1981) and Duberman and Bendixen (1984) argue repeated false alarms can produce stressful responses such as annoyance and irritation together with habituation to the sound of the alarms.

False alarms can be seen as presenting a major problem because they can be seen as a threat to patient care, if the staff do become increasingly irritated by them, or habituate to them. The staff may respond by turning the alarm system off altogether. Deliberately de-activating alarms was demonstrated in a study conducted by McIntyre (1985). A questionnaire was used to determine the attitudes of Canadian anaesthetists to auditory alarms in the operating room. The questionnaire did not directly ask the anaesthetists how many false alarms sounded, but instead asked whether they routinely activated audible alarms; whether they deliberately disabled auditory alarms at the beginning of a case; and their reasons for such actions. The answers to these questions were considered by McIntyre (1985) to be the most important information obtained from the study. Because of the unacceptably high number of false alarms 460 out of a total of 789 anaesthetists (58%) stated that they deliberately deactivated an alarm at the beginning of the case.

Disabling auditory warnings is a serious cause for concern because a facility that should be providing important information to an operator is not being utilised and may in fact be a dangerous practice. As Raison et al. (1968) argue, a false sense of security may be created if other members of staff believe that the alarm is operational, and this is probably worse than having no alarms at all. Nevertheless, as the following studies indicate there is a high percentage of false alarms occurring in high dependency units, and
subsequently it is important that alarm sounds are not irritating and
distracting to the staff working in these areas.

There are several studies that have looked at the incidence of false
alarms. The frequency of alarms in the Intensive Care Unit were
investigated by O'Carroll (1986) and by Koski et al. (1990). In both
studies the staff on the unit recorded the frequency of the alarms
over a period of time. The staff recorded the frequency of alarms in
the I.C.U. over a three week period in the O'Carroll (1986) study. It
was found that out of a total of 1455 auditory warnings only eight
(0.549%) signified a potentially life threatening problem while the
remainder were recorded as false alarms.

In a similar study by Koski et al. (1990), conducted in Finland, nurses
who were considered to be experienced in postoperative care
recorded and categorised the alarms that sounded during a 26 hour
period on ten cardiac patients. Out of 1307 alarms that sounded, 139
(10.6%) were categorised by the nurses as being significant, which
necessitated the nurse checking the patient's condition and
implementing treatment.

It is likely that the recording of alarms undertaken in the above
studies were retrospective, particularly if the staff were extremely
busy for example, during a critical incident. This might mean that
information regarding the actual occurrence of alarms could be
forgotten or even incorrect. Another consideration is that the staff
involved would need to be highly motivated to record each alarm, as
undoubtedly they would have a demanding workload to contend with
already.

One way of overcoming these problems is to use independent
observers as in two studies undertaken by Raison et al. (1968). The
studies investigated the cause of the numerous false alarms that
occurred during the development of a cardiovascular monitoring
system. The studies specifically examined the number of alarms
from the E.C.G. monitor and initially found that many of the alarms
resulted from accidental disconnection of the electrodes and from
staff or patients inadvertently touching the electrodes. The
observers recorded each visual signal (the audible signals were suppressed although no reason for this was given) over 44 hours during an eight day period. During this period a total of 333 alarms were recorded, none of which required medical intervention. The causes of the alarms were patient movement and staff contact with the patient or from lead disconnection. By recalibrating the conditions under which an alarm would occur the second study demonstrated far fewer false alarms. Over a 92 hour period only eight false alarms occurred.

Kestin et al. (1988) also used an independent observer to record the frequency of alarms during surgery. The observer recorded each alarm that sounded during 50 surgical cases. The findings showed that on average 10 alarms per patient sounded with a mean frequency of one alarm every 4.5 minutes. The results showed that 75% of the alarms that sounded were spurious, which the authors categorised as those alarms caused by patient movement, interference or mechanical problems, with only 3% of the alarms that sounded indicating a risk to the patient. The remaining 22% of alarms occurred when the equipment registered a change above the upper pre-determined limits (which did not constitute a risk to the patient).

Another study in which an independent observer investigated the occurrence of alarms was conducted by Stanton (1992) in a Coronary Care Unit (C.C.U.). Stanton concluded that there were a number of alarms (37.6%) that were activated unnecessarily, which were caused by artefact, such as the patient eating or getting out of bed.

The results of these studies would appear to indicate that there are an unacceptably high number of false alarms occurring in the O.R., the recovery room, and the I.C.U. It should be emphasised that irrespective of how many false alarms there are, it is unacceptable to disable the alarms by turning them off, as there will be relatively infrequent occasions when an alarm is signalling a real emergency. Cooper et al. (1986) concluded that when ventilator alarms were disabled, either deliberately or inadvertently, it was overall the most important contributory factor in injury. In an analysis of
anaesthetic mishaps by Davis (1989), it was found that the majority of critical incidents were due to mistakes made by the operator, such as the failure of the operator to cope with acute yet common situations, or lapses in vigilance, and stated that mechanical or machine failures were not indicated in the study.

There has been considerable investigation into the development of 'intelligent' alarm systems (e.g. Fukui and Masuzawa 1989) that are able to distinguish between potentially dangerous situations and non-threatening causes such as artefact. However these are engineering concerns and beyond the scope of this review. However, such research should decrease the number of false alarms and misses occurring. Even though there are many false alarms the equipment, as previously discussed should not have the alarm system switched off, for ten years ECRI have recommended that "nondefeatable" alarms should be used for critical variables, i.e. alarms that cannot be turned off either visually or audibly for an indefinite period. If the alarm cannot be silenced for more than a few seconds or minutes, it would seem logical to incorporate a sound that is not going to irritate and unduly annoy the staff.

1.1.3.3 Noise

'Noise' is usually considered to be an unwanted, irritating and interfering stimulus (e.g. Kryter 1972). While in some situations 'noise' becomes background sounds that individuals habituate to, in other situations it becomes a contaminant that interferes with communication, concentration and can increase the error rate in an ongoing task. Baker (1984) comments that sounds that are high pitched, intermittent, and of a long duration are more annoying. Annoyance is a response to a noise rather than a dimension of auditory experience, and it is related to the physical characteristics of sound, the activities that are disturbed by the noise, physiological reactions to noise and the meaning of the noise to the individual (Miller 1974).

Smith (1991) examined the effects of moderate intensity noise on cognitive vigilance tasks, a focused attention task and a search task.
The reactions of subjects were recorded in continuous freefield noise in two conditions, quiet (60 dB) and loud (85 dB). The findings showed that the response time was faster when the presence of a letter was known in advance, as in the focused attention task and the difference between the two tasks was greater in the noise condition than in the quiet condition. The study concluded that performance of tasks involving the monitoring of cognitive information may be impaired by moderate density (70-80dB) noise.

It was also demonstrated by Smith (1990) that noise also has an effect on specific aspects of a task in a dual-task situation, particularly in memory tasks although noise had no effect on the other task which involved estimating the relative proportion of two classes of events. Smith (1990) concluded that the effects of noise in cognitive dual-task situations depends on the specific combination of tasks.

Noise is a constant source of sensory overload in the I.C.U. environment (Baker 1984). Because of the patient's condition and the need for constant monitoring the average noise level may be relatively high over a 24 hour time period. Noise levels in the I.C.U. have been recorded at levels ranging from 45 to 85 dB and sources of the noises vary from the monitors alarming to the talking and laughing of the staff. The threshold of tolerable sound is lower for ill persons than for those who are in good health (Turner et al. 1975). Hansell (1984) showed that patients perceived non-verbal signals as offensive, frustrating and confusing because they could not attach a meaning to the sounds. It has been found that the staff contribute to the overall noise, for example Aiken (1982) found that patient cries, staff movements, telephones and the communication system were responsible for the majority of noises over 70 dB. The staff were unaware of how much noise they created, the increase in the noise being directly related to the number of staff on duty, particularly if they were in a confined space.

Noble (1982) speculated that equipment would produce the most noise, yet her results showed that the noise from the monitors and equipment became part of the background noise as people became
accustomed to the rhythm and routine noises. Noble suggests that verbal communication, which is neither rhythmic nor routine but unexpected and varied from day to day, is more disturbing or arousing to individuals, particularly if the words are muffled and cannot be heard distinctively by the individual.

Along these lines it could also be argued that while the noise of the equipment is rhythmic and merges into the background, auditory warnings at present do not. Hilton (1976) investigated factors that disturbed patients' sleep by looking at the changes shown in the E.E.G. near waking. The results showed that the most disturbing factors were non-patient centred noises (54%) with the highest proportion created by the staff (22%). Environmental noises were the second largest factor with equipment, machinery and alarm noises rating third with 3%. As Baker (1984) comments, it would be interesting to compare these results with subjective ratings of which noise sources patients perceive to be the most psychologically distressing.

1.1.3.4 Loudness of Auditory Warnings

Auditory warnings are supposedly designed to cut through speech and background noise and command people's attention. However the sounds used in the auditory warnings are usually loud and jarring and prevent communication just at the time when it is essential (Patterson 1982). Anaesthetists divide their time at work looking at the patient, the surgical field and also performing other unspecified tasks some distance away from the patient, making it impossible for them to observe the patient or the monitors continuously (McIntyre 1985). This suggests that information from monitors must be presented audibly and possibly visually as well to attract the attention of the anaesthetist, when the need arises. However, as Patterson (1985) states the alarms are all too often installed with a 'better safe than sorry' attitude by the manufacturers resulting in warnings which are 'too loud, too strident and too insistent.' Schmidt and Baysinger (1986) support this view, stating that most audible alarms are loud, produce continuous noxious signals that cannot be adjusted or silenced and that anaesthetists waste time silencing these alarms. They suggest that during an emergency it
may be more effective to have a problem indicated by a "pleasant sound". Hedley-Whyte (1988) adds his support to the argument by stating that the alarms are many times too loud, disturbing the surgeon and aggravating the rest of the staff.

The ECRI (1987) report suggests using a priority based system of alarms that are easily distinguishable, but as with other suggestions made in the report does not discuss which types of sounds to use in order to produce an effective hierarchy of priority alarms which are easily distinguishable from each other. New problems are then created in that manufacturers are left to decide which type of sound to use without any supporting research into what makes a sound appear urgent.

Kerr (1985) states that many medical auditory warnings are continuous or rapidly intermittent tones of constant, and usually high pitch. The loudness is often uncontrollable and may startle the staff and the patients, the initial reaction being to turn the alarm off before attending to the patient. Kerr (1985) also argues that the high pitched sounds are difficult to locate. In order to avoid litigation claims manufacturers perhaps make an attempt to protect themselves by installing shrill, intrusive alarms that make it impossible for staff to claim that they did not hear the alarm above the general noise of the work area. As the ECRI (1987) report concludes, in an attempt to prevent death of patients by misuse of alarms, agreement should be reached upon the design and use of alarms. This can only be successful if the appropriate research is undertaken and the results implemented into auditory warning design.

While there is a preponderance of subjective opinions and anecdotal evidence regarding problems associated with auditory warnings in critical care units, there are nonetheless, several studies that have investigated specific problems which will be discussed in the following section.
1.1.4 Attempts to Remedy the Problems of The Auditory Warnings.

1.1.4.1 Design of New Auditory Warnings

There have been considerable attempts and suggestions to redress some of the problems regarding auditory alarms in the hospital environment. For example, McIntyre and Nelson (1989) stated that objectives for an I.C.U. alarm system should warn staff when immediate attention to a patient is necessary, indicate the location of where the action is required, and provide information about the nature of the intervention necessary. McIntyre and Nelson examined the use of verbal signals in the I.C.U. environment. The results showed that if one message occurred performance was perfect, however, the likelihood of error increased with the number of signals.

Kerr (1985) suggested that each sound should have the same meaning wherever it is encountered in the hospital, presumably this principle would also need to be applied between hospitals as well which implies some kind of agreement between manufacturers. As discussed previously, the auditory signals produced for medical equipment are designed by manufacturers in an arbitrary manner and may both resemble those with entirely different significance or differ from those with a similar meaning, which can obviously result in confusion for the staff who may be required to learn a new set of sounds each time they move to a new environment, which is obviously time consuming.

However, Kerr (1985) further argues that once the sounds are learnt by the staff, the differences between the sounds produced by the different equipment is believed by the staff to be valuable as it provides information regarding the problem about to be encountered.

1.1.4.2 How Many Auditory Alarms?

Miller (1956) described the limit of human information processing as being 7 plus or minus 2 items. Whilst individuals can readily identify hundreds of everyday sounds (e.g. Ballas 1993), Patterson
reported that even in unstressed conditions humans do not learn and remember the significance of more than eight arbitrary sounds. A study by Patterson and Milroy (1980) showed that subjects find it easy to learn and retain 4 to 6 aircraft auditory warnings but when the warnings exceeded 6, the subjects found them more difficult to learn and the signals more often became confused. McIntyre (1985) and Sanders and McCormick (1993) recommend that there should be no more than four to six alarms in any given setting. Based on these proposals, Kerr (1985) suggests that the number of alarms for medical equipment should therefore be limited to an absolute maximum of ten and recommended six alarm sounds. The alarm sounds would be based on the risk to the patient and the required response necessary from the staff. Each sound would have a fast (emergency level) and a slow (cautionary level) version, with the six problem areas being hypoxia, ventilator, cardiovascular, artificial perfusion, drug administration and thermal risk.

For auditory warnings to be of optimal use in alerting the operator before irrevocable damage occurs, Kerr (1985) suggests that it should be possible to distinguish the pattern of events between the change in the variable being measured which produces the auditory warning (alarm event) and the change which produces damage to the patient (damaging event). Kerr further suggests classifying alarms into two types, immediate and anticipatory, depending upon whether the pathway between the alarm event and the damaging event involves a few or many steps, for example if the oxygen supply is cut, the pathway to brain damage occurring is very short. Warning signals that indicate conditions that require responses of varying urgency are also proposed, these levels of urgency are classified as Emergency, Cautionary and Alerting although, as Kerr acknowledges, there is little agreement about what separates one category of conditions from another.

Kerr's (1985) suggestion of an optimum number of 6-10 auditory warnings is broadly supported by one of the very few empirical studies to be conducted in the clinical environment of the recovery and operating rooms of a large hospital. Momtahan and Tansley (1989) showed that staff consistently overestimated the number of
warnings that they believed they would be able to remember. In one study, for example, the nursing staff remembered 8 out of 21 alarms while the anaesthetists remembered only 5 out of 21. However, as Momtahan points out in the clinical setting staff use other cues such as visual indicators to localise a particular alarm. The research supports Kerr's (1985) suggestion that some of the alarm sounds were so similar that they were easily confused, and also found that some of the sounds were masked by other pieces of hospital equipment.

1.1.4.3 The Nature of Warning Sounds

Ideally, the signal produced by the alarm system should not only indicate that something is wrong, but also give some idea of what the problem is. Patterson's (1982) guidelines for auditory warnings were developed for use on aircraft flight decks. The guidelines show that it is possible to design warnings that do not startle, are non-aversive and are informative. The guidelines describe the spectral characteristics required to make a pulsive sound that is distinctive and is not masked by other background noises. Briefly the recommendations made by Patterson are as follows:

i) The pitch of the warning sound should be between 150 and 1000 Hz.
ii) The signals should have at least four prominent frequency components, this ensures that the chances of masking by other sounds is minimised;
iii) a minimum of change to pitch and sound quality under masking conditions;
iv) the number of different distinctive signal codes that may be generated are maximised.
v) Signals with harmonically regular frequency components should be used rather than inharmonic spectra
vi) Lower priority warnings should have most of their energy in the first five harmonics while higher priority signals that require immediate action should have more energy in harmonics 6 to 10. High priority signals can be made very distinctive by incorporating a small number of inharmonic components.
vii) The prominent frequency components for signals should be in the range from 1000 to 4000 Hz.

viii) Finally, the rapid Glides (100 msecs) in the fundamental frequency of a signal (providing the above constraints are not violated) can be an effective way to signal urgency and command listener attention.

1.1.4.4 Perceived Urgency of Auditory Warnings.

Another important conclusion in Mottahan and Tansley's (1989) study involved the lack of a relationship between the medical urgency of the alarm and the psychoacoustic urgency, or the urgency implicit within the sounds themselves. In other words a non-urgent medical situation (e.g. a syringe pump) has a relatively urgent sound, while an urgent medical situation, such as a problem with the ventilator for example, has a less urgent sound.

It would appear intuitive that alarms which are signalling an urgent situation should sound more urgent than a less important alarms. This problem could be resolved by applying some of the principles discovered in a series of experiments conducted by Edworthy et al. (1991) and Hellier and Edworthy (1990). Although some of the work was undertaken originally for use in the high workload environment of helicopter cockpits, the results are potentially applicable to warnings designed for any situation (e.g. Edworthy 1994).

The results of the studies demonstrated that some acoustic parameters are more important in terms of their effect on perceived urgency than others. Sound parameters such as the temporal, spectral (timbre) and melodic characteristics were manipulated in order to find how changes in them produced changes in subjective perceived urgency. The most consistent effect in the series of experiments involved the parameters affecting the temporal qualities of the sounds. Although significant effects were found for the melodic parameters these were less consistent. Hellier and Edworthy (1990) conducted a series of experiments to examine the relationship between changes in three temporal parameters (speed,
unit of repetition and length) and perceived urgency, the subjects having to rank and rate the perceived urgency. The results showed that an increase in all three parameters individually increased the perceived urgency of the stimuli.

Momtahan (1990) similarly examined the effects of certain sound parameters on perceived urgency. In the first experiment, using a paired-comparison paradigm subjects had to judge which levels of different parameters sounded more urgent. Some of the sound parameters were the same as those used by Edworthy et al. (1991), which include interpulse interval length, fundamental frequency and envelope shape. Other parameters used by Momtahan included amplitude and frequency modulation rates, spectral shape (the distribution of the energy amongst the harmonics) and number of harmonics. The results showed that as in the Edworthy et al. (1991) and Hellier and Edworthy (1990) studies, stimuli with a faster tempo (shorter interpulse intervals) and stimuli with a higher fundamental frequency were perceived as being more urgent than slow stimuli and a low fundamental frequency.

Momtahan (1990) also supported Patterson's (1982) proposal that stimuli with more harmonics would be perceived as being more urgent than stimuli with fewer harmonics. However, contrary to Patterson's earlier assertions that low priority warnings should have more energy in the first five harmonics and high priority warnings have relatively more energy in harmonics 6-10, Momtahan found that the more urgent sounds were composed of harmonics with more energy in the higher and lower harmonics rather than in the middle. Momtahan also found that a rapid glide would attract attention, her results showed that a steady pulse was more urgent than any of the other frequency glides studied, a different finding to that of Patterson (1982). Similar results to Momtahan were also found by Edworthy et al. (1991) in that a steady pulse is more urgent sounding than an irregular pulse. Momtahan also suggests that there should be a distinction between the perceived urgency and the distinctiveness of an auditory signal. She comments that while the two perceptions are not mutually exclusive they can not be assumed to be the same. While frequency glides, frequency modulation and amplitude increase
the distinctiveness of a signal, they do not necessarily increase the perceived urgency.

Momtahan (1990) also investigated the effect of loudness on perceived urgency, the supposition being that a loud sound would be perceived as being more urgent than quiet sounds. The results did show that louder sounding signals were judged as more urgent than softer signals and it was concluded that while louder alarms were not suitable for use in an environments such as a hospital, alarms could be louder to attract attention if the problem is not attended to within a reasonable length of time. Empirical knowledge of the effects discussed above could be applied to a set of auditory warnings with varying degrees of urgency (Edworthy et al. 1991). This could solve the problems and confusion that occur when several alarms sound simultaneously. Staff would be able to identify and attend to the most urgent condition first.

1.1.5 Recommendations For Medical Equipment Alarms

A report submitted jointly by the Institute of Sound and Vibration Research and The MRC Applied Psychology Unit, (Patterson et al. 1986) proposes a rationalised warnings system for use in I.C.U. and O.R. Existing hospital auditory warnings were found, on the basis of Patterson's 1982 guidelines, to be unsuitable to form the basis of new warning sounds. The reasons given were that when analysed the existing warnings typically showed one main spectral peak, whereas Patterson's guidelines recommend that the sound should have at least four of the first ten harmonics to fall in the frequency region 0.5-4.0 kHz. The temporal patterns of the existing warning sounds were also shown to be unsuitable as they consisted of a simple continuous sound or an off-on pattern. It was found that confusion often occurs with aircraft warnings when the temporal patterns are similar. The report's final criticism of the existing hospital warnings is that they come on at full volume, which is considered unnecessary due to the possibility of startle reaction in both the staff and the patients.

The sounds of the new auditory warnings suggested by Patterson et al. (1986) for use in hospitals are described as resembling atonal
melodies, with a syncopated rhythm, somewhat similar to bird calls, each alarm having its own characteristic set of sound pulses, referred to as 'bursts', sequence of pitch changes and temporal pattern. The urgency is indicated by increasing the speed rather than by increasing the loudness and the sounds will not be played continuously, but have several seconds of silence to allow the staff to communicate rather than the immediate reaction being to turn the alarm off.

The alarm sounds will be differentiated from one another in three ways: the harmonic structure; the pattern of loudness and pitch; and the time interval between the notes. The pitch, intensity and speed of the burst are used to vary the perceived urgency of the warning sound. Therefore, if a situation requires attention the alarm will sound, the first burst played at a pitch and speed that indicates moderate urgency and at a level that is clearly audible but not excessive as determined by the background noise in the environment. This first burst should attract the attention of the staff and immediately convey the message of the warning sound. The burst is then repeated after a 1 or 2 seconds gap. After the second burst it is highly likely that the warning will have conveyed its message so repetition of the burst at the most urgent level could be unnecessarily irritating. The pitch, level and speed of the burst are lowered to reduce its perceived urgency and it is played every 4 seconds or so in this non-urgent form in order to allow time for communication or the necessary action to be taken by the staff. If the condition initiating the warning sound is not corrected after a certain length of time, the warning is then repeated at its most urgent, that is, a high pitch, high pulse rate and at a level that overrides any ongoing speech and commands immediate attention.

The set of alarms are composed of two levels, emergency and cautionary and an 'information available' signal, with seven categories; general, oxygenation, ventilation, cardiovascular, artificial perfusion, drug administration and temperature. The pulses for the cautionary signals are 200ms long while the urgent signals are shorter bursts of pulses of 150ms. The emergency signals have delayed or more harmonics than the cautionary signal
and are also differentiated by the number of sound pulses that occur in a 2 second interval.

Stanford et al. (1985) developed a set of eight auditory alarms which were design to be less irritating than the existing alarms on anaesthetic equipment, and less susceptible to masking by background noises. Each pulse of sound used could be likened to vowel components (e.g. low i; high i; low a; high a; low u; high u.) The length of each pulse is 50, 100 or 150 msec in length, so an auditory warning would be composed of 6 or 7 of these pulses, each warning being 500 msecs long using different combinations of the pulses. Stanford et al. (1985) found that the signals were detected with 93% accuracy within the background noise of the O.R.

In a second study, Stanford et al. (1988) compared affective response to 7 existing hospital alarms and the 8 proposed alarms developed by Stanford et al. (1985). The results showed that there was a higher percentage of positive responses to the new alarms which presumably indicates that the sounds produced were found to be less irritating to the staff.

In summary, when a situation requires attention, the new auditory warnings proposed by Patterson et al. (1986) will present a moderately urgent sound initially, which is then silent to allow time for conversation, only returning to interrupt forcefully if the situation is not attended to. The notes are rounded to cause the minimum of 'startle' reaction and the sounds will be easier to localise because both high and low frequency harmonics are included.

Patterson (1982) suggests that a burst that is designed to convey a high level of urgency would have a fast irregular pattern, a rising pitch contour and an amplitude envelope which ends at its maximum sound level. A low level sound would have a slow, regular tempo with a falling pitch contour and its amplitude envelope which ends below the maximum sound level. While there has been considerable investigation into the design of a set of warnings for areas such as the I.C.U. and the O.R, there have been no clinical trials to evaluate their effectiveness.
1.1.6 Clinical Trials of Proposed Auditory Warnings

North American Drager, who manufacture anaesthetic equipment and machines for I.C.U. such as ventilators drafted a standard in the early 1980's to organise alarm messages in their medical equipment. The standard categorised alarm sounds according to the needed response, warning, cautionary or advisory. With a continuous repeating pattern for the urgent condition, an intermittent repeating pattern for the cautionary condition and a single tone (or in some instances no sound, just a visual message) for the advisory condition. However the company,

"disagreed to identify the cause for an adverse condition by a specific sound pattern. The latter approach may be acceptable for pilots but in our opinion is not acceptable for hospital personnel." (P. Schreiber, President, North American Drager, Personal Communication 1991).

However, no reason is given why it would be unacceptable for adverse conditions to be identified by a specific sound. The priority system and original sounds designed by Drager have now been utilised in over 10,000 pieces of equipment.

While Drager have not conducted any formal studies on the success of the alarm system they state that they have had many unsolicited, positive comments regarding the system and that as other companies are now copying this approach then it has to be right.

Schreiber and Schreiber (1987) state that the alarms should occur before the vital signs of the patient become critical, giving the operator time to identify and correct the problem, that the alarm messages should identify the problem as soon as possible, the operator's attention being directed towards the source of the problem not towards the investigation or management of the alarm sound or alarm message, and finally that the design of the alarms should not be irritating to the personnel. They have attempted to incorporate this philosophy into the design of their equipment by having a centralised display on the monitors with multiple
information that is clearly displayed. The auditory warnings are centralised on one loudspeaker with the temporary silencing of alarms possible from one central control. To minimise confusion during a situation when several alarms sound simultaneously, only the most urgent condition is heard.

While Drager have attempted to rationalise and prioritise alarm sounds in hospitals, the underpinning of the system is based on the view that because the alarm system works in aircraft, it will therefore be successful in hospitals. There are obvious differences between aircraft and hospitals in their requirements for auditory warnings. Aircraft systems provide information to a pilot in a fixed position, with a reasonable constant level of background noise. This is not always the case in hospitals, where a signal loud enough to be heard during the day would be far too loud at night. As there have been no formal clinical trials within the hospital environment to support Patterson's recommendations, it would be interesting to see how this attempt to introduce a more systematic approach to auditory warnings actually works with a formal survey conducted instead of the usual anecdotes and opinions that characterise this area of research. There have been no corresponding follow-up studies to evaluate the design of the auditory warnings or any other problems incurred by the staff. Satisfaction is denoted by hearsay - is it possible that unsatisfactory remarks are ignored?

1.1.7 How Much Monitoring?

Medical staff are divided about the optimal amount of monitoring necessary many of the arguments based on personal belief and anecdotal evidence. Obviously, the more monitors used, the greater number of alarms there will be. The American Society of Anaesthesiologists in 1986 recommended standards for basic patient monitoring, which includes, arterial blood pressure, ECG, an oxygen analyser, and a ventilator disconnect alarm. Pulse oximetry, capnography, and spirometry are recommended but not compulsory. Similar levels of monitoring are recommended in Great Britain and also in Australia (Cass et al. 1988). However, this impressive list of equipment does not take into account the physical and
psychological limitations of the operator in efficiently observing all these monitors. As will be discussed further in this chapter, it is generally recognised that there is a limit to the amount of information the human can process at any given time (e.g. Kahneman 1973).

Are auditory alarms during monitoring of any benefit to patients? Eichhorn (1989) reports that in an investigation of over a million anaesthetics administered in the U.S.A. between 1976 and 1988 the accident rate was reduced in monitored patients and that in analysis of individual patients, most mishaps can be identified early enough to prevent major incidents occurring. However, Orkin (1989) questions the significance of these findings arguing that further research is necessary to clarify whether the monitors and in particular those recommended are really effective in reducing the incidence of major anaesthetic incidents. Hamilton (1986) states that there is too much emphasis on monitoring and a danger that attention may be diverted away from the patient and subsequently decrease the quality of care.

McCarthy (1985) conducted a questionnaire survey amongst I.C.U. nurses that showed that the majority of nurses viewed the increase in new technology as helpful and increasing the skill requirements of their job, although some expressed concern about using the new equipment. The nurses also expressed concern about the extent to which monitoring equipment takes attention away from the patient, with 77% stating that their work was more impersonal and 66% who felt that they had become technicians because of the increase in monitoring equipment, rather than a carer. McCarthy reports that the impact of new technology on nursing staff can have both positive (i.e. increasing the level of skill required for the job) and also negative effects (i.e. interfere with the relationship between the nurse and the patient). Block (1986) postulates that while there is no evidence to show that monitors do actually improve patient care, as there is little risk associated with non invasive monitors, such as ECG electrodes, they should be used if there is the possibility that useful information may be provided, or in other words the more monitoring the better. If a critical incident should then occur, this can be
attributed to a lack of training and education on the part of the operator, not to negligence. Block's argument, however, seems to be based on the possibility of litigation rather than the actual need to monitor patients. It could be considered unethical to have patients connected to every type of monitoring device when it may not be necessary.

The arguments as to whether this increase in monitoring is beneficial to patients can perhaps be best illustrated by an investigation into litigation claims by Davis (1989). He questioned the extensive use of cardiovascular monitoring, which, although indicating that a patient has cardiac dysfunction, does not alarm in sufficient time to prevent permanent brain damage from occurring. It could be suggested that to be of optimal use to an operator, a monitor and its alarm system should warn of impending damage that may occur to the patient if corrective action is not taken, not alarm when it is either too late to prevent damage, or when irrevocable damage has already occurred.

There has been a rapid increase in medical technology and it is difficult to determine whether there are more demands on the cognitive abilities of staff working in an environment where monitors process and display considerable information. In the I.C.U. environment the monitors provide information regarding the condition of the patient. Such equipment is also overspilling into general wards and there is a need to ensure that the systems used actually assist the staff in caring for the patient rather than overloading them further. As Eggemeier (1988) argues in areas of high technology with complex systems the need to assess the load imposed on operator processing capacities is particularly important.

1.1.8 Sustained Attention

In many tasks undertaken by operators it is necessary to maintain attention over a period of time whilst monitoring critical signals. In multi task environments such as the I.C.U. tasks are very complex and there are many other variables in the environment. Some studies have attempted to remedy the gap between the laboratory and the
real-world environment and have shown that the vigilance decrement in comparable tasks is similar (e.g. Craig 1984, Parasuraman et al. 1987). An example often cited to describe sustained attention or vigilance is that of radar operators scanning screens to detect a signal. Sustained attention can also be applied to staff in an I.C.U. scanning monitors and equipment in order to check a patient's condition. With increases in technology there is a subsequent increase in the importance of staff sustaining attention by monitoring and assimilating the information produced by often complex equipment.

The majority of laboratory studies investigating vigilance and sustained attention have shown that prolonged monitoring results in failures or delays in detecting signals that would have been correctly reported at the beginning of the observation period. Mackworth (1957) observed radar operators and recorded a decrement in the performance of the operators after 30 minutes on a detection task. He stated that vigilance in monitoring tasks requires a state of maximum physiological and psychological readiness to react so as to be able to detect and respond to small stimulus changes over long periods of time. Other studies have also demonstrated that as time progresses, operators become less efficient at detecting visual or auditory signals (e.g. Stroh 1971, Mackie 1977, Warm 1984).

However, some studies have argued that the efficiency in detecting critical signals depends on the level and complexity of the signal used in the task. For example, in a two hour simulated air traffic control task, Thackray and Touchstone (1989) found there were no errors when participants had to detect a simple signal in which three altitude numbers changed to three 'Xs'. For the complex signals two aircraft on one flight path were given identical numbers, the two aircraft either moving towards each other (potential conflict), or they were not. Thackray and Touchstone found that there were three times the number of complex signals missed in the second hour as in the first hour and that the latencies were longer with increases in the information processing demands of the tasks.
In some conditions task manipulation did not produce a decrement in sustaining attention and in some studies an improvement was demonstrated in detecting signals over time as demonstrated by Warm et al. (1984).

1.1.9 Do Monitors Aid Vigilance?

There has been very little research regarding vigilance and sustained attention in the hospital environment. Gravenstein and Weinger (1986) define vigilance as it applies to anaesthetists as follows:

'a state of clinical awareness whereby dangerous conditions are anticipated or recognised and promptly corrected.' (page 145).

Several studies have debated about the degree to which monitors and human vigilance of the patient should be combined. For example, Klein and Moyes (1985) argue that the monitors should be;

'used to augment and not replace the time honoured skills of inspection, palpation, percussion and auscultation' (page 410).

These skills of observation require considerable attention, motivation and concentration especially when performed over a long period of time, together with the complex signals emitted by the monitoring equipment. ECRI (1987) states that although monitors and alarms cannot replace consistent clinical vigilance and alertness, if integrated together in an 'appropriate' way, the alarms should be instrumental in preventing patient injury and death. The report does not, however, explain the 'appropriate' way that human vigilance and monitoring equipment should be combined efficiently, leaving individual operators to make arbitrary judgements.

Hyman and Drinker (1983) similarly argue that although alarms should not replace the operators' observation of the patient's condition, they can be of particular use, for example during periods of inattention due to fatigue, or stress or when there are numerous pieces of equipment which would be impossible to monitor effectively. Kestin et al. (1988) concluded in their study that the
anaesthetist's attention and ability to integrate information remains an important way of monitoring patients. All of the above mentioned studies or commentaries are rather vague in defining how vigilance and monitoring should be optimally integrated to provide the most efficient monitoring of patients.

It has also been argued that monitoring equipment and alarms cause anaesthetists to relax their vigilance and lull them into a false sense of security (Edge and Braude 1985). Hyman and Drinker (1983) also urge caution in over-reliance being placed on the alarms as a substitute for the operator's attention as problems would occur if the alarm failed or was switched off, either inadvertently or purposely. Gravenstein and Weinger (1986) similarly state:

'Intraoperative electronic instruments and automated alarms can lull clinicians into a false sense of security. These devices may trick with artefacts, distract with irrelevant data or rob the anaesthetist of the motivation to observe, record and appreciate the clinical data' (page 145).

In concluding this section on sustained attention and in particular its relevance in the hospital environment, it can be seen that there is very little clinical research into the optimal relationship between the use of monitors, alarms, and human vigilance, the majority of the literature expressing the opinions and views of the authors. There would appear to be a need for research regarding the vigilance of staff in hospital environments such as the I.C.U where there is an increasing amount of technology used to provide information regarding the patients condition. The main problems can be summarised as follows. If there is over-reliance on the auditory warnings signalling an adverse situation, then there is the possibility of serious consequences if the alarms fail or are turned off. Studies, such as those conducted by Mackworth have shown that decrements in monitoring tasks increase after about 30 minutes, yet other studies have argued that this decrement is not found if the operator is presented with complex signals over a long period of time.
1.2 Information Processing

The concept of individuals possessing a limited potential to undertake two or more tasks simultaneously is central to the many discussions regarding multi-task environments. The concept derives from communications engineering and the theory of information developed by Shannon and Weaver (1949). A pivotal tenet of information theory is that of the communication channel which exists between the two communicating points. It is defined by its ability to send information between sender and receiver and it is characterised by a number of parameters, the most important in terms of the mental capabilities of an operator in a multi-task environment, is the capacity of the channels. Shannon and Weaver (1949) postulated that channels could vary in their capabilities. A channel demonstrates its full capacity if there is no reduction of information between sender and receiver and degradation of channel capacity are measured as decrements in information transmission. Information theory proposes that information could be seen as a commodity that could be manipulated, transmitted and transformed.

There were many attempts to use the information theory model to demonstrate the limits of human performance (e.g. Crawford 1979, Rault 1976) and to apply information theory to human information processing which was likened to a communication channel that processed messages and transmitted information from a stimulus. However, the use of information theory in the analysis and prediction of an individual’s mental capabilities have been limited although they have substantially influenced further research into workload.

1.2.1 Divided Attention

In certain situations individuals are capable of dividing their attention and processing information from more than one source at the same time. However, in other conditions individuals often find it difficult to perform two tasks simultaneously. An important characteristic of attention is its selectivity and the ability to shift focus from one stimuli to another when it is necessary to perform more than one task simultaneously.
When two or more tasks are to be undertaken at once an operator may use the time available efficiently to undertake the tasks by switching rapidly from one task to another. It has been argued that the performance of two or more tasks is sequential, involving very rapid switching between the tasks (e.g. Schweickert and Boggs 1984). Perfect time-sharing is said to occur when two reasonably difficult tasks are performed concurrently with no decrement in performance to either task (e.g. Allport et al. 1972). However, Wickens (1991) argues that in a situation where a time constraint occurred, for example, if an operator had to perform two five minute tasks in seven minutes, then the operator would change from successive time-sharing to concurrent processing, or in other words, undertake the two tasks together.

Pashler (1994) found in laboratory experiments that there is a considerable delay (the psychological refractory period) in the response time to a second task even with very simple dual-tasks activities. Damas and Lintern (1981) and Wickens (1976) amongst others have shown that there is a decrement in the dual-task performance involving complex motor tasks relative to the base-line single task performance. Furthermore, it has been found that the combination of certain tasks can be undertaken together more successfully than other combinations of tasks (Wickens 1989).

There are two major classes of models that have developed to explain the decrement that occurs in the performance of two tasks in comparison to the performance of a single task. The first group of models to be discussed are the postponement or the 'bottleneck' theories, where during a crucial stage in the processing of each task a common mechanism is occupied. This means that the critical stages of the second task cannot begin until the critical stages of the first task have been completed. These theories include single channel models (e.g. Kahneman 1973, Broadbent 1971) and the psychological refractory period models (e.g. Welford 1952, Pashler and Johnson 1989, Pashler 1994).

The second group of theories involves capacity or multiple resource models (e.g. Navon and Gopher 1979, Norman and Bobrow 1975,
Wickens 1980, 1984). It is assumed that resources can be divided between tasks and that processing on tasks can proceed simultaneously. It is suggested that there are fewer resources available to each task and the tasks therefore proceed at a reduced rate. Performance of tasks depends partly on the amount of resources available for each task. Hence, postponement models assume that processes undertaken in each task are performed serially, while capacity models conjecture that all stages of processing can proceed simultaneously but at a reduced rate.

At a practical level interference between the performance of two or more tasks has important implications for operators such as air traffic controllers and flight crews, using complex technology in multi-task situations. As McCann and Johnson (1992) state, a better understanding of what happens when tasks are performed in multi-task situations could help to improve the performance and efficiency of operators in such environments.

1.2.2 Time Sharing

Effective time-sharing and resource allocation together with an appropriate strategy could enhance performance. For example, an experienced, expert operator may have a very efficient time-sharing strategy that results in an automated performance of certain aspects of a task. Damos and Wickens (1980) demonstrated that dual-task training was generalisable from one set of tasks to another and that there was better performance and a more efficient strategy in task integration than in task alternation. They suggest that most time-sharing skills seem to be applicable to a specific combination of tasks and may not be generic.

Lintern and Wickens (1991) suggest that it may be advantageous to develop and train operators in general time-sharing skills or strategies in multi-task environments. Training for one task alone is not always applicable as it is in the combination of the tasks that the difficulty for the operator results which could have important implications for the staff in a multi-task environments.
The assumption that strategies in allocating and switching resources can contribute to improved time-sharing performance has been demonstrated in various studies (e.g. Schneider and Fisk 1982, Schneider and Shiffrin 1977). Gabriel and Burrows (1968) also demonstrated that visual scanning is an important component of attention switching. They trained pilots to spend more time scanning activities that were happening outside the plane rather than spending all their time scanning the equipment inside the plane. This resulted in an improvement in pilots locating out-of-cockpit targets and there was no decrement in their activities inside the cockpit (e.g. detecting equipment malfunctions). Attention switching may be a critical aspect of efficient multiple task performance and may relate to the ability to prioritise the appropriateness of the multiple sources of information (Schneider and Detweiler 1988).

1.2.3 Postponement (Bottleneck) Theories

An important consequence of the idea of an individual having a limited information processing capability resulted in the assumption that different tasks imposed different loads on the operator and the concept of 'spare capacity'. If certain tasks did not use all the available capacity then there was the possibility that the residual capacity could be measured.

The first group of early theories involved the single channel model. Broadbent (1971) proposed that the human operator has a single channel, a limited-capacity central processor that is required to interpret stimulus information which is then be used in task performance. If two tasks were to be undertaken at the same time, the central processor would engage in processing information for the first task and then after a delay, the information for the second task would be processed. All information had to pass through one channel and wait its turn to be processed. A resource could be allocated to one task but could not be shared between tasks. Other early research (e.g. Deutsch and Deutsch 1963, Welford 1967) also proposed a single information channel through which all information must pass in order to be processed.
Pashler and O'Brien (1993) suggest that the clearest evidence for a bottleneck is the psychological refractory period (PRP) which occurs in even very simple dual tasks. The PRP occurs when subjects are presented with two stimuli in quick succession each of which requires a fast, distinct response. The typical finding is that the response to the second stimulus is delayed, as compared with when the same stimulus is presented alone. Thus PRP refers to the increase in the reaction time to the second of two successive response signals. Two stimuli are presented which occur at different times. The interval between the signals is decreased (this is called stimulus onset asynchrony (SOA)) and the subject makes a separate speeded response to each stimulus. A decrease in SOA results in an equal increase in the second reaction time and this is known as the PRP effect.

A single channel explanation of PRP was first proposed by Welford (1952) and has received considerable support (e.g. McCann and Johnson 1992, Pashler 1984; Pashler and Johnson 1989). According to the single channel models PRP effects are caused by a bottleneck. Two reactions can proceed in parallel until they reach the bottleneck and the second reaction must wait until the first reaction has cleared the bottleneck (Osman and Moore 1993). If the SOA is short the second reaction is more closely preceded by the first and must therefore wait longer before taking its turn.

PRP is very robust and occurs with very simple stimuli. It is not eliminated with practice and occurs whether the tasks are in the same or different modalities. PRP appears to exist for all studies that involve a pair of concurrent tasks (e.g. see Pashler 1993 for review). Some models of attention (e.g. Broadbent 1958, 1971) have attributed the same mechanisms responsible for PRP to selective attention. More recently PRP effects have been used to study the effects of interference between tasks (e.g. Pashler 1984; Pashler and Johnson 1989).

Much of the research regarding single channel models focuses on where the 'bottleneck' occurs, and there is considerable disagreement amongst researchers. For example, Broadbent (1958) in the filter
theory of attention proposed that the bottleneck occurred during the perceptual or identification stage of the stimulus. Other researchers (e.g. McCann and Johnson 1992; Pashler 1984; Pashler and Johnson 1989) have argued that the bottleneck occurs at a more central stage involving the decision or stimulus response translation process, while others such as Keele (1973) suggest that the bottleneck occurs at the level of response initiation.

There are however, some alternative views of the PRP effect. First, PRP could occur because of competition between two tasks for limited resources (e.g. Kahneman 1973, McCleod 1977). The processing of two tasks could proceed in parallel but each task depends on the availability of resources. If the first task is designated as being of primary importance it receives all the available resources at the expense of the second task. Second, that there is a response conflict (e.g. Herman and Kantowitz 1970, Kantowitz 1974). As with limited capacity models the processing of two tasks could proceed in parallel but may result in conflicting response tendencies. A third suggestion is the inability to prepare fully for both tasks as suggested by De Jong (1993) and Gottsdanker (1979). The processing of the second task is slowed down not by the actual performance of the first task, but by the preparation to perform the first. The second reaction could be postponed until preparation for the first is complete or the processing could be slowed by the resource sharing during the preparation of the responses.

Moreover, despite claims that the PRP phenomenon is found in all dual task experiments, Greenwald and Shulman (1973) demonstrated that there was no evidence of 'refractoriness' and in fact found that the responses to both cues were as rapid as when either task was performed alone.

Norman and Bobrow (1975) introduced the concept of performance-resource-function. They postulated that an individual has a finite number of processing devices that they termed 'resources'. The concept of resources has also become central to models and theories regarding workload. Kantowitz and Knight (1976) stated that
a cognitive resource is a theoretical construct. Resources are characterised by two general properties. These are that their deployment is under voluntary control, and they are scarce. If resources are allocated to one task to improve its efficiency a second task will decrease in efficiency. Norman and Bobrow (1975) postulated that if single task performance was generally better than when two tasks were performed, it could be because the tasks are sharing resources that were previously being exclusively used for one task. There must therefore be some function that relates the quality of performance to the quantity of resources invested in the task. In other words, the performance of the task is positively related to the amount of resources available to complete the task satisfactorily.

1.2.4 Single Resources Model

Kahneman (1973) proposed a general purpose capacity that was competed for and allocated amongst all the cerebral processes. Resources were seen as a single, undifferentiated pool of energising forces necessary for all tasks and mental activities. Kahneman (1973) postulated that if an operator fails to perform two tasks effectively, this is due to a shortage of resources. Decrements in performance are due to the demands of two concurrent activities exceeding the total demands of the two tasks. If all the resources of an operator are being utilised to complete one task, and a second task then demands resources, these will be shared out between the tasks regardless of whether the two activities share any common characteristics. Kahneman (1973) states that,

'Interference is non-specific and depends only on the demands of both tasks' (page 11).

However, ever though the amount of resources available at any one time is finite, Kahneman states that the limit can vary with the level of arousal (heart rate, pupil dilation etc.) of the operator. For example several experiments conducted by Kahneman and colleagues demonstrated that when a task was very difficult there was a corresponding increase in the size of the pupil (e.g. Kahneman et al.)
1968). However, as Allport (1980) argues, Kahneman does not explain the character of the relationship between 'arousal' and 'capacity'. As Gopher and Donchin (1986) state, the single channel model did not fare well in experimental situations because performance on some tasks interferes with one task but not with another, for example, performance on a mental arithmetic task is not impaired when undertaken jointly with a tracking task. These results are inconsistent with a large 'pool' of generic resources.

1.2.5 Multiple Resources Model

The weakness of the single channel model led to the development of the multiple channel model (e.g. Navon and Gopher 1979; Polson and Friedman 1988; Wickens 1980, 1984, Allport 1980). The capacity models propose that individuals have a number of processing mechanisms each having its own resource functions.

Wickens (1992) has described resources as consisting of three dichotomous dimensions,

'...two stage-defined resources (early verses late processes), two modality defined resources (auditory verses visual encoding), and two resources defined by processing codes (spatial verses verbal processing)' (page 375).

That hemispheres of the brain are involved in processing information pertinent with regard to research on the degree of functional specialisation of the cerebral hemispheres in sensorimotor and cognitive activities (e.g. Friedman and Polson 1981). It is purported that spatial processing and left handed control occurs predominantly in the right hemisphere, while verbal processing, speech response and right handed control is dominated by the left hemisphere (Springer and Deutsch 1985). Friedman and Polson (1981) linked hemispheric specialisation with dual task interference and suggested a 'multiple resources' model in which the left and right hemispheres provide two mutually inaccessible and finite collections of resources. Therefore, if tasks are undertaken that utilise resources
from each hemisphere there will be little interference between the
tasks.

Closely related to the hypothesis of multiple resources is the work
of Baddeley in which a 'spatial' and a 'verbal' working memory
system has been identified (e.g. Baddeley 1986, Baddeley and
Liberman 1980).

Neurophysiological literature also adds support to the notion that
there might be highly specific, independent resources within the
cognitive system. One example illustrating this are those patients
who suffer from aphasia. Aphasia is a broad term for speech and
language dysfunction that can occur after damage to specific areas
of the brain. There are many different types of aphasia, depending on
where the damage to the brain occurs. If there is damage to the
frontal regions of the brain, in particular Broca's area for example,
an individual will have difficult in speaking words, even if the motor
areas used for speech are not damaged. The individual will be able to
comprehend other people's speech, but their own speech will be very
hesitant and is often called telegraphic or agrammatic speech (e.g.
Springer and Deutsch 1985). Other examples include disorders of
reading and writing such as alexia with agraphia, which occurs when
there is damage to the angular gyrus. In this condition a patient can
write a sentence, but is unable to subsequently read the same
sentence back.

The important tenet of the multiple resources model is the
assumption that the resources are divided into groups so that
processing on two tasks can often proceed simultaneously. Because
fewer resources are available for each additional task, the
processing of two or more tasks subsequently proceeds at a slower
rate than for a single task. However, two tasks can be performed
satisfactorily provided the tasks require separate, and not similar,
resources.

The multiple resources model postulates that the processing of two
tasks is dependent on graded resources that are used in differing
quantities, the greater the quantity available, the more efficient the
processing. Wickens and Liu (1988) state that the model describes the concurrent processing of two tasks and how they will interfere with each other, first as a result of their difficulty (resource demand) and second of their shared processing mechanisms (resource composition). Resource demand is described as occurring when one task becomes more demanding in terms of difficulty during a dual-task situation, which will either result in increasing interference with the second task, or the primary task (i.e. the task of interest) itself will suffer a decrease in performance. Resource composition refers to the performance-demand decrement described above which will occur if the tasks share processing structures or resources.

As described previously, one of the reasons why the single resource model (e.g. Kahneman 1973) was unsuccessful in empirical settings was because of the many examples of certain tasks that could be performed successfully together. For example, Vidulich (1988) and Wickens (1980) demonstrated that the combination of verbal (vocal) and spatial (manual) tasks, together with a tracking task, was more effectively shared with a verbal response than with a manual response.

The multiple resources theory would therefore appear to propose that if tasks are similar, greater interference will occur. Wickens (1992) further suggests that two difficult tasks that use different processes or resources may involve less interference than two relatively easy tasks that are within the same dimension. Interference occurs if the resources are required within a stage of processing, for example, spatial memory and spatial perception. Less interference occurs if the resources are required between stages, for example, spatial perception and a manual response.

If tasks are similar, resources are activated for one task and are then activated for another similar task, confusion or what has called 'outcome conflict' (Navon 1984, Navon and Miller 1987) can then occur. An example of the type of interference and confusion that can occur can be clearly demonstrated by using the example of the Stroop task (Stroop 1935). In the Stroop Task the semantic characteristics
of a word (for example the name of the colour 'red') interferes with the subjects' ability to verbally report, as quickly as possible, the colour of the ink, which may be, for example, the colour 'green' in which the word 'red' was printed. This type of confusion has since become known as Stroop interference. However, as Wickens (1991) points out the Stroop task is not a dual-task. It does, however, provide a good demonstration of the confusion that can occur when a single task involves very similar features and this could be what occurs when there is confusion between two similar tasks.

Korteling (1993) further suggests that tasks that appear similar at a superficial level, can be manipulated to be more dissimilar or almost the same. For example, if participants are asked to identify one of two messages or to shadow two simultaneous messages, the results show a decrement if the semantic similarity is increased between the two messages. Thus, it is suggested, interference may be reduced by decreasing the similarity of the characteristics of the tasks.

Hence, there appear to be two considerations regarding the similarity between tasks. First increased similarity may decrease the performance of two tasks because of confusion and interference between those tasks. For example, Hirst and Kalmar (1987) found that performance was better for one mental arithmetic task and one spelling task than for two mental arithmetic tasks or two spelling tasks. Secondly, increased similarity may improve the performance by co-operation between the tasks. This could be because when simultaneous tasks are undertaken there may be a resemblance between aspects of the tasks, such as the spatial or temporal properties. Therefore, the competition for resources could be less than, for example, when tasks are completely different and each different task may require considerable resources. When the tasks are similar there may be common information processing routes that both tasks share. For example Klapp (1981) and Peters (1981) found that the performance of two time-sharing rhythmic tasks were better when the rhythms were the same as opposed to different rhythms. Wickens (1984, 1989) states that these differences in
aspects of the task could be an important determinant of dual-task performance.

A dimensional representation of the multiple resources model is shown in Figure 1.1. The concept of the model illustrated in Figure 1.1 as described by Wickens (1984) is that task demands can be placed in a space according to whether they are verbal or spatial the input forms are visual or auditory and their output form manual or vocal. As Wickens et al. (1984) point out a further consideration is that some tasks may require a mixture of resources. However, the important factor would appear to be that the closer together two tasks or interface channels are in a space, the more they draw on the same attentional resources. They therefore become more difficult to timeshare and to undertake two or more tasks concurrently.

The dimensions of the multiple resources model shown in Figure 1.1 are discussed in more detail below. The dimensions involved in the model are, first, processing stages, second, processing modalities and third processing codes.
Figure 1.1: The Proposed Structure of Processing Resources. (From Processing Resources in Attention C.D.Wickens 1984 in R Parasuraman and R Davies (Eds) Varieties of Attention, New York Academic Press).
1.2.5.1 Processing Stages

In the processing stages there are two dimensions, perceptual and cognitive processes, for example thinking, which can be shared effectively with the selection and implementation of overt manual and vocal response processes (Wickens 1991). The resources used for perceptual and central processing are the same, but are different to those resources used to select and execute the response. An example given by Wickens (1992) is of an air traffic controller who would be able to acknowledge the change in the positions of aircraft either manually or vocally whilst at the same time could retain a precise mental model of the airspace, which Wickens states is a perceptual-cognitive demand.

1.2.5.2 Perceptual Modalities

In the second dimension, the perception modalities, it is postulated that separate resources are used for auditory stimuli and visual stimuli. There are many studies that have shown that tasks using different modalities are in general performed better than tasks which share the same modalities. For example, Rollins and Hendricks (1980) demonstrated that simultaneous auditory messages were difficult to process, and it was shown to be more effective if one message was displayed visually.

1.2.5.3 Processing Codes

The processing codes are involved in the third dimension and include auditory versus visual perception and vocal versus manual responses. Spatial and analogue information is assumed to use different resources to those processing verbal and linguistic information. It is suggested by Wickens (1992) that the code related resources may to some degree be determined by the functions of the cerebral hemispheres, as discussed above.

Gopher and Donchin (1986) argue that the multiple resources model described by Wickens (1980) neglects to discuss the relationship between, 'energetic and structural elements' (page 41.17).
These factors are incorporated in the cognitive-energetic stage model proposed by Sanders (1983) and Gopher and Sanders (1984) which developed from data on the effects of stressors, such as sleep deprivation, and on performance in choice reaction time tasks.

However, as discussed previously, the concept of workload is multifaceted and as Wickens (1984, 1989) concedes, is possible that one theory is inadequate to explain multi-task performance and that several different mechanisms may operate.

1.2.6 Task Difficulty and Practice

It has been postulated that task difficulty influences the amount of resources available. For example, an important consideration is that in novel, complex tasks resources will be involved in learning, coordinating and executing the various components of that task (Wickens 1991). It is also suggested that there is a performance decrement in tasks that are practised less than other tasks. Continued practice in either a dual-task or multi-task situation often leads to improved performance.

Initially, some tasks appear very difficult and all attentional resources are utilised. This can perhaps be illustrated by using the example of learning to drive a car or learning about the functions of a ventilator. However, with practice driving and checking the information on the ventilator, at least to a certain extent, becomes automatic. As the practice continues performance improves and becomes more automatic, requiring fewer and fewer cognitive resources. It is well documented that the development of automation in skill learning will decrease the amount of interference with a concurrent task (Bahrick and Shelly 1958, Schneider and Fisk 1982).

It is suggested that the early learning of a task is inefficient, particularly if the resources required in a multi-task environment are similar (e.g. requiring verbal resources). Sweller et al. (1990) demonstrated that complex problem solving interfered with the resources required for learning. This assumption would therefore appear to have important implications for the results of laboratory-
based experiments in which participants are relatively inexperienced in undertaking certain tasks and may use resources learning how to undertake the task.

1.2.7 Automated and Controlled Processing

Fisk and Schneider (1981) propose that human performance involves two qualitatively different types of information processing. The two classes are referred to as 'controlled' and 'automated' processing (e.g. Schneider and Schiffrin 1977). Controlled processing is slow and limited by short-term memory capacity. Automated processing on the other hand is described as being very fast, is not limited by short term memory capacity and involves very little effort from the individual.

For an experienced operator in multi-task situations there may well be a combination of automated and controlled processing occurring. It has also been suggested that high mental workload situations involve those situations where a high degree of 'controlled' processing is required and in order to alleviate and lessen the mental load of these situations, training could be adapted to assist individuals to develop an automated processing of particular tasks (Schneider and Fisk 1982).

To summarise, as discussed in the preceding section there are several explanations regarding the decrement in performance when individuals undertake two tasks concurrently. Two main classes of models have been proposed to explain the interference between two tasks. First the 'bottleneck' or postponement models including the PRP effect, propose that processes in each task are handled serially and that decrements in performance are due to a bottleneck occurring when the first task is being processed, and processing of the second task is delayed. The second group of models incorporate capacity or multiple resources models (e.g. Navon and Gopher 1979; Norman and Bobrow 1975; Wickens 1980, 1984) which propose that all stages of processing are parallel.
However, as Wickens (1984, 1989) states, it is possible that single theories are inadequate in explaining multi-task performance, and that several different mechanisms may operate. He suggests that while some aspects of information processing are serial, there are others that are processed in parallel, such as talking and driving. Serial processing may occur when the demands imposed by one task become excessive and cannot be shared with another task and some sort of sequential sampling or 'swapping' between stimuli then occurs. The individual has to determine which is the most important stimulus to concentrate on. Wickens emphasises that parallel processing does not imply perfect time sharing but that the information processing system is engaged in handling information for two tasks simultaneously and that the resources are not exceeded.

As discussed by Wickens (1999) there are many environments in which operators have to undertake two or more activities concurrently, perceive different stimuli, assimilate information, make responses and process feedback. For example, pilots, nuclear plant operators who may be attempting to diagnose a fault and I.C.U. staff caring for a critically ill patient who is attached to various items of equipment and monitors. In all these examples, if an operator becomes overloaded with complex tasks, then mistakes could ensue and lives put into jeopardy. It is therefore important to attempt to understand the limits of attention. Research has shown that if information is presented in different modalities then generally there is better performance than if tasks are performed in the same modalities.

1.2.8 Workload

In many applied environments the mental effort of operators, how they perform in multi-task situations and whether they are overloaded by the demands of the tasks have been studied under the generic heading of workload. The studies are generally concerned with the interaction of the operator, the task and the type of system or equipment being used.
Over the past decade there has been considerable research into mental workload (e.g. O'Donnell and Eggemeier 1986, Wierwille and Williges 1980) particularly in the aviation industry. However, despite the considerable amount of literature concerning the evaluation and prediction of mental workload there is no consensus on a definition. As Gopher and Donchin (1986) observe, it is a difficult, multidimensional, complex concept to define precisely. Many researchers also agree that workload is multidimensional (e.g. Yeh and Wickens 1988), which includes behavioural, performance, physiological and subjective components (Johannsen 1979). Nevertheless, there appears to be considerable disagreement about what these factors are and how they can be measured (Eisen and Hendy 1987, Linton et al. 1989). Hart (1986) describes it as a general term that evaluates the 'cost' for the human operator accomplishing task requirements. The 'cost' could be reflected in the depletion of attentional or response resources, the inability to complete additional activities, stress, fatigue and performance decrements.

Hart and Wickens (1990) define mental workload as; 'The effort invested by the operator into task performance; workload arises from the interaction between a particular task and the performer.' (page 258).

Experienced operators do learn to handle overload by using various strategies. These may include estimating an answer or a correct response rather than providing a solution, delaying the performance of less important tasks or leaving the performance of the cognitive task until there is sufficient time to perform it. However, by using any of these strategies described above the operator is, in effect, overcoming an inadequate system. If the system requires strategies like this to be undertaken by the operator then it is an indication that the system has not been well designed.

In summary it could be argued that because the concept of workload is multifaceted, developing a general technology to measure it is a significant problem as there are so many aspects to consider; the situation, time scales and time pressure vary from seconds to
minutes to hours. Other factors to consider are the diverse influences that affect the performance of an operator, such as the amount of training they have received for the task or how much practice they have had at a task. Psychological influences include factors such as motivation, acute or long term fatigue caused by acute mental effort. As Jex (1988) argues it is difficult to define motivation or fatigue let alone measure or predict their influence on mental workload.

1.2.9 Comparison With Physical Workload

In comparing physical workload and mental workload Rohmert (1987) states that, like mental workload, physical workload is also measured in terms of human capabilities, in that there are limitations in an operator's ability to undertake certain physical tasks, for example, extremely heavy manual tasks.

Physical workload is relatively easy to measure in the amount of energy expended in kilocalories and oxygen consumption. There are definite boundaries between light physical workload and heavy physical workload. Light physical workload is defined and measured as an expenditure of energy less than 2.5 kilocalories/minute and oxygen consumption under 0.5 litres/minute. Very heavy physical workload is defined as expenditures exceeding 12.5 kilocal/minute and 2.5 litres/minute (Kantowitz and Casper 1988). An operator undertaking sustained levels of high physical workload will suffer fatigue and a performance decrement.

However, workload cannot be measured in exclusively physical terms. Rohmert (1987) defined different aspects that need to be considered when evaluating workload. These include evaluating the following: i) whether the task involves heavy dynamic muscular work or static muscular work, ii) whether the task involves mental effort with the emphasis on concentration, or unvaried work in monotonous surroundings, iii) additional environmental factors such as temperature, ventilation, heat, noise and stress.
In areas such as aviation, physical exhaustion is no longer seen as a major cause of concern, it is more likely that an operator will become mentally overloaded.

1.2.10 Mental Workload

There has been considerable research into the mental effort expended by operators in the military and in aviation. The measures are generally obtained to evaluate the effects of different systems on the human operator, or to quantify the effects of individual differences in abilities or training of operators working with a particular system (Hart and Wickens 1990). It is considered important to assess the various aspects involved in aviation (such as cockpit design, aircrew evaluation and mission effectiveness) because of the implications for safety, crew size and the effects of cockpit automation (Kantowitz and Casper 1988). Mental overload may result in pilot error with the possible culmination of an accident. Design engineers in environments such as aviation have for some time realised that a system, for example, a new tracking device, must be effectively designed in order not to overburden the operator (Linton et al. 1989).

Mental workload evaluation is also assessed in the military establishment and it is considered as being important to establish whether an operator is able to successfully carry out the required mission in battlefield conditions which consist of numerous tasks and where stress and the enemy threat are at a maximum (Linton et al. 1989).

It could be argued that the evaluation of workload should be considered important in the hospital environment, particularly areas such as the I.C.U. First because of the specific stressors in that environment which may affect the staff's ability to deal with a heavy workload even if performance is satisfactory under normal working conditions. Second because of the increasing amount of information technology in the hospital environment, which mirrors the earlier increases in aviation and the military.
Although workload is a complex phenomenon there are a number of practical measures that have been developed. Some of the measures discussed below are used primarily to predict workload early in the design process of a new system prior to its introduction in the working environment. Other measures of workload are often used to examine where performance could be improved in existing systems and include primary task performance, secondary task performance, subjective rating scales and physiological measures. Each of these methods are purported to quantify different aspects of workload experienced by different operators performing a wide variety of tasks in very different environments. The measures are also said to be sensitive, objective, diagnostic and practical (Hart and Wickens 1990). Some of the techniques are described in the following section which are,

1) Task Analysis Methods
2) Primary Task Performance.
3) Secondary (subsidiary) task performance
4) Subjective measures
5) Physiological measures.

1.2.10.1 Task Analytic Measures

Task analytic measures are a commonly used preliminary technique to estimate workload particularly in the military aircraft environment by analysing the tasks undertaken and determining what occurs during a mission scenario. Each general mission requirement is broken down into segments or functions or operator tasks. Each task is then divided into detailed operator requirements. The final section is defined as the operator actions necessary to complete a particular task. Some approaches break down the categories into whether the right hand or left hand will be used during the task while other models (e.g. Aldrich and Szabo 1986) have attempted to incorporate the cognitive components involved in the task. McCracken and Aldrich (1984) developed an influential task analytic model that was not time based and was one of the first models to identify behavioural dimensions. The dimensions which are said to contribute to overall workload demands are visual, auditory,
cognitive and psychomotor. The ratings are assigned values of 1 to 7. McCracken and Aldrich (1984) state that level 7 was, "judged to be the upper boundary of human attention in any single mode" (page 19). However, there is no explanation as to how these weighting scales were derived and also no specific reason given for assuming that a workload rating of 7 would correspond to operator overload.

Hill et al. (1987) stated that task analysis could be used to discover any situations in which the time available is less than the time required to undertake the necessary tasks. During this initial stage it could be hypothesised that the operator is overloaded and will have difficulty performing the required tasks satisfactorily. Another type of task analysis examines individual task components and examines the compatibility and conflict between them (for example, the W/ndex model North and Riley 1989).

1.2.10.2 Primary Task

Primary task measures assess workload by measuring the performance of the task. It is assumed that as workload increases performance will deteriorate (e.g. O'Donnell and Eggemeier 1986). As Wickens (1992) maintains, when a system or an operator is being evaluated, the performance on the system of interest should be examined first. Measuring performance is an essential component of workload analysis, and it is difficult to interpret workload measures without knowing the level of performance the operator was supposed to achieve. There are three types of primary task performance measures,

i) Accuracy (correct responses, control error compared to a target value)

ii) Speed (response time measured in seconds or fractions of seconds)

iii) Number (how many tasks or task components were completed correctly within an interval of time).

As discussed previously, one reason suggested for performance decrements is that the resources necessary to perform tasks are only
available in finite amounts. If sufficient resources are available then performance of the task should be unaffected. As the difficulty of either a single task or concurrent tasks increases, more resources are required to maintain performance and a decrement in performance of the task will occur. Some tasks have been found to interfere with each other more than others.

It is assumed that decrements in primary task performance indicate higher workload, however the actual relationship between workload and performance is more complex. There is an important limitation regarding the measurement of performance of the operator by recording the number of errors made. As Morris and Rouse (1988) comment, mistakes can occur during slack periods as well as during busy periods.

It is argued that measuring performance is not an adequate measure of workload because an operator may use different strategies to maintain primary task performance. For example, an operator may increase the effort required to maintain performance, which subsequently increases the difficulty of the task for that operator, but no errors are recorded. However, this increased difficulty for the operator is not observable. Task difficulty does not only involve the physical, observable composition of the task. Covert tasks, for example planning, problem solving or decision making, may be equally demanding for the operator's resources as the observable activities.

Another strategy that operators may employ when task demands vary over time and operators is to trade off performance and workload against each other. O'Donnell and Eggemeier (1986) state that for relatively easy tasks consistent performance is maintained over a range of difficulty levels; that for moderately difficult tasks performance deteriorates as task demands increase; and for very difficult tasks operators may ignore some tasks and maintain a consistent level of errors on critical or essential tasks, even in the face of increasing demands, so performance decreases but workload is kept constant. When the task demands vary operators may decrease or increase their effort to maintain a constant level of performance (Hart and Wickens 1990). An operator may just
concentrate on the essential, critical tasks and perhaps complete the non-essential tasks when there are lower levels of workload, thereby maintaining the performance in high workload periods.

Casali and Wierwille (1983) similarly suggest that in situations of excessive task demand, important tasks that need to be dealt with immediately are the focus of an operator's attention, while other tasks are subsequently time shared or ignored all together.

A further factor that influences performance is time, if an operator had unlimited time then all tasks could be completed. However, there is usually a time limit involved in completing most tasks and an operator may choose to emphasise speed at the expense of accuracy or vice versa.

Gopher and Donchin (1986) concluded in their study that measurement of primary task performance was a poor indicator of mental workload because it did not reflect the resource investment of the operator. For example, as previously discussed, if a task becomes more difficult the operator may invest more resources to maintain performance. However, as Hart and Wickens (1990) comment although primary task measures should be used with caution when evaluating workload, they must be obtained to determine whether or not the operator can accomplished the task. Otherwise measures such as subjective ratings or physiological measures are difficult to interpret without this information.

1.2.10.3 Secondary Task Measures

During the performance of a task that is within an operator's capabilities, if that task increases in difficulty the operator may then exert more effort to maintain accurate performance of that task. Secondary tasks were developed as an alternative approach as the operator is given an additional task to complete together with the primary task. The operator is instructed to continue with the primary task and to undertake the secondary task, which is purported to utilise the 'reserve capacity' of the operator. Ogden et al. (1979) propose that the secondary task paradigm is highly similar to a dual-
task paradigm the difference being that in the dual-task situation there is no emphasis on any difference in the performance of either task. A secondary task should impose a sufficient additional load so as to exceed the operator's capabilities and the level of performance on the secondary tasks will decrease as the difficulty of the primary task increases.

A critical aspect of the secondary task paradigm is to determine what is an acceptable performance on the secondary task. For example, if performance is acceptable on both the primary and secondary task then it could be suggested that the operator is not overloaded and has sufficient resources to complete both tasks.

There are many types of secondary tasks used in workload assessment. As Hart and Wickens (1990) point out many of the tasks were originally designed to investigate psychological theories such as human performance, memory and attention, while other tasks are designed to reflect the real-world environment. The most commonly used tasks are monitoring, tracking and memory tasks. Other tasks include mental mathematics, time estimation paradigms and simple reaction time. Some of the tasks will be briefly discussed below.

i) Choice reaction time tasks involve the presentation of more than one stimulus that requires a different response for each stimulus, for example, pushing different buttons for the correct response. The stimulus can be presented using the auditory or visual modalities. Bortolussi et al. (1986) found there was an increased in response time and in errors when the primary task's difficulty increased. Choice reaction time tasks are said to impose central processing demands (e.g. Fisk et al. 1982).

ii) Tracking tasks use visual stimulation and usually involve a continuous manual response depending on the type of task. O'Donnell and Eggemeier (1986) state that central processing resources and motor demands are utilised. An example of a tracking task used as a secondary task is demonstrated in a study by Wickens and Liu (1988) together with a spatial decision task and a verbal decision task.
iii) **Monitoring tasks** are characterised by the requirement to detect the occurrence of a stimulus from among other stimuli, sometimes among several alternative stimuli and impose demands on perceptual resources. An example of the use of a monitoring task as a secondary task is demonstrated in a study by Wierwille *et al.* (1985), in which pilots had to maintain a steady course in a mild cross wind while detecting a varying number of emergency and warning lights. Wierwille *et al.* (1985) found that manipulation of the secondary task resulted in reduced performance of the secondary task, while performance of the primary task was relatively stable.

iv) **Memory tasks** can involve either short term or long term specific memory requirements and as O'Donnell and Eggemeier (1986) postulate impose demands on central processing resources. One of the most common memory tasks used as a secondary task is the Sternberg (1969) memory search paradigm. The Sternberg task involves subjects retaining in memory either two (easy task) or five (hard task) letters, which can be presented either visually or verbally, and require a vocal response. The Sternberg task has the potential to permit central processing effects to be discriminated from stimulus encoding and responses and is frequently used in studies of multiple resources theory (e.g. Wickens and Derrick 1981).

v) **Mental arithmetic** can involve various forms of mental mathematics, usually addition tasks, but also subtraction and multiplication and according to O'Donnell and Eggemeier (1986) draws heavily on central processing resources. The difficulty of the task can be manipulated by increasing the number of digits or number of operations that must be completed in a calculation (e.g. Roscoe and Kraus 1971), or by decreasing the time in which the subject must complete the calculations.

vi) **Time estimation tasks** and **interval production tasks** are related. In time estimation subjects are asked to assess retrospectively how much time has passed, the premise being that during conditions of high workload, an interval of time is generally underestimated.
In interval production tasks the subjects produces a series of regular intervals by performing a motor response at a specific rate. Typically, this may involve tapping in intervals of 10 seconds the premise being that the busier an operator is the less attention or resources will be available to judge the time. Hart (1986) postulated that time estimation techniques are an effective measure of operator workload, little instrumentation or training are necessary and the procedure is non-intrusive. Hart and Wickens (1990) state that the task imposes central processing demands and manual response resources.

However, there are issues that arise when implementing a secondary task that need to be considered, particularly in an operational environment. If the primary task requirements are high the operator may abandon the secondary task in order to maintain an acceptable performance on the primary task. Hart and Wickens (1990) state that in some experimental situations the secondary tasks are seen as unimportant or as uninteresting and are subsequently ignored or dropped by the operator or by participants in the laboratory. One solution to this problem is to use embedded tasks which are designed to appear as if they are part of or important to the primary task. It has been demonstrated that this strategy improves the likelihood that the tasks will be performed.

It has been argued (e.g. Gopher and Donchin 1986) that it is important to ensure that when two tasks are combined they do not change their nature, and that the dual task situation does not have different properties from the single task situation. In many dual-task situations many tasks are not viewed as independent, subjects often integrate the tasks to reduce demands. Conversely, it could be suggested that in the real-world environment there are few tasks that occur in isolation, and one of the interesting features of these studies may be what occurs when tasks are combined rather than when tasks are undertaken separately. This has been the main focus of research for example for researchers such as Gopher and Sanders (1984)
Subjective Measures

In an applied setting, asking the opinion of operators about a particular system or the proposed design of a similar system is one of the oldest and most extensively employed ways of predicting and assessing workload. Subjective rating scales are widely used for several reasons. They are practical in the working environment, easy to implement and to score, the financial cost is low, there is usually a ready supply of subject matter experts (SME) and are generally acceptable to the operators being assessed (Williges and Wierwille 1979). If they are administered correctly, they do not necessarily disrupt the primary task (Eggemeier 1980) which can be a problem with other techniques, particularly secondary task measurements. Rating scales are also said to be sensitive to changes in the mental load and to have a certain amount of face validity (e.g. Skipper et al. 1986). However, eliciting individual opinions has many associated problems. For example, individuals may express an opinion that they consider to be socially acceptable rather than admit to a negative opinion. Because of such problems there have been numerous attempts to formalise expert opinion.

Subjective rating scales have considerable support as an important component in evaluating workload (Eggemeier 1980). This is based on the premise that if an operator 'feels loaded and effortful', then he or she is loaded and effortful regardless of what performance measures may otherwise suggest (Johannsen et al. 1979 p 105). Skipper et al. (1986) also argue that an operator undertaking a task is best able to judge the mental demands of that task, and the perceived demand could be rated on a scale representing mental load. This would appear to suggest that in general, an operator's direct perception or estimation of their feelings regarding the effort involved in a task or combination of tasks could provide a sensitive and reliable indication of workload.

There are two methods of obtaining subjective reports, either by rating scales or by questionnaires and interviews. Rating scales generally consist of an ordered sequence of response categories. The
scale categories define the agreement between the stimuli (workload experience) and the responses (rated levels). Questionnaires and interviews depend on written reports and the data obtained is usually more qualitative.

Although the terminology of different rating scales varies generally, the operator is asked to rate how difficult they feel a task has been either during or after that task has been completed. The rank order of ratings are usually stable across raters, although the absolute values of the ratings can show high variability.

The characteristics of the rating scales depend on whether they are ordinal, nominal, ratio or interval. Another consideration is the dimension of the scale. In workload prediction and assessment in the aircraft industry, where several scales have been developed the scales used are unidimensional or multidimensional, and some are presented as hierarchical decision tree formats. The method of choice depends on what the scale is intending to measure.

Unidimensional scales used in evaluating workload measure only one aspect or attribute and often use numeric values such as 1-7 or 1-100. Descriptive words are also used such as low, moderate etc. One of the problems with a unidimensional scale is that results could suggest that there is only one single attribute of workload that can be rated and identified.

An example of a hierarchical scale is the Modified Cooper-Harper scale (Wierwille et al. 1986) which allows pilots to record their impressions of workload (or the difficulty of the task). However, although the decision tree format allows the evaluation process to be separated into a series of explicit decisions, is easy to implement, and can be scored by the operator without causing interference to the primary task, Hart and Wickens (1990) argue that the scales have not received extensive evaluation.

Multidimensional scale measures are purported to measure more than one dimension concurrently. Commonly-used multidimensional scales include the NASA Task Load Index (Hart and Staveland 1987)
which has six dimensions, while the Subjective Workload Technique (SWAT, Reid and Nygren, 1988) uses three dimensions. Multidimensional scales have an advantage over the unidimensional scales in that by separating the scales, qualitatively different effects (e.g., stress vs. cognitive load) can be identified.

The advantages of subjective rating scales are that they can provide valuable information concerning the operator's perception of their workload experience in specific tasks or activities.

In general, rating scales have been developed in the aviation environment to measure pilot workload. In many instances the scale is very specific to pilot activities, and would require modification and validation before being used in other environments. In some instances individual assessment procedures have been developed for a specific application and therefore have not been subject to the validation which would be required to recommend their general application.

As Hart and Wickens (1990) suggest there often appears to be no direct relationship between the values on rating scales and actual measurable events, and the intervals between the items on the scales may not be equal. Thus most scales provide ordinal information and index relative, rather than absolute, measurements. Williges and Wierwille (1979) also argue that there is relatively little evidence that some of the rating scales are based on psychometric theory.

Ratings can be obtained during the performance of a task or after a task has been completed. If the former method is used there may be interference with performance of the primary task, while if they are completed after the operator has finished a task, important information may have been forgotten. Rating scales only record the rater's memory of what was generally experienced across time and are therefore insensitive to variations over time and are subject to rater biases (Hart and Wickens, 1990).
1.2.10.5 Physiological Measures

Physiological measures include variables such as changes in the heart rate, respiration, blood pressure, analysis of body fluids, galvanic skin responses, electrical activity of the brain (event related potentials, ERP) which can be measured using electrodes placed on the scalp, pupil diameter, eye movement and eye blinks. As Wilson and O'Donnell (1988) state it would appear intuitive that undertaking a task would require an expenditure of physiological effort. The physiological information could be obtained unobtrusively, is sensitive to variations over time and indicates objective, physical and emotional responses that occur with increases in workload, while some measures accurately reflect cognitive load. The predominant problem with physiological measures is that there has been only limited success in correlating the data with other measures of workload.

Different physiological measures have been shown to be sensitive to different types of workload. For example, Wierwille and Connor (1983) demonstrated that heart rate is sensitive to stress and physical effort, while measurement of event related potentials reflects mental effort. It has been demonstrated in laboratory conditions that pupil dilatation and changes in heart rate are sensitive to the cognitive demands of certain tasks (e.g. Mulder 1986).

While some measures are relatively easy to obtain in a laboratory environment, they are more difficult to record accurately in a field environment. An example of this difficulty is shown in the physiological measurement of pupil diameter. Pupil diameter has been shown to be an accurate indicator of increased mental load, as demonstrated by Beatty (1982). For example, the diameter of the pupil increases as the mental effort expended increases in tasks such as sentence comprehension and visual tasks involving comparison of letter pairs. However, in the field environment pupil diameter is difficult to assess because of factors such as overhead fluorescent lighting and in general the technique of measuring pupil diameter is limited to laboratory studies.
1.2.11 Comparison and Dissociation Between Measures of Workload

O'Donnell and Eggemeier (1986) state that techniques from each of the above categories have been used with a variety of success. Overall, the literature regarding research into workload appears to suggest that specific combinations of measures appear to provide an accurate indication of workload. Other factors that need to be considered when comparing the techniques are the validity and reliability of each technique together with the sensitivity and intrusiveness of a technique. The sensitivity refers to the capability of a technique to reflect differences in the levels of processing capacity expenditure that are associated with performance of a task or the combination of tasks. Intrusiveness refers to the technique's interference and subsequent degradation of the primary task.

For complex tasks interpretation of performance is complicated when many measures are available; if different measures are taken they can suggest very different levels of workload, for example one measure may suggest high workload while another measure will shown a low workload rating. As the units of measurement, frequency of occurrence and priority vary across task components, it is not easy to provide a single figure to as a performance measure.

A comparison of the available techniques described in the previous sections was conducted by Casali and Wierwille (1983) in which pilots were given different communication tasks that varied in difficulty. The study compared sixteen different workload estimation techniques to establish the sensitivity and the intrusiveness of the technique. The techniques used included i) opinion, (rating scale) measures, ii) spare mental capacity (secondary task) measures, iii) physiological measures and iv) primary task measures.

The rating scales examined included the modified Cooper-Harper Scale (Wierwille and Casali 1983), the ratings obtained immediately after each flight simulation. The secondary tasks used were a time estimation task and a tapping rhythm task. The physiological measures included respiration rate, heart rate, eye fixations and
eyeblinks. The primary task involved control movements of the pilot, errors of omission and commission regarding the individual call sign for their aircraft, and communication response time.

The study concluded that some measures were more sensitive to an increase in the difficulty of the communication task. The modified Cooper-Harper Scale demonstrated a perceived increase in the mental effort expended, the ratings significantly increasing between high and low load, but not between medium and high load. In the secondary task measures the tapping rhythm measure was not sensitive to changes in the communication load although the time estimation technique did reflect the increase in the primary communication task. Amongst the physiological measures only the pupil diameter reflected changes in task difficulty although Casali and Wierwille suggest that this result should be treated with caution as they did not use the most sophisticated equipment available. The primary task measures of the pilot's control of the aircraft were not degraded by the introduction of the communication task, but errors of omission and commission increased with task difficulty and the response time increased. Casali and Wierwille suggest that often in the aircraft there are long periods of silence between communications and the pilots are not as alert to responding and answering these as they are during conditions when they are presented almost continuously. The rating scales were shown to be effective for the highly trained population of pilots, and the secondary task of time estimation was demonstrated as being sensitive to the increase in the communication load, however the secondary task would not be as easy to implement in actual flight conditions as the rating scale and the physiological measures were not generally responsive to the increase in the task.

Other studies also report inconsistent results obtained from the different measures. For example, Wierwille et al. (1985) also examined 16 measures of mental workload and recorded divergent sensitivity to increases in task difficulty. The measures included a similar variety of primary and secondary task measures, subjective opinion measures and physiological measures as described in the
1983 study by Casali and Wierwille. In this study only one measure, (one of the of primary task measures, meditational reaction time for correctly answering questions) reflected the change in mental load.

Cox et al. (1983) used subjective behavioural and physiological measures to investigate the performance of military bomb disposal experts under stress, half of whom had been decorated for gallantry. The subjects were required to discriminate auditory signals and received an electric shock for incorrect answers. The subjective assessments of both groups did not differ before, during or after stress. However the physiological measure of heart rate differed between the two groups. The decorated officers had lower heart rate levels than the non-decorated officers. In this example of the dissimilarity between measures, physiological measures were shown to be more sensitive than the objective measure.

Yeh and Wickens (1988) have proposed a theory of dissociation (the inconsistent evaluations) between the various measures of mental effort based on the multiple resources model. There are many factors that may affect similarity between measures, for example, if two visual tasks are being undertaken there is a physical barrier in that the eyes cannot look in two directions at the same time.

Yeh and Wickens (1988) postulate that time-sharing between tasks imposes a heavy demand on working memory, depending on the amount of information that is held in working memory. Subjective perception can reflect the demand on working memory and other factors such as the amount of stress the individual perceives. Yeh and Wickens suggest that there are four sources of dissociation between measures, which are, i) when there is either overload or underload ii) increased resource investment, iii) time-sharing requirements iv) competition for common resources.

1.9 Summary of Workload Measures

Workload could be described as a hypothetical construct (MacCorquadale and Meehl 1948) which involves "terms that are not wholly reducible to empirical terms; they refer to processes or
entities that are not directly observable although in principle they need not be unobservable' (page 104).

By defining workload in terms of the limits of the operator's capacity this demonstrates a close association with research concerned with sustained attention and dual-tasking research. The areas of research are concerned with examining the limitations of individuals when performing two or more tasks, even though there is often the apparent ability to perform both tasks well.

As the concept of workload is defined as a multi-dimensional, multifaceted concept that is difficult to define, it is unlikely that one dimension will be of any use in obtaining a representative meaningful measurement. As Gopher and Donchin (1986) observe, often the measures that are obtained are well suited to the needs of the design engineers for a particular system, for example in the aircraft industry.

Mental workload varies not only because of differences in the system used, but also with which particular operator is using that system. As Linton et al. (1989) state, workload reflects relative, rather than absolute, individual states which depends on both external demands and the internal capabilities of an individual. Individuals differ qualitatively and quantitatively in their response to workload. There are many different kinds of task demands and corresponding internal capabilities and capacities to handle these demands. Individuals differ in the amount of capabilities which they possess and their strategies for employing them.

Workload and its relationship to operator performance is dependent on the strategies used by the operator, and, as discussed previously, the level of motivation, fatigue, and the time of day or night, environmental features and the design of the equipment. Siegel and Wolf (1969) also argue that stress has been shown to affect problem solving abilities, for example under mild stress the time taken to resolve a problem is sometimes reduced, but there is a limit to the ability to solve a problem faster and once that limit is reached the ability of the operator to resolve a problem deteriorates.
One of the criticisms of current workload prediction techniques is that many of the experimenters do not use well established techniques of measuring subjective workload and instead use an arbitrary method which is neither valid nor reliable (Markou 1991).

It would seem that a systematic, empirical evaluation of operator workload in areas such as the I.C.U. and O.R. should be undertaken with as much enthusiasm as it has been in the aircraft industry due to the rapid increase in the use of integrated monitor displays and intelligent alarm systems. These systems were supposedly developed to reduce the work of the staff and in order to see whether in fact this is the case, workload must be empirically measured. If workload is increased by additional tasks imposed by the increase in the number of monitors the operator's ability to respond in an emergency might be impaired because of the use of the 'reserve capacity'. As in aviation both excessively high or low workload may contribute to accidents occurring (Kantowitz and Casper 1988). Hart and Wickens (1990) similarly state that decrements in performance can occur if workload is too high or too low. Block (1986) postulated that if anaesthetists and nurses have too few monitors to look at they may become bored and their vigilance decreased. Block is assuming that monitoring equipment automatically decreases workload, whereas there is no evidence to support this assumption and he offers no empirical evidence to support this claim. Monitoring equipment alone does not constitute workload, nurses and anaesthetists have many other tasks to attend to other than solely looking at the monitors.

Since the 1970s there has been a great deal of research regarding workload but there appears to be a considerable divide between the theoretical components and the applied interest (Brookings and Damos 1991). Studies that look at the workload of nurses and pilots are typically atheoretical and are concerned with an immediate problem (usually of a financial nature, such as staffing levels).
1.2.12 Workload in the Hospital Environment

McCarthy (1985) developed a subjective workload scale to examine the impact of new technology on hospital staff although there is no indication as to whether the scale was validated. Nevertheless, the scale provides a starting point in assessing the stress and demands made upon the nursing staff in I.C.U. by the increase in monitoring equipment. As Hart and Wickens (1990) argue, face validity of a rating scale alone is not sufficient, actual validity needs to be measured because the degree to which subjective measures can account for variance in the performance remains unknown. There is a need for empirical evidence in the medical environment to define workload patterns in different situations in an attempt to identify the part that workload contributes when, for example, a critical incident occurs.

A study by Gaba and Lee (1990) is one of the very few studies that have empirically attempted to measure workload of anaesthetists, using a secondary task involving a mathematical problem together with a retrospective, subjective assessment of the workload. The results showed that excess response time (above baseline) correlated with activities like 'manual task', such as preparing the patient for anaesthetic. Subjective ratings confirmed that induction of a patient prior to anaesthesia and emergence from anaesthesia were the highest periods of workload, with attention described as the highest cognitive requirement.

Gaba and Lee (1990) argued that primary task measurement is not feasible in anaesthesia because there is no objective measure of performance for the task of administering anaesthesia. However, as Wickens (1992) argues, if a system or an operator is being evaluated then the operator's performance should be examined prior to other measures being undertaken. Physiological measures, such as measuring heart rate and heart rate variability would be feasible during anaesthesia using standard ambulatory ECG recording, but measuring evoked potentials and pupil diameter would be difficult on a anaesthetist who is constantly moving in an environment with substantial electromagnetic interference. Gaba and Lee state that
the secondary task paradigm offers a combination of objectivity, practicality and unobtrusiveness that makes it suitable for the O.R. environment. They concede that the secondary task paradigm selected (mental arithmetic) was not completely successful as when the anaesthetist was busy with a manual task the secondary task could not be completed. The authors further suggest that in future studies a vocal response could be used, but this too could create problems, as one of the tasks undertaken by the anaesthetists is instructing colleagues.

In concluding the section concerning workload it would appear that workload is best defined as a multi-faceted, multidimensional construct and it is possible that 'workload' cannot be defined by one unique representative measure. An individual brings to a task a wide range of different motivations and abilities which may be due to practice or training and some aspects of the task may have become automated. There are many different methods of measuring workload and often there is dissociation between the measures. Both the studies described above that have attempted to measure workload in the hospital environment (Gaba and Lee 1990, McCarthy 1985) have done so without, in the first instance, attempting to define the primary task, which, as Hart and Wickens (1990) state, is essential in order to obtain a meaningful measure of workload.

1.2.13 Rationale for Current Research

As Eggemeier (1988) argues in areas where there are large amounts of technology and complex systems there is an urgent need to assess the load imposed on the processing capacities of the operator.

As Casali and Wierwille (1983) discuss, pilots often deal with the tasks that require immediate attention while other tasks are either time-shared or ignored altogether. If a pilot is cognitively overloaded significant pilot errors could occur which then result in an accident.

There has been considerable research into pilot workload in an attempt towards easing the load of the pilot and to detect tasks that
CHAPTER 2: A Series of Experiments Examining Auditory Confusion Between Hospital Alarms

2.1 Introduction

Concern has been expressed for many years (e.g. Kerr and Hayes 1983, Samuel 1986) over the many alarms on the I.C.U. and other critical care areas in hospitals. One of the problems associated with a large number of auditory warnings is the fact that confusions could occur. An example is cited in a study conducted by Cooper and Couvillon (1983) in which they reported on breathing system disconnections. A physician, discussing the reasons for a fatality, stated that the ventilator alarm had been confused with the electrocardiogram (E.C.G.) alarm. In the same report many of the staff expressed the view that several ventilator alarms sound like other alarms in the I.C.U.

Seriously ill patients in I.C.U. will be attached to many monitors and pieces of equipment, the sophisticated technology used to monitor the patients condition and to inform the staff of any changes in that condition. For example, critically ill patients could be attached to a ventilator, a humidifier, and a pulse oxymeter (to measure the oxygen content of the patients blood). It would not be unusual for these patients to have two or more syringe pumps controlling the rate of drug administration and several infusion pumps which regulate drugs and control fluid input, an electrocardiogram (E.C.G.) monitor which displays the heart rate and a dynamap to measure blood pressure.

Each piece of equipment has its own alarm and in some cases one piece of equipment may have different sounds depending on the type of problem. For example, a particular make of pulse oxymeter has 'rev-up' noises, calibration noises and a different noise again should the oxygen saturation of the patient drop below a certain pre-determined figure. It is therefore possible for a patient to have twenty or so alarms emanating from the various pieces of equipment they are attached to. If there are two or three critically ill patients on the unit then the number of alarm sounds is consequently increased. As technology develops and more sophisticated monitors
identification of a set of auditory warnings currently in use in the I.C.U. Derriford Hospital, Plymouth. Thus the first set of experiments looks simply at the learning and retention of a set of hospital alarms without any further context or secondary tasks. However, as discussed by Kerr (1985), the many problems regarding auditory warnings cannot be addressed in isolation and it is therefore necessary to establish what kind of activities are undertaken by the staff in the I.C.U. environment when alarms occur. To this end, two observational studies were undertaken, one using a video camera and the other being a direct observation of the staff in order to identify the kinds of tasks that are being undertaken when auditory alarms are activated.

In a second set of experiments, participants were again required to learn and retain a set of auditory warnings, but representations of the tasks were introduced into these studies in order to examine whether there was an effect on the primary task, the identification of the auditory sounds.
interfere with each other, there has been almost non-existent research in the hospital environment, particularly areas where technology is increasing rapidly such as the I.C.U. With the increase in the number of monitors there is a concomitant increase in the number of auditory warnings.

Stanton and Booth (1990) suggest that the psychology of alarms can be split into three main areas: specification, detection and diagnosis. Specification of the most appropriate system of alarms for a specific environment should occur when an analysis of the tasks to be undertaken by the operator has been performed. This would then provide information regarding situations that require attention and alarms could then be presented in a way that supports the operator's tasks so that problems can be detected and diagnosed efficiently.

There has been considerable investigation into the actual warning sounds that equipment should produce so that the urgency of the situation is matched by the urgency of the alarm. Although these psychoacoustic design principles were originally for use in the aircraft industry they are purported to be applicable in any high workload environment (Edworthy et al. 1991). There is some evidence to show that the manipulation of perceived urgency is a design improvement. For example, recent studies have shown that increased psychoacoustic urgency of alarms decreases reaction time (Haas and Casali, in press). In attempting to understand behavioural responses to auditory warnings in the hospital environment and in particular high stress areas such as I.C.U. there are many variables that need to be taken into account, some of which have been discussed above.

In many of the measures of workload described above it seems to be important to firstly look at the system of interest. As Wickens (1992) states an essential component in assessing the mental load on an operator is to examine the performance of operators on the primary task.

One of the aims of the current research is to investigate the psychological dimension of confusion between alarm sounds. The experiments investigate, amongst other issues, the correct
and equipment are produced, each one having some type of auditory warning, the numbers of sounds in critical care areas will increase.

The problem is further complicated, as previously mentioned in Chapter 1, by the fact that there are currently no standards agreed between manufacturers on auditory warnings for medical equipment. Therefore, the same piece of equipment manufactured by different companies may have different alarms. Individual items of equipment are obtained at various times with the emphasis on function and cost and not on compatibility with existing equipment (Wiklund and Hoffman 1988). Conversely, equipment for different functions, made by the same manufacturer, may possess similar alarms even though their functions are quite different. For staff working on the Unit the large number of alarms is a potential source of confusion. This confusion may arise not only because of the number of alarms but also because many of the alarms sound acoustically similar. For example, some of the warnings are continuous high pitched tones. As pitch judgement tends to be a relative, rather than an absolute, judgement and information is lost very quickly about the absolute values of pitches (Deutsch 1978), then the vast majority of hospital staff who do not possess absolute pitch are at a disadvantage in discriminating warnings of this sort.

Another reason why auditory warnings in the I.C.U. could be confused is because alarms are often inappropriate in terms of their 'urgency mapping' (Momtahan and Tansley 1989). There is generally no relationship between the urgency of a medical situation and the perceived urgency of the alarm which signals that condition (Edworthy 1994a). An example of this is the food pump alarm, which is in use on the I.C.U. at Derriford Hospital, Plymouth: this piece of equipment has an extremely loud, urgent sounding alarm. In contrast the ventilators have quieter, less urgent sounding alarms. However, in terms of importance the ventilator is more essential in maintaining the life support of the patient than the food pump. While experienced staff may have learned which sound is more important, new staff may not and therefore could assume that the food pump, because it sounds so urgent, is signalling an urgent problem that needs attention immediately.
There are other environments similar to the I.C.U. where many auditory warnings are used, for example on the flight deck of an aircraft or in a nuclear power plant. Patterson (1982) reported that flight crews of commercial aircraft complained that the large numbers of warnings in the cockpit are confusing. DuRoss (1978) demonstrated that pilots of commercial aircraft made errors when asked to identify tape recordings of the warnings from their own flight decks. In a similar study Momtahan et al. (1993) assessed the identification of auditory warnings in use in an I.C.U. and Operating Theatre of a Canadian Hospital. The staff involved in the monitoring of the alarms in these areas were asked to individually identify the alarms. The results were similar to the DuRoss study, staff could not recognise all the alarms and made errors distinguishing them. (Correct identification was 39% by I.C.U. nurses and 40% by anaesthetists and O.R. technicians).

Although the error rate by the pilots was not high, it was considered by DuRoss (1978) important to determine the severity of the problem and the sources of the confusions. This could be because of the possibility of a fatality occurring, such as that reported in the Cooper and Couvillon (1983) study. The question was asked by Patterson and Milroy (1980) whether it was intrinsically difficult to learn and remember auditory warnings, or whether the confusions were due for example to acoustic similarity amongst the warnings.

Webster et al. (1973) showed that the confusion rate for listeners trying to learn a sound using a paired comparison technique was inversely related to the number of dimensions on which that sound varied (fundamental frequency; wave shape; number of formants and formant frequency). In other words, the more dimensions along which sounds varied the easier it was to differentiate between them in a paired-comparison task.

The investigation of confusions between auditory warnings would seem to be a laboratory-based task from which principles about confusion could then be generalised to the working environment. However, caution is needed here if such generalisations are to be made. For example, in the study by Patterson and Milroy (1980) the participants were required to learn the set of aircraft warnings in
both a cumulative and a paired-associate learning paradigm and the results showed that the learning rate slowed significantly after five or six warnings had been committed to memory.

However, experimentation on the learning and retention of warnings may be compromised by the techniques of experimental psychology themselves. As Stellmack and Dye (1993) state, in everyday noise environments there are many cues that are simultaneously present and therefore contribute to the identification of a sound, and under laboratory conditions these cues can be isolated or removed to investigate their effect on identification. Memory for warnings may therefore be better than experiments suggest, but the methods of experimental psychology dictate that information which may be of use in the real environment is eliminated in the laboratory.

There are many examples in the acoustic literature of very specific sounds being used in laboratory studies, one such example being a study conducted by Weber (1989) who investigated the detection/recognition theorem using complex auditory stimuli. The detection/recognition paradigm predicts that an observer would be able to correctly identify an individual sound that was presented from a group of sounds. This assumption derives from the presupposition that each individual signal can be detected by an independent process and that all possible signals are equally detectable (Green et al. 1977). The independent process however assumes that all possible signals are sufficiently separated along one or more dimensions (e.g. time, frequency etc.), so that there is no influence of one sound upon any other sound. The stimuli described by Weber (1989) used four tonal patterns that varied over the nine conditions of the experiment. For example, the sinusoids within a pattern were inside of a half-octave interval. The results of the study showed that when the sounds were similar (e.g. when identification depended on identifying the four patterns that were composed of the same seven frequencies presented in different orders) participants were unable to identify them as easily as when there were differences between the sounds (e.g. when the frequencies were within separate and discrete frequency regions.) These results would suggest that even in very tightly controlled
sounds used in the laboratory environment, identification is relatively easier with specific cues such as frequency.

Generally, therefore, in laboratory based studies, the stimuli would all be equalised for loudness, length and so on in order to isolate specific acoustic qualities which could cause confusion. Whilst it is the convention to control for such features in the name of correct experimental psychology, it is important to consider how individuals might discriminate between auditory warnings in practice, and to allow these cues into the laboratory. For example, it is possible that the most important cues used by people in differentiating between warnings might be differences in loudness and length, with more subtle acoustic cues perhaps being only of secondary importance.

An experiment by Patterson and Milroy (1980) examined how quickly participants could learn and identify a set of ten aircraft warnings. The results showed that participants found it easy to identify five or six warnings, however after that, the rate of acquisition decreased, with the learning of subsequent sounds becoming more difficult. The participants, after initially learning the sounds during the first session, returned one week later and were asked to re-identify the same set of sounds. The results showed that during the return session the participants achieved almost perfect performance. After only a few trials all the participants had learnt all ten warnings and once learnt, they were remembered. The study also identified consistent confusions and non-confusions and concluded that a similarity of temporal characteristics, in particular repetition rate, can cause confusion.

Three experiments are reported in this chapter which investigate the contrast between warnings as they are typically heard in the working environment, and as they would be heard in the experimental laboratory-based setting. A third experiment was conducted in which the sounds of the warnings were changed to ensure that the confusions were acoustic and not semantic. The hypotheses being that the ‘natural’ sounds would be easier to learn than the
'controlled' sounds and that any confusions occurring would remain constant in the 'neutral' names experiment.

2.2 Method

2.2.1 Participants

All participants were first year undergraduate students from the Psychology Department, University of Plymouth. Twelve participants undertook Experiment 1 (11 females and one male with a mean age of 21.6 years, ranging from 18 to 39 years). Twelve participants undertook Experiment 2 (12 females with a mean age of 23.7 ranging from 18 to 35 years). Twelve participants undertook Experiment 3 (11 females and one male with a mean age of 24.2 ranging from 18 to 33 years). No participant undertook more than one experiment.

Prior to the start of the experiment the participants were asked whether they had any problems with their hearing. No replies in the positive were received. The participants' hearing was not formally tested.

2.2.2 Apparatus/stimuli

A total of 24 auditory warnings were recorded from warnings currently in use in the Intensive Care Unit at Derriford Hospital, Plymouth. A Sony 'Datman' (Digital Audio Tape) with a microphone was used to make the recordings. The microphone was always held at the same distance (1 metre) from the alarm-emitting device. The majority of warnings were recorded in an empty side ward on the I.C.U.. Where possible, individual items of equipment were taken into the room and the warnings recorded for an interval of 10 seconds in the form that they would normally be heard. The name of the warning was then spoken into the microphone by the experimenter. When it was not possible to take items of equipment into the side-room, for example the Drager ventilator which was in long term use, the warnings were recorded in situ. Several of the sounds could not be used because of excessive background noise and were subsequently not used in the study reported in this chapter. A subset of twelve warnings was selected from the original 24 warnings,
the criteria for selection being that there was no background noise, and no long gaps within the sounds during their normal cycle. These sounds were then transferred on to an Archimedes A4000 computer using the sound sampling program, 'Armadeus'.

Two versions of the set of twelve sounds were produced. One set, used in Experiment 1, was retained in exactly the same form in which they had been recorded from the hospital ward. Consequently in this set some warnings were louder than others although they had all been recorded from the same distance, some were longer, and some repeated whereas other did not. This set of sounds is referred to as the 'natural' warnings. In cases where the warning, once activated, did not automatically switch off at some later point, the progression of the warning was curtailed after ten seconds.

The second set of stimuli were standardised so that they were all of the same length and the same loudness in a way more typical of experimental psychology methods. In order to achieve this, warning loudness was boosted or reduced so that it was in the middle loudness range of the 'natural' sounds. If a sound cycle was repeated only one cycle was used for the set of stimuli, in other words, there were no gaps within the sounds. This set of warnings is referred to as the 'controlled' warnings.

The functions of the equipment are briefly described in Table 2.1 and the warnings themselves are described in Table 2.2 and Table 2.3. Table 2.2 shows the equipment with which the warnings are normally associated and the names used for these warnings in the three experiments. Table 2.3 gives a brief acoustic description of each warning.

The names given to the warnings were also different depending upon the experiment. In Experiments 1 and 2, the names used were those associated with the warnings in their normal usage, in a few cases these names were changed if the terms were obscure or duplicated. Thus, in these two experiments the names were the same, but the stimuli differed between the experiments. In Experiment 3, as Table 2.2 shows, the names of the stimuli were altered so that they possessed names without any medical or other connotations. This was to ensure that any confusions that occurred were acoustic and
not semantic. The 'police alphabet' system was used in Experiment 3, and these sounds are referred to as 'neutral names'. Therefore in Experiment 1 and 2 the names were the same and the sounds different, while in Experiments 2 and 3 the sounds were the same and the names different.
<table>
<thead>
<tr>
<th>Equipment</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ventilator</td>
<td>Life support machine breathes for patient alarm occurs if pressure drops below a pre-determined limit, if the airway tube connected to the patient becomes detached or obstructed</td>
</tr>
<tr>
<td>Pulse Oxymeter</td>
<td>External non-invasive clip attached to a patients finger to determine oxygen perfusion to the peripheries the rationale being that if the peripheries are well perfused then vital organs, such as the brain and heart, will also be well perfused. Alarms if the oxygen saturation of the patient drops or if the clip falls of either the patients finger or toe.</td>
</tr>
<tr>
<td>Syringe Pump</td>
<td>Administers drugs to the patient via an intravenous line. Alarms if the syringe becomes empty, or if the tubing is obstructed.</td>
</tr>
<tr>
<td>Infusion Pump</td>
<td>Administers fluid to the patient via an intravenous or a central line also allows staff to administer intravenous drugs. Alarms if tubing is obstructed or unplugged from the mains</td>
</tr>
<tr>
<td>E.C.G.</td>
<td>Connected to the patients chest to record heart rate and rhythm, pulse rate and blood pressure. Alarms when heart rate or blood pressure falls below a pre-determined level, can also alarm if the patient has severe cardiac abnormalities, such as arrhythmia or due to artefact such as the patient moving.</td>
</tr>
<tr>
<td>Humidifier</td>
<td>Container of water which is attached to the ventilator to ensure the oxygen administered to the patient is humidified alarms when empty</td>
</tr>
</tbody>
</table>

Table 2.1: Description of Equipment
<table>
<thead>
<tr>
<th>Sounds</th>
<th>Experiments 1 &amp; 2 Names</th>
<th>Experiment 3 Names</th>
<th>Equipment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Aircall</td>
<td>Alpha</td>
<td>Erica ventilator</td>
</tr>
<tr>
<td>2</td>
<td>Bleep</td>
<td>Bravo</td>
<td>Doctor's bleep</td>
</tr>
<tr>
<td>3</td>
<td>Heart rate</td>
<td>Delta</td>
<td>Electrocardiogram</td>
</tr>
<tr>
<td>4</td>
<td>Fire</td>
<td>Echo</td>
<td>Doctor's 'crash' bleep</td>
</tr>
<tr>
<td>5</td>
<td>Humidifier</td>
<td>Foxtrot</td>
<td>Fisher - Pago humidifier</td>
</tr>
<tr>
<td>6</td>
<td>Infusion</td>
<td>Golf</td>
<td>3M 200 infusion pump</td>
</tr>
<tr>
<td>7</td>
<td>Oxymeter</td>
<td>Hotel</td>
<td>Simed pulse oxymeter</td>
</tr>
<tr>
<td>8</td>
<td>Perfusion</td>
<td>Kilo</td>
<td>Brompton ventilator (2)</td>
</tr>
<tr>
<td>9</td>
<td>Pulse</td>
<td>Lima</td>
<td>Pulse oxymeter 505</td>
</tr>
<tr>
<td>10</td>
<td>Syringe</td>
<td>Sierra</td>
<td>Syringe driver</td>
</tr>
<tr>
<td>11</td>
<td>Ventilator</td>
<td>Tango</td>
<td>Brompton ventilator (1)</td>
</tr>
<tr>
<td>12</td>
<td>Respiratory</td>
<td>Uniform</td>
<td>Vickers syringe pump</td>
</tr>
</tbody>
</table>

Table 2.2: Stimuli
<table>
<thead>
<tr>
<th>Sound</th>
<th>Description of sound:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Two tone pattern; first tone = 0.5 secs at 500 Hz, second tone = 1.0 secs at 250 Hz (including a short gap of 0.2 secs)</td>
</tr>
<tr>
<td>2</td>
<td>Total sound lasts for 5 seconds; tone 0.7 secs at 280Hz followed by a repeating two-tone pattern at 2600Hz and 2000Hz. Sound then repeats after a gap of 7 secs</td>
</tr>
<tr>
<td>3</td>
<td>1000Hz continuous tone</td>
</tr>
<tr>
<td>4</td>
<td>Tone of 1 second at 260Hz, followed by burst of short tones each lasting 0.25 secs at 2600Hz</td>
</tr>
<tr>
<td>5</td>
<td>Repeated 4 pulse pattern =0.6 secs; first two pulses at 1000Hz; second two pulses at 930Hz; each individual pulse = 0.15 secs</td>
</tr>
<tr>
<td>6</td>
<td>&quot;Beethoven's 5th&quot;; 3 pulses at 280 Hz + 1 at 220 Hz, each pulse = 0.15 secs; total length =0.6 secs + 0.6 sec gap</td>
</tr>
<tr>
<td>7</td>
<td>Two pulses lasting 0.5 secs at 260Hz and 275Hz, with a pause between pulses of the same time interval (0.5 secs)</td>
</tr>
<tr>
<td>8</td>
<td>Repeated tone lasting 0.5 secs at 1760Hz; 1 second pause between tones</td>
</tr>
<tr>
<td>9</td>
<td>Two tone repeating pattern; first tone = 0.25 secs at 400Hz; second tone = 0.25 secs at 310Hz</td>
</tr>
<tr>
<td>10</td>
<td>Complex pattern of long tones (1 second) and very short pulses (100ms) all at 3000Hz</td>
</tr>
<tr>
<td>11</td>
<td>Repeated tone lasting 0.35 secs at 2200Hz; pause of 0.35 secs between each tone</td>
</tr>
<tr>
<td>12</td>
<td>Continuous tone at 2350Hz with strident higher harmonics</td>
</tr>
</tbody>
</table>

Table 2.3: Description of Stimuli
2.2.3 Procedure

In order to elicit a clearer picture of auditory confusion, three experiments were conducted in which the stimuli and procedures were almost identical except for small, but significant, differences. Three groups (12 participants in each group) learned either the 'natural' sounds, the 'controlled' sounds or the 'neutral-named' sounds in a paired-associate learning paradigm. It was explained to the participants that the experiment was concerned with the learning and retention of auditory warnings.

The stimuli were then tested in one of the three experimental conditions, for which the procedure was identical. In each experiment there were two experimental sessions. In the first session, the learning phase, the set of twelve sounds were learnt in a paired-associate learning paradigm. Each of the twelve sounds was presented to the subject via a sound sampling program on the Archimedes A4000 microcomputer. The name of each warning was told to the subject by the experimenter and the name was also printed on a card which was placed in front of the participant on the table when the corresponding sound was played. When all twelve warnings had been presented, the participants then heard each sound in a random order and were required to name the sound. If named incorrectly, correct feedback was given immediately. The presentation of the sounds in random order was repeated twelve times, so participants were required to name each sound on twelve occasions.

In the second session, the return phase, the participants were asked to return after one week and the recognition and retention of the warnings was tested, using the same 12 sounds. The participants were reminded of the 12 sounds in a similar manner to the first session. Each of the twelve sounds was presented to the participant via the Archimedes computer. The name of each warning was again told to the participant by the experimenter and the card with the name printed on was placed in front of the participant on the table when the corresponding sound was played. The twelve sounds were
then presented in a random order in twelve blocks, until all sounds were identified correctly.

2.2.4 Results

2.2.4.1 Introduction

Three sets of data were collected for each participant. These were;

i) The number of trials taken to identify the sounds correctly in the learning phase and the return phase;

ii) The total number of times a sound was identified correctly;

iii) The response given when a sound was identified incorrectly.

The results are presented in the following order. First the number of trials, second the correct responses to the sounds, and finally the significant confusions.

2.2.4.2 Number of Trials

In order to provide a gross measure of the relative merits of presenting warnings in the 'natural' state versus their 'controlled' condition together with comparing the effects when the different neutral names were used the number of trials taken for each participant to correctly identify all twelve sounds was recorded.

The mean number of trials to correctly identify all sounds for each experiment (learning and return phases combined), for each phase (learning and return) and for each experiment are shown in Tables 2.4 to 2.6.

The mean number of trials to correctly identify the sounds for each experiment (Table 2.4) showed that participants took fewer trials to recognise all the sounds in Experiment 1 (natural sounds) than in both Experiment 2 (controlled sounds) and Experiment 3 (neutral names). The participants in Experiment 3 required more trials to correctly identify all the sounds.

The mean number of trials taken to reach criterion for each phase (learning and return) is shown in Table 2.5. The number of trials in
the return phase is less than the number of trials in the learning phase for all three experiments.

The mean number of trials shown in Table 2.6 for experiment and phase demonstrates that each experiment appeared to have followed a similar pattern, that fewer trials were needed to reach criteria in the return phase of the experiments than were needed in the learning phase. Participants in Experiment 1 in both phases needed fewer trials to reach criterion than Experiments 2 and 3, while both phases in Experiment 3 required more trials than Experiment 1 and Experiment 2 to reach criterion.

A 2-way mixed analysis of variance [Experiment: (3 levels, between subjects)] by Phase: [2 levels, learning and return (within subjects)] was performed on the number of trials participants took to reach the criterion of correctly identifying all twelve sounds in each phase of the three experiments. As repeated measures were used, the figure for Huynh-Feldt (H-F) Epsilon correction was calculated as these give more conservative degrees of freedom (Howell 1987). The summary Anova table is shown in Table 2.7.

As shown in the summary Anova table the main effect of experiment was significant \( (F(2,33)=6.14 \ p=<0.01) \). The main effect of phase was significant \( (F(1,22)=100.2 \ p=<0.001) \). There was no interaction between experiment and phase \( (F(2,22)=0.225 \ p=>0.05) \).

A Post Hoc test (Newman Keuls) was conducted on the results for experiment which showed that the mean number of trials required to reach criterion was significantly lower for Experiment 1 in comparison to Experiment 3 \( (p=<0.01) \). Experiment 1 and Experiment 2 did not differ significantly from one another and there was no significant difference between Experiment 2 and Experiment 3. The implications of these results are, as previously suggested, the participants undertaking Experiment 1 ('natural' sounds) found the sounds easier to learn than the participants in Experiment 2 ('controlled' sounds) and the participants in Experiment 3 ('neutral names') where many of the cues including the original names of the equipment had been removed.
The main effect for phase is due to fewer trials being required by the participants to correctly identify the sounds in the return phase than in the learning phase. This indicates that the participants were able to retain the names of the sounds correctly during the week between the learning phase and the return phase of the experiment in all three conditions.
<table>
<thead>
<tr>
<th>Experiment</th>
<th>Mean</th>
<th>St Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.875</td>
<td>3.481</td>
</tr>
<tr>
<td>2</td>
<td>7.292</td>
<td>3.712</td>
</tr>
<tr>
<td>3</td>
<td>8.208</td>
<td>3.514</td>
</tr>
</tbody>
</table>

Table 2.4: Number of Trials for Experiment (learning and return phases)

<table>
<thead>
<tr>
<th>Phase</th>
<th>Mean</th>
<th>St Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning</td>
<td>9.833</td>
<td>2.58</td>
</tr>
<tr>
<td>Return</td>
<td>4.417</td>
<td>2.298</td>
</tr>
</tbody>
</table>

Table 2.5: Mean Number of Trials for Phase (all experiments)

<table>
<thead>
<tr>
<th>Phase</th>
<th>Mean</th>
<th>St Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning Exp 1</td>
<td>8.500</td>
<td>2.939</td>
</tr>
<tr>
<td>Learning Exp 2</td>
<td>9.833</td>
<td>2.657</td>
</tr>
<tr>
<td>Learning Exp 3</td>
<td>11.167</td>
<td>1.267</td>
</tr>
<tr>
<td>Return Exp 1</td>
<td>3.250</td>
<td>1.288</td>
</tr>
<tr>
<td>Return Exp 2</td>
<td>4.750</td>
<td>2.768</td>
</tr>
<tr>
<td>Return Exp 3</td>
<td>5.250</td>
<td>2.261</td>
</tr>
</tbody>
</table>

Table 2.6: Mean Number of Trials for Experiment and Phase

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-Value</th>
<th>P-Value</th>
<th>G-G</th>
<th>H-F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment</td>
<td>2</td>
<td>66.33</td>
<td>33.167</td>
<td>6.147</td>
<td>.0054</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject (group)</td>
<td>33</td>
<td>178.042</td>
<td>5.395</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase</td>
<td>1</td>
<td>528.125</td>
<td>528.125</td>
<td>100.02</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>Phase*Experiment</td>
<td>2</td>
<td>2.333</td>
<td>1.167</td>
<td>0.225</td>
<td>0.7997</td>
<td>0.7997</td>
<td>0.7997</td>
</tr>
<tr>
<td>Phase*Subject</td>
<td>22</td>
<td>171.042</td>
<td>5.183</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.7: Summary Two-Way Anova Table for Number of Trials
2.2.4.3 Correct Responses

During each trial the participants' response to the individual sounds was recorded by the experimenter. The following section focuses on the correct response made for each sound, for example when sound 1 was called sound 1, sound 2 correctly identified as sound 2 and so on.

The correct responses for the total of each sound for the three experiments were compared to examine whether participants were able to learn the 'natural' sounds, with fewer errors than the 'controlled' sounds. To ensure that any confusions were acoustic rather than semantic, a comparison was made between the two experiments in which the names were changed but the sounds were the same.

Statistical comparisons were made between Experiment 1 (natural sounds) and Experiment 2 (controlled sounds) as the names of the sounds were the same (see Table 2.1), and between Experiment 2 ('controlled' sounds) and Experiment 3 ('neutral' names) as the sounds were the same. Statistical comparisons were not undertaken between the correct responses of Experiment 1 ('natural' sounds) and Experiment 3 ('neutral' names) as they differed in two important ways. The first difference was the sounds themselves and the second difference was the names given to the sounds.

The majority of the 'natural' sounds had the names of the medical equipment they originated from, while the neutral names were picked from the police alphabet. However, some comparison has been included between the means of the correct responses for Experiment 1 and Experiment 3 as they illustrate very well the considerable difference in participants' ability to identify the sounds correctly in each experiment, with more correct responses in Experiment 1 and fewer correct responses in Experiment 3. The mean number of correct responses for the learning phase of each experiment are shown in Table 2.8. and the means for the return phase are shown in Table 2.9.
<table>
<thead>
<tr>
<th>Sounds</th>
<th>Experiment 1</th>
<th></th>
<th>Experiment 2</th>
<th></th>
<th>Experiment 3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>St Dev</td>
<td>mean</td>
<td>St Dev</td>
<td>mean</td>
<td>St Dev</td>
</tr>
<tr>
<td>sound 1</td>
<td>9.50</td>
<td>1.62</td>
<td>7.75</td>
<td>2.73</td>
<td>5.75</td>
<td>3.27</td>
</tr>
<tr>
<td>sound 2</td>
<td>9.25</td>
<td>2.09</td>
<td>7.75</td>
<td>2.81</td>
<td>5.16</td>
<td>2.36</td>
</tr>
<tr>
<td>sound 3</td>
<td>10.58</td>
<td>1.67</td>
<td>9.50</td>
<td>2.84</td>
<td>7.58</td>
<td>2.87</td>
</tr>
<tr>
<td>sound 4</td>
<td>9.33</td>
<td>2.14</td>
<td>9.00</td>
<td>2.86</td>
<td>6.58</td>
<td>3.47</td>
</tr>
<tr>
<td>sound 5</td>
<td>8.75</td>
<td>3.52</td>
<td>8.41</td>
<td>3.17</td>
<td>6.66</td>
<td>3.11</td>
</tr>
<tr>
<td>sound 6</td>
<td>10.25</td>
<td>2.59</td>
<td>7.50</td>
<td>3.52</td>
<td>7.75</td>
<td>3.84</td>
</tr>
<tr>
<td>sound 7</td>
<td>9.50</td>
<td>2.32</td>
<td>7.41</td>
<td>2.02</td>
<td>7.50</td>
<td>2.51</td>
</tr>
<tr>
<td>sound 8</td>
<td>8.42</td>
<td>2.61</td>
<td>6.50</td>
<td>2.51</td>
<td>6.92</td>
<td>2.99</td>
</tr>
<tr>
<td>sound 9</td>
<td>8.58</td>
<td>2.53</td>
<td>6.92</td>
<td>3.28</td>
<td>8.16</td>
<td>3.92</td>
</tr>
<tr>
<td>sound 10</td>
<td>10.16</td>
<td>2.82</td>
<td>8.50</td>
<td>2.71</td>
<td>10.66</td>
<td>1.72</td>
</tr>
<tr>
<td>sound 11</td>
<td>7.75</td>
<td>3.54</td>
<td>6.08</td>
<td>2.67</td>
<td>7.83</td>
<td>3.78</td>
</tr>
<tr>
<td>sound 12</td>
<td>9.83</td>
<td>2.44</td>
<td>9.33</td>
<td>1.82</td>
<td>8.00</td>
<td>3.01</td>
</tr>
</tbody>
</table>

Table 2.8: Mean Number of Correct Responses for the Learning Phase of Experiments 1, 2 and 3
| Sounds | Experiment 1 | | | Experiment 2 | | | Experiment 3 | | |
|---------|-------------|---|---|----------------|---|---|----------------|---|
|         | mean | St Dev |       | mean | St Dev |       | mean | St Dev |       |
| sound 1 | 12.00 | 0.0  |       | 10.75 | 1.21  |       | 10.50 | 1.44  |       |
| sound 2 | 11.66 | 1.15 |       | 11.25 | 0.96  |       | 10.91 | 1.08  |       |
| sound 3 | 11.33 | 0.77 |       | 11.58 | 0.66  |       | 11.41 | 0.66  |       |
| sound 4 | 11.75 | 0.45 |       | 11.91 | 0.28  |       | 10.58 | 1.50  |       |
| sound 5 | 11.58 | 0.90 |       | 11.33 | 0.98  |       | 10.16 | 2.16  |       |
| sound 6 | 11.83 | 0.38 |       | 11.08 | 1.16  |       | 11.00 | 1.47  |       |
| sound 7 | 11.41 | 0.79 |       | 11.08 | 1.50  |       | 11.50 | 0.67  |       |
| sound 8 | 10.91 | 1.08 |       | 10.33 | 1.72  |       | 10.98 | 1.44  |       |
| sound 9 | 11.08 | 0.79 |       | 10.50 | 2.11  |       | 10.66 | 2.18  |       |
| sound 10| 11.75 | 0.62 |       | 10.66 | 1.87  |       | 11.91 | 0.28  |       |
| sound 11| 11.41 | 1.16 |       | 9.75  | 2.59  |       | 11.25 | 1.05  |       |
| sound 12| 11.67 | 0.49 |       | 11.83 | 0.38  |       | 11.75 | 0.45  |       |

Table 2.9: Mean Number of Correct Responses for the Return Phase of Each Experiment.
Two three-way analyses of variance [Experiment (two levels) by Phase (two levels) by Sound (twelve levels)] were conducted on the correct responses obtained in Experiments 1 and 2, and Experiments 2 and 3. The summary Anova Table for Experiments 1 and 2 is shown in Table 2.10. As repeated measures were used the figure for Huynh-Feldt (H-F) Epsilon correction was calculated as these give more conservative degrees of freedom (Howell 1987). The Epsilon correction is shown in Table 2.11

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of sq</th>
<th>Mean Square</th>
<th>F-Value</th>
<th>P-Value</th>
<th>H-F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expt</td>
<td>1</td>
<td>142.007</td>
<td>142.007</td>
<td>4.609</td>
<td>0.0431</td>
<td></td>
</tr>
<tr>
<td>Subject (Group)</td>
<td>22</td>
<td>677.903</td>
<td>30.814</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase</td>
<td>1</td>
<td>1029.340</td>
<td>1029.340</td>
<td>58.911</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>Phase x Experiment</td>
<td>1</td>
<td>31.174</td>
<td>31.174</td>
<td>1.784</td>
<td>0.1953</td>
<td>0.1953</td>
</tr>
<tr>
<td>Phase x Subject (Group)</td>
<td>22</td>
<td>384.403</td>
<td>17.473</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sounds</td>
<td>11</td>
<td>208.660</td>
<td>18.969</td>
<td>6.633</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>Sounds x Experiment</td>
<td>11</td>
<td>46.826</td>
<td>4.257</td>
<td>1.488</td>
<td>0.1362</td>
<td>0.1393</td>
</tr>
<tr>
<td>Sounds x Subject (Group)</td>
<td>242</td>
<td>692.097</td>
<td>2.860</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase x Sounds</td>
<td>11</td>
<td>48.910</td>
<td>4.446</td>
<td>2.157</td>
<td>0.0174</td>
<td>0.0278</td>
</tr>
<tr>
<td>Phase x Sounds x Exp</td>
<td>11</td>
<td>13.410</td>
<td>1.219</td>
<td>0.591</td>
<td>0.8351</td>
<td>0.7994</td>
</tr>
<tr>
<td>Phase<em>Sounds</em>Exp (Group)</td>
<td>242</td>
<td>498.764</td>
<td>2.061</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.10: Summary Table of Three-Way Anova Correct Responses Experiment 1/2

<table>
<thead>
<tr>
<th></th>
<th>H-F Epsilon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase</td>
<td>1.048</td>
</tr>
<tr>
<td>Sounds</td>
<td>.967</td>
</tr>
<tr>
<td>Phase*sounds</td>
<td>.799</td>
</tr>
</tbody>
</table>

Table 2.11: Adjustment of Degrees of Freedom (H-F)
The results show a significant interaction between phase and sounds \( (F(11,242)=2.157 \ p<0.05) \). The interaction is shown in Figure 2.1 and the means of the two phases (which combine the results of Experiment 1 and Experiment 2) are shown in Table 2.12. There was no significant interaction between phase and experiment \( (F(1,22)=1.784 \ p>0.05) \), sounds and experiment \( (F(11,242)=1.488 \ p>0.05) \) or Experiment and phase and sound \( (F(11,242)=1.219 \ p>0.05) \).

A post hoc (Tukey hsd) was conducted on the significant interaction between phase and sounds. The Tukey hsd statistic calculated, in which a difference between means equal or greater than \( q=2.2586 \ (p<0.05) \) is judged significant showed that only three sounds did not have a significantly higher correct response rate in the return phase than in the learning phase. These were sound 3 (learning mean 10.04; return mean 11.45); sound 10 (learning mean 9.33; return mean 11.21) and sound 12 (learning mean 9.58; return mean 11.75). All other sounds produced a significant difference between the number of correct responses produced in the two phases.

The results of the follow-up analysis indicates that the typical pattern for the correct response in each phase would show a considerably higher number of correct responses in the return phase than in the learning phase. However, with the results combined over the two experiments, sounds 3, 10 and 12 had a higher correct response rate in the learning phase as well as in the return phase.
Figure 2.1: Interaction of Phase and Correct Response to Sound (Experiment 1/2)
<table>
<thead>
<tr>
<th>Learning</th>
<th>Mean</th>
<th>St Dev</th>
<th>Return</th>
<th>Mean</th>
<th>St Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sound 1</td>
<td>8.625</td>
<td>2.374</td>
<td>Sound 1</td>
<td>11.375</td>
<td>1.056</td>
</tr>
<tr>
<td>Sound 2</td>
<td>8.375</td>
<td>2.584</td>
<td>Sound 2</td>
<td>11.458</td>
<td>1.062</td>
</tr>
<tr>
<td>Sound 3</td>
<td>10.042</td>
<td>2.349</td>
<td>Sound 3</td>
<td>11.458</td>
<td>0.721</td>
</tr>
<tr>
<td>Sound 4</td>
<td>9.167</td>
<td>2.479</td>
<td>Sound 4</td>
<td>11.833</td>
<td>0.381</td>
</tr>
<tr>
<td>Sound 5</td>
<td>8.583</td>
<td>3.283</td>
<td>Sound 5</td>
<td>11.458</td>
<td>0.932</td>
</tr>
<tr>
<td>Sound 6</td>
<td>8.875</td>
<td>3.340</td>
<td>Sound 6</td>
<td>11.458</td>
<td>0.932</td>
</tr>
<tr>
<td>Sound 7</td>
<td>8.458</td>
<td>2.377</td>
<td>Sound 7</td>
<td>11.250</td>
<td>1.189</td>
</tr>
<tr>
<td>Sound 8</td>
<td>7.458</td>
<td>2.686</td>
<td>Sound 8</td>
<td>10.625</td>
<td>1.439</td>
</tr>
<tr>
<td>Sound 9</td>
<td>7.750</td>
<td>2.996</td>
<td>Sound 9</td>
<td>10.792</td>
<td>1.587</td>
</tr>
<tr>
<td>Sound 10</td>
<td>9.333</td>
<td>2.839</td>
<td>Sound 10</td>
<td>11.208</td>
<td>1.474</td>
</tr>
<tr>
<td>Sound 11</td>
<td>6.917</td>
<td>3.189</td>
<td>Sound 11</td>
<td>10.583</td>
<td>2.145</td>
</tr>
<tr>
<td>Sound 12</td>
<td>9.583</td>
<td>2.125</td>
<td>Sound 12</td>
<td>11.750</td>
<td>0.442</td>
</tr>
</tbody>
</table>

Table 2.12: Mean Number of Correct Responses for Sounds in Each Phase (Experiments 1/2)
There was a significant main effect for experiment (F(1,22)=4.609 p=<0.05). The means are shown in Table 2.13. There is also a significant main effect for phase (F(1,22)=58.911 p=<0.001). The means are shown in Table 2.14. There was a main effect for sounds (F(11,242)=6.633 p=<0.001). The means are shown in Table 2.15.

A post Hoc test (Tukey hsd) was performed on the main effect of sounds to elicit which sounds differed significantly from one another. The results showed fewer correct responses were made for sound 11 (mean 8.75) than sound 3 (mean 10.75), sound 4 (mean 10.5), sound 6 (mean 10.16), sound 10 (mean 10.27) and sound 12 (mean 10.66). (Tukey ps=<0.01)

Sound 8 (mean 9.04) had significantly fewer correct responses (ps=<0.01) than sounds 3 (mean 10.75); sound 4 (mean 10.5) and sound 12 (mean 10.67). Sound 9 (mean 9.27) produced significantly fewer correct responses (ps=<0.01) than sounds 3 (mean 10.75) and sound 12 (mean 10.66). Although no other significant effects between the sounds were found, in Experiments 1 and 2, sound 3 overall appears to have produced the highest number of correct responses (mean 10.75) while sound 11 appears to have produced the lowest number of correct responses (8.75).
<table>
<thead>
<tr>
<th>Experiment</th>
<th>Mean</th>
<th>St Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 1</td>
<td>10.431</td>
<td>2.212</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>9.438</td>
<td>2.786</td>
</tr>
</tbody>
</table>

Table 2.13: Mean Number of Correct Responses for Main Effect of Experiment (Experiment 1/2)

<table>
<thead>
<tr>
<th>Phase</th>
<th>Mean</th>
<th>St Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning</td>
<td>8.597</td>
<td>2.828</td>
</tr>
<tr>
<td>Return</td>
<td>11.271</td>
<td>1.251</td>
</tr>
</tbody>
</table>

Table 2.14: Mean Number of Correct Responses for Phase (Experiments 1/2)

<table>
<thead>
<tr>
<th>Sounds</th>
<th>Mean</th>
<th>St Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sound 1</td>
<td>10.00</td>
<td>2.288</td>
</tr>
<tr>
<td>Sound 2</td>
<td>9.92</td>
<td>2.500</td>
</tr>
<tr>
<td>Sound 3</td>
<td>10.75</td>
<td>1.862</td>
</tr>
<tr>
<td>Sound 4</td>
<td>10.50</td>
<td>2.212</td>
</tr>
<tr>
<td>Sound 5</td>
<td>10.02</td>
<td>2.794</td>
</tr>
<tr>
<td>Sound 6</td>
<td>10.16</td>
<td>2.755</td>
</tr>
<tr>
<td>Sound 7</td>
<td>9.85</td>
<td>2.334</td>
</tr>
<tr>
<td>Sound 8</td>
<td>9.04</td>
<td>2.665</td>
</tr>
<tr>
<td>Sound 9</td>
<td>9.27</td>
<td>2.826</td>
</tr>
<tr>
<td>Sound 10</td>
<td>10.27</td>
<td>2.430</td>
</tr>
<tr>
<td>Sound 11</td>
<td>8.75</td>
<td>3.265</td>
</tr>
<tr>
<td>Sound 12</td>
<td>10.66</td>
<td>1.872</td>
</tr>
</tbody>
</table>

Table 2.15: Mean Number of Correct Responses for Sounds (Experiment 1/2)
The summary three-way Anova [Experiment (two levels) by Phase (two levels) by Sound (twelve levels)] for the correct responses obtained in Experiments 2 and 3 is shown in Table 2.16. The figures for the H-F Epsilon correction are shown in Table 2.17.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of Sq</th>
<th>Mean Square</th>
<th>F-Value</th>
<th>P-Value</th>
<th>H-F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expt</td>
<td>1</td>
<td>9.507</td>
<td>9.507</td>
<td>0.387</td>
<td>0.5405</td>
<td></td>
</tr>
<tr>
<td>Subject (Group)</td>
<td>22</td>
<td>540.986</td>
<td>24.590</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase</td>
<td>1</td>
<td>1633.50</td>
<td>1633.51</td>
<td>119.29</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>Phase x Experiment</td>
<td>1</td>
<td>7.563</td>
<td>7.563</td>
<td>0.552</td>
<td>0.4653</td>
<td>0.4653</td>
</tr>
<tr>
<td>Phase x Subject (Group)</td>
<td>22</td>
<td>301.264</td>
<td>13.694</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sounds</td>
<td>11</td>
<td>219.410</td>
<td>19.946</td>
<td>4.051</td>
<td>0.0001</td>
<td>0.0002</td>
</tr>
<tr>
<td>Sounds x Experiment</td>
<td>11</td>
<td>187.410</td>
<td>17.037</td>
<td>3.460</td>
<td>0.0002</td>
<td>0.0002</td>
</tr>
<tr>
<td>Sounds x Subject (Group)</td>
<td>242</td>
<td>1191.68</td>
<td>4.924</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase x Sounds</td>
<td>11</td>
<td>66.660</td>
<td>6.060</td>
<td>1.813</td>
<td>0.0524</td>
<td>0.0599</td>
</tr>
<tr>
<td>Phase x Sounds x Exp</td>
<td>11</td>
<td>40.271</td>
<td>3.661</td>
<td>1.095</td>
<td>0.3654</td>
<td>0.3667</td>
</tr>
<tr>
<td>Phase x Sounds x Exp x Group</td>
<td>242</td>
<td>808.736</td>
<td>3.342</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.16: Summary Table of Three-Way Anova Correct Responses Experiment 2/3

<table>
<thead>
<tr>
<th></th>
<th>H-F Epsilon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase</td>
<td>1.048</td>
</tr>
<tr>
<td>Sounds</td>
<td>.997</td>
</tr>
<tr>
<td>Phase x Sounds</td>
<td>.905</td>
</tr>
</tbody>
</table>

Table 2.17: Adjustment of Degrees of Freedom (H-F)

There was a significant interaction between sounds and experiment (F(11,242)=3.460 p=<0.01) the means are shown in Table 2.18 and the interaction is illustrated in Figure 2.2.

Figure 2.2 seems to show that the change in the names in Experiment 3 had an effect on the correct recognition of the sounds. It can be seen that for sound 1, sound 2, sound 3, sound 4, sound 5, sound 9 and sound 12 the participants produced more correct responses in
Experiment 2 where the sounds had the names of medical equipment than in Experiment 3. However, for sound 6, sound 7, sound 8, sound 10 and sound 11 the correct responses seemed to be slightly higher in Experiment 3, the 'neutral names' condition. However, follow-up analysis (Tukey hsd) showed that none of these comparisons were significant (p>0.05). The comparisons that were significant were different sounds in different conditions, so, for example, there was a significant difference in the means of sound 11 in Experiment 2 and sound 10 in Experiment 3.
Figure 2.2: Interaction Between Experiment and Sound (Experiment 2/3)
<table>
<thead>
<tr>
<th>Experiment 2</th>
<th>Mean</th>
<th>St Dev</th>
<th>Experiment 3</th>
<th>Mean</th>
<th>St Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>sound 1</td>
<td>9.25</td>
<td>2.57</td>
<td>Sound 1</td>
<td>8.12</td>
<td>3.46</td>
</tr>
<tr>
<td>Sound 2</td>
<td>9.37</td>
<td>2.81</td>
<td>Sound 2</td>
<td>8.04</td>
<td>3.44</td>
</tr>
<tr>
<td>Sound 3</td>
<td>10.54</td>
<td>2.28</td>
<td>Sound 3</td>
<td>9.50</td>
<td>2.82</td>
</tr>
<tr>
<td>Sound 4</td>
<td>10.46</td>
<td>2.48</td>
<td>Sound 4</td>
<td>8.58</td>
<td>3.32</td>
</tr>
<tr>
<td>Sound 5</td>
<td>9.87</td>
<td>2.74</td>
<td>Sound 5</td>
<td>8.42</td>
<td>3.17</td>
</tr>
<tr>
<td>Sound 6</td>
<td>9.29</td>
<td>3.15</td>
<td>Sound 6</td>
<td>9.37</td>
<td>3.29</td>
</tr>
<tr>
<td>Sound 7</td>
<td>9.25</td>
<td>2.55</td>
<td>Sound 7</td>
<td>9.50</td>
<td>2.72</td>
</tr>
<tr>
<td>Sound 8</td>
<td>8.41</td>
<td>2.87</td>
<td>Sound 8</td>
<td>8.50</td>
<td>2.81</td>
</tr>
<tr>
<td>Sound 9</td>
<td>8.70</td>
<td>3.26</td>
<td>Sound 9</td>
<td>9.41</td>
<td>3.36</td>
</tr>
<tr>
<td>Sound 10</td>
<td>9.58</td>
<td>2.53</td>
<td>Sound 10</td>
<td>11.29</td>
<td>1.36</td>
</tr>
<tr>
<td>Sound 11</td>
<td>7.92</td>
<td>3.18</td>
<td>Sound 11</td>
<td>9.54</td>
<td>3.23</td>
</tr>
<tr>
<td>Sound 12</td>
<td>10.58</td>
<td>1.82</td>
<td>Sound 12</td>
<td>9.87</td>
<td>2.85</td>
</tr>
</tbody>
</table>

Table 2.18: Mean Number of Correct Responses for Sounds for Experiment 2 and Experiment 3
The interaction between phase and sounds approached significance (F11,242)=1.813 p=>0.05. The means are shown in Table 2.19 and illustrated in Figure 2.3. There were no other significant interactions.

The interaction between the number of correct responses and phase that was approaching significance in Experiments 2 and 3 was also analysed using Tukey (hsd). These results showed that all but one of the responses in the learning phase were significantly lower than in the return phase (ps=<0.01). Sound 10 was not significantly different (mean learning phase 11.29; mean return phase 9.58). This indicates that for the majority of sounds there was a considerable difference between the number of correct responses obtained during the learning phase and return phase, the pattern suggesting that correct responses were substantially higher in the return phase.
Figure 2.3: Interaction Between Correct Responses and Phase (Experiment 2/3)
<table>
<thead>
<tr>
<th>Learning</th>
<th>Mean</th>
<th>St Dev</th>
<th>Return</th>
<th>Mean</th>
<th>St Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>sound 1</td>
<td>6.75</td>
<td>3.12</td>
<td>sound 1</td>
<td>10.62</td>
<td>1.31</td>
</tr>
<tr>
<td>sound 2</td>
<td>6.33</td>
<td>2.81</td>
<td>sound 2</td>
<td>11.08</td>
<td>1.01</td>
</tr>
<tr>
<td>sound 3</td>
<td>8.54</td>
<td>2.96</td>
<td>sound 3</td>
<td>11.50</td>
<td>0.65</td>
</tr>
<tr>
<td>sound 4</td>
<td>7.79</td>
<td>3.35</td>
<td>sound 4</td>
<td>11.25</td>
<td>1.26</td>
</tr>
<tr>
<td>sound 5</td>
<td>7.54</td>
<td>3.20</td>
<td>sound 5</td>
<td>10.75</td>
<td>1.75</td>
</tr>
<tr>
<td>sound 6</td>
<td>7.62</td>
<td>3.61</td>
<td>sound 6</td>
<td>11.04</td>
<td>1.30</td>
</tr>
<tr>
<td>sound 7</td>
<td>7.45</td>
<td>2.22</td>
<td>sound 7</td>
<td>11.29</td>
<td>1.16</td>
</tr>
<tr>
<td>sound 8</td>
<td>6.70</td>
<td>2.71</td>
<td>sound 8</td>
<td>10.20</td>
<td>1.56</td>
</tr>
<tr>
<td>sound 9</td>
<td>7.54</td>
<td>3.59</td>
<td>sound 9</td>
<td>10.58</td>
<td>2.10</td>
</tr>
<tr>
<td>sound 10</td>
<td>9.58</td>
<td>2.48</td>
<td>sound 10</td>
<td>11.29</td>
<td>1.45</td>
</tr>
<tr>
<td>sound 11</td>
<td>6.95</td>
<td>3.33</td>
<td>sound 11</td>
<td>10.50</td>
<td>2.08</td>
</tr>
<tr>
<td>sound 12</td>
<td>8.66</td>
<td>2.53</td>
<td>sound 12</td>
<td>11.79</td>
<td>0.41</td>
</tr>
</tbody>
</table>

Table 2.19: Mean Number of Correct Responses to Sounds in Each Phase (Experiments 2/3)
There was no significant main effect for Experiment (F(1,22)=0.387, p=>0.05). However, there was a significant main effect of phase (F(1,22)=119.288, p=<0.001). The means are shown in Table 2.20. There was also a significant main effect for Sounds (F(11,242)=4.051, p=<0.001). The means are shown in Table 2.21.

A post hoc test (Tukey hsd) was performed on the main effect of sounds in Experiments 2 and 3 to elicit which sounds produced significantly fewer correct responses. The results showed that sound 8 (mean 8.45) had significantly fewer correct responses (p=<0.01) than sounds 10 (mean 10.43) and 12 (mean 10.23). Sound 1 (mean 8.68) had significantly less correct responses (p=<0.01) than sound 10 (mean 10.43). Although these were the only significant differences between the correct responses for the sounds the overall the sound with the highest number of correct responses in Experiments 2 and 3 was sound 10 (mean 10.43) and the sound with the least number of correct responses was sound 2 (mean 8.71).

<table>
<thead>
<tr>
<th>Phase</th>
<th>Mean</th>
<th>St Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning</td>
<td>7.625</td>
<td>0.183</td>
</tr>
<tr>
<td>Return</td>
<td>10.993</td>
<td>0.086</td>
</tr>
</tbody>
</table>

Table 2.20: Mean Number of Correct Responses for Phase (Experiments 2/3)
<table>
<thead>
<tr>
<th>Sounds</th>
<th>Mean</th>
<th>St Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sound 1</td>
<td>8.68</td>
<td>3.07</td>
</tr>
<tr>
<td>Sound 2</td>
<td>8.71</td>
<td>3.18</td>
</tr>
<tr>
<td>Sound 3</td>
<td>10.02</td>
<td>2.59</td>
</tr>
<tr>
<td>Sound 4</td>
<td>9.52</td>
<td>3.05</td>
</tr>
<tr>
<td>Sound 5</td>
<td>9.14</td>
<td>3.02</td>
</tr>
<tr>
<td>Sound 6</td>
<td>9.33</td>
<td>3.19</td>
</tr>
<tr>
<td>Sound 7</td>
<td>9.37</td>
<td>2.61</td>
</tr>
<tr>
<td>Sound 8</td>
<td>8.45</td>
<td>2.81</td>
</tr>
<tr>
<td>Sound 9</td>
<td>9.06</td>
<td>3.29</td>
</tr>
<tr>
<td>Sound 10</td>
<td>10.43</td>
<td>2.19</td>
</tr>
<tr>
<td>Sound 11</td>
<td>8.73</td>
<td>3.27</td>
</tr>
<tr>
<td>Sound 12</td>
<td>10.23</td>
<td>2.39</td>
</tr>
</tbody>
</table>

Table 2.21: Mean Number of Correct Responses for Sounds (Experiments 2/3)
2.2.4.4 Confusions

The final set of results for this set of experiments examines the errors made by participants when identifying the sounds and at the significant confusions made between the sounds. Using a method of analysis based on the multinomial distribution (Howell 1987) which also takes into account response bias, significant confusions were isolated. The significant confusions between warnings in both learning and return phase of both experiments are shown in Table 2.22. The number of highly significant confusions (p=<0.01) is relatively small, and these are shown in the table. The number of confusions at a lower level of significance (p=<0.05) is considerably higher, particularly for the learning phase of Experiment 1 which can be seen in Tables 2.24 to 2.29. (In the return phases of the experiments there were very few sounds that were identified incorrectly, therefore only the errors, as opposed to the zero scores were analysed).

The pattern of confusion between experiments is also slightly different, for example it can be seen that sound 8 is consistently confused with sound 11 in all phases of the experiments except for the return phase of Experiment 2.
There were also significant non-confusions (P<0.01) that occurred in the learning phase of Experiments 2 and 3. This occurred when there was no error, not even as a guess, for a particular pair of sounds. In Experiment 2 the non-confusions were sound 6 which was never called sound 10; sound 8 which was never called sound 5 and sounds 8 and 9 which were not confused with sound 12. In Experiment 3 the significant non-confusions were sound 3 which was not called sound 10, and sound 11 which was not called sound 12. This suggests that despite the lack of cues for Experiments 2 and 3, participants had identified for example, that sound 6 was definitely not sound 10 and even if the participants made a guess at the response when sound 6 was played, it would not be for sound 10. The significant non-confusions are shown in Table 2.23.

<table>
<thead>
<tr>
<th>sound played</th>
<th>Exp 1 Learning</th>
<th>Exp 1 Return</th>
<th>Exp 2 Learning</th>
<th>Exp 2 Return</th>
<th>Exp 3 Learning</th>
<th>Exp 3 Return</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7, 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>2, 8</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>9</td>
<td>9, 10</td>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td></td>
<td>1, 7</td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td>6</td>
<td>6</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>11</td>
<td>7, 11</td>
<td>10, 11</td>
<td></td>
<td>1, 11</td>
<td>11</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>5</td>
<td></td>
<td>7</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>2</td>
<td>1, 8</td>
<td>8</td>
</tr>
<tr>
<td>12</td>
<td>3, 11</td>
<td>3, 11</td>
<td></td>
<td>3</td>
<td></td>
<td>3</td>
</tr>
</tbody>
</table>

Table 2.22: Significant Confusions (p<0.01) Between Warnings
<table>
<thead>
<tr>
<th>Sound Played</th>
<th>Exp 1 Learning</th>
<th>Exp 1 Return</th>
<th>Exp 2 Learning</th>
<th>Exp 2 Return</th>
<th>Exp 3 Learning</th>
<th>Exp 3 Return</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5,12</td>
<td>9</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.23: Significant Non-Confusions (p<0.01) Between Warnings
Sound Identified As:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>*</td>
<td>0.1360</td>
<td>0.0794</td>
<td>0.5494</td>
<td>0.5414</td>
<td>0.9406</td>
<td>0.0090</td>
<td>0.1080</td>
<td>0.9406</td>
<td>0.4270</td>
<td>0.0048</td>
<td>0.1536</td>
</tr>
<tr>
<td>2</td>
<td>0.1133</td>
<td>*</td>
<td>0.0564</td>
<td>0.0942</td>
<td>0.5832</td>
<td>0.3496</td>
<td>0.7644</td>
<td>0.1622</td>
<td>0.0710</td>
<td>0.0744</td>
<td>0.9014</td>
<td>0.1174</td>
</tr>
<tr>
<td>3</td>
<td>0.4578</td>
<td>0.4214</td>
<td>*</td>
<td>0.4368</td>
<td>0.4316</td>
<td>0.3650</td>
<td>0.9888</td>
<td>0.3438</td>
<td>0.3694</td>
<td>0.3608</td>
<td>0.5466</td>
<td>0.0001</td>
</tr>
<tr>
<td>4</td>
<td>0.5122</td>
<td>0.0014</td>
<td>0.2988</td>
<td>*</td>
<td>0.1175</td>
<td>0.8466</td>
<td>0.3476</td>
<td>0.0034</td>
<td>0.6642</td>
<td>0.0830</td>
<td>0.4372</td>
<td>0.1284</td>
</tr>
<tr>
<td>5</td>
<td>0.3294</td>
<td>0.0600</td>
<td>0.4626</td>
<td>0.2978</td>
<td>*</td>
<td>0.9594</td>
<td>0.2990</td>
<td>0.0464</td>
<td>0.5496</td>
<td>0.0128</td>
<td>0.6734</td>
<td>0.0704</td>
</tr>
<tr>
<td>6</td>
<td>0.9666</td>
<td>0.8988</td>
<td>0.2030</td>
<td>0.3062</td>
<td>0.0001</td>
<td>*</td>
<td>0.3254</td>
<td>0.7728</td>
<td>0.8030</td>
<td>0.2420</td>
<td>0.8742</td>
<td>0.3244</td>
</tr>
<tr>
<td>7</td>
<td>0.5656</td>
<td>0.5178</td>
<td>0.7906</td>
<td>0.1358</td>
<td>0.5334</td>
<td>0.2688</td>
<td>*</td>
<td>0.8814</td>
<td>0.0080</td>
<td>0.9096</td>
<td>0.5160</td>
<td>0.1476</td>
</tr>
<tr>
<td>8</td>
<td>0.0492</td>
<td>0.5608</td>
<td>0.8718</td>
<td>0.2180</td>
<td>0.5686</td>
<td>0.1564</td>
<td>0.0840</td>
<td>*</td>
<td>0.1772</td>
<td>0.4270</td>
<td>0.0010</td>
<td>0.2239</td>
</tr>
<tr>
<td>9</td>
<td>0.1984</td>
<td>0.0730</td>
<td>0.3948</td>
<td>0.0900</td>
<td>0.0900</td>
<td>0.9482</td>
<td>0.8906</td>
<td>0.8796</td>
<td>*</td>
<td>0.4870</td>
<td>0.5952</td>
<td>0.0574</td>
</tr>
<tr>
<td>10</td>
<td>0.3014</td>
<td>0.8460</td>
<td>0.3512</td>
<td>0.8628</td>
<td>0.8628</td>
<td>0.0896</td>
<td>0.7124</td>
<td>0.1014</td>
<td>0.7502</td>
<td>*</td>
<td>0.5960</td>
<td>0.2968</td>
</tr>
<tr>
<td>11</td>
<td>0.0256</td>
<td>0.7444</td>
<td>0.2402</td>
<td>0.7688</td>
<td>0.7566</td>
<td>0.2748</td>
<td>0.9314</td>
<td>0.0100</td>
<td>0.0656</td>
<td>0.6420</td>
<td>*</td>
<td>0.0248</td>
</tr>
<tr>
<td>12</td>
<td>0.2200</td>
<td>0.6762</td>
<td>0.0001</td>
<td>0.2012</td>
<td>0.2012</td>
<td>0.1568</td>
<td>0.5476</td>
<td>0.1432</td>
<td>0.5906</td>
<td>0.5688</td>
<td>0.0002</td>
<td>*</td>
</tr>
</tbody>
</table>

Table 2.24: Experiment 1 (Learning Phase) Significant Confusions (P<=0.01) (shown in bold borders) Approaching Significance (p<=0.05 shown in double borders)
Table 2.25: Experiment 2 Learning Phase: Significant Confusions (P<0.01 shown in bold borders)
Approaching Significance (p<0.05 shown in double borders)
### Table 2.26: Experiment 3 Learning Phase Significant Confusions (P<=0.01 shown in bold borders) Approaching Significance (p<=0.05 shown in double borders)

<table>
<thead>
<tr>
<th>Sound Played</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.0</td>
<td>0.0</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
<td>0.5</td>
<td>0.6</td>
<td>0.7</td>
<td>0.8</td>
<td>0.9</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>0.0126</td>
<td>0.1950</td>
<td>0.1560</td>
<td>0.1344</td>
<td>0.2264</td>
<td>0.0228</td>
<td>0.6362</td>
<td>0.4752</td>
<td>0.6346</td>
<td>0.5062</td>
<td>0.0120</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1454</td>
<td>0.8702</td>
<td>0.0001</td>
<td>0.5152</td>
<td>0.7500</td>
<td>0.9466</td>
<td>0.0472</td>
<td>0.1782</td>
<td>0.0078</td>
<td>0.6958</td>
<td>0.0314</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.5752</td>
<td>0.1330</td>
<td>* 0.0136</td>
<td>0.0228</td>
<td>0.0190</td>
<td>0.4566</td>
<td>0.4846</td>
<td>0.1018</td>
<td>0.0106</td>
<td>0.9062</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0990</td>
<td>0.0001</td>
<td>0.2950</td>
<td>* 0.1108</td>
<td>0.5262</td>
<td>0.0760</td>
<td>0.1406</td>
<td>0.3804</td>
<td>0.2800</td>
<td>0.8288</td>
<td>0.0226</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0782</td>
<td>0.9098</td>
<td>0.3404</td>
<td>0.1570</td>
<td>* 0.2412</td>
<td>0.0900</td>
<td>0.9206</td>
<td>0.0546</td>
<td>0.1262</td>
<td>0.0268</td>
<td>0.2932</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0520</td>
<td>0.0230</td>
<td>0.6558</td>
<td>0.0444</td>
<td>0.6962</td>
<td>* 0.9774</td>
<td>0.0268</td>
<td>0.4610</td>
<td>0.0940</td>
<td>0.3474</td>
<td>0.0838</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1630</td>
<td>0.4352</td>
<td>0.0808</td>
<td>0.5926</td>
<td>0.4352</td>
<td>0.8456</td>
<td>* 0.7362</td>
<td>0.6906</td>
<td>0.9454</td>
<td>0.0442</td>
<td>0.8954</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0024</td>
<td>0.9684</td>
<td>0.1534</td>
<td>0.5188</td>
<td>0.3624</td>
<td>0.8962</td>
<td>0.3174</td>
<td>* 0.0084</td>
<td>0.0386</td>
<td>0.0001</td>
<td>0.3290</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.8420</td>
<td>0.0380</td>
<td>0.7604</td>
<td>0.8770</td>
<td>0.0380</td>
<td>0.5444</td>
<td>0.6720</td>
<td>0.8414</td>
<td>* 0.4038</td>
<td>0.2258</td>
<td>0.8298</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.3432</td>
<td>0.2608</td>
<td>0.8862</td>
<td>0.8498</td>
<td>0.7454</td>
<td>0.7830</td>
<td>0.6922</td>
<td>0.1194</td>
<td>0.8128</td>
<td>* 0.1946</td>
<td>0.3916</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0001</td>
<td>0.0782</td>
<td>0.2812</td>
<td>0.1012</td>
<td>0.0276</td>
<td>0.1310</td>
<td>0.0340</td>
<td>0.0001</td>
<td>0.1168</td>
<td>0.0126</td>
<td>* 0.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.8632</td>
<td>0.1942</td>
<td>0.0001</td>
<td>0.0214</td>
<td>0.0342</td>
<td>0.4532</td>
<td>0.5674</td>
<td>0.0914</td>
<td>0.2014</td>
<td>0.0170</td>
<td>0.1570</td>
<td></td>
</tr>
</tbody>
</table>
Table 2.27: Experiment 1 (Return Phase) Significant Confusions (P=<0.01) (shown in bold borders) Approaching Significance (p=<0.05) (shown in double borders)
### Table 2.28: Experiment 2 (Return Phase) Significant Confusions (P<0.01) (shown in bold borders)
Approaching Significance (p<0.05) (shown in double borders)
Table 2.29: Experiment 3 Return Phase Significant Confusions (P=<0.01 shown in bold borders) Approaching Significance (p=<0.05 shown in double borders)
2.3 Discussion

The results of this set of three experiments have shown that participants found the natural sounds easier to identified correctly than, firstly, those which had been controlled for loudness and length in line with experimental protocol, and secondly, those sounds which used the controlled warnings and had neutral names. This may explain why Patterson and Milroy (1980) found that participants could only learn five or six warnings. If the warnings are controlled in some way then cues that are normally used in the environment for remembering sounds are lost and this therefore makes them difficult to remember. The results of the present set of experiments show that participants are in fact quite good at learning and remembering up to twelve warnings with relatively few confusions. As was found in the Patterson and Milroy (1980) study, all participants eventually learned the set of 12 warnings within one hour, many in a much shorter time. The participants required fewer trials in the return phases of the experiments, with Experiment 1, the natural sounds requiring significantly fewer trials than the sounds with the neutral names. The participants retained that knowledge for a period of one week, when they returned to undertake the second phase of the experiment, many sounds being remembered by the participants with no errors and a ceiling effect occurred.

These results may have some relevance to the number of warnings recommended by Patterson (1982) for use in the working environment. He suggests that the optimum number of warnings should be no more than five or six. This suggestion is further supported by Kerr (1985) for use in I.C.U. and other critical care areas.

Overall, a consistent pattern did not emerge for one sound that was either the easiest or the most difficult sound to correctly identify. During the learning phase of both Experiments 1 and 2 sound 11 produced the lowest number of correct responses and sound 3 the highest number of correct responses. In Experiment 3 sound 2 produced the lowest number of correct responses and sound 10 the highest number of correct responses. In the return phases sound 8
produced the lowest number of correct responses in both Experiment 1 and 3 while sound 11 had fewer correct responses in Experiment 2. The highest number of correct responses in Experiment 1 was for sound 1 with all subjects remembering the sound with no errors. In Experiment 2 sound 4 had the highest number of correct responses and for Experiment 3 sound 10 had the highest number of correct responses.

When the experiments were compared, Experiments 1/2 showed that overall sound 3 produced the highest number of correct responses, and sound 11 produced the lowest number of correct responses. It was shown that sounds 3, 10 and 12 produced a high correct response rate in the learning phase as well as in the return phase. Sound 10 also produced a high correct response rate in both phases of Experiments 2/3.

Although sound 3 produced the highest number of correct responses in some of the conditions it was also one of the sounds that was consistently confused with sound 12. This would appear to indicate that the overall properties of the sound, a continuous tone, was relatively easy for subjects to remember, but was likewise easily confused with a similar, continuous tone. The confusions between the sounds are considered in greater detail below.

The results of the present experiments appear to suggest that a considerable number of warnings can be learned by individuals but that acoustic properties of the sounds can lead to confusion. As Patterson and Milroy (1980) concluded in their study, temporal characteristics, particularly similarity of repetition rate in warnings, cause confusion. This was also found in the present study. For example, sound 8 was consistently heard as sound 11 in all but one of the experimental phases, the sounds having a similar on/off repetitive sound. Sound 11 was confused with sound 8 in Experiment 1 in both the learning and return phase and also in both conditions of Experiment 3. However, there was a difference in the timing; sound 11 was much faster with approximately 15 beats in a 10 second period while sound 8 was slower with 6 beats in a 10 second period, sound 8 being about 2.5 times slower than sound 11.
The fact that the confusion was asymmetrical suggests that participants could identify the faster tempo when sound 11 was heard, but when they heard sound 8, they could not distinguish between the tempi.

Other studies have investigated the effect of tempo on identification of sounds. Pollack (1990) found that detection and discrimination in an auditory periodicities task was poorer at very low periodicities. Monahan and Hirsh (1990) found that absolute discrimination was poorer the longer the intervals within a tone i.e. the slower the tempo.

Most of the confusions that occurred in the present study were asymmetric, the only symmetrical confusions in the learning phase of all experiments was between sound 3 and sound 12, and sound 8 and sound 11 in Experiment 1 and Experiment 3, but not Experiment 2. The asymmetrical confusions occurred more frequently. For example sound 4 was frequently named as sound 2 but not vice versa (2 was not named as 4 in the same way).

Asymmetries also occur in visual perception, for example, Dawson and Harshman (1986) examined asymmetries in visual confusion matrices. They found for example, that the letter 'Q' was more frequently mistaken for the letter 'O', than the letter 'O' was for the letter 'Q'. It has been suggested (e.g. Navon 1981) that the global features and the overall shape of a stimulus, such as its roundness or straightness, are perceived initially before perception of local features, such as individual lines etc. This is referred to as the global-to-local processing hypothesis. An alternative viewpoint is the local-to-global processing hypothesis (e.g. Treisman and Gelade 1980) in which detailed information regarding the stimuli is processed first. In this theory, letters that have several features (e.g. the letter 'R') are more likely to be confused with a letter that has fewer features (e.g. the letter 'P') because of a failure to detect a specific feature. The confusion does not occur in reverse processing hypothesis (e.g. the letter 'P' being mistaken for the letter 'R') as frequently, because missing a particular feature would not transform a 'P' into the letter 'R'.

112
The research on asymmetrical confusions in visual perception may help interpret the asymmetrical confusions that occurred in the current study. In the example discussed above, sound 4 was frequently named as sound 2 but not vice versa.

Sound 2 is a sound lasting for 5 seconds, the tone at 0.7 secs is at 280Hz which is then followed by a repeating two-tone pattern at 2600Hz and 2000Hz. The whole sound then repeats after a gap of several seconds. Sound 4 begins with a tone of 1 second at 260Hz, followed by burst of short tones each lasting 0.25 secs at 2600Hz. The alphabetic confusion research may provide a framework for addressing the issue of asymmetrical auditory confusions found in the current study. One problem would be establishing the simplicity or complexity of a sound. The sounds could be analysed in Bits of information, for example sound 4 has more tones whereas sound 2 has more pitches. However, this analysis would be complex and is beyond the scope of the research reported in the current study.

Another feature of acoustic confusion that occurred in the present study was that of pitch. As discussed previously, there was a symmetrical confusion that occurred between sound 3 and sound 12 in all three experiments. Both sounds are a continuous tone but sound 12 has a notably higher pitch than sound 3. It is interesting to note that although they are approximately a tenth apart in pitch, and when heard together are very clearly discriminable, participants confuse them when they are heard with longer time intervals between them (the time intervals would be much greater in practice than in the experiment). This confusion shows that participants find it difficult to discriminate between warnings on the basis of pitch alone, as Deutsch (1978) stated, as well as sharing other characteristics such as temporal pattern, which in this example is a continuous sound with no temporal characteristics.

Some of the confusions are between warnings which begin with a long tone and continue in a completely different fashion (for example, warnings 2, 4, and 8), and those which have a succession of
very short pulses but differ substantially in other ways (warnings 10, 5, and 6).

These results suggest that the basis on which confusions can occur might be related to the sorts of labels participants use to describe, and memorise, the warnings rather than to their precise acoustic qualities. For example, some of the confusions we have found are more readily understood if we assume that participants used the label 'high-pitched, continuous tone' to memorise warnings 3 and 12, and 'regular on/off' for warnings 8 and 11. For warnings which are acoustically quite different, simply the label 'complex' or 'musical' might have caused confusion.

The change in the names did appear to have a detrimental effect on the identification of the sounds. This is indicated by the higher number of trials required by participants to correctly identify the sounds in Experiment 3 which also produced fewer correct responses than in either Experiment 1 or Experiment 2, Experiments 2 and 3 having the same sounds. When participants had completed the return phase of the experiments, they were asked how they had remembered the sounds and the responses were categorised into two groups, imagery and musical or acoustic cues. Many of the participants used a combination of these mnemonics.

The equipment or warning names were directly used by participants to help them memorise the sounds. For example, the name 'Oxygen' reminded some participants of oxygen 'leaking' or 'dripping' into a patient. 'Syringe' was remembered as a soft 'ss' sound and also as a tiny sound like a needle. 'Pulse' was remembered as someone running and out of breath and therefore with an increase in pulse rate. Some of the names, for example 'Aircall' (sound 1), or 'Kilo' (sound 1) were used by the participants to mimic the sound by fitting the syllables of the name to the sound made by the alarm. Many of the names used in Experiments 1 and 2 evoked visual images, for example 'Humidifier' suggested to one subject a hot sunny day with people running around while 'Infusion' reminded another subject of roadworks. Participants in Experiment 3 seemed to take longer to develop a mnemonic strategy with the neutral names which might
explain why the correct response rate was lower in Experiment 3 than for Experiments 1 and 2. The same kind of images were not evoked by the neutral names, for example participants in Experiment 3 did not use any imagery for names like Uniform and Alpha, as no images are immediately apparent, and this could explain why the results were worse for Experiment 3 in comparison to Experiments 1 and 2.

The musical and acoustic mnemonics used were similar for all three experiments. For example some participants described sound 11 as a piercing sound, other sounds were described as going from high to low or low to high, differences in pitch were used as a mnemonic strategy, and many participants recognised the 'Beethoven's Fifth Symphony' theme of sound 6.

One of the aims of the studies was to isolate the significant confusions between the warnings when learned in the laboratory under ideal conditions, in order to provide a baseline for comparison when warnings are learnt and heard under conditions in which other tasks are being performed at the same time (reported in Chapter 5). In the I.C.U. the identification of alarms would not occur in isolation. It would therefore appear relevant to examine whether identification of the sounds becomes more difficult and whether confusions between sounds change or increase when subjects are also undertaking tasks. In order to ascertain the types of tasks to use in the laboratory study a video study was undertaken to examine the activities of staff undertaken on the I.C.U., which is reported in the following study.
CHAPTER 3 An Exploratory Video Study to Examine Activities and Auditory Warnings on The Intensive Care Unit

3.1 Introduction.

The three experiments described in the previous chapter demonstrate how certain auditory sounds, that appear to be acoustically quite different, can nonetheless be confused when they are being identified in the laboratory environment. When aspects of the sounds, such as the name, were manipulated, many of the confusions endured throughout all three conditions. In the set of experiments reported, participants were asked to simply identify the auditory warnings. However, in the environment of the I.C.U. the identification of alarms does not occur in isolation. It would therefore appear relevant to examine whether identification of the sounds becomes more difficult and whether confusions between sounds change or increase when participants are also undertaking tasks. To determine which tasks should be used in a further laboratory study, it would be pertinent to use tasks that bear some resemblance to the tasks undertaken in the I.C.U. environment.

Although several studies have examined problems regarding alarms in critical care areas (e.g. O'Carroll 1986, Kestin et al. 1988, Koski et al. 1990, and Momtahan et al. 1993), they do not relate the occurrence of the alarms to either the activities of the patient or the staff. Stanton (1992) investigated the occurrence of alarms on coronary care unit (C.C.U.) over a three day period. In this study it was found that certain alarms (such as the E.C.G. alarm) were directly related to patient activity (such as washing and eating). The responses by the staff to the alarms were also reported which included checking the patient, changing the infusion and checking the E.C.G. heart-rate thresholds. However, activities being undertaken by the staff when alarms occurred were not reported.

As there appeared to be limited information regarding the relationship between the activities of staff and their responses
to auditory warnings that occur on the I.C.U., an exploratory observation study was undertaken using a video camera at the I.C.U. Derriford Hospital, Plymouth. The primary objectives of the study were to investigate the activities being undertaken by staff when the alarms occurred. Tasks could then be developed to incorporate into a future laboratory experiment (reported in Chapter 6) to examine the effect on identification of the set of auditory warnings whilst also undertaking the tasks. Further aims of the video study were to examine the frequency and duration of alarms and to determine whether the behavioural responses of the staff to the alarms changed when the unit was quiet, compared with when it was busy.

3.1.1 Observation Studies

Observation, either by observers who record the information, or by using a video camera, is used as a method of collecting overt behavioural data in laboratory, clinical and naturalistic settings. Direct observation is often used in studies involving children. For example, the assessment of children with a behavioural problem which needs to be investigated in relation to environmental factors (Cairns and Green 1979), or assessing parent and infant interactions (Baird et al. 1992), is often undertaken through direct observation. Direct and systematic observation has also been the primary methodology for assessing the social interaction of young children (Odam and Ogawa 1992).

In deciding whether direct observation is the appropriate research method to use for collecting behavioural data as opposed to other methods, for example, controlled laboratory experimentation, standardised tests and questionnaires (Sackett 1978) and self report or rating by others (Kerlinger 1973), there are important issues that need to be considered.

The distinguishing feature of direct observation is that the information is obtained directly rather than through the report of the individual. The experimenter can see what is done, rather than what is said is done and for these reasons it is suggested that
direct observation is usually more accurate and of a higher quality than recalled information (e.g. Cone 1982, Smith 1981).

Asking people to report about their behaviour in a specific situation or about their beliefs and attitudes could be inappropriate in some circumstances, for example, in the case of young children where the questions would probably not be understood (e.g. Smith 1981). This type of research is very dependent upon an individual's subjective memory, the questions are usually responded to when the subject is distant from the situation and concurrent stresses that influences what they actually do or say (Sackett 1978). The subject's ability, their willingness to respond and pressure of time may influence how accurately an individual replies to the questions. For example, some familiar, everyday activities that are taken for granted by an individual are often not reported when individuals are asked about their behaviour in everyday circumstances. Nevertheless, these activities would be recorded by an observer (e.g. Moser 1971). Often these factors mean that answers to the questions are unreliable. The fallibility of an individual's memory could cause reported data to be seriously distorted in a way that the direct observation of events, as they occur, would avoid. For example, many people report spending the majority of their time talking, however research has shown that less than 15% of daily activities involves verbal communication with a considerably higher percentage of non-verbal behaviour occurring (e.g. Birdwhistell 1970).

Some of the problems discussed above indicate that there are situations in which questioning or informal interviews are not applicable and direct observation then becomes an obvious way to proceed as suggested by Moser and Kalton (1971) and Sackett (1978). Although there are some disadvantages in using observation techniques, which are discussed in more detail later, if the research question involves complex behaviours or activities that need to be studied in an everyday life environment, then direct observational methods would appear to be the research method of choice. By undertaking an observation
study, and in particular, when the behaviour or activities are recorded on video tape, an accurate and permanent record of events is obtained (Johnson and Pennypacker 1980).

3.1.2 Sampling Strategy

Often the behaviours or activities of interest that are to be observed do not occur at regular intervals across time, which is pertinent to the occurrence of auditory warnings in an I.C.U. environment. If such activities occur infrequently then it is necessary to have longer observational periods so that these events can then be accurately recorded and ensure a representative picture concerning the pattern of behaviours. Continuous, real-time sampling may be appropriate if, for example, certain alarms, occurred at random only two or three times during a 24 hour period.

An important consideration when undertaking an observation study is to ensure that behaviours and activities are represented accurately. Some behaviours may occur frequently and have a long or variable duration, while others are momentary and may have a rapid on/off duration. Other behaviours or activities may occur less frequently and have either a long or a short duration. If a large number of categories are also to be recorded then a combination of the duration of activities described above may occur and real-time, continuous sampling may then be the only way to ensure representative sampling of all the behaviours.

As Sackett (1978) suggests observation data is usually analysed using five basic measures.

i) Examining the absolute frequency of events or behaviour. Frequency ascertains how often certain categories of activities or behaviours occur during an observation period;

ii) Relative frequency (probability). Sackett (1978) discusses how if the relative frequency of events or behaviours is calculated a difference can be shown even when the absolute frequencies for separate observation sessions are the same. Consequently, if frequencies alone were used as the measure, no
difference would be detected. The relative frequency is calculated by dividing each individual frequency by the total number of events;

iii) Absolute duration of events or behaviour indicates the length of time the event or behaviour lasts for and is usually measured in seconds.

iv) The relative duration of events (percentage total duration) measures the chances of seeing each event or behaviour at any randomly chosen second of observation. These scores are obtained by dividing each individual duration by the total session length.

v) Comparison between duration and frequency, for example, one activity may happen frequently, but not last very long, while another behaviour can occur less frequently but when it does occur it lasts for a relatively long time. This implies the need to observe for adequate periods of time to ensure that infrequent and very brief events are recorded.

By recording the activities using a video camera, long periods of continuous, real-time data can be obtained which should encapsulate all of the above categories, that is, the frequent or infrequent occurrence, together with long or short durations of activities and auditory warnings.

Recording activities as they occur in a long, continuous and uninterrupted time period together with an indication of the time that elapsed both during and between events is fairly complicated and is usually undertaken using a mechanical recording device, such as a video camera as used in the current study.

3.1.3 Video Studies

There are many studies that report the advantages of observing using audio or video recording machines over direct observation using observers (e.g. Johnson and Pennypacker 1980). These advantages include obtaining a permanent, accurate and detailed record of the events which can assist memory. With a permanent record such as a video tape, the researcher's interpretation of
for example, particular behaviours, can be confirmed by other investigators who can also analyse the tape. By increasing the observer agreement the reliability of the study is also increased.

Nevertheless, there are disadvantages of using a video camera which are discussed by Weick (1968) who states that by using certain types of lenses on a video camera the perspective can be foreshortened, for example, if a group of people are standing far apart, on film they appear to be close together. This particular disadvantage in using these types of lenses may have important implications when considering them for a study in an environment such as the I.C.U. For example, equipment and monitors may appear very close together and therefore when analysing the tape would be difficult to identify.

Another disadvantage is the time taken to analyse taped data. Typically, the field investigation is highly complex and behavioural events that may only occupy minutes in time may take many times that long to describe (Barker and Wright 1955). Wohlstein and McPhail (1979) reported that the time required for them to produce, code and analyse their tape records was 200 hours for each minute of film record. Sackett (1978) also warns that the real danger is spending an inordinate amount of time analysing the tapes.

If recording equipment is used it can become obtrusive and its presence may subsequently have an effect on subject behaviour.

3.1.4 Reactivity

One major problem with obtrusive observation is that it can be reactive, which means that the behaviour of the participants is influenced by the presence of either observers or the recording equipment (Webb et al. 1981). One example of subject reaction is reported by Belk et al. (1988) who found that when researchers integrated a video camera into the participant-observer phase of a study, some of the subjects would approach them and ask to be filmed.
It has been suggested that when people are being observed they tend to present more positive behaviours to perhaps exhibit a more favourable self-image (Roberts and Renzaglia 1965). Other studies have postulated that adults become involved in more positive play with children (Zegiob et al. 1975) and participants perform more altruistic acts when being video-taped (Samph 1969). However it is impossible to know whether these events would have occurred had there been no observers or video camera present!

Other potential sources of reactivity include how conspicuous the observer and/or recording equipment are; interaction of physical characteristics, for example age or gender of the observer or subject, and the rationale given for undertaking the observation study (Johnson and Bolstad 1973).

There are many ways suggested as to how reactivity could be minimised thereby acquiring an accurate picture of the behaviours of interest. These include the personal appearance and tact of the observers (Haynes 1978), their professional status (Wallace 1976) and the relationship, particularly that of trust, that develops with the participants (Weick 1968).

The observers must be able to articulate what they are trying to do and to show a genuine interest in what is going on. It is important to develop a rapport with the participants and an ability to interpret and describe what is observed accurately (Johnson and Pennypacker 1980).

Other ways of minimising reactivity are suggested by Foster et al. (1988). They postulate that participants should agree to be observed, but be unaware of the exact times that the observations will occur. The observers should take care to ensure they are not in the way of ongoing activities and encourage the participants to act naturally which could be achieved by providing an acceptable and non-threatening rationale for the study. The participants should also be reassured that
while the results obtained from the data collected may have important consequences, it will not affect their job.

It is difficult to assess how successful all these tactics are as there appears to be very little empirical evidence (Foster et al. 1988), to suggest that if a combination of the above suggestions are used, there will be a positive difference in the results, i.e. that the participants behave naturally. One way to examine if employing such tactics does improve the results would be to compare the same situation using direct observation and covert, surreptitious recording techniques. However, recording without subjects' consent is considered unethical and raises issues such as informed consent of the participants and invasion of the participant's privacy. Studies of this nature are rarely, if at all, undertaken nowadays.

In order to achieve many of the above suggestions to decrease reactivity, the study should involve an adequate period of investigation and not a quick, superficial visit as this could result in a seriously biased and unrepresentative picture (Moser 1971). If there are problems with the approach and ability of observers it could cause participants to feel suspicious (Johnson and Pennypacker 1980), and can also create effects such as heightened paranoia, hostility, uncertainty, together with changes in the participants' verbal behaviour and behavioural responses (Weick 1968). It could be suggested that to a certain extent the success of an observational study depends on how unobtrusive the observers or the recording equipment are.

The study reported in this chapter involved the use of a video camera which could record continuously for eight hours thereby obtaining a substantial amount of data. A single video camera was used to record the equipment and monitors in use at one bed on the I.C.U. Derriford Hospital, Plymouth.
3.2 The Current Study

3.2.1 Introduction

The primary aim of the video study was to enumerate the types of tasks undertaken by the staff in the I.C.U., the frequency and duration of alarms, and the behavioural responses by the staff to auditory alarms in the I.C.U. using an unobtrusive observational technique. The aim was to categorise the concrete activities that occur on the unit into similar, representative abstract tasks to determine which kind of tasks could be used in the prospective laboratory based experiment. It was thought that the initial approach of using a video recorder would avoid observer intrusiveness and reduce reactivity by decreasing the participant’s awareness of being observed. The video recording technique was also considered the method of choice because the films would provide a permanent record of the data to be analysed and that substantial, real time recording would ensure a sufficient range of activities and alarms would be included.

The study had to pass both the University of Plymouth’s ethics committee and Derriford Hospital’s ethics committee. A guarantee was given that the complete anonymity of both staff and patients would be protected. It was discussed with the staff that if relatives questioned the purpose of the video camera it would be strongly emphasised that it was the monitoring equipment that was being filmed and not the patient. It was agreed unconditionally that if patients or relatives were upset by the presence of the video camera, then the staff were to stop the filming and remove the camera.

3.2.2 Method

3.2.2.1 Procedure

A Panasonic WVP 100E video camera was used, which had a sensitive microphone attached as an integral part of the machine.
The camera used could record for a continuous eight hour period. The video camera was installed in a ward with four beds in the I.C.U., Derriford Hospital, Plymouth, and which is illustrated in Figure 3.1. Although there were single rooms on the I.C.U. these were primarily used for patients who were dying, children, and those who were at risk from infection. These patients were often surrounded by distraught relatives and it was considered unethical and inappropriate to undertake a video study in these rooms.

Figure 3.1: Diagram of the Four Bed Unit (not to scale)

In the early stages of the study, a wide angled lens was used to film all four beds. However, there were problems using a wide angled lens, which not only distorted the size of the four-bedded bay, but also obscured the individual detail of the equipment. Another problem transpired when the curtains were pulled around one or more of the beds, which meant that nothing could be observed for those beds. Therefore, to enable a clear view of the equipment one camera was used with the focus on one bed space.

The criteria used to determine which bed to film depended on how much equipment the patient was attached to, the patient with the
majority of equipment being selected. The video camera was attached to a high stand and positioned inside the curtains so that all the equipment could be seen. This position also ensured that the video camera was relatively unobtrusive and not in the way. The view of the equipment was checked using a small television monitor positioned on a nearby trolley.

Filming was undertaken at random times for continuous eight hour periods with a timing mechanism used for night-time, late evening and early morning recording. A representative, 24 hour time period was included in the filming, and was obtained over a 4 day, non-sequential period. The time periods filmed were as follows and it can be observed that there was some overlap in the intervals taped.

Day 1: midday to 20.00 hours  
Day 2: 16.00 hours to midnight  
Day 3: midnight to 08.00 hours  
Day 4 08.00 hours to 16.00 hours

The total amount of filming was thirty-two hours, although two of the tapes could not be analysed visually because the camera had inadvertently been moved by the staff almost at the beginning of the recording sessions. (One tape recorded the end of the bed whilst the other recorded eight hours of curtain.) However, the sounds on the unit were recorded onto the tape as the microphone picked up all the alarms on the I.C.U., not just those at the bed space being filmed.

Prior to the data being analysed, each eight hour long-play tape was re-recorded onto two four hour tapes. Each video was then time coded and the faces of the patients, where visible, were blanked out to ensure anonymity.
3.2.3 Results

3.2.3.1 Introduction

There were some methodological problems with the video study, for example, when the camera was accidentally moved, the activities of the staff were not recorded. Stanton (1992) reported that certain activities by patients, such as eating and washing, initiated alarms such as the E.C.G. alarm for example. However, in the current study the relationship between the activities of the staff and the activation of the alarms could only be inferred. For example a specific activity, such as physiotherapy, appeared to activate the ventilator alarm but the evidence to support this claim was limited because the physiotherapist and staff nurses would often be out of view of the camera and could only be heard and the activities could not be observed. When activities were being undertaken and the alarms were activated, the activities could be only be seen clearly on the tape on a few occasions and consequently the data regarding activities and behavioural responses to the alarms were limited and incomplete.

Nevertheless, as all the sounds on the unit were recorded by the microphone the results from the video films are presented as a descriptive 'audio picture' of the auditory warnings and miscellaneous sounds that occurred over the total 32 hours filming in the I.C.U.

Each video was watched in real time and each time an alarm or a miscellaneous sound occurred three pieces of information were recorded. The possible alarms were from the following pieces of equipment; ventilator, pulse oximeter, syringe pump, infusion pump, electrocardiogram (E.C.G.) and humidifier. The miscellaneous sounds were as follows; telephone, doorbell, doctors bleep; pulmonary arterial flotation catheter (P.A.F.). The pulmonary arterial flotation catheter was included under 'miscellaneous' as the sounds recorded occurred when the
keyboard was used to type in or request information and a 'beep' sounded each time a key was depressed.

1) Every alarm was identified;
2) Every miscellaneous sound as identified;
3) The onset/offset time of each alarm and each miscellaneous sound was noted.

A few of the alarms from the video tape were difficult to identify and certain sounds were initially confused by the experimenter. To help with identification, a recording of all the sounds was made (together with the spoken name after each alarm), and was used when analysing the tapes. In this way, if a particular sound was difficult to identify, the tape could be played until the sound was recognised.

The results section examines the frequency (the number of times) an alarm or miscellaneous sound occurred during the observation study and the duration (the length of time) for which an alarm or miscellaneous sound occurred. The relative frequency and the relative duration of alarms and miscellaneous sounds are also discussed. The results section is presented as follows:

i) The absolute frequency of alarms and miscellaneous sounds;
ii) The relative frequency (probability) of alarms and miscellaneous sounds, is presented as a subset of the alarms and miscellaneous sounds and their relative frequency to other alarms or miscellaneous sounds;
iii) The absolute duration of alarms and miscellaneous sounds;
iv) The relative duration of alarms and miscellaneous sounds; is presented first, as a percent of the total duration of the observational sessions and second, as a subset of the alarms and miscellaneous sounds and their relative duration to other alarms and miscellaneous sounds.
v) Comparison between the duration and frequency of the sounds. These results are presented to distinguish between activities or alarms which occur frequently, but do not last very long, and
other activities or alarms which may occur less frequently but last for relatively longer periods of time.

3.2.3.2 Absolute Frequency of Alarms and Miscellaneous Sounds

The frequency of every alarm and miscellaneous sound was obtained by writing down the name of each sound heard on the video tape. Table 1 shows the total number of sounds recorded for each session during the observation study. The number of sounds ranged from 94 (08.00hrs-16.00hrs) to 129 (Midday-20.00 hrs). A chi-square analysis conducted on the totals shows that there was no significant difference ($X^2 (df 3) = 7.073$) between the number of sounds recorded for each recording session during the observation study. This would indicate that each of the eight hour sessions recorded during the observational study were comparable in terms of the number of alarms that occurred.

<table>
<thead>
<tr>
<th>Day</th>
<th>Midday-20.00 hrs</th>
<th>16.00hrs-midnight</th>
<th>midnight-08.00hrs</th>
<th>08.00hrs-16.00hrs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Sounds</td>
<td>129</td>
<td>97</td>
<td>110</td>
<td>94</td>
</tr>
</tbody>
</table>

Table 3.1: Total Frequency of Sounds Recorded During the Observation Study

Although the numbers of alarms for each session are similar, Figure 3.2 illustrates that the overall total for each individual alarm and miscellaneous sound differs. It can be seen in Figure 3.2 that altogether the ventilator alarm occurs most frequently (99 times), while in comparison the humidifier occurs the least number of times (5) during the observation study.
Figure 3.2: Absolute Frequency for All Sounds Recorded During the Observation Study.

However this pattern is not constant for each session recorded during the observation study. Figure 3.3 illustrates the comparison between the totals for each alarm and miscellaneous sound for each session recorded during the observation study. The frequency of alarms and miscellaneous sounds differs for each of the time periods recorded, for example, Day 1 (midday-20.00 hours) the ventilator alarm (34), pulse oxymeter (36) and telephone (31) sound frequently while the pulmonary artery flotation catheter was not heard at all. The ventilator also sounded frequently during Day 2 (16.00- midnight) and Day 4 (08.00-16.00 hours). In comparison, during the recording on day 3 (midnight-08.00 hours) the most frequent sound occurring was the pulmonary artery flotation catheter (66) with the ventilator only occurring 8 times. The absolute frequency for the individual alarms and miscellaneous sounds that occurred during each observation session are shown in Tables 3.2 and 3.3.
Figure 3.3: Comparisons of the Absolute Frequency of Alarms and Miscellaneous Sounds Occurring During the Observation Study.
<table>
<thead>
<tr>
<th>Measure</th>
<th>Day</th>
<th>Ventilator pump</th>
<th>Pulse oxymeter</th>
<th>Syringe pump</th>
<th>Infusion pump</th>
<th>E.C.G.</th>
<th>Humidifier</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute</td>
<td>1</td>
<td>34</td>
<td>36</td>
<td>10</td>
<td>10</td>
<td>1</td>
<td>0</td>
<td>91</td>
</tr>
<tr>
<td>Frequency</td>
<td>2</td>
<td>26</td>
<td>1</td>
<td>22</td>
<td>12</td>
<td>9</td>
<td>1</td>
<td>71</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>6</td>
<td>7</td>
<td>15</td>
<td>3</td>
<td>1</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>31</td>
<td>6</td>
<td>16</td>
<td>10</td>
<td>1</td>
<td>0</td>
<td>64</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.2: Absolute Frequency (Number) of Individual Alarms

<table>
<thead>
<tr>
<th>Measure</th>
<th>Day</th>
<th>Telephone</th>
<th>Doorbell</th>
<th>Bleep</th>
<th>P.A.F</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute</td>
<td>1</td>
<td>31</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>38</td>
</tr>
<tr>
<td>Frequency</td>
<td>2</td>
<td>26</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>26</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>66</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>13</td>
<td>0</td>
<td>1</td>
<td>16</td>
<td>30</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.3: Absolute Frequency (Number) of Miscellaneous Sounds
3.2.3.3 Frequency of Alarms and Miscellaneous Sounds Occurring During 15 Minute Modified Frequency Scores.

The data for each session recorded during the observation study is presented as modified frequency scores (Sackett 1978). The data is presented sequentially for each eight hour period. However, the frequency of each alarm or miscellaneous sound is not plotted because of the length of each observation period. Instead, the total number of alarms and miscellaneous sounds that occurred within a 15 minute interval is presented. As described by Sackett (1978) the time interval chosen depends upon the study and may be quite arbitrary. Sackett gives the example of a 300 second session that might be broken down into 20 x15 second modified frequency scores. The modified frequency score decided upon for the current study as the most appropriate time interval to present the data was 15 minute intervals within each eight hour session. This resulted in a total of 32 time periods per session.

Figures 3.4 to 3.7 illustrate each eight hour recording session divided into 15 minute intervals. The total number of sounds that occurred within the 15 minute interval is shown. It can be seen that the alarms and sounds are not activated in a periodic pattern, but occur in clusters. The graphs demonstrate that, as Stanton (1992) found, on some occasions no alarms occur, whilst on other occasions the situation changes rapidly and several alarms occurred during one 15 minute interval. An example is illustrated during the time period midnight to 08.00 hours (Figure 3.6) the majority of sounds occurred during a 30 minute period from 00.15 to 00.45, and thereafter there are less than five sounds recorded per 15 minute interval. During the period 05.00 hours until 06.00 hours, for example, there were no sounds recorded.
Figure 3.4: Frequency of Sounds During the Time Period 12 midday to 20.00 hours (15 minute intervals)
Figure 3.5: Frequency of Sounds During the Time Period 16.00 hours to Midnight (15 minute intervals)
Figure 3.6: Frequency of Alarms and Miscellaneous Sounds During the Time Period Midnight to 08.00 hours (15 minute intervals)
Figure 3.7: Frequency of Alarms and Miscellaneous Sounds During the Time Period 08.00 hours to 16.00 hours (15 minute intervals)
Whilst a fifteen minute modified frequency score is useful to demonstrate the frequency of the alarms for the total of each observation session, it is a fairly large interval and does not demonstrate which individual alarms are occurring. A smaller, one hour interval score (between 15.00 hours and 16.00 hours) is presented in Figure 3.8 during which a total of 38 sounds occurred. (During this one hour period physiotherapy was being undertaken by a physiotherapist and a staff nurse.)

During the hour selected every alarm and miscellaneous sound was noted each time it occurred. During the 1 hour period selected each alarm and miscellaneous sound did not occur more frequently than at 1 minute intervals (although at other times in the study the interval between alarms were seconds, rather than minutes). During the hour selected it can also be seen that the duration of each alarm was less than 1 minute, the syringe pump having the maximum duration of 15 seconds.

Each individual alarm and its duration recorded during the 1 hour period is demonstrated in Figure 3.8. The one hour period was selected to demonstrate the diversity of the types of alarms that occurred during the observation study and the variable duration of the alarms and miscellaneous sounds illustrated. Figure 3.8 also shows that there were occasions when two or three alarms occurred within the same one minute interval, (for example the ventilator, pulse oxymeter and infusion pump at 38 minutes) and for some periods no alarms or miscellaneous sounds were recorded.
Figure 3.7: Frequency and Duration of Sounds During a 1 Hour Period (15.00 to 16.00)
3.2.3.4 Relative Frequency (Probability) of Sounds Occurring.

The relative frequency demonstrates the probability of each subset of alarms and miscellaneous sounds occurring in comparison to other alarms and miscellaneous sounds. The relative frequency (probability) of a sound occurring was calculated by dividing the frequency of each alarm and miscellaneous sounds by the total frequency of all sounds. The relative frequency for alarms is shown in Table 3.4 and the relative frequency for the miscellaneous sounds is shown in Table 3.5.

Sackett (1978) states that if the probability (relative frequency) of events is calculated, this can show a difference when scores are the same during different time periods. For example, during the observation session on Day 1 the absolute frequency of the ventilator alarm was 34 (Table 3.2) and the relative frequency in comparison to all other sounds that occurred during Day 1 was 26% (Table 3.4). This can be compared with Day 2 when the absolute frequency of the ventilator alarm was 26 (Table 3.2) and the relative frequency was again 26% (Table 3.4). Hence on both days there was a 26% chance that if an alarm was heard at random, it would be the ventilator alarm.

The relative frequency for the totals for each category are also shown in Tables 3.4 and 3.5 to demonstrate the probability of either an alarm or a miscellaneous sound occurring. For example on Day 2 (Table 3.4) if a sound were heard at random, there was a 73% chance that it would be an alarm, while on Day 3 there was only a 36% chance it would be an alarm (Table 3.4) but a 63% chance that it would be a miscellaneous sound (Table 3.5).

These findings would appear to demonstrate the randomness of alarms and miscellaneous sounds that occurred during the period of the observation study, despite there being no significant
The difference between the total number of alarms recorded as the four sessions observed were shown to be homogeneous.
<table>
<thead>
<tr>
<th>Measure</th>
<th>Day</th>
<th>Ventilator</th>
<th>Pulse oxymeter</th>
<th>Syringe pump</th>
<th>Infusion pump</th>
<th>E.C.G.</th>
<th>Humidifier</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative</td>
<td>1</td>
<td>26.3</td>
<td>27.9</td>
<td>7.7</td>
<td>7.7</td>
<td>0.7</td>
<td>0</td>
<td>70.5</td>
</tr>
<tr>
<td>Frequency</td>
<td>2</td>
<td>26.8</td>
<td>1.0</td>
<td>22.6</td>
<td>12.3</td>
<td>9.2</td>
<td>1.0</td>
<td>73.1</td>
</tr>
<tr>
<td>(%)</td>
<td>3</td>
<td>7.2</td>
<td>5.4</td>
<td>6.3</td>
<td>13.6</td>
<td>2.7</td>
<td>0.9</td>
<td>36.3</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>32.9</td>
<td>6.3</td>
<td>17.0</td>
<td>10.6</td>
<td>1.0</td>
<td>0</td>
<td>68.0</td>
</tr>
</tbody>
</table>

Table 3.4: Relative Frequency (Probability: Frequency/Total Number of Sounds x100) of Individual Alarms Occurring.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Day</th>
<th>Telephone</th>
<th>Doorbell</th>
<th>Bleep</th>
<th>P.A.F</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative</td>
<td>1</td>
<td>24.0</td>
<td>3.8</td>
<td>1</td>
<td>0</td>
<td>29.4</td>
</tr>
<tr>
<td>Frequency</td>
<td>2</td>
<td>26.8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>26.8</td>
</tr>
<tr>
<td>(%)</td>
<td>3</td>
<td>3.8</td>
<td>0</td>
<td>0</td>
<td>60.0</td>
<td>63.6</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>13.8</td>
<td>0</td>
<td>1.0</td>
<td>17.0</td>
<td>31.9</td>
</tr>
</tbody>
</table>

Table 3.5: Relative Frequency (Probability: Frequency/Total Number of Sounds x100) of Individual Miscellaneous Sounds Occurring
3.2.3.5 Absolute Duration of the Alarms and Sounds.

The duration of the alarms and miscellaneous sounds were measured in seconds, using the time coding on the re-recorded 4 hour tapes. The duration began with the onset of the sound and ended with the offset. The sound of the pulmonary artery flotation catheter occurred as a rapid succession of 'beeps', each of which lasted less than a second. As Sackett (1978) proposes, if individual behaviours or events occur so rapidly that they cannot be individually recorded they should be treated as one event. Subsequently, when a long, consecutive group of 'beeps' occurred which could not be individually timed they were treated as one single sound.

The absolute duration of each alarm recorded during the observation study is shown in Table 3.6 and the absolute duration of each miscellaneous sound in Table 3.7. The mean of each alarm is included to demonstrate the length of time for which each alarm and miscellaneous sound occurred during the observation study. The standard deviation and the range are included to illustrate how each duration varied, (unlike frequency, for example, which is a single, discrete event). An example of the variation in the range of the duration of the alarms is illustrated by the infusion pump (Table 3.6), in which the range is 1 to 95 seconds and the average time the alarm was activated was 8.3 seconds (standard deviation 15.176). The tables also illustrate that during the period of the observation study, overall the pulse oxymeter had the shortest mean (3.53 seconds) and the syringe pump the longest mean (9.61 seconds).

Table 3.8 demonstrates the absolute duration of the individual alarms recorded during each observation session and Table 3.9 shows the absolute duration for the individual miscellaneous sounds recorded during each observation session. For example, during the observation session on Day 1, the ventilator alarm was activated for a total of 145 seconds.
### Table 3.6: Total Duration of Alarms Recorded During the Observation Study

<table>
<thead>
<tr>
<th>Duration (seconds)</th>
<th>Ventilator</th>
<th>Pulse oxymeter</th>
<th>Syringe pump</th>
<th>Infusion Pump</th>
<th>E.C.G.</th>
<th>Humidifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>451</td>
<td>173</td>
<td>519</td>
<td>348</td>
<td>108</td>
<td>2</td>
</tr>
<tr>
<td>Mean</td>
<td>4.79</td>
<td>3.60</td>
<td>9.61</td>
<td>8.48</td>
<td>7.71</td>
<td>1</td>
</tr>
<tr>
<td>St Dev</td>
<td>5.19</td>
<td>4.16</td>
<td>10.05</td>
<td>15.32</td>
<td>5.13</td>
<td>0</td>
</tr>
<tr>
<td>Range</td>
<td>1 to 43</td>
<td>1 to 16</td>
<td>1 to 55</td>
<td>1 to 95</td>
<td>1 to 18</td>
<td>1</td>
</tr>
</tbody>
</table>

### Table 3.7: Total Duration of Miscellaneous Sounds Recorded During the Observation Study

<table>
<thead>
<tr>
<th>Duration (seconds)</th>
<th>Telephone</th>
<th>Doorbell</th>
<th>Bleep</th>
<th>P.A.F.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>326</td>
<td>5</td>
<td>18</td>
<td>194</td>
</tr>
<tr>
<td>Mean</td>
<td>4.657</td>
<td>1</td>
<td>6</td>
<td>4.85</td>
</tr>
<tr>
<td>St Dev</td>
<td>4.775</td>
<td>0</td>
<td>2</td>
<td>9.87</td>
</tr>
<tr>
<td>Range</td>
<td>1 to 26</td>
<td>1</td>
<td>4 to 8</td>
<td>1 to 49</td>
</tr>
<tr>
<td>Measure</td>
<td>Day</td>
<td>Ventilator</td>
<td>Pulse oxymeter</td>
<td>Syringe pump</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----</td>
<td>------------</td>
<td>----------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Duration</td>
<td>1</td>
<td>145</td>
<td>87</td>
<td>61</td>
</tr>
<tr>
<td>(seconds)</td>
<td>2</td>
<td>70</td>
<td>4</td>
<td>142</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>139</td>
<td>36</td>
<td>212</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>97</td>
<td>46</td>
<td>104</td>
</tr>
</tbody>
</table>

Table 3.8: Absolute Duration (Seconds) for Alarms During the Each Observation Session.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Day</th>
<th>Telephone</th>
<th>Doorbell</th>
<th>Bleep</th>
<th>P.A.F</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration</td>
<td>1</td>
<td>97</td>
<td>5</td>
<td>10</td>
<td>0</td>
<td>112</td>
</tr>
<tr>
<td>(seconds)</td>
<td>2</td>
<td>143</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>143</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>61</td>
<td>0</td>
<td>8</td>
<td>123</td>
<td>192</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>25</td>
<td>0</td>
<td>0</td>
<td>71</td>
<td>96</td>
</tr>
</tbody>
</table>

Table 3.9: Absolute Duration (Seconds) for Miscellaneous Sounds During Each Observation Session.
3.2.3.6 Relative Duration (Percent Total Duration) of All Sounds.

The relative duration for each sound in relation to all other sounds that occurred during each of the observation session was calculated by dividing the individual duration of each alarm and miscellaneous sound by the total duration of all sounds occurring.

<table>
<thead>
<tr>
<th>Day</th>
<th>Total Duration Alarms</th>
<th>Total Duration Miscellaneous Sounds</th>
<th>Total Duration All Sounds (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>383</td>
<td>112</td>
<td>495</td>
</tr>
<tr>
<td>2</td>
<td>322</td>
<td>143</td>
<td>465</td>
</tr>
<tr>
<td>3</td>
<td>443</td>
<td>192</td>
<td>635</td>
</tr>
<tr>
<td>4</td>
<td>452</td>
<td>96</td>
<td>548</td>
</tr>
</tbody>
</table>

Table 3.10: Absolute Duration of Alarms and Miscellaneous Sounds (Seconds).

The relative duration (percent total duration for all sounds) for the alarms is shown in Table 3.11 and the miscellaneous sounds are shown in Table 3.12.
<table>
<thead>
<tr>
<th>Measure</th>
<th>Day</th>
<th>Ventilator</th>
<th>Pulse oxymeter</th>
<th>Syringe pump</th>
<th>Infusion pump</th>
<th>E.C.G.</th>
<th>Humidifier</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Total</td>
<td>1</td>
<td>29.2</td>
<td>17.5</td>
<td>12.3</td>
<td>18.1</td>
<td>0</td>
<td>0</td>
<td>77.3</td>
</tr>
<tr>
<td>Duration</td>
<td>2</td>
<td>15</td>
<td>0.86</td>
<td>30.5</td>
<td>6.0</td>
<td>16.5</td>
<td>0.2</td>
<td>69.2</td>
</tr>
<tr>
<td>(seconds)</td>
<td>3</td>
<td>21.8</td>
<td>5.6</td>
<td>33.3</td>
<td>7.2</td>
<td>1.5</td>
<td>0</td>
<td>69.7</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>17.7</td>
<td>8.3</td>
<td>18.9</td>
<td>33.5</td>
<td>3.6</td>
<td>0.1</td>
<td>82.7</td>
</tr>
</tbody>
</table>

Table 3.11: Relative Duration (seconds) of Alarms (Duration/Total Duration of All Sounds)

<table>
<thead>
<tr>
<th>Measure</th>
<th>Day</th>
<th>telephone</th>
<th>doorbell</th>
<th>bleep</th>
<th>P.A.F</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Total</td>
<td>1</td>
<td>19.5</td>
<td>1.0</td>
<td>2.0</td>
<td>0</td>
<td>22.6</td>
</tr>
<tr>
<td>Duration</td>
<td>2</td>
<td>30.7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>30.7</td>
</tr>
<tr>
<td>(seconds)</td>
<td>3</td>
<td>9.6</td>
<td>0</td>
<td>1.2</td>
<td>19.3</td>
<td>30.2</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4.5</td>
<td>0</td>
<td>0</td>
<td>12.9</td>
<td>17.5</td>
</tr>
</tbody>
</table>

Table 3.12: Relative Duration (seconds) of Miscellaneous Sounds (Duration/Total Duration of All Sounds)
3.2.3.7 Relative Duration (Percent Total Duration) of Total Session Time.

The relative duration (percent total duration of total session time) as described by Sackett (1978) was calculated by dividing the duration of each individual sound by the total session time.

The individual durations were divided by the total observation period for each day (8 hours) which was converted into seconds (8x60x60=28800 seconds). The calculations for the total observation period differ from the calculations described above (relative duration of each sound to total duration all sounds) as it examines the possibility of hearing an individual sound at any random second of the total observation session. For example, during Day 1 (Table 3.13) the ventilator alarm occurred for 0.5% of the 8 hour observation period. The syringe pump on Day 3 (Table 3.14) alarmed for 0.73% of the 8 hour time period. The total column in Table 3.13 shows that overall the alarms sounded for 1.32% of the total 8 hour session for Day 1, which indicates that for 98.68% of the time on Day 1 there were no alarms recorded. The results of these calculations are interesting, as they demonstrate that all the sounds occur for only a brief period of time during the total observation study.
<table>
<thead>
<tr>
<th>Measure</th>
<th>Day</th>
<th>Ventilator</th>
<th>Pulse oxymeter</th>
<th>Syringe pump</th>
<th>Infusion pump</th>
<th>E.C.G.</th>
<th>Humidifier</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Total</td>
<td>1</td>
<td>0.50</td>
<td>0.30</td>
<td>0.21</td>
<td>0.31</td>
<td>0</td>
<td>0</td>
<td>1.32</td>
</tr>
<tr>
<td>Session</td>
<td>2</td>
<td>0.24</td>
<td>0.01</td>
<td>0.49</td>
<td>0.09</td>
<td>0.26</td>
<td>0.003</td>
<td>1.11</td>
</tr>
<tr>
<td>Time</td>
<td>3</td>
<td>0.48</td>
<td>0.12</td>
<td>0.73</td>
<td>0.15</td>
<td>0.03</td>
<td>0</td>
<td>1.54</td>
</tr>
<tr>
<td>(seconds)</td>
<td>4</td>
<td>0.33</td>
<td>0.15</td>
<td>0.36</td>
<td>0.64</td>
<td>0.06</td>
<td>0.003</td>
<td>1.56</td>
</tr>
</tbody>
</table>

Table 3.13: Relative Duration (Percentage) of Alarms (Duration/Total Session Time).

<table>
<thead>
<tr>
<th>Measure</th>
<th>Day</th>
<th>Telephone</th>
<th>Doorbell</th>
<th>Bleep</th>
<th>P.A.F</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Total</td>
<td>1</td>
<td>0.33</td>
<td>0.17</td>
<td>0.03</td>
<td>0</td>
<td>0.38</td>
</tr>
<tr>
<td>Session</td>
<td>2</td>
<td>0.49</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.49</td>
</tr>
<tr>
<td>Time</td>
<td>3</td>
<td>0.21</td>
<td>0</td>
<td>0.02</td>
<td>0.42</td>
<td>0.66</td>
</tr>
<tr>
<td>(seconds)</td>
<td>4</td>
<td>0.08</td>
<td>0</td>
<td>0</td>
<td>0.24</td>
<td>0.33</td>
</tr>
</tbody>
</table>

Table 3.14: Percentage Duration of Sounds (Duration/Total Session Time).
3.2.3.7 Comparison Between Frequency and Duration Data.

The final section of the results examines the comparison between frequency and duration of alarms and miscellaneous sounds. In the introduction to this chapter it was proposed that an important issue to consider when undertaking an observation study is to ensure that behaviours and activities are represented accurately. It was discussed how some behaviours may occur frequently and have a long or variable duration, while others are momentary and may have a rapid on/off duration. Other behaviours or activities may occur less frequently and have either long or short durations. The important factor is that the observation session must be long enough to observe the momentary events with rapid on/off durations.

Table 3.15 shows the comparison between the frequency and duration of alarms and Table 3.16 demonstrates the comparison between the frequency and duration for the miscellaneous sounds. The observation sessions did include infrequently occurring alarms and miscellaneous sounds such as the humidifier alarm, the doorbell and the bleep. The humidifier alarm as well as occurring infrequently also had a very brief duration. It was activated twice, and the absolute duration was 2 seconds. The ventilator alarm had the highest absolute frequency overall (99) but not the longest duration (451 seconds). The longest absolute duration is shown to have been the syringe pump with a total of 519 seconds but a total frequency count of only 55 recorded occurrences.

These results would seem to add support to the argument that an adequate period of observation is very important in order to record infrequent alarms with brief durations such as the humidifier and that within a total observation period of 32 hours it was recorded as occurring only twice and with a very rapid on/off duration of 2 seconds.
<table>
<thead>
<tr>
<th>Measure</th>
<th>Ventilator</th>
<th>Pulse oxymeter</th>
<th>Syringe pump</th>
<th>Infusion pump</th>
<th>E.C.G.</th>
<th>Humidifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration (seconds)</td>
<td>451</td>
<td>173</td>
<td>519</td>
<td>348</td>
<td>108</td>
<td>2</td>
</tr>
<tr>
<td>Absolute Frequency</td>
<td>99</td>
<td>49</td>
<td>55</td>
<td>47</td>
<td>14</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 3.15: Comparison Between Frequency and Duration of Alarms

<table>
<thead>
<tr>
<th>Measure</th>
<th>Telephone</th>
<th>Doorbell</th>
<th>Bleep</th>
<th>P.A.F.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration (seconds)</td>
<td>326</td>
<td>5</td>
<td>18</td>
<td>194</td>
</tr>
<tr>
<td>Absolute Frequency</td>
<td>74</td>
<td>5</td>
<td>3</td>
<td>82</td>
</tr>
</tbody>
</table>

Table 3.16: Comparison Between Frequency and Duration of Miscellaneous Sounds
3.3 Discussion

Although there were various methodological problems with the video study, which will be discussed in detail later, an accurate audio representation of the frequency and duration of alarms and miscellaneous sounds that occurred during the observation period was obtained. However, the data regarding the activities and behavioural responses of the staff when alarms occurred was unsatisfactory because for the majority of the time the responses and activities could not be seen and subsequently one of the aims of the video study, to establish the types of tasks being undertaken by the staff when alarms were activated, was not accomplished.

The microphone on the camera clearly recorded all the noises on the unit, including the telephone and the doorbell, which were some distance from the four bedded unit and the camera's microphone. As all the sounds were clearly recorded onto all the tapes the overall presentation of sounds that occurred on the unit during the period of the observation study could be examined. The results have therefore been presented as frequency and duration of the auditory warnings during continuous 8 hour periods for a total of 32 hours over 4 non-consecutive days. The four sessions observed were shown to be homogeneous in that there were no statistical differences between the number of alarms that occurred during each period recorded. However, there were differences in the types of alarms that occurred during each session and it is probable that these differences are linked to the activities of the staff and procedures being undertaken by staff with the patient.

The results of the frequency data demonstrate that during the period of the observation study there was not a predictable pattern for the occurrence of alarms. Some alarms were activated more frequently than others, for example, overall, the ventilator alarm occurred frequently, while the humidifier and the E.C.G. monitor alarm were activated less frequently overall. The pattern varied from session to session. For example during
Day 3 the ventilator was heard 8 times while the pulmonary artery flotation catheter was heard 66 times in comparison to Days 1 and 2, when the pulmonary artery flotation catheter was not heard at all.

As Stanton (1992) reported in his study on a C.C.U. there were periods when there were no alarms or miscellaneous sounds recorded. The frequency data for the current study also suggests that sounds appear to occur in groups with intervals when no sounds occur. An example is illustrated during the time period midnight to 08.00 hours when the majority of sounds occurred during a 30 minute period from 00.15 to 00.45 and thereafter there were less than five sounds recorded per 15 minute interval. During the period 05.00 hours until 06.00 hours for example, there were no sounds recorded until early morning when the lights were turned up and staff began undertaking various procedures with the patient.

O'Carroll (1986) reported that there was little variation in the overall frequency of alarms occurring during the day or night. In the current study in the example described above during the time period midnight to 08.00 hours the majority of alarms occurred at the beginning of the observation period and there were long periods of up to 1 hour when no alarms were recorded. However, for this particular session recorded one reason may have been that there were no activities being undertaken with the patients during the night and early hours of the morning (the video showed that the lights were turned down and the unit was generally very quiet during the later part of the night and early hours of the morning). Whether or not activities are undertaken by staff during the night would, of course, depend on the condition of the patient. If a seriously ill patient needed various procedures throughout the night then it is possible that the pattern of alarms would follow a similar daytime pattern. This, however, was not observed in the current study.
As in the current study, O'Carroll (1986) reported that the ventilator alarm occurred more frequently than other alarms. One explanation for this may be that although there was a delay time of approximately 2 minutes before the alarm sounds again after having been silenced, the ventilator alarm cannot be turned off unlike other alarms such as the E.C.G. alarm. It was observed from the tapes in the current study that occasionally the alarms on the E.C.G. monitors were inactivated, particularly if the patient had a persistent cardiac arrhythmia which would entail the monitors alarming continuously. This interpretation may also explain the fact that in the current study the E.C.G. is not a frequently sounding alarm.

The ventilator alarm also appeared to be activated frequently when the physiotherapist was undertaking physiotherapy with the patient. However, because this could not be directly observed from the tape (because the physiotherapist and trained staff were often out of view of the camera and could only be heard,) this relationship between the presence of the physiotherapist and the ventilator alarm sounding is inferential.

The absolute duration data for all of the categories, either alarms or miscellaneous sounds, demonstrated that the overall duration for all sounds was generally very short. The longest duration recorded during the observation study was for the infusion pump (95 seconds). The means for the duration of the alarms range from 3.60 seconds for the pulse oxymeter to 9.61 seconds for the syringe pump. The difference in these durations might be explained by the fact that a pulse oxymeter alarm can be silenced more quickly than a syringe pump, possibly because the problems presented by a syringe pump take longer to determine than those of a pulse oxymeter. The mean durations for the miscellaneous sounds, were also brief, ranging from 4.65 seconds for the telephone to 6 seconds for the doctor's bleep.

The relative duration of alarms and miscellaneous sounds that occurred during the total session time also demonstrates how briefly the alarms occurred during each 8 hour session. For
example the ventilator alarm during Day 1 sounded for 0.50% of the total session time. Similarly, the pulse oxymeter was heard for 0.30% of the 8 hour session. This implies that although alarms would appear to occur quite frequently if looking at Figures 3.4 to 3.7 for example, their total duration was very brief. In most instances they only occurred for a few seconds.

As the sounds occurred for a brief period of time the observation sessions of 8 hours appears to have been an adequate time sampling period. This is also demonstrated in the comparisons between frequency and duration data as alarms and miscellaneous sounds with brief durations such as the humidifier which occurred only twice and with a very rapid on/off duration of 2 seconds, were recorded during a total observation period of 32 hours.

As already discussed, although one of the aims of the video study was to observe the activities of the staff and their behavioural responses when alarms sounded, this objective was not achieved satisfactorily. There were several reasons for this, including the unintentional movement of the camera which resulted in the end of a bed or the curtains being filmed for several hours. Although a wide-angled lens was experimented with during the preliminary stages of the study this proved to be unacceptable as the equipment could not be viewed satisfactorily, and as the curtains were often pulled around a patient's bed when a particular procedure necessitated the privacy of the patient, nothing could then be observed for that bed space.

When procedures such as turning the patient or when drugs were being administered via the syringe or infusion pumps, the view of the equipment was often obscured by the staff as they undertook these procedures, so if staff did respond to alarms the response could not be seen on these occasions. This meant that it was difficult to observe the activities and behavioural responses of the staff when the alarms occurred, although they were visible on some occasions.
On the few occasions when behavioural responses were observed on the tape, it could be seen that the staff immediately silenced the alarms. It could be suggested that one of the reasons why staff silenced the alarms so rapidly, is the annoyance and irritation caused by many of the sounds as is widely reported in the literature concerning alarms (e.g. Schmidt and Baysinger 1986, Samuel 1986, Edworthy 1994) and as discussed in Chapter 1. Many of the staff involved in the current study expressed irritation at the disagreeable sound of many of the alarms.

Unwanted noise may also cause distress to the patients (Turner et al. 1975). The psychologically harmful effects of noise on performance are well documented and it could add to the stressors of an already stressful job. Noise may also distract individuals and may mask other important sounds. A non-essential noise such as the 'beep' from the pulmonary flotation artery catheter every time data is entered, is an unnecessary addition to the noise on the I.C.U.

Which alarms occur appear to depend to some extent on the activities being undertaken by staff. In a similar way Stanton (1992) found that certain activities by patients, such as eating and washing, initiated alarms, for example the E.C.G. alarm. However, in the current study the relationship between the activities of the staff and the activation of the alarms could only be implied, as in the example of physiotherapy and the ventilator alarm described above. This was because when activities were being undertaken and alarms were activated, the activities could only be seen clearly on the tape on a few occasions, and consequently the data regarding activities and behavioural responses to the alarms was limited.

There may be various reasons to explain the presence of staff when an alarm is activated. One reason, for example, may be artefact, if a patient is moving and the electrodes from the E.C.G. monitor fall off this would then cause the monitor to alarm. Another reason may be a medical problem, an example being the syringe pump or infusion pump may have no drugs or fluid left in
them and the staff may be replacing either the infusion or replacement drugs in the syringe pump.

As discussed in the introduction to this chapter the time taken to analyse video data (e.g. Sackett 1978; Wohlstein and McPhail 1979) is often extremely time-consuming. This was certainly true in the current study as each video was analysed in real-time even when there was no appropriate picture on the tape, as the sounds could be clearly heard. Another time consuming activity involved the immediate identification of a sound, for example, if a sound occurred for as briefly as 1 second it was occasionally difficult to identify the sound instantaneously. Identification would then entail identifying these sound by listening to a pre-recorded tape of the sounds from all the various items of equipment (with the spoken name after each sound on the tape) in order to correctly identify them.

There is the possibility that the presence of the video camera altered the responses of the staff they may, for example, have answered the alarms more quickly than usual. However, the staff did not know when the timer on the video camera was set by the experimenter to start filming as suggested by Foster et al. (1988) as a way of minimising reactivity, so it is feasible to infer that the staff did forget about the camera, particularly if they did not know when it was filming. The camera remained on the ward in between filming so it is also possible that the staff habituated to its presence. Feedback from the staff after completion of the study suggests that they did appear to forget about the video. A good rapport was developed with the staff, who were very interested in the rationale for the study and had many positive suggestions and opinions to offer.

While the video study presents a representative 'audio picture' of a non-consecutive 32 hour period, it was not possible to establish which events and activities were occurring on the unit when the alarms sounded. For example, as already discussed, it could be inferred that when the physiotherapist was on the unit the ventilator alarm sounded considerable more than when
physiotherapy was not being undertaken. However as this relationship could not directly be observed from the tape it was decided to conduct an observation study in which two observers directly observed the activities of the staff and corresponding alarms on the I.C.U. which will be described in the following chapter.

Other studies have found that video studies can sometimes yield results that are less effective than direct, 'live' observation. For example, Bench, Hoffman and Wilson (1974) and Bench and Wilson (1975, 1976) compared 'live' versus video recording on neonates and infants and found that little or no information was lost using live observation, which they suggest may in fact be superior despite the opportunity for repeated observation available when video recording is used. Sackett (1978) argues that even in a restricted setting, a video tape can miss much that could be observed live. This view is also supported in a study by Kihlgren et al. (1993) who video recorded interactions between caregivers and demented patients. One of the behaviours to be observed was 'eye contact' which the authors of the paper state could not be observed adequately as only one camera was used. It is possible that if 'live' observation had been undertaken the behavioural measure could have been observed. In the present study activities and behaviours could not be adequately observed. Therefore, the direct observation study reported in the following chapter will concentrate on activities of staff and the relationship of those activities when alarms are activated. Detailed information about duration, which in general occurred in seconds, and frequency, which demonstrated in general that alarms occur at random, have been obtained in sufficient detail during the current study.
CHAPTER 4 The Development and Implementation of a Direct Observation Procedure for Use in the I.C.U.

4.1 Introduction

The primary objectives of the video study described in the preceding chapter were first to identify activities being undertaken by the nursing staff on the I.C.U. when auditory warnings were activated, and secondly to determine the frequency and duration of the auditory warnings therein. However, due to the methodological problems encountered during the video study the study was limited in its success in isolating the tasks undertaken by the staff. In order to identify the activities it was therefore decided to undertake a direct observation study of the I.C.U. using two observers.

The specific objective of the current observation study was to determine the types of tasks the staff on the I.C.U. carried out in their day-to-day routine, and which alarms occurred during these tasks in order to ascertain more about the activities of staff than the video study revealed. The aim was to categorise the concrete activities that occur on the unit so that similar, representative abstract tasks could be used in the laboratory-based study reported in Chapter 5.

In using individual observers to record data instead of a video camera it would be impractical to observe every antecedent event, the behaviour of interest and the consequences of that behaviour, that occur in a complex situation. Very large amounts of information would be impossible to record accurately and consequently, before commencing an observational study, it should be decided how to answer the research questions. For example, which behaviours or activities should be observed, how the behaviours or activities are to be sampled, how to ensure that the behaviours or activities are measured accurately and possibly most importantly, how the data is to be analysed and interpreted (Johnson and Pennypacker, 1980). These issues are discussed in the following sections.
4.2 Which Observational Technique to Use: Participant or Passive Observation?

As discussed in the previous chapter, the presence of recording equipment may influence the behaviour that is being observed. Similarly, in some situations the presence of one or more observers could also affect the behaviour being observed (Moser 1971, Kazdin 1982). If an observer could blend into the environment being studied then the participants should be less aware of being observed and a more natural, authentic and less self-conscious picture should emerge. This type of observation is known as participant observation and the observer joins in the daily life of a community, village or workplace and watches what happens, how individuals behave and engage in conversations with them to find out their reactions to and interpretations of the events that have occurred. In an open community such as a factory, town or village observers can blend in and may not be noticed, while in a closed community, such as a family, it would be impossible for them not to be noticed.

One disadvantage of participant observation is that depending on the role the observer adopts, an understanding of the total situation or environment can be restricted. This was described by Riley (1963) as bias-viewpoint effect, and an example relevant to the I.C.U. environment would be if the participant observer was a patient, the viewpoint would be considerably different to that of a member of the nursing staff and different again to that of a physiotherapist. A single participant observer would be unable to become involved in all the different levels necessary to acquire a complete overview of the workplace or community being studied.

Another disadvantage of participant observation is that if the observer's role involves them becoming well integrated into the community then everyday, regular events may become taken for granted and only unusual events recorded. This is an important factor in favour of using observers from outside the environment.
being studied for, at least at the start of the research programme, the events will be seen dispassionately (e.g. Moser 1971).

An alternative method to participant observation is passive, direct observation that involves observers recording behaviour, events, interactions between the behaviours and sequencing of events with no interaction with the participants. If observers attempt to be as unobtrusive as possible, participants appear to adapt quickly to the observer's presence (Kerlinger 1973). Other researchers have also argued that observers appear to have little effect on the situations they observe (e.g. Heynes and Lippitt 1954). The problems of obtrusive observation techniques are discussed in more detail later in this chapter.

4.2.1 Summary

It was decided that participant observation would be inappropriate for the current study. One reason for this decision was that it would be difficult for an observer to completely integrate in the I.C.U. environment as a member of staff. The first priority for all staff is to care for the patient. It is difficult to imagine how an observer, whose first priority would be the collection of data, rather than the well-being of the patient, could integrate recording the activities of the staff, the occurrence of alarms and also participate in the care of a patient. The data would probably have to be recorded retrospectively, with the associated problems of forgetting information. Because of these problems, passive, direct observation was chosen as the methodology of choice.

4.3 Deciding to use Direct Passive Observation: What to Observe?

In most day to day environments there are a considerable number of activities and behaviours occurring, therefore, it is necessary to establish which behaviours or activities should be observed, the most appropriate way to observe the selected behaviours or activities, and the length of the observational sessions (Bramlett and Barnett 1993).
4.3.1 Informal and Systematic Observation

Sommer and Sommer (1980) suggest that there are two methods of sampling for use during observation studies, casual or informal observation and systematic observation. Many observation studies begin with informal observations that are undertaken without pre-arranged categories or scoring systems prior to the systematic phase of a study being undertaken. The casual phase involves watching behaviours to understand what occurs in particular situations and, for example, making notes with a pad and pencil. The casual observation stage is also useful for training new observers and habituating participants to the presence of observers (Rosenblum 1978).

The descriptive, casual observations help to develop ideas about the categories into which particular behaviours or activities should be classified for the systematic observation phase. For example, Bayles (1950) developed his list of categories for analysing small group discussion by, in the first instance, watching how the groups conversed.

In the current study in order to classify the activities and alarms that occurred on the I.C.U. into categories, an informal, casual observation phase was undertaken together with a formal, systematic observation phase.

4.4 Classification of Data.

It is necessary to classify the data into categories so that it can then be analysed. Systematic coding systems have been shown to be the most statistically sound method for observational research (Spain and Hollenbeck 1975). Bakeman (1978) suggests that there are four main ways in which data can be represented.
Type I Data is event-based and sequential. This means that behaviours are defined so that only one behaviour can occur at a time (mutually exclusive categories), and the duration of the event or behaviour is not recorded. The observer records behaviours or activities that can only occur sequentially, not together and disregards their duration.

Type II Data consists of activities or behaviours that can occur together concurrently, the behaviours or activities are not mutually exclusive. The duration of the events or behaviours is not recorded. The observer records behaviours or activities, some of which occur together and the durations of the behaviours or activities are disregarded. Bakeman suggests that this data type represents the essence of everyday life but is difficult to analyse statistically because potentially there are so many behaviours to record.

Type III Data is time-based sequential data in which the categories are mutually exclusive and the duration of each behaviour or event is recorded. The observer records behaviours or events that can only occur one at a time. The duration of the behaviours or events is usually recorded within a time interval.

Type IV Data is time-based and comprises of categories that can occur together and overlap. The duration of behaviours or activities is also recorded. This data type is very flexible and is characteristically is used for studies investigating a large number of activities.

For the current study Type IV data was collected in which alarms and the activities of the staff could occur within the same recording interval. For example, if the ventilator alarm occurred, this did not preclude the occurrence of the syringe pump or the E.C.G. alarm occurring. This data type was chosen specifically because one of the aims of the current study was to obtain an overview of the activities, many of which can occur at the same time, together with occurrence of auditory warnings that occurred on the I.C.U. The duration of events was recorded in two minute intervals.
4.4.1 Molecular and Molar Categories

The categories can further be divided into molecular and molar descriptions of the behaviour being observed. Molecular categories are very specific definitions of each type of behaviour that occurs, for example if a study was examining the prevalence of aggression in the playground, a specific incidence of aggression would be defined, e.g. hitting, punching, pushing and so on. One disadvantage of using a molecular taxonomy is that judgements and interpretations are often involved because a particular response is not always obvious; for example was the behaviour in the playground a push or a punch?

Molar categories of behaviour combine a number of actions into one category. For example aggression would be one category of playground behaviour and that category would include all incidences of aggressive behaviour. The observers would need to be clear about which category a certain behaviour belonged to and again one of the problems with molar categories is that a certain amount of interpretation and judgement may be used (Sackett 1978).

When using a molar categorisation it is important to ensure that the categories are neither too vague nor too specific. If categories are too vague without adequate specification of exactly what the observer should be recording, this may result in problems for the observer. For example different observers may have different interpretations of the same behaviour. Categories that are too specific will decrease ambiguity and uncertainty but may then be too rigid, inflexible, and even trivial (Kerlinger 1973). The data type used depends on the research question being addressed and whichever type of category is used it must define accurately and specifically what is to be recorded. This minimises idiosyncratic interpretations and discrepancies between the explicit written definition and the implicit working definition used by the observer in the field (Foster et al. 1988).
When defining the categories or behaviours to be observed another consideration is that of how many should be recorded. Some researchers suggest that 8 categories is the maximum number that should be used (e.g. Mash and McElwee, 1974) while others suggest as many as 14 can be reliably recorded (e.g. Frame 1979). The actual number of categories probably depends on the research program and the situation in which it is being used. The number of categories can be simplified, for example, by combining categories as in a molar categorisation of behaviours or activities.

Because on an average I.C.U. there are many activities, the current study used a molar classification of categories which combined a number of activities into one category. A similar strategy was used to classify the alarms, because equipment such as ventilators and syringe pumps is produced by different manufactures they subsequently have different alarms. Therefore, the same types of equipment (e.g. all syringe pumps) were classified into one molar category, as in the video study. Using a molar procedure to group the activities and alarms also simplified the number of categories for the observers to record.

4.4.2 Focal Sampling

There are different ways in which the behaviour or activities of interest can be recorded. When observing a situation that involves more than one individual, one subject could be observed at a time. This method is known as focal subject sampling (Sackett 1978). All the behaviours and interactions of the target individual are recorded during an observational period and all other behaviours are ignored. Sackett suggests that this method generates representative behaviours for the individual and the group as a whole when;

i) Each individual in the group acts as the focal subject;
ii) The identity and behaviours of each interaction with the focal individual are recorded;
iii) There are a large number of scoring sessions for each focal individual.
iv) The behaviours are fairly stable over observational sessions.

Sackett further suggests that the method is not suitable if the individuals in the group continually interact together and that simultaneous observation of all the individuals needs to be undertaken as often happens in the I.C.U. If there are more than a few categories to score, it then becomes very difficult for the observer to record the behaviours accurately. An obvious solution to the problems of large numbers of categories plus several participants would be to video-tape the sessions (which was attempted in the first observation study, see Chapter 3).

4.4.3 Summary

In the current study it was decided to undertake both a casual observation phase and a systematic observation phase. For the systematic observation phase Type 1V data was chosen because the specific aim of the observation study was to examine which activities occurred when the auditory warnings were activated on the I.C.U. As there are many different activities and behaviours occurring on the I.C.U. a molar classification of events was used so that there was a minimum number of categories to record, thereby increasing the reliability of the study. A molar classification was also used for the alarms. As the staff in the I.C.U. continually interact together and simultaneous observation of all the individuals needed to be undertaken together with several categories to record, focal sampling was not undertaken.

4.5 Deciding on a Sampling Strategy

When conducting observational research there are different temporal sampling strategies that can be employed. These include a continuous period of recording of behaviours occurring in real-time as described in the previous chapter, interval recording in which the data is collected in successive time intervals with only the first occurrence of the behaviour being recorded, and event recording in which the activities or behaviours are recorded as they occur. To guarantee that behaviours are represented accurately
depends on which type of sampling strategy is used, for example if the data is recorded on video tape, as was undertaken in the video study described in the previous chapter, then long periods of real-time activities and behaviours can be recorded. However, this is not so easy if the data is being recorded by hand.

If real time continuous sampling is not practical then a sufficiently large sample of behaviour should be recorded by the observers to try to obtain a representative overview of the behaviours occurring. Generally, the less a particular behaviour of interest occurs, the longer the observational session should be as incomplete observations contribute towards sampling error (Johnson and Pennypacker 1980, Moser 1971) and infrequent behaviours or activities will not be observed.

4.5.1 Real-time Sampling

As discussed in the previous chapter, activities and behaviours of interest that are to be recorded do not always necessarily occur at regular intervals across time. Some behaviours occur infrequently and may have either a long or variable duration, while other activities are momentary and may have a rapid on/off duration.

Real time sampling was not considered suitable for the current study for various reasons. For example, when there were distressed relatives on the ward, it was difficult to remain unobtrusive and the observers felt it was inappropriate to continue recording data on the ward. It also proved difficult to maintain the high level of concentration necessary to observe each activity, and to record both the activities and the occurrence of the alarms. Other circumstances that precluded the use of continuous real-time sampling were, for example, when the curtains were pulled around one or more of the beds, making observation infeasible. Other strategies were therefore considered and are described below.
4.5.2 Time Sampling

The recording of behaviours or events during a pre-determined time period is known as time sampling. Often the pattern of this type of study follows the following format. The duration of an observational phase is usually brief, sometimes less than one minute, while the non-observational phase could be 10 minutes or more. An example is given by Kerlinger (1973) in which observation of certain behaviours is sampled in either a systematic or random time period. If the behaviours were selected in a random way then an example could be the following scenario. The observations could occur for five minutes at any time during a five hour period with a total of six five-minute sessions. If the sampling is systematic then the five minute observations will occur at pre-determined, fixed times during the five hour period.

The use of this type of sampling strategy increases the probability of obtaining representative samples of behaviour, but only if the behaviours happen comparatively frequently. Behaviours that occur infrequently may not be sampled. Some of the disadvantages of time sampling are lack of continuity, lack of adequate context and lack of naturalness (Kerlinger 1973).

4.5.3 Event sampling

In event sampling the observer records the occurrence of a specific behaviour when it occurs (e.g. temper tantrums). The observer must either know when such events are going to occur or wait until they do occur. This method is particularly suitable for participant observers. However, event sampling usually involves only one target behaviour so other information is lost, for example the relationship with other behaviours that may occur and, if the time of occurrence is not recorded, temporal patterns cannot be determined (Cone and Foster 1982). However, in event sampling the recording of the behaviour of interest has continuity and is therefore more realistic than time sampling, for example if the occurrence of certain behaviours is infrequent. The advantage of
event sampling is that the behaviours occur in natural lifelike situations, unlike time sampling, and is advantageous if a rough guide to the frequency of certain behaviours is required.

4.5.4 Interval recording

The method of choice for the current study was interval recording of the activities and alarms that occurred. The purpose of interval sampling is usually to estimate the frequency or duration of events rather than to provide an exact record of their occurrence in real time (Sackett 1978). Observations are recorded in sessions that approximate real time recording. In interval recording the observation period is divided into arbitrary lengths that are usually brief, ranging from 10 seconds to 2 minutes (e.g. Sellltiz et al. 1959). The observers then record whether the behaviour occurred during that time. The observers have a cueing device, for example a stopwatch, to indicate the beginning and the end of the observation period. Advantages of interval recording are that several behaviours can be recorded simultaneously and observer agreement can be pinpointed.

While interval sampling is useful in certain conditions (e.g. if behaviours change so rapidly that the observers cannot keep up with them), one of the problems with this method is that the specified interval may include any number of occurrences of the particular behaviour, whereas only one is recorded and therefore gives no clear interpretation of the actual frequency or duration of the behaviour categories.

While Johnson and Pennypacker (1980) quite rightly argue that ideal observation is continuous and complete;

'...in science convenience cannot be defended at the expense of full and precise measurement of subject matter' (page 150)

It could be argued, however, that this statement does not consider the fact that after a considerable period of observation, the observer could succumb to boredom or fatigue that could then contribute to observer unreliability (Smith 1981).
The advantages of using a modified time sampling strategy are summarised by Sackett (1978), who states that the demands on observer energy and concentration are less. This then suggests that there will be better reliability, particularly if the categories have been well defined. The major disadvantage of time sampling in comparison with real time sampling is that infrequent behaviours may be missed (Sackett 1979).

4.5.5 Summary

In conclusion to this discussion of the various temporal aspects that can be employed during an observational study, it could be concluded that the ideal goal of observation is for continuous and complete sampling that results in less opportunity for sampling error to occur. By increasing both the frequency and the duration of observational periods this subsequently results in increased accuracy and precision (Johnson and Pennypacker 1980; Odom and Ogawa 1992). It should be emphasised that if the behaviour of interest occurs at low frequencies it cannot be accurately observed without continuous periods of observation of considerable duration. The sampling strategy and the time intervals used in an observation period depend upon the nature of the study and the questions being asked.

In the current study, real-time, continuous recording of events would have been problematic. It would have been difficult to remain unobtrusive all the time and to have maintained the high levels of concentration necessary to watch and record all the activities occurring. As has been discussed, interval recording approximates real-time recording and has some advantages over real-time recording, such as being less demanding on observer concentration and consequently this was the sampling strategy of choice to record the activities of the staff and the auditory warnings in the current study.
4.6 Some Disadvantages of Using Direct Observational Techniques

Direct observation of behaviour using human observers has some advantages and some disadvantages over other methods of collecting behavioural data. It can be very time consuming and in situations where several observers are used it can also be expensive. There is considerable disagreement between researchers regarding the presence of observers and their effects on behaviour. Some researchers argue that the presence of observers appears to have little effect on the situations they observe (e.g. Heynes and Lippitt 1954). Others argue that the introduction of observers and/or recording devices into an environment to collect data does have a significant effect on the behaviour observed, and can affect the external validity, (the extent to which the findings can be generalised to other situations or contexts, e.g. Haynes and Horn, 1982), and the internal validity (Nelson et al. 1978) of the data. Weick (1968) suggests caution, as stable behaviour in the presence of observers does not necessarily mean typical behaviour.

4.6.1 Reactivity

As discussed in the previous chapter, reactivity can occur when the presence of recording equipment or researchers observing behaviour and activities may inadvertently affect that behaviour. Many of the suggestions as to how reactivity could be decreased were also utilised in the current study. A propitious relationship had been established with the staff on the I.C.U. during the video study and as proposed by Johnson and Pennypacker (1980) a rapport had developed with the staff. As in the video study, the observers were careful to ensure they were not in the way of ongoing activities and the staff were encouraged to act naturally with the observers explaining the rationale for the study. Another method of minimising reactivity is the habituation of the participants to the presence of the observers. This was achieved on the I.C.U. as the staff were accustomed to observer TM being on the unit to set up the video camera for the various recording sessions.
4.6.2 Observer Bias

When observer bias occurs, the way the observer perceives the situation is distorted. This can occur when behaviour is categorised inappropriately because of the prior expectancies, beliefs and preconceptions of how people normally behave by the observer. An example is given by Selltiz et al. (1959) in which he suggests that many observers have a similar cultural background and may share similar perceptions of certain events. Selltiz et al. suggest that if a street gang was being observed for the amount of aggression or friendliness shown between the group members, the observer's viewpoint may have an important influence. To the observer, what appears as overt aggression may, to the group members who have grown up in this culture, be good-natured messing about. Different observers can also have different interpretations of the same behaviour (Moser 1971, Bramlett and Barnett 1993). With human observers there is always the problem that data are collected inaccurately or inconsistently and it is suggest by some researchers that adequate training and practice can overcome this to a certain degree but not completely in some circumstances (e.g. Selltiz et al. 1959).

4.6.3 Observer drift

Observer drift occurs when observer consistency and accuracy decrease as a function of idiosyncratic modifications made by observers following training and they drift away from the intended, original meaning of the code. If observers observe together they can agree on a definition of the code and drift from the intended meaning together. Therefore it is better if the observers do not watch each other as they code, the recording of behaviours should be done independently and discussion of disagreements should occur after the recording session (Reid 1982).
4.6.4 Inter-observer reliability

In order to assess the impact that the above problems have on the reliability of the data, Odam and Ogawa (1992) suggest that researchers would be well advised to include procedures for measuring observer accuracy, agreement and controlling for observer drift in their methodology. Gellert (1954) stated that data will have higher reliability if there are fewer categories for coding the data and greater precision in defining the categories and therefore less inference in making classifications.

Data that is unreliable because of systematic error, bias or observer drift decreases the likelihood of detecting a relationship between variables (Sutcliffe, 1980). To ensure that the data collected is correct there should be periodic assessment of the quality of the data (Hartmann 1982). One of the most common methods of assessing bias and measuring the reliability of observational data is to compare the recordings made between two or more observers. A related term, observer accuracy, refers to the comparisons between an observer and an established criterion. It has been argued that observer accuracy assessments should be used in preference to inter-observer reliability (e.g. Johnson and Pennypacker 1980).

To measure the agreement between observers, independent observations are undertaken early in the study by each observer. The data is recorded separately for each category without communicating with other observers. The scores are then compared for agreement on the occurrence or non-occurrence of behaviours or activities. Such comparisons can reveal ambiguities and overlaps in scoring categories and that is why the comparisons should be made before the main body of data are collected, not after. The categories may not be reliable and the value of the data would be diminished (Sommer and Sommer 1980).
Many researchers who discuss the reliability of observational data agree that the concept of 'reliability' is complex in that it includes the accuracy and stability of measurements as well as the actual conditions in which the observations are made (e.g. Weick 1968; Johnson and Bolstad 1973). There is also considerable debate in many studies about which statistical methods should be used. The general confusion concerning the terminology and concepts of reliability results in the misuse of reliability statistics. Additionally, as Odom and Ogawa (1992) point out many authors fail to report any measure of reliability at all.

When reliability measures are calculated, those most frequently used are percentage agreement, with Kappa coefficients used less frequently as discussed by Odom and Ogawa (1992) and Hollenbeck (1978). Some authors state that results of at least 80% agreement (e.g. Page and Iwata 1986) are considered sufficient to suggest that the obtained data are accurate and reliable. If poor agreement is obtained, i.e. less than 80%, then there will be less confidence when interpreting the data.

However, it is important to note that percentage agreement between the total of the observers scores does not evaluate the accuracy or stability of the measurements, or, in other words, while it is common for observers to agree on the total number of observations there may be considerable disagreement between each individual score (Hollenbeck 1978).

As a set of pre-established criteria suitable for use in the current study did not appear to be available, observer accuracy, the comparisons between an observer and an established criterion, could not be undertaken. Therefore, in the current study in order to investigate the reliability of the observations recorded by the two observers, each individual observation was compared to examine whether an activity or an alarm had occurred or not occurred. This procedure was undertaken during the pilot study as well as in the study proper. The phi coefficient (Siegel and Castellan 1988) was then used to calculate the degree of congruency between the
observers. As is common practice in many observation studies the percentage agreement for the total number of observations was also calculated.

To conclude, a direct, non-participant observation strategy was used which involved two observers. The categories for the activities and the alarms were not mutually exclusive and were recorded using an interval sampling schedule of 2-minute intervals. In order to examine the inter-observer reliability two sets of calculations were performed, first, the phi coefficient to determine the agreement for individual observations, and second, the percentage for the total agreement of the observers.

4.7 The Current Study

4.7.1 Introduction

The aims of the study described in this chapter were to establish which activities and behaviours occurred on the I.C.U. and to record the frequency of the occurrence of alarms. This was done in order to determine activities and to develop tasks for use in laboratory-based experiments.

As there were methodological problems with the video study, (as described in Chapter 3), whilst the frequency and duration of the alarms were established, the analysis of activities undertaken by staff when auditory warning were activated was not accomplished. It was therefore decided that two observers would undertake a direct observational study at the I.C.U. Derriford Hospital. The study was conducted in the four bedded bay on the unit, the same room used in the video study (see Figure 3.1 Chapter 3).
4.7.2 Method

4.7.2.1 Development of the Coding Strategy.

The observational study involved two observers, the experimenter (TM) and a colleague (DR), a developmental psychologist with experience in observational techniques from the Psychology Department, University of Plymouth. The author, (TM) had previous I.C.U. nursing experience.

Ideally, when an observational study is undertaken the categories should be used from pre-existing systems where possible (Foster et al. 1988). There are some general categorisations in the nursing audit and workload literature, but these are specifically designed with a singular aim in mind such as predicting manpower requirements (e.g. Devine et al. 1993) and by examining the dependency of patients (e.g. Bell and Storey 1984, Proctor and Hunt 1994).

Although there are some published accounts of observation studies being undertaken in the ward environment, (e.g. Balogh 1991, Reid 1991) there are few suitable, readily defined categories of behaviour available. For example Balogh (1991) conducted an observation study that was part of a larger study in a psychiatric hospital and had some similar categories of activities as the classifications used in the current study. However, there is a considerable difference between an I.C.U. and a psychiatric ward and the emphasis of the psychiatric categories of behaviour was biased toward the type of care a psychiatric patient would require, for example 'therapy' and 'E.C.T'.

The ABICUS (Aberdeen Intensive Care Unit System) project (Gilhooly et al. 1991) is another example of a classification system of activities on the I.C.U. The project is involved in the instigation of a computerised system for recording information regarding the patient’s condition. However, the classifications used are very
specific and are patient focused. For example, one category is for 'external analysis' which included biochemistry, blood samples and urine samples sent to the laboratory (Green, C. 1994 personal communication) and these classifications were not considered general enough for the current study.

Stanton (1992) used an observation form specifically for recording details concerning alarms adapted from Kragt and Bonten (1983). However, the form was too detailed to record data in the current study. For example, the information required included recording the time of the alarm, which alarm sounded, the type of alarm (whether auditory or visual), details of the action taken, the urgency of situation the alarm was signalling and any additional comments to be recorded.

4.7.2.2 The Informal Observation Phase

As there did not appear to be a suitable pre-existing group of categories that could be used in the current study to observe the activities of staff when alarms occurred, in order to develop the categories of behaviour that were to be recorded the first phase of the study involved informal, casual observation as described by Rosenblum (1978) and Sommer and Sommer (1980). During this phase the behaviours and events that would eventually be recorded in the systematic phase of the study were selected and defined (Hawkins 1982, Johnson and Pennypacker 1980). Every procedure that was undertaken by one member of the trained staff with each patient in the four-bedded bay was recorded.

During this initial phase the time of day that the ward was visited was varied. The routines of different nurses were recorded onto a notepad during varying periods of activity, which included quiet periods when the patients required minimum care, ranging through moderately busy periods to very busy periods.

It was during this preliminary stage that a relationship was developed with the staff and an attempt made to habituate the staff to the presence of the observers on the I.C.U., thereby reducing any
reactivity that might occur. Although the study focused on the activities of the trained nursing staff, there were many other staff on the unit, for example doctors, the receptionist, student nurses, and physiotherapists. In order to establish a rapport with the staff, a brief explanation of the study was given. A meeting was arranged which coincided with one of the afternoon training sessions to explain the purpose of the study and the two observers met with as many of the staff as possible. Care was taken to emphasise that no value judgements were being made on the activities being observed, for example, whether the behaviour was appropriate or not. The staff were also told that the main emphasis of the study was to observe the impact on the trained staff, of the equipment and alarms monitoring the patients. This was seen by the staff as a positive rationale for the study, thereby decreasing any reactivity that may occur (Johnson and Bolstad 1973, Foster et al. 1988).

The overall study was also frequently discussed in an informal way during the coffee breaks or lunch and when the unit was quiet. This helped to alleviate any fears that the staff may have had regarding the use of the data. The staff were again reassured that there would be no serious or important consequences for them in terms of their jobs as a result of the observational study, that the names of staff would not be used in the study and the information would not be given to anyone in 'management'.

The trained staff nurses were asked during these informal talks what their routine with patients consisted of. Together with the casual observations, a base line of the day to day routine care of a patient was established. This assisted in forming an overall picture of what happened on the ward and further developed the relationship with the staff by involving them in the development of the procedures and by inviting their expert opinion as suggested by Hawkins (1982) for example, to establish the content validity of what was being measured.

During the casual phase the decisions as to which behaviours should be included into which categories were considered. In order to record systematically the physical workload of the trained nursing
staff it was necessary to develop a checklist. An initial starting point was to list the most prevalent occurrences in the ward routine.

A molar categorisation of behaviours as discussed by Sackett (1978) and Kerlinger (1973) was adopted for the current study in which a broadly defined number of actions were combined into one category and subsequently a variety of behaviours were observed under one rubric, similar behaviours were grouped together to form the categories.

The final checklist consisted of fifteen discriminable categories. There was also a space for any activity or event that could not be coded into any of the existing categories. Any such activity or event may have inadvertently have been excluded from the categories, perhaps because it had not occurred during the casual observation period. An example of the final checklist for the physical workload used in the systematic phase of the study is shown in Figure 4.1. The final categories described below evolved over a considerable period of time to ensure clear, precise, unambiguous definitions. Some categories (such as 'admission') were not modified throughout the development of the checklist while others required some redefinition. How each category was developed and what the definitive categories incorporated will now be briefly described.

During the informal period of observation it was observed that there were many aspects of talking that occurred on the I.C.U. and these were categorised into teaching, staff talking, talking to the patient, staff handover and talking to visitors.

**Teaching** involved the third year student nurses who were present on the unit at the time of the observational study. (At the time of writing student nurses no longer work on the I.C.U. at Derriford.) Senior staff also instructed more junior staff or new members of staff about particular procedures or the benefits and actions of certain drugs, for example.
Staff talking involved different members of the staff, often the physiotherapists, the doctors and the nursing staff talking more informally about the patient's condition. This category usually consisted of talk that was related to work, however occasionally the staff would talk between themselves about more personal matters, for example their children.

Talking to the patient involved the trained staff talking to the patient.

Staff-handover occurred when the shifts changed over, for example at 13.30. The staff who had been on duty during the morning shift informed the newly arrived staff about the condition of the patient they were to look after.

Talking to visitors involved the staff talking to the visitors of the patient.

Observation.
Initially, during the informal phase 'observation' had been classified into two separate categories, based to a certain degree on information from the trained staff about their routine. The first category, informal observation, occurred when the staff generally checked the condition of the patient and the second more formal category, routine observation, occurred regularly for each patient at four hourly intervals (6am, 10am, 14.00hrs, 18.00hrs and so on). However, it was decided to combine both categories as the staff were constantly looking and checking the monitors and the patient's condition.

The Observation category included all the incidence of observation conducted by the trained nurse regarding the patient's condition. These included procedures such as looking at the monitors and recording the readings onto charts, recording information about the condition of the patient such as checking the urine output from the catheter drainage system and recording the figures onto the patients charts.
The Drugs category included the calculation, preparation and administration of drugs to the patient. In many instances it is necessary for the staff to calculate the dosage of drugs administered to the patients. For example, an asthmatic patient may be prescribed 3mls of 0.2% salbutamol to be given by way of a nebulizer. The salbutamol held by pharmacy has a strength of 0.5%. The nurse must calculate the correct amount of salbutamol and also the correct amount of water in which the drug is mixed. The correct dosage is calculated using the following formula.

\[
\text{Amount of stock required} = \frac{\text{strength required} \times \text{total volume}}{\text{stock strength required}}
\]

\[
=\frac{0.2\% \times 3\text{ml}}{0.5\%} = 1.2 \text{ ml} \, 0.5\% \text{ salbutamol required}
\]

For some intravenous drugs, factors such as the weight of the patient has to be considered. An example of one such intravenous drug calculation used on the I.C.U. is demonstrated below.

250 mg of the drug dobutamine is made up to 50 mls with either normal saline or dextrose solution. The total volume is, therefore, 50 mls (5 mg of dobutamine per ml of solution). The infusion of the drug is then commenced at 2.5 micrograms per kilogram of patient bodyweight per minute. The drug is titrated according to the patients blood pressure which must maintain a systolic pressure of 100 mmhg.

The drugs category also included any activity that involved the intravenous fluids being administered to the patient. Intravenous infusions are administered to patients over a number of hours, for example, a litre of fluid can be given over an eight hour period. A formula is used for calculating the number of drips per minute depending on the tubing (the giving set) used to administer the fluid, for example some giving sets administer 20 drops per ml for normal saline or dextrose solution and 15 drops per ml for blood. An example of the formula used to calculate the number of drops per minute is demonstrated below.
Example of an intravenous infusion formula used:

Amount required x drops
no of hours x minutes

\[
1000 \times \frac{20}{8} = 42 \text{ drops per minute}
\]

The number of drops are then regulated using an infusion pump.

Physiotherapy involved physiotherapy being undertaken with the patient by one or two physiotherapists often together with the nurse involved with the care of the patient. Physiotherapy included activities such as passive exercise (movement of the limbs for example) of an unconscious patient or suction for a patient on a ventilator who may have excessive secretions in their lungs which they were unable to remove, by coughing, themselves.

Doctor's visit involved the doctors concerned with the patient's care visiting the patient to assess their condition, talking to patients and other staff and examining the charts. The visit could involve one doctor (usually the senior house officer) or several doctors including the consultant, registrar, senior house officers, medical students etc.

Admission involved a specific procedure which included the preparation before and during the arrival of a new patient. Depending on the condition of the patient there could be several staff involved in stabilising the condition of the patient and connecting up the various items of monitoring equipment. If the patient was being transferred from theatre for example, the procedure would involve handover from the theatre staff, if the patient were conscious the I.C.U. staff would talk to the patient, if relatives were accompanying the patient, then depending on the condition of the patient the various procedures would be explained.
Discharge included the preparation during and before the discharge of a patient to a general ward.

X-ray involved the presence of radiographers and a portable x-ray machine for a patient who required an urgent x-ray on the unit.

E.C.G. involved the presence of an E.C.G. (electrocardiogram) technician for a patient who required an urgent E.C.G. on the unit.

Answering the telephone. The telephone rang on many occasions on the I.C.U. and was often answered by the ward clerk. However, if the ward clerk was busy elsewhere it was necessary for one of the staff on the unit to answer the phone. Sometimes a doctor or nurse was at the nurses station and they would answer the phone. However, the final category involved one of the members of staff from the four-bedded bay going to the nurses station to answer the telephone.

The second checklist listed the auditory warnings of the monitors and equipment in current use in the I.C.U. Derriford. The alarm checklist had a total of seven categories which include the ventilator, syringe pump, pulse oxymeter, infusion pump, E.C.G. monitor, humidifier. Miscellaneous sounds, such as the bleeps and telephone, were also included. An example of the auditory sounds checklist is shown in Figure 4.2.

During the informal observation phase the observers also learnt the alarm sounds. This was done independently by each observer first, when a piece of equipment alarmed the sounds were learnt on the I.C.U., and also by using the auditory warnings on the Archimedes computer that were to be used for the laboratory-based experiments (see Chapter 6).

The sampling strategy of interval recording was established and this involved recording during 2 minute periods for sessions that totalled approximately 28 minutes. This time schedule was eventually selected as the total time period because of the concentration required by the experimenters to observe and record...
each activity accurately. Interval recording was therefore used to record the activities of the staff and the occurrence of the auditory warnings. When the alarms sounded the time on the stopwatch was jotted down on the auditory warning checklist next to the appropriate name. This was later compared with the workload that was occurring at the concurrent time and in this way it was hoped to establish what was happening on the ward when the alarms occurred.

4.7.2.3 The Development of the Systematic Observation Phase

During the informal phase of the investigation the checklist underwent several refinements, as initially there were too many categories to record accurately. While it would have been undesirable to omit any relevant behaviours, there are practical constraints on trying to observe a large number of categories and there had to be a realistic compromise to their number.

As discussed above one of the aims of the current study was to examine the relationship between the frequency of auditory warnings and the physical workload of the trained staff. In order to examine that relationship it was felt that sufficient examples of all the frequent behaviours that occurred on the I.C.U. were represented on the final checklist.

Before the main study was undertaken, the observers practised using the two checklists together with the stopwatches. The practice sessions were short and were undertaken during quiet periods on the I.C.U. Again, after the practice sessions, any problems were discussed, as in, for example Bayles (1950) study in which the observers discussed problems and methods of coding the pilot study systems after they had been conducted in order to increase the reliability of the study.
<table>
<thead>
<tr>
<th>ACTIVITY</th>
<th>2MINS</th>
<th>2MINS</th>
<th>2MINS</th>
<th>2MINS</th>
<th>2MINS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teaching</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Staff talking</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Talking to patient</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Talking to visitors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Staff handover</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drugs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physiotherapy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dr's visit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Admission</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discharge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X-Ray</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E.C.G.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Answering phone</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4.1: Physical Workload Checklist
<table>
<thead>
<tr>
<th>EQUIPMENT</th>
<th>TIME &amp; BEDNUMBER</th>
</tr>
</thead>
<tbody>
<tr>
<td>VENTILATOR</td>
<td></td>
</tr>
<tr>
<td>SYRINGE PUMP</td>
<td></td>
</tr>
<tr>
<td>PULSE OXYMETER</td>
<td></td>
</tr>
<tr>
<td>INFUSION PUMP</td>
<td></td>
</tr>
<tr>
<td>E.C.G MONITOR</td>
<td></td>
</tr>
<tr>
<td>HUMIDIFIER</td>
<td></td>
</tr>
<tr>
<td>TELEPHONE</td>
<td></td>
</tr>
<tr>
<td>BLEEP</td>
<td></td>
</tr>
<tr>
<td>ANY OTHER SOUNDS</td>
<td></td>
</tr>
</tbody>
</table>

Figure 4.2: Auditory Sounds Checklist
4.7.2.4 Implementation of the Systematic Phase

When the final checklist had been completed a preliminary study was undertaken which involved the observers independently observing the activities and the data then being compared to examine whether disagreements had occurred and if they had, why. Discovering and correcting inadequacies in the definitions early in the study improves observer reliability (Foster et al. 1988). As discussed previously, during the development of the systematic phase some of the categories in the present study were not precise enough and this was subsequently corrected and the degree of agreement between the two observers improved as a result. The observations were conducted at random times in order to obtain representative sampling.

The two-minute time scale was used to monitor the activities of the trained staff as this time period allowed adequate time to record the procedures. A stopwatch was used to time the intervals. A shorter time limit would have made it difficult to observe what was occurring at each bed space. The first occurrence of each activity that occurred at each bed was recorded in two minute intervals under one of the headings on the activity checklist by noting down the bed number. The time of the first occurrence of each alarm that occurred was also noted. Initially, the observers attempted to record the duration of each alarm, but this proved difficult as often the alarm sounded for only a few seconds. Subsequently, the time on the stopwatch and the bed number were recorded on the alarm checklist and this was later compared with the activity checklist when the data was being analysed.

The total period of time for the observations varied depending on the status of the unit, whether it was very quiet or busy, or whether the observations had to stop due to particular activities on the unit. The sessions varied from 1 to 2 sessions (28 minutes and 56 minutes respectively). The total amount of time for the observation study proper was 5.13 hours. For the systematic phase of the study it was decided to undertake the observations during
random periods between 13.00 hours and 14.30 hours. It was decided to visit at this time because it had been determined during the informal stage of the observation that this time was usually busy on the unit. The afternoon staff came onto the ward and the morning staff handed over to them, observations were undertaken, drugs were often given during this period and the physiotherapists often visited as well. As one of the aims of the study was to observe what activities were occurring when alarms were activated, it seemed appropriate to visit the ward when there would be activities occurring, rather than in a more random way when the unit might have been quite with few activities happening.

Each activity was recorded by writing the number of the bed where the behaviour had occurred in the appropriate space on the checklist. The numbering of the beds had been discussed before commencing the preliminary study and was the same numbering system used by the staff on the unit and also in the video study (Chapter 3). In the four-bedded ward, the beds were numbered by the staff as bed one and subsequently in a clockwise direction to bed four.

The procedure for the systematic phase of the study was undertaken in the following way. It was noted, prior to the start of the observational session, how many trained staff were on duty, how many trainee staff, and how many patients there were in total in the four-bedded bay. As in the preliminary study the observations were conducted independently and the observers were as inobtrusive as possible with no interference in the ongoing activities and no interaction with participants, in an attempt to avoid observer effect as suggested by Reid (1982).

The activities and procedures being undertaken by the staff together with any alarms that occurred were recorded every two minutes in an identical manner to the preliminary study, in blocks so as to avoid observer fatigue.
4.7.3 Results

4.7.3.1 Introduction

The findings of the observation study are presented as follows; first, the degree of agreement between the observers. Secondly, the sequence of activities and alarms that occurred together within each 2 minute interval are presented. The final section discusses the differing amounts of physical workload undertaken by trained staff during the periods of observation.

The degree of agreement between the observers was examined by calculating the phi coefficient and the total percentage agreement of the occurrence or non occurrence of an activity or an alarm, recorded by each observer is given. It will be shown that the degree of agreement between the observers is very high (see Tables 4.2 and 4.3 and Figures 4.3 and 4.4). For convenience, therefore, the scores of one observer (TM) were used to describe the sequencing and organisation of the activities and alarms that occurred on the I.C.U. during the period of the observation study.

4.7.3.2 Inter-observer reliability

In order to determine the strength of agreement between the two observers' data the phi coefficient was calculated. The phi coefficient measures the extent of association between two variables which can only have one of two values, for example, 'yes' or 'no' (Siegel and Castellan 1988). For each observer every observation recorded in the current study was coded '0' for the non-occurrence of a behaviour and '1' when the behaviour occurred. The data was then arranged in a 2 by 2 contingency table, an example of which is shown in Table 4.1.
Table 4.1. An Example of the Contingency Table used in the Calculation for the Degree of Agreement Between Observers.

In Table 4.1 'A' denotes agreement between the two observers that an event did not occur, both observers having scored a 0. 'B' denotes Observer TM recording the non-occurrence of an event, (O) while observer DR. has recorded an event as occurring (1). 'C' illustrates Observer TM recording the occurrence of an event, (1) while observer DR has recorded the non-occurrence of the event (O). Finally 'D' denotes both observers agreeing that an event has taken place.

When the data for each activity and each alarm had been organised into separate contingency tables, the Phi coefficient ($\phi$) was calculated to examine the congruity of the observations made by each observer. The results of the phi coefficient can range from zero, (which indicates a poor degree of association between observers), to one, (which shows complete agreement between the observers).

As discussed in the introduction to this chapter, many studies determine observer reliability by calculating the percentage for the overall agreement and a figure of 80% is considered an adequate degree of agreement. However, as Hollenbeck (1978) argues percentage agreement between the total of the observers scores does not evaluate the accuracy or stability of the measurements. While it is common for observers to agree on the total number of
observations there may be considerable disagreement between each individual score and the phi coefficient indicates the strength of agreement for each individual score for observers. In the current study both the phi coefficient and the percentage agreement (shown in parenthesis) was calculated to demonstrate overall agreement.

The results of the phi coefficient for the degree of agreement between the two observers are shown in Table 4.2 for the activities and Table 4.3 for the alarms. (Two categories, portable x-ray and an electrocardiogram being undertaken by a technician, did not occur during the period of observation and therefore these categories were not included in the data analysis). The absolute frequency of activities recorded by the two observers during the observation study are shown in Figure 4.3 and the absolute frequency for the alarms is shown in Figure 4.4. Figures 4.3 and 4.4 illustrate how many activities and alarms occurred during the study and also demonstrate the similarity between the two sets of observations.

Tables 4.2 and 4.3 establish that the degree of agreement for the majority of the activities and alarms is close to 1, which as discussed earlier, demonstrates a high degree of agreement between the observers. For example, the phi coefficient for the observation of the administration of drugs is 0.9, for the admission of a patient, the phi coefficient is 0.96. Two categories, staff talking and staff handover had relatively lower agreement (\( r_0 = 0.63 \) and 0.59 respectively) though the percentage agreement was above 80%. The phi coefficient may have been slightly lower for the talking categories than for the other activities as occasionally it was difficult for the observers to determine exactly what the staff were talking about if there was a degree of background noise for example. There was complete agreement between the two observers for three of the alarms that occurred. These were the ventilator alarm, the humidifier alarm and the E.C.G. alarm (\( r_0 = 1 \)).

The percentage agreements for all the categories were above the generally accepted criterion for inter-observer reliability of 80%.
<table>
<thead>
<tr>
<th>Activity</th>
<th>$r_0$</th>
<th>%</th>
<th>Activity</th>
<th>$r_0$</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teaching</td>
<td>0.80</td>
<td>(94%)</td>
<td>Physiotherapy</td>
<td>0.95</td>
<td>(98%)</td>
</tr>
<tr>
<td>Staff Talking</td>
<td>0.63</td>
<td>(87%)</td>
<td>Dr's Visit</td>
<td>0.78</td>
<td>(96%)</td>
</tr>
<tr>
<td>Talking to Patient</td>
<td>0.84</td>
<td>(94%)</td>
<td>Admission</td>
<td>0.96</td>
<td>(99%)</td>
</tr>
<tr>
<td>Staff Handover</td>
<td>0.59</td>
<td>(96%)</td>
<td>Discharge</td>
<td>0.78</td>
<td>(97%)</td>
</tr>
<tr>
<td>Talking to Visitors</td>
<td>0.71</td>
<td>(96%)</td>
<td>Answering Phone</td>
<td>0.96</td>
<td>(99%)</td>
</tr>
<tr>
<td>Observations</td>
<td>0.77</td>
<td>(89%)</td>
<td>Answering Alarms</td>
<td>0.90</td>
<td>(96%)</td>
</tr>
<tr>
<td>Drugs</td>
<td>0.90</td>
<td>(96%)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.2: Inter Observer Agreement for Activities Showing the Phi Coefficient and Percentage Agreement (shown in parenthesis)
Figure 4.3: Total Number of Activities Recorded by the Two Observers (TM and DR)
### Table 4.3: Inter Observer Agreement for Alarms Showing the Phi Coefficient and Percentage Agreement.

<table>
<thead>
<tr>
<th>Alarm</th>
<th>$r_\phi$</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ventilator</td>
<td>1.0</td>
<td>(100%)</td>
</tr>
<tr>
<td>Pulse Oxymeter</td>
<td>0.86</td>
<td>(98%)</td>
</tr>
<tr>
<td>Syringe Pump</td>
<td>0.95</td>
<td>(99%)</td>
</tr>
<tr>
<td>Infusion Pump</td>
<td>0.88</td>
<td>(99%)</td>
</tr>
<tr>
<td>E.C.G. Monitor</td>
<td>1.0</td>
<td>(100%)</td>
</tr>
<tr>
<td>Humidifier</td>
<td>1.0</td>
<td>(100%)</td>
</tr>
<tr>
<td>Telephone</td>
<td>0.96</td>
<td>(99%)</td>
</tr>
</tbody>
</table>
Figure 4.4: Total Number of Alarms Recorded by the Two Observers (TM and DR)
4.7.3.3 Frequency of Activities and Auditory Warnings.

From Figures 4.3 and 4.4 it can be seen that overall during the observation study there were differences in the number of times certain categories of activities occurred. For example observation was the most frequently recorded activity occurring 177 times, while admission was comparatively less frequent, occurring only 10 times during the observational sessions.

In the video study it was shown that the ventilator alarm was the most frequently occurring alarm overall. In comparison the most frequently sounding alarm in the direct observation study was the pulse oxymeter. During the direct observation study alarms that had been heard infrequently during the video study, such as the humidifier and the E.C.G. monitor, were heard and recorded by the observers on several occasions. However, throughout the period of the direct observation study the pulmonary artery flotation catheter was not used to monitor the condition of any patients and was therefore not heard at all. Two other miscellaneous sounds were also not heard during the direct observation study which were the doctors bleep and the doorbell.

During the total observational study 819 activities were recorded (once during each two minute interval) and the proportion for each activity out of the total activities is shown in Table 4.4. The category 'answering alarms' is included as when an alarm occurred during the observation study the staff would respond to the alarm by either silencing it or turning it off. It can be seen that the activity 'answering alarms' (0.14) constitutes a relatively high proportion of the total number of activities undertaken by the trained staff during the observation study.
<table>
<thead>
<tr>
<th>Activity</th>
<th>Number of Occurrences</th>
<th>Proportion of Occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teaching</td>
<td>83</td>
<td>0.10</td>
</tr>
<tr>
<td>Staff talking</td>
<td>117</td>
<td>0.14</td>
</tr>
<tr>
<td>Talking to patient</td>
<td>99</td>
<td>0.12</td>
</tr>
<tr>
<td>Staff Handover</td>
<td>29</td>
<td>0.03</td>
</tr>
<tr>
<td>Talking to visitors</td>
<td>31</td>
<td>0.03</td>
</tr>
<tr>
<td>Observation</td>
<td>177</td>
<td>0.22</td>
</tr>
<tr>
<td>Drugs</td>
<td>106</td>
<td>0.12</td>
</tr>
<tr>
<td>Physiotherapy</td>
<td>69</td>
<td>0.08</td>
</tr>
<tr>
<td>Dr's visit</td>
<td>36</td>
<td>0.04</td>
</tr>
<tr>
<td>Admission</td>
<td>10</td>
<td>0.01</td>
</tr>
<tr>
<td>Discharge</td>
<td>36</td>
<td>0.04</td>
</tr>
<tr>
<td>Answering phone</td>
<td>26</td>
<td>0.03</td>
</tr>
<tr>
<td>Answering Alarms</td>
<td>116</td>
<td>0.14</td>
</tr>
</tbody>
</table>

Table 4.4: Proportion of Each Activity as a Proportion of the Total Activities.

During the non-sequential observational period of 5 and a half hours the alarms sounded in total 116 times. Table 4.5 shows the proportion that the individual alarms sounded as a proportion of the total number of alarms sounding.
Table 4.5: Proportion of Each Alarm as a Proportion of the Total Number of Alarms

<table>
<thead>
<tr>
<th>Equipment</th>
<th>total number of alarms occurring</th>
<th>proportion of total alarm sounding time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ventilator</td>
<td>20</td>
<td>0.17</td>
</tr>
<tr>
<td>Pulse oxymeter</td>
<td>36</td>
<td>0.31</td>
</tr>
<tr>
<td>Syringe pump</td>
<td>20</td>
<td>0.17</td>
</tr>
<tr>
<td>Infusion pump</td>
<td>24</td>
<td>0.21</td>
</tr>
<tr>
<td>E.C.G. monitor</td>
<td>11</td>
<td>0.09</td>
</tr>
<tr>
<td>Humidifier</td>
<td>5</td>
<td>0.04</td>
</tr>
</tbody>
</table>

4.7.3.4 Sequence and Organisation of Activities and Alarms

One of the aims of the observation study was to establish the activities being undertaken by staff when auditory warnings were activated. Therefore, the following section presents the total number of times each activity was recorded during the observation study when an alarm was activated and the results are shown in Table 4.6.
<table>
<thead>
<tr>
<th>ACTIVITIES</th>
<th>Ventilator</th>
<th>Pulse Oxymeter</th>
<th>Syringe</th>
<th>Infusion Pump</th>
<th>E.C.G.</th>
<th>Humidifier</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Activity involving patient</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Teaching</td>
<td>10</td>
<td>0</td>
<td>2</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>Staff talking</td>
<td>9</td>
<td>8</td>
<td>6</td>
<td>8</td>
<td>4</td>
<td>1</td>
<td>36</td>
</tr>
<tr>
<td>Talking to patient</td>
<td>0</td>
<td>12</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>4</td>
<td>27</td>
</tr>
<tr>
<td>Staff Handover</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>Talking to Visitors</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>Observation</td>
<td>2</td>
<td>26</td>
<td>10</td>
<td>8</td>
<td>3</td>
<td>2</td>
<td>51</td>
</tr>
<tr>
<td>Drugs</td>
<td>4</td>
<td>2</td>
<td>16</td>
<td>19</td>
<td>3</td>
<td>0</td>
<td>44</td>
</tr>
<tr>
<td>Physiotherapy</td>
<td>14</td>
<td>0</td>
<td>1</td>
<td>7</td>
<td>0</td>
<td>5</td>
<td>27</td>
</tr>
<tr>
<td>Dr's Visit</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Admission</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Discharge</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Answering Phone</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 4.6: The Total Number of Activities Occurring When Each Alarm was Activated.
As Stanton (1992) observed, certain activities performed by a patient (such as washing or eating) activated certain alarms (e.g. the E.C.G. alarm). In the analysis of the tapes from the video study (Chapter 3), it was suggested that some activities, such as physiotherapy, activated the ventilator alarm but the evidence to support this claim was limited as, for example, the physiotherapist was often out of view of the camera and could only be heard, consequently her activities could not be seen. In the current study, however, it can be implied by the results shown in Table 4.6 that certain activities undertaken by the staff activate certain alarms.

During the observation study the ventilator alarm sounded in total 20 times. During 14 of the 20 occurrences (70%) when the alarm was activated, physiotherapy was being undertaken with the patient. The ventilator also sounded 10 times (50%) when teaching was being conducted and also 9 times (45%) when the staff were talking. However, these events were not mutually exclusive and can be seen to occur concurrently in many instances (see Figures 4.5 to 4.15).

Also during the observational period the pulse oxymeter alarm occurred in total 36 times. On 26 (72%) of those occasions observation of the patient was being undertaken by a trained member of staff. The syringe pump was activated 20 times in total, 16 (80%) times when drugs being administered. Similarly, the number of times the infusion pump alarms were activated totalled 20 and during 19 of the 20 (95%) times the alarm sounded, drugs were being intravenously administered.

It can also be seen from Table 4.6 that staff were often involved in various aspect of talking, whether it be teaching, staff talking, talking to the patient, staff handover or talking to visitors. If all the different categories of talking are combined together to create one category of 'talking' it can be seen that the highest number of alarms occur (103) when talking of one sort or another is being undertaken. Talking often occurred when the staff were undertaking other procedures, for example observation or the
administration of drugs, and they engaged in conversation with either the patient, visitors or other members of staff.

Table 4.6 however, does not show which activities were occurring concurrently within each of the two minute intervals, and as the activities were not mutually exclusive, more than one activity could occur during each two minute interval recorded. To examine which activities and alarms occurred together in each of the two minute intervals and to enable the extrapolation of the concrete findings of the observational data into more abstract tasks for use in the proposed laboratory experiments, it was necessary to examine in greater detail the types of activities which occur when the alarms are activated. Therefore, the activities and alarms recorded for each observation session are presented in Figures 4.5 to 4.15.

The observation data recorded by observer TM for each day and each bed are presented in Figures 4.5 to 4.15. The bed spaces for each figure were numbered clockwise 1 to 4 as they are on the I.C.U. (see Chapter 3, Figure 3.1). Beds that were not occupied by a patient are shown as an empty bed (for example Figure 4.6b). The horizontal axis represents the time in cumulative units of two minutes. The time periods observed varied from 28 minutes in total (for example, Figure 4.5) to 56 minutes (for example, Figures 4.6 and 4.7). As previously discussed, the time period observed depended on various factors for example, if the curtains were pulled around one or more beds observation then became infeasible, also if there were relatives visiting a patient the observers felt it inappropriate to remain on the unit collecting data.

The graphs illustrated in Figures 4.5 to 4.15 demonstrate the activities and auditory warnings that occurred within each 2 minute period. As discussed in the introduction to this chapter, it is standard practice with interval recording to record the first occurrence of each activity and each alarm during the time interval. However, many of the activities, for example physiotherapy (e.g. Figure 4.8b), occurred in blocks and were continuous throughout the observation session.
Figure 4.5: Alarms and Activities Day 1

Figure 4.5a: Bed 1

Figure 4.5b: Bed 2

Figure 4.5c: Bed 4

Figure 4.5d: Bed 3
Figure 4.6: Activities and Alarms Day 2(i)

Figure 4.6a Bed 1

Figure 4.6b Bed 2

Figure 4.6c Bed 4

Figure 4.6d Bed 3
Figure 4.7: Alarms and Activities Day 2(ii)
Figure 4.8: Alarms and Activities Day 3

Figure 4.8a: Bed 1

Figure 4.8b Bed 2

Figure 4.8c: Bed 4

Figure 4.8d: Bed 3
Figure 4.9: Alarms and Activities Day 4

Figure 4.9a: Bed 1

Figure 4.9b: Bed 2

Figure 4.9c: Bed 4

Figure 4.9d: Bed 3
Figure 4.10: Alarms and Activities Day 5

Figure 4.10a: Bed 1

Figure 4.10b: Bed 2

Figure 4.10c: Bed 4

Figure 4.10d: Bed 3
Figure 4.11: Alarms and Activities Day 6

Figure 4.11a: Bed 1

Figure 4.11b: Bed 2

Figure 4.11c: Bed 4

Figure 4.11d: Bed 3
Figure 4.12: Alarms and Activities Day 7(i)
Figure 4.13: Alarms and Activities Day 7(ii)

Figure 4.13a: Bed 1

Figure 4.13b: Bed 2

Figure 4.13c: Bed 4

Figure 4.13d: Bed 3
Figure 4.14: Alarms and Activities Day 8(i)

Figure 4.14a: Bed 1
- Answering Phone
- Dr's Visit
- Talking to Patient
- Staff Talking

Figure 4.14b: Bed 2
- Answering Phone
- Dr's Visit
- Observations
- Talking to Patient
- Staff Talking

Figure 4.14c: Bed 4
- Answering Phone
- Dr's Visit
- Observations
- Talking to Patient
- Staff Talking

Figure 4.14d: Bed 3
- Humidifier
- Infusion Pump
- Physiotherapy
- Drugs
- Observation
- Talking to Patient
- Staff Talking
- Teaching

Time
Figure 4.15 Alarms and Activities

Figure 4.15a: Bed 1

Figure 4.15b: Bed 2

Figure 4.15c: Bed 4

Figure 4.15d: Bed 3
To clarify the most frequently occurring activities that occurred concurrently with alarms being activated within each two minute interval (illustrated in Figures 4.5 to 4.15), the combinations of alarms and activities are presented in Tables 4.7 to 4.10. In some instances the same combinations of activities and alarms occurred several times, for example in Figure 4.8d the ventilator alarm occurred in the same two minute interval as two activities, physiotherapy and talking, thirteen times. This would be very repetitive if each example were shown in the tables. However, it is necessary to establish how frequently particular combinations of activities occurred together with an alarm being activated in order to reproduce similar combinations in the laboratory based experiments. Therefore, for convenience, one example of all similar combinations is shown in the tables, with the total number of times it occurred at the relevant bed space presented in the final column of the tables (entitled total number of episodes). For the example described above and shown in Table 4.6, the ventilator alarm occurring with physiotherapy and talking, the figure in the column 'total number of episodes' is 13.

As shown in Table 4.6 and illustrated in Figure 4.12c there was one instance during the observation study when one alarm, the E.C.G alarm, was activated without any corresponding activities being undertaken by the staff. There could be different reasons why this occurred. Prior to the alarm being activated, there were various activities being undertaken at the bed. These included observation and talking, including talking to the patient. The alarm could have sounded for two main reasons, firstly due to artefact for example, if the patient was moving about and a lead from the E.C.G. machine fell off then the alarm would be activated. A second reason could be a change in the patient's condition, for example if the heart rate became very slow or irregular then the alarm would have been activated.

There were 24 occasions (total number of episodes) during the observation study when one alarm and one activity were recorded together within the two minute observation period and the
combinations are shown in Table 4.7. There were 7 occasions (29\%) when drugs were being calculated or administered, 8 occasions (33\%) when observation was being undertaken, four occurrences (16\%) of talking, two occurrences (8\%) of physiotherapy and three occasions (12\%) when a patient was being discharged.

One alarm together with two activities occurred in total 67 times during the observation study. This was the largest category of combinations observed during the observation study and is shown in Table 4.8. The majority of the combinations, (57 occurrences, 85\%) involved the generic category of talking together with a specific procedure being undertaken. For example, observation occurred 20 times (30\%) together with talking; drugs and talking occurred 14 times (20\%) and physiotherapy and talking occurred on 24 occasions (25\%). On seven occasions (10\%) drugs and observation occurred in the same 2 minute interval as when an alarm was activated.

Table 4.9 shows the 17 occasions during which an alarm occurred whilst three activities were being undertaken in each two minute interval. The prevalent combination of activities in this category was ten occasions (59\%) when drugs, observation and talking were happening.

There were fewer occasions when four activities were being undertaken with one alarm being activated within the two minute interval and the four combinations are shown in Table 4.10. It can be seen that three out of the four combinations are depicted in Figure 4.9a (which was the busiest bed observed during the study and will be discussed in more detail in the next section). It can also be observed from Table 4.10 that on three occasions, three out of four of the predominant activities were drugs, observation and talking occurring within the same two minute interval.

There were three occasions during the observation study when two alarms were recorded during the same two minute interval. The pulse oxymeter and the ventilator alarm can be seen in Figure 4.8c to occur within the same interval that observation is being conducted. Figure 4.9a shows the syringe pump and pulse oxymeter
occurring within the same interval on two occasions. Firstly, when observation and talking are occurring and secondly, when drugs observation and talking occur. (The activities were scored as occurring once in the following tables).

<table>
<thead>
<tr>
<th>Figure</th>
<th>Alarm</th>
<th>Activity</th>
<th>Number of episodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>5a</td>
<td>infusion pump</td>
<td>drugs</td>
<td>1</td>
</tr>
<tr>
<td>7c</td>
<td>pulse oxymeter</td>
<td>observation</td>
<td>2</td>
</tr>
<tr>
<td>7d</td>
<td>infusion pump</td>
<td>physiotherapy</td>
<td>1</td>
</tr>
<tr>
<td>8b</td>
<td>ventilator</td>
<td>physiotherapy</td>
<td>1</td>
</tr>
<tr>
<td>8c</td>
<td>pulse oxymeter</td>
<td>observation</td>
<td>1</td>
</tr>
<tr>
<td>8c</td>
<td>ventilator</td>
<td>observation</td>
<td>2</td>
</tr>
<tr>
<td>8c</td>
<td>ventilator</td>
<td>drugs</td>
<td>4</td>
</tr>
<tr>
<td>9c</td>
<td>syringe pump</td>
<td>drugs</td>
<td>2</td>
</tr>
<tr>
<td>11b</td>
<td>E.C.G.</td>
<td>talking</td>
<td>3</td>
</tr>
<tr>
<td>13d</td>
<td>E.C.G.</td>
<td>talking</td>
<td>1</td>
</tr>
<tr>
<td>15a</td>
<td>pulse oxymeter</td>
<td>observations</td>
<td>3</td>
</tr>
<tr>
<td>15d</td>
<td>pulse oxymeter</td>
<td>discharge</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 4.7: One Activity Occurring With One Alarm During Each 2 Minute Interval
<table>
<thead>
<tr>
<th>Figure</th>
<th>Alarm</th>
<th>Activities</th>
<th>Episodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>5a</td>
<td>syringe pump</td>
<td>drugs</td>
<td>observation</td>
</tr>
<tr>
<td>6a</td>
<td>syringe pump</td>
<td>drugs</td>
<td>talking</td>
</tr>
<tr>
<td>6d</td>
<td>infusion pump</td>
<td>observation</td>
<td>talking</td>
</tr>
<tr>
<td>6d</td>
<td>infusion pump</td>
<td>drugs</td>
<td>talking</td>
</tr>
<tr>
<td>6d</td>
<td>infusion pump</td>
<td>dr's visit</td>
<td>drugs</td>
</tr>
<tr>
<td>7c</td>
<td>pulse oxymeter</td>
<td>observation</td>
<td>talking</td>
</tr>
<tr>
<td>7d</td>
<td>infusion pump</td>
<td>physiotherapy</td>
<td>talking</td>
</tr>
<tr>
<td>8b</td>
<td>ventilator</td>
<td>physiotherapy</td>
<td>talking</td>
</tr>
<tr>
<td>9a</td>
<td>syringe pump</td>
<td>dr's visit</td>
<td>talking</td>
</tr>
<tr>
<td>9a</td>
<td>syringe pump</td>
<td>observation</td>
<td>talking</td>
</tr>
<tr>
<td>9a</td>
<td>infusion pump</td>
<td>drugs</td>
<td>observation</td>
</tr>
<tr>
<td>9c</td>
<td>syringe pump</td>
<td>drugs</td>
<td>observation</td>
</tr>
<tr>
<td>11b</td>
<td>E.C.G.</td>
<td>answering phone</td>
<td>talking</td>
</tr>
<tr>
<td>11c</td>
<td>E.C.G.</td>
<td>observation</td>
<td>talking</td>
</tr>
<tr>
<td>12d</td>
<td>E.C.G.</td>
<td>drugs</td>
<td>talking</td>
</tr>
<tr>
<td>12d</td>
<td>E.C.G.</td>
<td>observation</td>
<td>talking</td>
</tr>
<tr>
<td>12d</td>
<td>syringe pump</td>
<td>drugs</td>
<td>observation</td>
</tr>
<tr>
<td>12d</td>
<td>syringe pump</td>
<td>drugs</td>
<td>talking</td>
</tr>
<tr>
<td>13a</td>
<td>syringe pump</td>
<td>drugs</td>
<td>observation</td>
</tr>
<tr>
<td>13c</td>
<td>syringe pump</td>
<td>drugs</td>
<td>observation</td>
</tr>
<tr>
<td>13d</td>
<td>pulse oxymeter</td>
<td>observation</td>
<td>talking</td>
</tr>
<tr>
<td>14b</td>
<td>pulse oxymeter</td>
<td>observation</td>
<td>talking</td>
</tr>
<tr>
<td>14d</td>
<td>humidifier</td>
<td>physiotherapy</td>
<td>observation</td>
</tr>
<tr>
<td>14d</td>
<td>humidifier</td>
<td>physiotherapy</td>
<td>talking</td>
</tr>
<tr>
<td>14d</td>
<td>infusion pump</td>
<td>drugs</td>
<td>talking</td>
</tr>
<tr>
<td>15a</td>
<td>pulse oxymeter</td>
<td>observation</td>
<td>talking</td>
</tr>
<tr>
<td>15b</td>
<td>pulse oxymeter</td>
<td>observation</td>
<td>talking</td>
</tr>
<tr>
<td>15c</td>
<td>infusion pump</td>
<td>drugs</td>
<td>talking</td>
</tr>
<tr>
<td>15d</td>
<td>infusion pump</td>
<td>physiotherapy</td>
<td>drugs</td>
</tr>
<tr>
<td>15d</td>
<td>infusion pump</td>
<td>drugs</td>
<td>talking</td>
</tr>
<tr>
<td>15d</td>
<td>pulse oxymeter</td>
<td>observation</td>
<td>talking</td>
</tr>
<tr>
<td>15d</td>
<td>pulse oxymeter</td>
<td>discharge</td>
<td>talking</td>
</tr>
</tbody>
</table>

Table 4.8: Two Activities Occurring With One Alarm During Each 2 Minute Interval
<table>
<thead>
<tr>
<th>Figure</th>
<th>Alarm</th>
<th>Activities</th>
<th>Episodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>5a</td>
<td>infusion pump</td>
<td>physiotherapy</td>
<td>drugs</td>
</tr>
<tr>
<td>6d</td>
<td>infusion pump</td>
<td>dr's visit</td>
<td>drugs</td>
</tr>
<tr>
<td>7d</td>
<td>infusion pump</td>
<td>physiotherapy</td>
<td>drugs</td>
</tr>
<tr>
<td>9a</td>
<td>syringe pump</td>
<td>dr's visit</td>
<td>drugs</td>
</tr>
<tr>
<td>9a</td>
<td>infusion pump</td>
<td>drugs</td>
<td>observation</td>
</tr>
<tr>
<td>10a</td>
<td>pulse oxymeter</td>
<td>drugs</td>
<td>observation</td>
</tr>
<tr>
<td>10a</td>
<td>syringe pump</td>
<td>drugs</td>
<td>observation</td>
</tr>
<tr>
<td>12d</td>
<td>syringe pump</td>
<td>drugs</td>
<td>observation</td>
</tr>
<tr>
<td>14 d</td>
<td>humidifier</td>
<td>physiotherapy</td>
<td>observation</td>
</tr>
<tr>
<td>14 d</td>
<td>infusion pump</td>
<td>drugs</td>
<td>observation</td>
</tr>
<tr>
<td>15c</td>
<td>infusion pump</td>
<td>physiotherapy</td>
<td>drugs</td>
</tr>
</tbody>
</table>

Table 4.9: Three Activities Occurring With One Alarm During Each 2 Minute Interval

217
<table>
<thead>
<tr>
<th>Figure</th>
<th>Alarm</th>
<th>Activities</th>
<th>No. of Episodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>7d</td>
<td>syringe</td>
<td>physiotherapy</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>pump</td>
<td>drugs</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>observation</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>talking</td>
<td></td>
</tr>
<tr>
<td>9a</td>
<td>syringe</td>
<td>answering</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>pump</td>
<td>dr's visit</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>observation</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>talking</td>
<td></td>
</tr>
<tr>
<td>9a</td>
<td>E.C.G.</td>
<td>dr's visit</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>drugs</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>observation</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>talking</td>
<td></td>
</tr>
<tr>
<td>9a</td>
<td>infusion</td>
<td>dr's visit</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>pump</td>
<td>drugs</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>observation</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>talking</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.10: Four Activities Occurring With One Alarm During Each 2 Minute Interval
### Table 4.11: Activities Occurring with Two Alarms Within the Same Two Minute Interval

<table>
<thead>
<tr>
<th>Figure</th>
<th>Alarms</th>
<th>Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>8c</td>
<td>pulse oxymeter</td>
<td>ventilator alarm</td>
</tr>
<tr>
<td>9a</td>
<td>syringe pump</td>
<td>pulse oxymeter</td>
</tr>
<tr>
<td>9a</td>
<td>syringe pump</td>
<td>pulse oxymeter</td>
</tr>
</tbody>
</table>

In summary, the prevalent activities that occurred within the same two minute interval as an alarm are as follows; when one activity and one alarm occurred as shown in Table 4.5 the two most frequently occurring activities were observation and drugs. The largest category of combinations involved one alarm together with two activities within the same two minute period. The predominant combinations in this group involved an aspect of talking together with either physiotherapy, observation or drugs. When one alarm occurred together with three activities it can be seen in Table 4.7 that the predominant combination of activities involved talking, observation and drugs.

During the four occasions recorded when four activities and one alarm occurred there were three combinations of activities when drugs, observation and talking occurring together with another activity (e.g. physiotherapy or dr's visit).

#### 4.7.3.5 Differences in Physical Workload

During the period of the observation study there were considerable differences in the amount of physical workload undertaken by the staff at each bed. In Figures 4.5 to 4.15 it can be seen that the number of activities undertaken by staff varies and illustrates busy and quiet periods (high and low periods of physical workload).
One of the busiest beds observed during the observation study is shown in Figure 4.9a. During the observation session there were two trained nurses and a third year student nurse present at the bed. Two of the patient's relatives were also present sitting by the bed. During the consecutive 28 minutes that the observational study was conducted, the following activities and alarms occurred. Teaching of the student nurse by one of the trained nurses; the staff talked between themselves, to the patient and to the visitors. One member of the trained staff took the patient's observations while the second administered intravenous drugs. The doctor visited the patient and one of the trained nurses also answered the telephone at the nurse's station. Ten alarms sounded in total during the 28 minutes. The pulse oxymeter sounded twice, the syringe pump four times, the infusion pump three times and the E.C.G. monitor alarm once.

There are additional examples of beds were several different activities were being undertaken during the observation session. For example Figures 4.7d, 4.10a and 4.14d. In comparison, there are also examples of less busy beds with fewer activities occurring, for example as shown in Figures 4.5c 4.7a and 4.10c.
4.8 Discussion

The results of the direct observation study have produced a detailed description of the activities being undertaken in an I.C.U. when alarms occur. By comparing the data recorded by the two independent observers the information has been shown to be accurate, with a high degree of agreement between the observers.

One of the primary aims of the current study was to identify the types of tasks undertaken by the staff in the I.C.U. and to extrapolate the concrete activities into similar, representative abstract tasks. The abstract tasks could then be used in a laboratory study in which the set of auditory sounds would be learnt and then recalled under conditions in which other tasks were being performed concurrently. The classification and sequence of the findings from the observation study facilitated determining the types and combinations of tasks that would represent activities undertaken on the I.C.U.

The results showed that the majority of alarms occurred within the same two minute interval with one or more activities. There was only one example during the current study when the E.C.G. alarm occurred and no corresponding activities by the staff were taking place.

During the observation study specific aspects of talking, (teaching, staff talking to one another or to patients or visitors) were recorded by the observers and it was shown that overall, a generic category of 'talking' was the most frequently occurring activity. Throughout the majority of combinations it was demonstrated that 'talking' was frequently one of the activities occurring in the combination. In the laboratory task, therefore, it would be appropriate to include an aspect of talking particularly within a combination of tasks, when participants are identifying the alarm sounds.
The single largest group of activities involved the observation of the patient by staff. During the period of the observation study staff constantly checked the patients condition, for example, by looking at the catheter bag to check the urine output, examining the equipment such as the infusion pumps or syringe pumps administering drugs and scanning the equipment that monitored the patient's condition, such as the E.C.G. machine or the ventilator. In determining the type of task to be used in the laboratory environment a visual scanning observation or vigilance task could be used to represent the activity of observing the patient.

A third category that occurred frequently during the observation study was the category drugs that involved the calculation preparation and primarily intravenous administration of drugs. Some aspects of the drugs category could not easily be reproduced in laboratory conditions such as the preparation or administration. However one component of the drug category involved the calculation of drug doses as described in the introduction to this chapter and this element could be represented in the laboratory by an arithmetic or calculation task.

While physiotherapy also occurred frequently during the observation study and was a prevalent activity in the various combinations of activities it would be difficult to find a suitable type of physical task to represent physiotherapy satisfactorily in laboratory conditions.

As has been discussed above, talking was present in many of the combinations of activities. Other prevalent combinations that occurred during the observation study included observation and drugs. A combination of these tasks could be used in the laboratory study.

As in the video study reported in the previous chapter, the findings of the direct observation study suggest that often activities are being undertaken by staff when alarms are activated. In the video study, it was implied that some alarms, for example the ventilator alarm, was activated when the physiotherapist was undertaking
treatment with the patient. However, this relationship could not be seen on the video tape. The findings from the current study suggest that in some instances the activities of the staff appear to initiate certain alarms. For example, when physiotherapy was being undertaken in the current study the ventilator alarm was often activated, when drugs were being administered the syringe pump and infusion pump alarms were activated and when observation was being undertaken the pulse oxymeter was often activated. There could be several explanations as to why this occurred, for example when physiotherapy was being undertaken usually the ventilator was disconnected to allow the staff to suction the patient. If the machine was purposely disconnected then it is reasonable to assume that the alarm would be activated, if the ventilator had accidentally become disconnected then a potential disaster could occur if the machine did not alarm to indicate the problem. Luckily, accidental disconnections appear to be rare occurrences, while the disconnection by the physiotherapists and trained staff on the I.C.U. are more common, everyday events.

Similarly, the infusion pump alarm is more likely to sound when drugs are being administered because the infusion is turned off briefly while the drugs are given. If an accidental, unforeseen obstruction to the tubing of the intravenous infusion occurred then again the alarm would be extremely important to alert staff to the problem. As has already been discussed, during the observation study the staff were constantly checking and looking at the equipment and therefore it could be suggested that problems are located and detected before they become serious issues.

During the observation study it was not possible for the observers to know precisely why an alarm was activated, for example, whether it was due to a medical emergency such as the ventilator being disconnected or the patient's heart stopping or, whether it was because of artefact, such as the patient moving and the monitor for example the pulse oxymeter, which is clipped onto a finger for example, falling off.
The sampling strategy used during the observation study appears to have been adequate as alarms that had been heard infrequently during the video study, such as the humidifier and the E.C.G. monitor, were heard and recorded by the observers on several occasions. However, throughout the period of the direct observation study the pulmonary artery flotation catheter was not used to monitor the condition of any patients and was therefore not heard at all. Two other miscellaneous sounds were also not heard during the direct observation study, the doctor's bleep and the doorbell. During the video study the pulmonary artery flotation catheter alarm had occurred frequently but it could be argued that the observation study was not sufficiently long enough to record infrequent events such as the doorbell and the bleep. Conversely, as the primary aim of the direct observation study was the association between the activities and alarms and during the direct observation study all the alarms had been recorded, even those that had occurred infrequently during the video study, and it was not imperative that three of the miscellaneous sounds were recorded.

The observation study has illustrated that there were considerable variations in the physical workload of the trained staff. When a bed was very busy, (for example, as shown in Figure 4.9a) there was a corresponding increase in the number of alarms. However, as discussed in Chapter 1 some complex tasks require little attention while others require considerable attention and although the physical tasks could be recorded easily during the period of the observation study, the mental effort invested by the staff in the various activities was not visible and therefore not obtained. For an experienced member of staff there is also the possibility that, to a certain degree, some aspects of a task may become automated.

As discussed in the introduction there are many considerations to take into account when undertaking an observational study and it is important to acknowledge that observational research can usually only tap a fraction of what is going on in a complex situation, and the goal of the specific research is to understand that fraction (Reid 1982). There are numerous methodological problems when undertaking an observational study in a hospital unit such as the
I.C.U. For example, when deciding which temporal sampling strategy to use, it became apparent that a continuous, real-time procedure would be impractical to use, firstly because of the sensitivity and ethical considerations, for example when distraught relatives were present it was inappropriate to remain in the four-bedded bay, and the observers retired to the staff room. Secondly, the amount of information being recorded by the observers required intense concentration that could only be sustained for short blocks (28 minutes) in order to ensure reliable accurate data collection.

Direct, passive observation was chosen in preference to participant observation. There have been examples of participant observation (e.g. O'Carroll 1986) in which the staff on the unit recorded each occurrence of an auditory warning. However, for a team of outside observers to completely integrate in the I.C.U. as a member of staff for example, would be difficult. The I.C.U. staff are a tight-knit community. Each member of staff has a specific role and there is a well-defined hierarchy. The first priority for all staff is caring for the patient. It is difficult to envisage how an observer, whose first priority would be the collection of the data, not the well-being of the patient, could integrate recording the behaviour of the staff and also when alarms sound unless it was recorded retrospectively. This is a problem reflected by nurse researchers such as Binnie (1988) and Reid (1991), who were confronted with the dilemma of whether or not to help with the work on the wards.

It was considered by the observers that bias and drift were avoided by involving as many staff as possible, in an informal way, to help classify the activities, which in the final checklist were easy to discriminate and the close degree of association between the observers scores would appear to substantiate this.

Another important problem to consider when undertaking an observational study is the influence of the observer's presence on the behaviour of the participants. It was felt by the observers undertaking the current observational study that many of the suggestions in the literature on how to reduce reactivity were practised. The staff had adapted to the video camera being on the
unit and were used to observer TM visiting the hospital to change the tapes and set up the camera for the next filming session.

Both observers spent a considerable amount of time on the unit prior to the start of the study in an attempt to habituate the staff to their presence so that the subsequent behaviour would be realistic and accurate, indicating what actually happens on the I.C.U. It was also felt that a very good rapport was developed with the staff, who were very interested in the rationale for the study and did not appear to feel threatened or worried by the consequences of the results.

During the current study the observers were careful not to interfere with the normal work on the ward. During an observational study undertaken at a psychiatric hospital (Balogh 1991) it was reported that staff felt the study had disrupted normal work on the ward and some staff felt an invasion of their privacy. However, in that study the staff were asked what they were doing rather than just being observed as in the current study. In the current study the observers were totally passive and non-intrusive and the staff appeared genuinely eager to co-operate and participate in the study.

However, there has been very little empirical research on reactivity and its effects are not well understood (Kazdin 1982). It is difficult to suggest an adequate solution to the problem if procedures are introduced to minimise obtrusiveness, this introduces other problems such as ethical considerations. It could be concluded that in the current study the presence of the observers and the participant's awareness of being observed did not alter their behaviour, the staff appeared to adapt quickly to the observer's presence.

There were specific problems to the I.C.U. environment regarding the current study. For example, when the nurses closed the curtains around the bed when performing procedures that were necessary to protect the privacy of the patient, it was impossible to record the activities of the nurses. However, as previously mentioned when this occurred the observers stopped recording the activities. The
recording was discontinued because although it would have been possible for the observers to speculate what was happening behind the curtains or indeed to ask the staff what they had been doing it was decided that it could produce inaccurate results due to inference rather than actually seeing what was happening.

Another problem was the learning and localisation of the alarms by the observers. Because the observers were positioned at one end of the four bedded bay, initially it was sometimes difficult to localise which bed the alarm was coming from and it was also difficult to know why the equipment was alaring. For example, it was difficult sometimes to know whether an alarm was signalling a genuine medical problem or not. For this reason the alarms were not categorised into 'false alarms' or 'medical emergencies', but were recorded into a generic category.

The disadvantage of using two-minute sampling sessions, as was previously discussed in the introduction to this chapter, is that only the first occurrence of each activity is recorded. However, this was necessary in order to observe all the activities that were occurring at each of the four beds. It was difficult at times during the study to time the two-minute intervals and to record the information on to the appropriate chart. If an observer must concurrently look at a stop watch and the subject, and then interrupt contact to record the data, there could be some loss of accuracy resulting from the incompatibility of all three responses.

If a longer observational study were to be undertaken a mechanical recording device, such as a Psion Organiser, would be used. This is more flexible than recording by hand (Weick 1968). The Psion Organiser is a hand held, lightweight portable instrument that allows the experimenter to collect information and data by clicking a button. The data is recorded in computer code, and is readily transferable to a computer for analysis. This could also result in longer observational sessions. It is also possible that the observer could video tape the session, with the observer holding the camera. This could be an improvement on the video study reported earlier, (see Chapter 3) when the video camera was in a fixed position, as
the observer could film the events of interest. However, an observer filming the activities of the staff could result in increased reactivity.

Foster et al. (1988) suggests that to collect checklist data a pocket sized device that clicks at the end of each two minute block, for example through an ear piece that can only be heard by the experimenter, could be used. This device then frees the observer from continually checking a watch and therefore reduces sampling errors of missing behaviours that occur when the observer is checking the stop-watch.

As discussed previously, Sackett (1978) postulates that the purpose of interval sampling is to estimate the frequency of events rather than to provide an exact record of their occurrence in real time and that a large range of behaviours can be recorded simultaneously. The interval recording technique worked well using the two-minute intervals for recording the physical workload of the staff as in this study there were a wide range of behaviours to observe and one of the aims of the study was to produce a summary of events that were happening when the auditory warnings occurred, rather than a precise record.

To conclude, as a result of the study it has become clearer which activities and combination of activities occur frequently on the I.C.U. when alarms are activated. Aspects of three of the predominant activities, talking, observation and drug calculation, could be utilised into the laboratory phase of this research. The following chapter describes representative abstract tasks that reproduce the underlying cognitive abilities used in the three tasks undertaken on the I.C.U.
CHAPTER 5 Experiment 4: An Examination of the Effect of the Introduction of Tasks on the Identification of Auditory Warnings.

5.1 Introduction

The results of the observation study reported in Chapter 4 demonstrated that when auditory warnings were activated on the I.C.U. there were specific tasks and combinations of tasks that occurred.

The single largest category recorded involved the activity 'observation'. During the period of the observation study staff were constantly checking the patient's condition, for example by examining equipment such as infusion pumps or syringe pumps and scanning the equipment that monitored the patient's condition, such as the ventilator, to ensure that the readings displayed on the equipment remained within the pre-set parameters. It was suggested in Chapter 4 that a similar task that could represent the activity 'observation' in the laboratory environment would be a visual scanning or vigilance task. Which would represent, at a cognitive level, the observation task involved.

Another activity that occurred frequently during the observation study was the category 'drugs' which involved the calculation, preparation and the administration of drugs which was, in general, intravenous. Some aspects of the drugs category could not easily be reproduced in laboratory conditions, such as the preparation or administration. However, one component of the drug category involved the calculation of drug doses and this element could be represented in the laboratory by an arithmetic or calculation task.

Whilst physiotherapy also occurred frequently during the observation study and was a prevalent activity in the various combinations of activities it was not included as a task in the laboratory experiments reported in this chapter. It was considered that a suitable type of task involving physical
activities such as those undertaken by physiotherapists could not be satisfactorily represented in the laboratory.

The most frequent activity overall recorded on the I.C.U. during the observation study involved one or more aspect of talking, for example either teaching, or talking to other staff, patients or visitors. The results of the observation study showed that talking often occurred in conjunction with other tasks such as physiotherapy, observation and drugs.

There are several objectives in the experiment reported in this chapter. These include first establishing whether participants could successfully identify the set of twelve auditory warnings whilst undertaking a concurrent task or series of tasks. The experiments reported in Chapter 2 showed that in all three conditions the participants learnt the set of twelve sounds and retained that knowledge for a period of one week when they returned to undertake the second phase of the experiment, many sounds being remembered by the participants with no errors. It was therefore decided in the current study to undertake the learning phase in the same way as in the previous set of experiments and to introduce the tasks in the return phase.

A second aim of the current study was to examine whether any confusions that occurred between the sounds would increase or change in comparison to confusions that occurred in the first set of experiments (reported in Chapter 2) as a result of participants in the current study having to undertake tasks whilst identifying the sounds.

A third aim of the study was to examine the secondary task performance as the tasks increased in difficulty. A further objective of the study was to determine whether the reaction time or latency of participants' response in identifying the sounds would increase as the tasks became more difficult.

Gopher and Sanders (1984) have suggested that there are few tasks that occur in isolation and one of the interesting features
of these studies is what occurs when tasks are combined, rather than when tasks are undertaken separately.

5.1.2 Background

As discussed in Chapter 1 a range of conclusions have been drawn in the literature with regard to dual-task experiments. For example some studies have demonstrated that the performance of two tasks undertaken together is significantly worse than when compared with the performance of one of the tasks in a single task condition (e.g. Broadbent 1971, Kahneman 1973). Other studies have demonstrated that two tasks can be performed with no decrement to either task (e.g. Allport et al. 1972, Wickens et al. 1983). As McCann and Johnston (1992) maintain, a practical consideration is that interference between tasks may severely limit an operator's ability to function competently.

However, an important point regarding performance of two or more tasks is whether the combination of certain tasks can be undertaken together more successfully than other combinations of tasks (Wickens 1989).

Traditionally, dual-task research examines resource availability. If the resources required for specific tasks are in limited supply and divided between two or more activities then the tasks will receive an insufficient supply of resources and there will a decrement in the performance of the tasks relative to single task baseline measures (e.g. Wickens 1991, Damos and Lintern 1981). This has generally been the result in many, but not all of the studies demonstrating the dual-task decrement. Intuitively, we might expect there to be some activities that can be undertaken together, such as listening to the radio and driving a car, while others cannot, for example reading and driving a car.

Time-sharing between tasks results in a loss of performance in one or both tasks because of the need to share resources between them. The amount of the decrement in the performance
of either task depends on whether common or separate resources are used. If a task shares common resources, it is argued, then performance will deteriorate more than if the task demands involve separate resources. Furthermore, if tasks are very similar confusion between the tasks may occur (see Chapter 1).

Performance may improve because the co-ordination or integration of the subtask is enhanced (e.g. Duncan 1979, Peters 1977). Korteling (1993) describes this as 'skill integration' or 'performance integration'. If the same components are present in sub-tasks, these may facilitate the integration of separate skills into higher order skills (Neumann 1987). Performance integration based on the relationship between the tasks could be attained as a way to overcome interference between tasks when multiple sources of information are present, thereby reducing attentional demands and improving task performance. Korteling (1993) and Neumann (1987) propose that if dual-tasks have a high degree of similarity the task may, in fact, be perceived as a single task. For example, piano playing involves two hands each of which simultaneously performs, yet to an accomplished player the task is perceived as a single task.

However, if tasks are completely independent and do not share compatible or common processing routes it may be impossible to combine actions or to integrate skills. In practical situations, such as driving a car, as Neumann (1987) describes some aspects of the task are related just as some aspects of the tasks are completely independent. Parasuraman et al. (1987) claim that successive tasks require more attentional resources than simultaneous tasks, making more demands on working memory.

There are many examples of concurrent skills in which tasks using different modalities can be performed with very little decrement in the performance of either task. For example, Gerver (1974) demonstrated that simultaneous translators could operate with an overlap of 75% between their speech output in one language, and the sentences they were listening to in another language.
Other studies that have investigated complex tasks such as speaking and piano playing (Allport et al. 1972), have shown that both tasks can be conducted concurrently without a decrement in the performance of either task.

Allport (1980) cites another example which demonstrates how some tasks can be performed successfully together while others cannot. Tierney (1973) demonstrated that certain pairs of tasks could be performed together with little or no mutual interference, even with minimal practice, while other tasks were unsuccessfully paired together. The successful pairing was a study involving a typist who, in the successful task, was able to type at normal speed copying from a visual text while at the same time repeating aloud from an unrelated text played over headphones. The unsuccessful task involved typing from an audio-recording while at the same time reading aloud, which severely impaired the performance of both tasks, even after three days of practice.

Another study that attempts to demonstrate the independence of cerebral processing resources is illustrated in a Japanese study (Hatano et al. 1977) in which highly trained abacus operators were able to answer general knowledge questions presented aurally while undertaking a long complicated mental arithmetic calculation without any decrement to performance. However, if a relatively simple mental arithmetic question were asked this often interfered with the mental calculation of the complex problem. Allport (1980) argues that the results of this study demonstrate that the performance of the complex mental calculations is functionally separate from the mechanisms of speech comprehension and information retrieval involved in the answering of relatively easy questions and does not involve a central resource.

However, as Allport acknowledges, some of the skills described above, such as piano playing, the complex mental arithmetic calculations by the abacus operators, and typing, differ from
laboratory experiments in that they have often been developed through extensive practice.

Nevertheless, laboratory experiments have also demonstrated that certain tasks can be undertaken together more easily than the combinations of other tasks.

5.1.3 Explanation of Decrement

As discussed above and in more detail in Chapter 1 there are different conclusions regarding performance of tasks in dual-task situations. Some studies appear to show that two independent tasks can be performed together with no decrement to performance, while other results suggest that even in tasks that appear to use completely independent resources they cannot be performed successfully together.

One suggestion as to why there is not a decrement when the task load is manipulated is proposed by Craig (1991). He argues that in some situations even the heaviest load would still be a manageable task because it is not exceeding capacity, and that decrements may occur only when there is an already demanding primary task and performance will subsequently suffer from any increase on that load.

The multiple resources literature also describes some tasks as being data-limited tasks (e.g. Norman and Bowbrow 1975) which are less complex, relatively easy tasks that may be well practised, or tasks in which an operator has become highly skilled in.

Wickens (1984) introduced the concept of difficulty insensitivity which is described as occurring when the increased demands of one task fail to cause any interference on the performance of a second task. An example is demonstrated in a study undertaken by Wickens and Lui (1988) in which the demands of a verbal decision task were increased but there was a decrease in the interference with a spatial tracking task. The
study concluded that increased task demands could lead to an increased mobilisation and availability of all resources within the information processing system in a similar way to if an operator is told to 'try harder'.

Wickens (1991) further suggests that progressively more and more resources will be allocated to maintain performance on a primary task at a near perfect level and fewer will be allocated to the concurrent task, therefore performance on the latter will deteriorate. Conversely, if resources to the primary and concurrent tasks stay fixed the difference between required and actual resources allocated to the primary task becomes greater as the resources are divided between all the tasks. Performance on the primary task will then show a progressively greater loss than when the task was performed alone, and it is possible that both these effects will be seen.

Studies have also been conducted that have instructed participants to vary the allocation of effort between two tasks in different proportions (e.g. Vidulich 1988, Wickens et al. 1983). In general the results showed that the performance of the two tasks trade off reciprocally as discussed by Wickens (1991). For example, participants transferred resources from one task and diverted them to improve the performance of a secondary task.

Damos (1991) states that one of the most problematic aspects of dual-task methodology concerns practice. Participants should be trained to a certain level, although it is difficult to determine what the correct level should be, and should reach a stable level of performance before the data is collected. Some researchers have given their subjects many hours of practice (e.g. Bradley 1969; Leek and Watson 1984) and still found their subjects performance was improving. However, other studies have suggested that if a task is practised alone and becomes automated then two tasks can be performed without any decrement (e.g. Schneider and Fisk 1982).
It has been argued by some (e.g. Gopher and Donchin 1986) that it is important to ensure that, when combined, two tasks do not change their nature, in that the dual-task situation does not have different properties than those present in the single task situation. In many dual-task situations tasks are not viewed as independent by the participants, who may integrate certain tasks in order to reduce demands, as happens in the real-world. Conversely, it could be argued that there are few tasks that occur in isolation particularly in the 'real-world' environment, and one of the interesting features of these studies may be what occurs when tasks are combined rather than when tasks are undertaken separately. As Eggemeier et al. (1992) claim, in general research involving workload has not investigated the complexity of multi-task environments, which are characterised by the imposition of many different tasks on the operator. Eggemeier et al. (1992) also argue that often in the laboratory environment the instructions given to participants emphasise uniformity of task performance. In a multi-task situation on the other hand operators are given the opportunity to vary the scheduling of their tasks.

However, Damos (1991) suggests a long-standing problem in traditional dual-task research has been the controlling of the participants' priorities. The participants may adopt individualistic between-task trade-offs that can increase the between-participants variability. In experiments examining general time-sharing abilities, if the individual participants decide on their own task priority it can conceal the presence of any such ability.

5.1.4 Operator workload

As discussed in Chapter 1 there has been considerable research to assess mental workload of the operator.
The four main categories of measurement used being primary performance, secondary task performance, physiological methods and subjective reports.

In the current study reported in this chapter, the methods of choice used to assess the performance of the participants were primary task measurement, secondary task measurement and subjective ratings.

As Wickens (1992) states, when a system or an operator is being evaluated, the performance on the system of interest should be examined first. Hart and Wickens (1990) suggest that primary task measures must be obtained to determine whether or not the operator can accomplish the task, otherwise additional measures are virtually impossible to interpret without this information. However, they also urge caution when evaluating workload using primary task measures alone. Measuring performance alone is not an adequate measure because an operator may use different strategies to maintain primary task performance. For example, an operator may increase the effort required to maintain performance which subsequently increases the difficulty of the task for that operator, but no errors are recorded. Moreover, this increased difficulty for the operator is not observable. Task difficulty does not only involve the physical, observable composition of the task. Covert tasks such as planning, problem solving or decision making may be as equally demanding of the operator’s resources as the observable activities.

A secondary task is performed when the operator is asked to undertake an additional task to the primary task. Whilst primary task measurement assesses the operator’s ability to perform the task of interest, it is postulated that the secondary task methodology should enable an index of workload to be derived from the ability to perform both the primary task and the secondary task by measuring the decrement in the performance of the secondary task (Eggemeier and Wilson 1991)
Ogden et al. (1979) propose that the secondary task paradigm is highly similar to a dual-task paradigm, the major difference being that in the dual-task situation there is usually no emphasis on any priority in the performance of the task. A secondary task should impose a sufficient additional load so as to exceed the operator's capabilities, and the level of performance on the secondary tasks will decrease as the difficulty of the primary task increases. Traditionally in secondary task scenarios, the difficulty of the primary task is manipulated. However, in some studies the focus of attention is the effect of the secondary task on the performance of the primary task.

There is disagreement, as discussed in Chapter 1, between some authors regarding the methodology used when investigating performance using secondary task measures. For example, O'Donnell and Eggemeier (1986), argue that the secondary task should not interfere with the primary task and performance should be maintained on the primary task, while others, for example, Gopher and Donchin (1986), argue that it is acceptable if the secondary task is intrusive on the performance of the primary task. Gopher and Donchin's view of the dual-task paradigm is that it can be used to assess the operator's ability to perform two tasks concurrently and, it could be suggested, to establish how an operator determines priorities when undertaking two or more tasks concurrently. These disagreements, however, appear to reflect problems discussed in the dual-task literature, for example by Damos (1991), in which participants decide on their own priorities in undertaking tasks and consequently mask any time-sharing abilities that may be present.

In the current study it was therefore decided to state categorically to the participants that the priority when undertaking two or more tasks was the identification of the auditory alarms. The participants were allowed to determine their own priorities regarding the secondary tasks.
Another problem with implementing secondary tasks, particularly in an operational environment, is that if the primary task requirements are high then the operator may abandon the secondary task in order to maintain an acceptable performance on the primary task. Hart and Wickens (1990) state that in some experimental situations the secondary tasks are seen as unimportant and are dropped by either the operators or by participants in the laboratory. One solution to this problem is to use embedded tasks which are designed to appear as if they are part of or important to the primary task. It has been demonstrated that this strategy improves the likelihood that the tasks will be performed.

As discussed in Chapter 1, subjective ratings are often employed in studies of workload to obtain the operators' perceptions regarding workload. In the current study subjective ratings were obtained from the participants regarding the difficulty of each phase of the Experiment. A unidimensional rating scale was used in which participants were asked to rate the difficulty of each phase on a scale of 1 to 100.

5.1.5 Summary

In summary, it is suggested that resources are finite and in general it would appear that there is better timesharing between tasks if separate rather than mutual resources are used. If the resources and strategies invested into two or more tasks are the same, performance will worsen when the tasks compete for common resources as opposed to different resources. Many studies, some of which have been discussed above, have shown that certain tasks can be performed together without any or very little mutual interference. It has also been argued that if the demand increases for one task there will be fewer resources available to undertake the second task adequately. In certain situations the tasks can be completed by rapidly switching between the tasks and in other situations features of the tasks mean that by necessity they must be undertaken concurrently,
for example if two five-minute tasks must be completed in seven minutes (Wickens 1991).

Multi-task performance would appear to be influenced by;
1) amount of invested resources which is influenced by the task difficulty/complexity, data quality and practice.
2) resource efficiency
3) time sharing strategies
4) competition for common resources.
5) if underload or overload capacity is exceeded.

Williges and Wierwille (1979) hypothesise that if a single task increases in difficulty there should be an accompanying degradation in the performance of that task. If such a change did occur this could then be used as a measure of workload. However, they also suggest that in order to maintain performance of that task, the operator may change strategy as the task increases in difficulty.

5.1.6 Multiple Resources Model

When performing a task on its own or in conjunction with other tasks, different mental operations must be conducted, for example, response, rehearsal, perception etc. and performance of each of these tasks requires some degree of the operator's limited resources. Two activities will demand more resources than a single activity. The multiple resources model proposes that certain tasks or activities will demand different resources to other tasks or activities. Timesharing, switching, confusion and co-operation are all important components of multiple task performance (Wickens 1992). The demand for resources is also determined by the difficulty of the task.

Shiffrin (1975) reviewed a series of experiments in which attention was apportioned over tasks using different sensory modalities (visual, auditory and spatial channels) or within the same modalities. Shiffrin concluded that when attentional capacity had to be distributed over different inputs in
comparison to just one input, there was no additional cost to the accuracy or the efficiency of the monitoring. Allport (1980) suggests that there is evidence that sensory patterns coming simultaneously from separate spatial locations or separate modalities can be identified without competing and without capacity limitations.

As discussed in Chapter 1 the multiple resources literature (e.g. Kantowitz and Knight 1976, Navon and Gopher 1979, Wickens 1980) suggests the existence of separate dimensions that define multiple resources. Although the resources in the model are described as separate it is possible that as suggested by Weinstein (1987) in some tasks participants may use either a verbal or a spatial strategy. Wickens et al. (1984) point out that some tasks may require a mixture of resources. However, the important factor would appear to be that the closer together two tasks or interface channels are in a space, the more they draw on the same attentional resources. It then becomes more difficult to timeshare the resources and to subsequently undertake two or more tasks concurrently.

The multiple resources theory proposes that instead of one supply of resources, individuals have several different capacities with resource properties. As discussed in Chapter 1 the dimensions involved in the multiple resources model proposed by Wickens (1980) first classify tasks within the processing stages, second the processing modalities and third the processing codes. These are discussed below.

5.1.6.1 Processing Stages

Wickens (1992) suggests that there are two dimensions of the processing stage, first, perceptual evaluation (e.g. input of information from a display) and the central, cognitive processes which include memory rehearsal of digits and words, logical problem solving, mental arithmetic decision making and transformations. The second dimension involves those resources
which encompass the selection and implementation of responses. Wickens (1992) states that evidence for this dichotomy is shown when the difficulty of responding in a task is manipulated, yet this variation in difficulty does not affect the concurrent performance of a perceptual task. For example, Shallice et al. (1985) demonstrated that dual-task performance on a series of tasks involving speech recognition (perception) and speech production (response) utilised different resources.

Therefore, it is suggested, resources that are used for perceptual and central processing are the same, but are different to those resources used to select and execute the response. Perceptual-cognitive activities include for example, display reading, information monitoring, voice comprehension, mental rotation, situational assessment, diagnosis or calculation. Resources related to the response processes include control manipulation, switch activation or voice command. Tasks with demands in the former category can effectively be shared with tasks in the latter category.

5.1.6.2 Perceptual Modalities.

In the second dimension, the perceptual modalities, it is postulated by Wickens (1991) that separate resources are used for auditory stimuli and visual stimuli. There are many studies that have shown that tasks using different modalities are in general performed better than tasks which share the same modalities. For example attention can be divided more efficiently between the eye and the ear when compared with two auditory channels or two visual channels. Parkes and Coleman (1990) demonstrated that when subjects were driving a simulated vehicle regarding high demands from visual resources, information regarding the route was more effectively presented in an auditory format than in a visual format.
5.1.6.3 Processing Codes

The processing codes consist of spatial and verbal processes and there is additional support for this dichotomy from the literature regarding processing in the cerebral hemispheres (e.g. Springer and Deutsch 1985). It is suggested that, in general, spatial and left-handed processing occurs in the right hemisphere whilst resources underlying verbal processing, speech response and right-handed control are in the left hemisphere.

In the processing code dimension, verbal processing utilises different resources to the processing of non-verbal material. There is a division between processes of perception and transformations in working memory, and vocal (oral) and manual (spatial) responses.

Wickens (1991) proposes that the division between verbal and spatial codes are relevant to three stages of information processing.

i) the perceptual representation of verbal (text or speech) versus non-verbal spatial orientations and analogue representations,

ii) the division is relevant to central processing and working memory, where for example, Baddeley (1986) has compared 'spatial' and 'verbal' working memory. The former is used to retain and rehearse visual, spatial or navigational information while the latter is said to retain words or digits, to perform mental arithmetic or to solve logical problems. In some tasks participants could use a verbal or a spatial strategy (e.g. Weinstein 1987)

iii) Spatial and analogue information is assumed to use different resources to those processing verbal and linguistic information.

Polson and Friedman (1988) demonstrated that verbal and spatial processes, whether operating in perception, working memory or during a response, depend on separate resources. The separation of spatial and verbal responses accounts for the high degree of
efficiency with which manual and vocal outputs can be effectively time-shared.

A study undertaken by Derrick (1988), which was based on the theoretical multiple resources model (e.g. Wickens 1980, 1984) is discussed in detail in order to present a comparison between the tasks used in Derrick’s study, and the tasks used in the current study.

In Derrick’s study four tasks were designed that he states according to the tenets of the multiple resources model, demand resources from different dimensions. There were two components, easy or difficult, to each task and the tasks were paired together in all possible combinations. Other measures used in the study to assess multidimensional performance were subjective measures and the physiological measure of heart rate variability. The tasks used were as follows;

i) A critical tracking task in which a cursor had to be kept on a horizontal line. Control for the easy version was with the left hand and when the task was time shared with itself, both hands were used. According to the multiple resources model this task should demand resources allocated to spatial codes, visual input modalities and manual response modalities.

ii) A visual search task in which participants searched for a target word 'NOW' that was embedded in an array of upper case letters and non-letters. Detection entailed the subject squeezing a joystick with the right hand when the correct word was detected. When the task was time shared with itself the task included searching for a second word, 'ONE' and squeezing the joystick with the right hand. The combination of resources for this task, according to Derrick, should include verbal codes of central processing, visual input modalities, and manual response modalities. Derrick states that perceptual codes are also involved in the processing of this task. McCarthy and Donchin (1981) showed that with a similar task and a
physiological measure, (evoked brain potentials) that perceptual processes were also involved.

Krose and Julesz (1990) suggest that there are two types of visual search task, first if the signals are very dissimilar the target signals are easy to detect. Second if the signals are very similar signal detection is more difficult. Derrick (1988) included non-letters in the task in order to make detection of the target word or words easier. In the easier task 15% non-letters were used while in the difficult task only letters were used.

iii) The auditory task in Derrick's study involved participants judging a tone burst and positioning an arrow on a vertical scale indicating the frequency of each tone heard. For the easy task judgement of four tones was required (400Hz, 480Hz, 650Hz and 850Hz). For the difficult task two tones were added (325Hz and 1000Hz). The participants were asked to position the arrow with their left hand. When the task was paired with itself, two joysticks and two scales were used and tones were played in both ears. Derrick (1988) states that performance of this task should utilise resources associated with spatial codes, auditory input modalities and manual response modalities. The judgement and control aspects of the tasks should also require perceptual and central processing modalities.

iv) An auditory Sternberg task (Sternberg 1969, 1975) was also used. The Sternberg task involved participants retaining in memory a set of either two (easy task) or five (hard task) letters, which were presented either visually or verbally, and required a vocal response. ('YES' if the letter came from the memory set or 'NO' if it didn't). When the task was paired with itself a set of letters was presented to each ear and 'YES' was the appropriate response only when the ear and the letter matched. Derrick (1988) states that the Sternberg task should demand resources for verbal codes, vocal response and auditory output. Mental rehearsing and matching of the letters would indicate the use of perceptual and central resources.
The results of the study were analysed in terms of performance decrements using the easy tasks as a baseline. The results of the study in general agreed with Wickens (1984) predictions that when tasks competed for common resources there was a greater performance decrement than when tasks utilised separate resources. The subjective ratings and resource-performance combinations dissociated in some of the conditions, for example the heart rate variability generally increased across the combinations of all the tasks (indicating increased effort) yet participants did not perceive (i.e. rate) the increased difficulty in the analogue manual task. However, Derrick concludes that this result was not unexpected, as similar conclusions of dissociation between measures had been found in other studies (e.g. Yeh and Wickens 1988).

As in Derrick's (1988) study, tasks for the current study were selected that, according to the multiple resources model, use separate resources and also reflected the type of tasks undertaken on the I.C.U. by the trained staff and which were identified in the observation study.

5.2 The Current study

5.2.1 Introduction

In this chapter an experiment is reported in which the main objective was to examine whether participants could correctly identify a previously learned set of sounds without any decrement in performance while concurrently undertaking a task or combination of tasks. A second aim of the study was to establish whether the number of confusions that occurred would increase or change with increases in the complexity of the secondary tasks, and whether certain tasks would interfere with the correct identification of the sounds.

A further aim of the study was to determine whether the reaction time or latency of participants' response in identifying
the sounds would increase as the tasks became more difficult. In the previous section the dual-task literature indicates that in general there is a decrement in the performance of two tasks undertaken concurrently in comparison to single task performance. However, as has been discussed this is not always the case. The multiple resources theory suggests that if tasks are dissimilar and use separate resources, then the tasks could be undertaken concurrently without any decrement in performance. Although as Craig (1991) suggests, if the tasks are difficult then capacity may be exceeded.

The participants in the current study were required to learn the same set of sounds as those used in the first three experiments (see Chapter 2) and to then identify them while undertaking a concurrent task or combination of tasks.

5.2.2. Experimental Overview

It would appear that in general using different stimulus modalities results in better multi-task performance than if using the same modality, and although a performance decrement is still found it is usually smaller. In considering the response modalities, traditional dual-task research uses one response modality, for example, hand or voice response, and usually the responses can be performed simultaneously. Some studies have used two separate response channels but have not required simultaneous responses they may, for example require a very infrequent response to one of the stimuli (e.g. Israel et al. 1980).

An attempt was made in the current study to select tasks that were representative of what happens in the 'real' world. A criticism of models used in some studies of selective attention is that the task situations used are rarely encountered in 'normal' interactions in the environment (e.g. Tipper et al. 1992).

Damos (1991) suggests that one of the most problematic areas of dual-task research is the selection of the task. She proposes that there are six characteristics of task combination which
should be considered. These include: the number of stimuli, the modality of the stimuli, whether there is a correlation between the stimuli, (which means that some aspects of one task can be used to predict a dimension or aspect of the second stimulus), the central processing requirement, the number of response channels and the modality of the response channels.

In the current study the tasks, which are described in greater detail below, were, in the first instance, paired with the primary task, as in the usual dual-task paradigm. The Experiment proceeded with a combination of secondary tasks together with the primary task. The tasks were not correlated as the stimuli were independent, the participants did not have to combine information from each of the tasks in order to respond correctly. The tasks had varying central processing requirements and the response channels were manual, apart from one, which was a vocal response.

The tasks used were selected on the basis of the findings of the observation study described in Chapter 4 and each task is related to the multiple resources model.

5.2.3 The Tasks

5.2.3.1 Mental Arithmetic Task

One of the activities that occurred frequently during the observation study was the category 'Drugs' which involved the calculation, preparation and administration of drugs. Some aspects of the drugs category could not easily be reproduced in laboratory conditions, such as the preparation or administration of drugs. For the current study the calculation of a drug dosage was selected as one of the tasks to be used in the laboratory study. (Some examples of the type of drug calculations undertaken on an I.C.U. are given in Chapter 4.)

A mental arithmetic calculation was used to represent the task of calculating drug dosages. The mental arithmetic task
selected was from the Human Assessment Laboratory (HAL), University of Plymouth which were developed for use by the Ministry of Defence in recruit selection. The tests, known as 'brackets', consisted of two A4 sheets of 72 calculations in total (36 on each sheet). The sheets were alternated throughout the experiment. Six of the participants had the first sheet of calculations in Phase 2, the second sheet of calculations in Phase 3 and the first sheet again in Phase 4. The remaining five participants were given the second sheet of calculations in Phase 2, the first sheet of calculations in Phase 3 and the second sheet of calculations again in Phase 4. Thus half of the participants saw the first sheet twice and half saw the second sheet twice. During each phase the participants began the calculations at the beginning of the relevant sheet. The participants were asked to calculate the problem in their head and then to tick the answer that was the larger of the two figures. By asking the participants to work out the calculations in their heads, they were thus holding the numbers in short-term working memory. The participants received no practice on the mental arithmetic task prior to performing it under the experimental conditions. The number of correct calculations for each subject was recorded.

An example of one of the 'bracket' calculations is shown below:

\[
\begin{align*}
42/7 & \quad (A) \\
27-(6\times3) & \quad (B)
\end{align*}
\]

(In the example given above A equals 6 and B equals 9, therefore the correct answer is B.)

A representation of the mental arithmetic task in relation to the multiple resources model proposed by Wickens is illustrated in Figure 5.1. The resources used, as suggested by the multiple resources literature, for the mental arithmetic task are shown in Figure 5.2.
Figure 5.1 Mental Arithmetic Task (Adapted from the Proposed Structure of Processing Resources. From Processing Resources in Attention C.D. Wickens 1984 in R Parasuraman and R Davies (Eds) Varieties of Attention, New York Academic Press).

<table>
<thead>
<tr>
<th>Perception</th>
<th>Central Processing</th>
<th>Response</th>
</tr>
</thead>
</table>
| VERBAL     | • logical problem solving  
             • mental arithmetic | • manual written response |
| • print    |                     |          |

Figure 5.2 Resources Used in Mental Arithmetic Task
5.2.3.2 Observation Task (F/E task)

In the previous chapter it was shown that the single most frequently occurring activity undertaken by the staff on the I.C.U. was that of 'Observation'. The trained staff were repeatedly checking the condition of the patient by scanning the patient's monitors and equipment to detect, for example, whether a patient's condition was deteriorating due perhaps to a sudden change in heart rate, which would be displayed numerically and also as an analogue display of the QRS complex, the waves displayed on the E.C.G monitor. A reduction in oxygen perfusion would be indicated numerically by the pulse oxymeter. The information from a ventilator, for example respiratory rate and PO$_2$ pressure, is also presented in a digital alphanumeric format. Blood pressure figures can also be presented in a digital format, for example, on the Dynamap machine. There is also information that is not presented via a monitor, for example drainage from a catheter bag or chest drain. It was suggested in Chapter 4 that an aspect of the observation task that could be represented in the laboratory environment would be a visual scanning or vigilance task.

When there is more than one critical event that can occur in a multi-task environment, signals may be missed because of the extra load on memory (e.g. Kidd and Micocci 1964, Childs 1976). The typical task used in most vigilance research requires the detection of one rather critical event that can occur at random from a single origin (Craig 1991). As discussed in Chapter 1 vigilance occurs when an operator has to look or listen over often lengthy periods of time for often infrequent events to occur. Staff who work in I.C.U. have to be vigilant to detect when sporadic, critical events occur, for example a rapid change from a pre-determined reading, and the cost of failing to detect these changes could be critical.

Another factor that affects the detection of visual signals is the knowledge of where the target signal will present. In studies in which participants have prior knowledge of a targets
position, for example in studies undertaken by Posner et al. (1982), the detection rate of the target signal increases notably. In the I.C.U environment, information regarding the patient is presented on a monitor in the same position, that is the target signal does not move and the position of the information is known to the staff. However, in order to assimilate information regarding the patient's condition the staff must look at several pieces of equipment.

In order to represent the activity of observation a vigilance task (the F and E task) was adapted by Dr Peter Dann (University of Plymouth) from Fischler et al. (1980) for use by Psychology undergraduate students. It was written as a B.B.C. Basic Program for use on a B.B.C. computer. The F and E task is based on the Theory of Signal Detection (Green and Swets 1966) whereby the observer accumulates sensory evidence and makes a judgement about it. The operator has to decide whether the signal is the wanted signal or not, and in this situation two types of error can occur; misses, where the operator fails to report or record the wanted signal, and false alarms when the wanted signal is reported although, in fact, it did not occur. If both kinds of error increase then performance will decrease.

The F/E task involved detecting the appearance of an upper case letter 'E' in a sequence of other upper case letters, the letter 'F'. The 'F's' appeared on the screen one after another filling up the screen. The letter 'E' was presented once within each run instead of an 'F'. The letters were all the same size (approximately size 16 Geneva font) the size remaining constant throughout the experiment. All the letters appeared in random position on the screen of the B.B.C. Computer. There were two programs which presented the letters on the screen at different speeds, fast and slow. The 'slow' program presented each letter at 3-second intervals, while the 'fast' presentation was at 0.9-second intervals. Each run lasted 30 seconds and there was a time limit allowed for detection of the letter 'E'. There were eight trials per sessions and the time limit varied each time a new trial commenced. The results for the fast and slow trials
were combined and the participants' scores were recorded as 'Early' (pressing the spacebar before the target letter appeared), 'Late' (pressing the spacebar after the signal had been presented) and 'Hit' (pressing the space bar at the correct time).

The F and E task was forced paced, with no accuracy feedback and with an immediate, simple response of pressing the space bar on the computer. The participants received no practice on the F/E observation task prior to the onset of the experiment. According to the multiple resources literature the task should demand resources from spatial codes, visual input and manual response with low perceptual and central resource demands. However, it is possible that there were higher cognitive demands when the participants were loading a new program into the B.B.C. computer. The observation task was predicted to result in poorer performance as the vigilance literature suggests that decrements in performance over even very short periods of time can occur.

A representation of the observation task with respect to the multiple resources model is shown in Figure 5.3 and a description of the resources used in the observation task is shown in Figure 5.4.
Figure: 5.3: Observation Task (Adapted from the Proposed Structure of Processing Resources. From Processing Resources in Attention C.D. Wickens 1984 in R Parasuraman and R Davies (Eds) Varieties of Attention, New York Academic Press).

<table>
<thead>
<tr>
<th>Processing codes</th>
<th>Central Processing</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPATIAL</td>
<td>• letter recognition</td>
<td>• manual press spacebar</td>
</tr>
<tr>
<td>• visual search</td>
<td>• display reading</td>
<td></td>
</tr>
</tbody>
</table>

Figure 5.4: Resources Used in the Observation Task.
Figure 5.3: Observation Task (Adapted from the Proposed Structure of Processing Resources. From Processing Resources in Attention C.D. Wickens 1984 in R Parasuraman and R Davies (Eds) Varieties of Attention, New York Academic Press).

Figure 5.4: Resources Used in the Observation Task.
5.2.3.3 Talking Task

Overall, the most frequent activities recorded during the observation study reported in Chapter 4 involved various aspects of talking, for example either teaching or talking to other staff, and talking to patients or visitors. It was also shown in the study that talking often occurred in conjunction with other tasks such as physiotherapy, observation, and drugs. It was therefore decided in the current laboratory study to use talking as one of the tasks.

The participants were asked to imagine that they were also instructing another member of staff. It would have been unrealistic to expect the participants, none of whom had any medical knowledge or experience, to actually role-play a medical scenario as it could have used all their resources to construct a sensible conversation. The topics of conversation were as follows,

i) directions to local landmarks in Plymouth, which were a) the Theatre Royal, b) Central Park and c) the milk department in the local Sainsbury store in the Armada Centre d) a long distance route travelled to either the parental home or to visit a friend.

ii) description of subject areas in the first year undergraduate course a) research methods, b) statistics.

The talking task was not formally measured. However, identical topics of conversation were conducted with each participant although the topics were asked in a random order. The participants were kept talking with the experimenter continuously prompting with questions to ensure that there were no periods of silence.

According to the multiple resources literature cognitive, perceptual and central demands are necessary to coherently answer questions. The demands of the tasks are likely to be greater than, for example, simply repeating a letter.
Therefore in this task that the verbal modality was used, although in questions that requested information regarding directions (for example to a local amenity) spatial representation would presumably also be utilised. Weinstein (1987) suggests that in some verbal tasks participants could use either a verbal or a spatial strategy. The verbal response modality in the talking task differed from the manual responses used in the other tasks.

The talking task with respect to the multiple resources model is illustrated in Figure 5.5. The resources used are illustrated in Figure 5.6.

![Diagram](Figure 5.5: Talking Task (Adapted from the Proposed Structure of Processing Resources. From Processing Resources in Attention C.D. Wickens 1984 in R Parasuraman and R Davies (Eds) Varieties of Attention, New York Academic Press)).
### Processing Codes

<table>
<thead>
<tr>
<th>VERBAL</th>
<th>Central Processing</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>• speech</td>
<td>• comprehension</td>
<td>• voice</td>
</tr>
<tr>
<td>• logical problem solving</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 5.6: Resources used in the Talking Task.

#### 5.2.3.4 Auditory Task

In the auditory task the participants were requested first, to learn the twelve sounds ('natural' sounds) which were played via the Archimedes computer. During the learning phase the name was told to the subject by the experimenter, and was also printed on a card which was placed in front of the subject on the table. When all twelve warnings had been presented, the participants then heard each sound in a random order and were required to name the sound. If named incorrectly, correct feedback was given immediately. This procedure was repeated twelve times, so that participants were required to name each warning on twelve occasions.

For the return phase of the Experiment a 'Basic' program was written by Professor J. Evans (University of Plymouth) for the Archimedes computer. The twelve sounds were recorded as modules and were incorporated into the 'Basic' program. The aim of the program was to present the sounds in a random order and at random times, to represent the conditions on the I.C.U. As demonstrated during the video study reported in Chapter 3, the frequency of occurrence of auditory alarms on the I.C.U. is unpredictable, and the duration was usually quite brief, consisting usually of only a few seconds.

Each of the twelve sounds was played 4 times. The sounds were played every 20 seconds plus or minus 6 seconds. During the return phases the participants were not given feedback as to whether the name chosen was correct or incorrect.
The names of the alarms were presented on the whole screen area of the Archimedes in the format illustrated in Figure 5.7.

<table>
<thead>
<tr>
<th>AIRCALL</th>
<th>OXYGEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLEEP</td>
<td>PERFUSION</td>
</tr>
<tr>
<td>HEARTMONITOR</td>
<td>PULSEMETER</td>
</tr>
<tr>
<td>FIRE</td>
<td>SYRINGE</td>
</tr>
<tr>
<td>HUMIDIFIER</td>
<td>VENTILATOR</td>
</tr>
<tr>
<td>INFUSION</td>
<td>RESPIRATORY</td>
</tr>
</tbody>
</table>

Figure 5.7: Representation of the Names Presented on the Archimedes Screen

The program recorded which sound had been played and the responses made by each subject. The program also recorded the time taken to respond to the sounds in milliseconds. It was hypothesised that some sounds would take longer to identify and would therefore have a longer latency. If participants did not respond to the sound they were 'timed-out' after 90 seconds.

The original idea had been to use a touchscreen so that when the alarm sounded the subject would be able simply to touch the screen with the pointer to indicate which alarm they thought had sounded. However, the software for the computer and the touchscreen were incompatible and the program was therefore modified to enable the participants to use the mousepointer and to click on the names of the alarms they thought to be correct.

The auditory task was forced paced with no feedback given regarding whether or not a correct response had been made for the name chosen.

According to the multiple resources model literature, the auditory discrimination task would use resources assigned to the auditory modality. As the participants retained the sounds in long term memory it could be suggested that perceptual and central cognitive resource demands were made. The response was a simple manual response which entailed participants using
the mouse to move the cursor on the computer screen and to click on the name of the alarm they thought had sounded. The auditory task compared with the multiple resources model is shown in Figure 5.8 and the resources used are shown in Figure 5.9.

Figure 5.8: Auditory Task (Adapted from the Proposed Structure of Processing Resources. From Processing Resources in Attention C.D. Wickens 1984 in R Parasuraman and R Davies (Eds) Varieties of Attention, New York Academic Press).
Figure 5.9: Resources Used in the Auditory Task.

5.2.3.5 Presentation of the Tasks in Each Phase of Experiment 4

Phase 1 Learning the auditory warnings.

Phase 2 i) mental arithmetic and identification of the sounds.
   ii) observation task and identification of the sounds.

Phase 3 mental arithmetic task plus the observation task and identification of the sounds. Phase 3 increased the complexity of the task by combining the mental arithmetic task and the observational task together with the identification of the sounds.

Phase 4 mental arithmetic and observation task plus talking task and identification of the sounds. Phase 4 further increased the complexity of the tasks, with the introduction of a talking task and also the identification of the sounds.

The primary task did not change in terms of difficulty or complexity, it was predicted that the combination of the secondary tasks would increase in the overall difficulty of the identification of the auditory sounds (the primary task). Based on the premise that increased task difficulty increases resource cost (Navon and Gopher 1980), the combination of tasks increased in the overall task difficulty and subsequently increased the amount of resources used by the participants to
complete the tasks. Yeh and Wickens (1988) propose that tasks are increased in difficulty by adding to the perceptual/central processing load and will be perceived as being more difficult. Derrick (1988) suggests that with increasing complexity, less feedback and greater time pressure should combine to increase perceptions of workload.

In order to determine whether the participants perceived that the workload had increased, subjective ratings regarding the difficulty of the tasks were obtained at the end of the experiment using a relatively simple method of asking the participants to rate on a scale of 1 to 100 how difficult they found each phase.
<table>
<thead>
<tr>
<th>Phase 1</th>
<th>Phase 2</th>
<th>Phase 2</th>
<th>Phase 3</th>
<th>Phase 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning sounds</td>
<td>Mental Arithmetic •verbal</td>
<td>Observation •spatial codes</td>
<td>Mental Arithmetic •verbal</td>
<td>Mental Arithmetic •verbal</td>
</tr>
<tr>
<td></td>
<td>codes •central processing</td>
<td>codes •peripheral</td>
<td>codes •central processing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>•manual response</td>
<td>processing •manual</td>
<td>processing •manual</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>response</td>
<td>response</td>
<td></td>
</tr>
<tr>
<td>Identify sounds</td>
<td>Identify sounds •verbal</td>
<td>Identify sounds •spatial</td>
<td>Observation •spatial codes</td>
<td>Observation •spatial codes</td>
</tr>
<tr>
<td></td>
<td>codes •central processing</td>
<td>codes •peripheral</td>
<td>codes •central processing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>•manual response</td>
<td>processing •manual</td>
<td>processing •manual</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>response</td>
<td>response</td>
<td></td>
</tr>
<tr>
<td>Identify sounds</td>
<td>Talking</td>
<td>Identify sounds •verbal</td>
<td>Talking</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>codes •central</td>
<td>codes •central</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>processing •manual</td>
<td>processing •manual</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>response</td>
<td>response</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.1: Combination of Resources Involved in the Tasks for Each Phase of Experiment 4.
5.2.4 Phase 1 Learning

5.2.4.1 Introduction

The learning phase of the experiment followed the same procedure described for the previous set of experiments (see Chapter 2). In the learning phase the set of twelve sounds were learnt in a paired-associate learning paradigm. First, each of the twelve sounds was played via the Archimedes computer, and the name was told to the subject by the experimenter. The name was also printed on a card which was placed in front of the subject on the table. When all twelve warnings had been presented, the participants then heard each sound in a random order and were required to name the sound. If named incorrectly, correct feedback was given immediately. This procedure was repeated twelve times, so that participants were required to name each warning on twelve occasions.

Because the results of the first set of experiments had shown the 'natural names' condition had been the easiest to learn, it seemed expedient in the current experiment for the participants to learn the sounds as quickly as possible as the focus of the current experiment was to examine any decrement in performance that occurred with the introduction of the tasks in the return phase.

After the participants had completed the learning phase they were asked to return one week later for the second phase of the study. However, one subject withdrew and the remaining 11 participants completed the remaining phases of the study. The participants were paid £7.00 for participating in the final two phases of the study.
5.2.4.2 Method

5.2.4.3 Participants

Four males and eight female participants undertook Phase 1. The participants were all members of the Psychology Department, University of Plymouth. Their ages ranged from 18-37 years. The participants were offered the choice of participation points or payment upon completion of the study. The participants had not undertaken any of the previous experiments. Prior to the start of the experiment the participants were asked whether they had any problems with their hearing. No replies in the positive were received. The participants' hearing was not formally tested.

5.2.4.4 Stimuli

The same twelve sounds used in the previous experiments (reported in Chapter 2) were used in the current experiment. The sounds were those used in the 'natural names' condition. The sounds remained in the same format in which they had been recorded from the I.C.U. Consequently, in this set some warnings were louder than others although they had all been recorded from the same distance, some were longer, and some repeated whereas other did not.

5.2.4.5 Procedure

The participants undertook the study one at a time in the auditory perception laboratory at the University of Plymouth and were given the following printed instructions:

"1) The aim of this phase of the experiment is to learn twelve sounds."
2) Firstly, you will hear the twelve sounds and I will tell you the name of each sound. To help you learn the names they are printed on the cards in front of you.

3) Twelve trials will then follow and you should try to match the sounds to the names. Say the name as soon as you know which sound it is.

4) Each sound will be played once in each trial and I will give you feedback as to whether or not the sound has been correctly identified.

5) Once all of the sounds have been correctly identified, or all twelve trials have been completed, that is the end of the first phase of the experiment."

The participants were then asked if they had any questions and the learning phase commenced. Each of the twelve sounds was played at random in a total of twelve trials. The name the participant said was recorded by the experimenter and the participants were told whether or not the answer they had given was correct. When the twelve trials had been completed the participants were asked to return one week later for the second phase.

5.2.5 The Return Phase

Phase two of Experiment 4 was again conducted in the auditory perception laboratory, University of Plymouth. Participants participated one at a time and were seated at a large desk. There were two computers on the desk, the names of the auditory warnings were presented on the Archimedes computer which was to the left with the screen turned to face the participant. A B.B.C. microcomputer with the observation task was to the right of the participant. There was room in front of the computers for the participant to write the answers to the mental arithmetic calculations. The experimenter sat unobtrusively behind the participants. A representation of the
layout of the tasks used in Experiment 4 is depicted in Figure 5.10.

As discussed above it was predicted that the combination of the tasks increased the overall task difficulty and consequently all the participants completed the tasks in the same order. However, as the phases were undertaken in the same order for all participants, there was the possibility that there would be an order effect.

The instructions to the participants were presented in the following order;

1) The participants were reminded of the names and the twelve sounds were played, the cards with the printed names were placed on the desk in front of the participant. The participants were asked to identify the sounds in a practice run.

2) The participants were told that the sounds would be played at random on the Archimedes computer and that they were
required to respond as quickly and as accurately as they could by clicking the mouse pointer on the name of the sound they thought had been played. The participants were shown an example by the experimenter. In order to control the participants' priorities, as reported by Damos (1991) and as discussed in the introduction to this chapter, the participants were told that their priority was to identify the auditory sounds. The participants were also informed that if they did not answer within 90 seconds they would be 'timed-out'.

The program was loaded by the experimenter and the command, 'subject number' appeared on the screen. When the participants number had been typed in the prompt, 'press spacebar or click mouse to continue' was shown on the screen. The participant was asked to carry out the command when they were ready to begin the experiment.

3) The participants were shown the mental arithmetic calculations and instructed to work through the calculations and to tick the answer with the highest value in each question. A pencil was provided.

4) The participants were asked if they had any questions. They then started the program on the Archimedes and began calculating the answers to the mental arithmetic. When an alarm sounded they responded by clicking with the mouse pointer. The program ran for approximately 20 minutes. The experimenter sat unobtrusively in the corner to ensure that the program ran correctly.

5) When the program had finished the experimenter collected the completed mental arithmetic calculations and wrote the subject number on the top. The Archimedes program was reloaded and a new subject number given. The instructions for the F and E observation task were then explained by the experimenter. The instructions for the task were also written on the white board in the experimental laboratory.
The participants were instructed: to "Load 'Program 1'"

When the program was loaded the participant was instructed to type in "Run" and the prompt "Subject number" appeared on the screen. The participant then typed in their number and for each subsequent trial typed in a sequential number prior to beginning the new trial. The participant was then asked to select either condition A/B (fast/slow), the conditions alternating thereafter. When they had selected a condition were then told to press the space bar when they were ready to begin. The instructions were then printed on the screen as follows:

"You will see a number of 'F's' appearing on the screen one after another filling up the screen. However at some point the letter 'E' will appear instead of an 'F'. As soon as you see the 'E' press the space bar as quickly as you can."

The participants responses was recorded on the computer under each individual participant's number.

6) The participants were again reminded that the priority was to identify the auditory sounds.

7) The participants were asked if they had any questions. They then clicked the space bar or mouse of the Archimedes computer when they were ready and pressed the space bar of the B.B.C. computer to commence the F and E program.

8) The program ran for approximately 20 minutes. The experimenter stayed in the room. When the program had finished the participants were thanked for their participation and asked to return one week later to complete Phase 3 and 4 of the experiment.

5.2.6 Phase 3 and Phase 4

Phase 3 and Phase 4 took place approximately one week after Phase 2 had been completed. The participants were again
reminded of the names and the twelve sounds were played, the cards with the printed names were placed on the desk in front of the participant. The participants were asked to identify the sounds in a practice run of the names and sounds.

The participants were reminded that the sounds would be played randomly on the Archimedes computer and that they were required to respond as quickly and as accurately as they could by clicking the mouse pointer on the name of the sound they thought had been played. The participants were reminded that their priority was to answer the auditory warnings.

It was explained that this phase of the experiment followed almost the same format as the previous phases, but this time the participant was required to identify the sounds, to complete the mental arithmetic calculations, and to undertake the observation tasks concurrently.

The participants were asked to imagine that they were a nurse on an I.C.U. and that they had to observe a patient's condition on monitor (i.e. that if an 'E' appeared on the screen of the B.B.C. computer, this could indicate a potentially life-threatening condition) at the same time they also had to calculate the doses of drugs to give the patient, which if they made mistakes could be dangerous, but most importantly their priority, which was clearly stated, was to correctly identify the alarms when they sounded on the computer.

The participants were then asked if they had any questions. The Archimedes program was loaded by the experimenter and the command 'subject number' appeared on the screen. When the participants number had been typed in the prompt, 'press spacebar or click mouse to continue' was shown on the screen. The participant was instructed to carry out the command when they were ready to commence the experiment and also to begin the Archimedes program by clicking the mouse or pressing the space bar when they were ready to commence. The F/E program was loaded by the participant in the same way reported in the
previous phase and the participant commenced the mental arithmetic task when ready.

The program ran for approximately 20 minutes.

Phase 4 took place immediately after Phase 3. The same scenario reported in phase 3 was again suggested to the participants. They were asked to imagine that they were a nurse on an intensive care unit and that they had to observe a patient's condition on the monitor (i.e. that if an 'E' appeared on the screen of the B.B.C. computer, this could indicate a potentially life-threatening condition) at the same time they also had to calculate the doses of drugs to give the patient, which if they made mistakes could be dangerous, but most importantly their priority was again to correctly identify which alarms were sounding on the Archimedes. In this phase however, there was the added task of talking, which is described in the introduction. Although the talking task was not formally measured, similar topics were used for each participant and was maintained at a conversational pace by the experimenter asking questions when a period of silence occurred.

Phase 4 ran for approximately 20 minutes when it was completed the participants were thanked for their participation in the Experiment and paid the £7.00. On completion of the Experiment the participants were debriefed and asked to rate on a scale of 1 to 100 how difficult they perceived each phase of the Experiment.
5.3 Results

5.3.1 Introduction

The results section for Experiment 4 is presented in the following sequence. First, the correct responses to the sounds in Phases 2, 3, and 4 are examined (Phase 1 is presented later). The second set of results to be presented examines the latency of the response to each sound by participants during each phase of the experiment. Performance on the secondary tasks is analysed and is presented in the third section of the results. The subjective ratings of the participants regarding how difficult they perceived each phase of the experiment to be are then presented. The significant confusions between the alarms for each phase are then examined. Comparisons are presented for how participants performed overall in the different phases of the experiment, and finally the learning phases of Experiment 1 and Experiment 4 are compared, as the same sounds were used for each experiment.

5.3.2 Correct Responses

A two-way (phase by correct response) repeated measures Anova was conducted on the participants untransformed correct response data for: Phase 2, (first the mental arithmetic task and identifying the sounds and second the observation task and identifying the sounds). Phase 3, (the mental arithmetic task and the observation task together with identifying the sounds), and finally, Phase 4 (the mental arithmetic task, the observation task, talking and identifying the sounds). The summary Anova table is shown in Table 5.2. As repeated measures were used the figure for Huynh-Feldt (H-F) Epsilon correction was calculated as these give more conservative degrees of freedom (Howell 1987) and the correction is shown in Table 5.3.

It can be observed from the summary Anova Table shown in Table 5.2 that there was a significant interaction between the phase
and the correct responses for the sounds ($F(33,330)=1.7651$ $p=<0.05$). The means are shown in Table 5.4 and the interaction is plotted in Figure 5.11.
<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-Value</th>
<th>P-Value</th>
<th>H.F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>10</td>
<td>81.5189</td>
<td>8.1518</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase</td>
<td>3</td>
<td>10.7026</td>
<td>3.5675</td>
<td>4.0602</td>
<td>0.0156</td>
<td>0.0156</td>
</tr>
<tr>
<td>phase*subject error</td>
<td>30</td>
<td>26.3598</td>
<td>0.8786</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sounds</td>
<td>11</td>
<td>71.8390</td>
<td>6.5308</td>
<td>3.267</td>
<td>0.0007</td>
<td>0.0056</td>
</tr>
<tr>
<td>Sounds*subject error</td>
<td>110</td>
<td>219.890</td>
<td>1.9990</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase*sounds</td>
<td>33</td>
<td>31.6837</td>
<td>0.9601</td>
<td>1.7651</td>
<td>0.0073</td>
<td>0.0196</td>
</tr>
<tr>
<td>Phase<em>sounds</em>subject error</td>
<td>330</td>
<td>179.503</td>
<td>0.5439</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.2: Summary Table for Two-Way Anova (Phase by Correct Responses to Sounds)

<table>
<thead>
<tr>
<th></th>
<th>H-F Epsilon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase</td>
<td>1.2292533</td>
</tr>
<tr>
<td>Sounds</td>
<td>0.6017118</td>
</tr>
<tr>
<td>Phase*sounds</td>
<td>0.6964519</td>
</tr>
</tbody>
</table>

Table 5.3: Adjustment of Degrees of Freedom (H-F)
5.3.2.1 Analysis of the Interaction.

The interaction is analysed in two ways, firstly by examining the differences in the correct responses to each sound between each phase of the experiment (for example by examining sound 1 in each phase to determine whether the number of correct responses to the sound differed in one particular phase), and secondly by examining the differences in the correct responses to the sounds within each individual phase of the experiment, (for example by examining sound 1 in Phase 3 to elicit whether it had either more or fewer correct responses than other sounds in Phase 3).

It can be seen from the interaction shown in Figure 5.11 that in general, for many of the sounds, the mean number of correct responses was very similar during each phase of Experiment 4. However, for sound 8, sound 10, sound 11 and sound 12 the mean number of correct responses were more disparate, i.e. not so closely grouped together. The lowest number of correct response for sound 8, 10 and 11 occurs in the mental arithmetic condition while for sound 12 the lowest number of correct responses occurred in Phase 4.

Within the individual phases of Experiment 4 it would appear that in general sound 6 had the highest number of correct responses, whilst sound 11 consistently has fewer correct responses than the other sounds.

In order to interpret the interaction and to determine whether the differences between the correct responses to the sounds were significantly different, simple interaction effects (Howell 1987, Winer 1971) were calculated. The simple interaction effects elicit whether there is a significant difference between the correct responses to the sounds, which was calculated first for the correct responses to the sounds between each phase of the experiment. However, the calculation for simple effects does not specify in which phase the difference in the
performance for the correct response to the sounds occurs and a post hoc test (Tukey hsd) must then be conducted to determine in which phases the sounds differ significantly.

Simple interaction effects were also calculated for the number of correct responses for each sound within each individual phase of Experiment 4.
Figure 5.11: Interaction of Phase and Correct Responses of Sounds
<table>
<thead>
<tr>
<th>Sound</th>
<th>Mental Arithmetic</th>
<th>Standard deviation</th>
<th>Observation</th>
<th>Standard deviation</th>
<th>Phase 3</th>
<th>Standard deviation</th>
<th>Phase 4</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>sound 1</td>
<td>3.18</td>
<td>1.17</td>
<td>3.36</td>
<td>1.03</td>
<td>3.54</td>
<td>.93</td>
<td>3.45</td>
<td>1.04</td>
</tr>
<tr>
<td>sound 2</td>
<td>3.45</td>
<td>.69</td>
<td>3.18</td>
<td>.98</td>
<td>3.64</td>
<td>.67</td>
<td>3.09</td>
<td>1.30</td>
</tr>
<tr>
<td>sound 3</td>
<td>3.45</td>
<td>.52</td>
<td>3.64</td>
<td>.67</td>
<td>3.54</td>
<td>.93</td>
<td>3.36</td>
<td>1.03</td>
</tr>
<tr>
<td>sound 4</td>
<td>3.45</td>
<td>.93</td>
<td>3.64</td>
<td>.67</td>
<td>3.64</td>
<td>.67</td>
<td>3.27</td>
<td>1.27</td>
</tr>
<tr>
<td>sound 5</td>
<td>3.54</td>
<td>.69</td>
<td>3.64</td>
<td>1.21</td>
<td>3.64</td>
<td>1.03</td>
<td>3.27</td>
<td>1.01</td>
</tr>
<tr>
<td>sound 6</td>
<td>3.72</td>
<td>.47</td>
<td>3.82</td>
<td>.60</td>
<td>3.73</td>
<td>.47</td>
<td>3.73</td>
<td>.65</td>
</tr>
<tr>
<td>sound 7</td>
<td>3.45</td>
<td>.82</td>
<td>3.64</td>
<td>.50</td>
<td>3.36</td>
<td>.81</td>
<td>3.73</td>
<td>.47</td>
</tr>
<tr>
<td>sound 8</td>
<td>2.64</td>
<td>1.43</td>
<td>2.73</td>
<td>1.49</td>
<td>3.18</td>
<td>1.40</td>
<td>3.54</td>
<td>.69</td>
</tr>
<tr>
<td>sound 9</td>
<td>3.27</td>
<td>1.27</td>
<td>3.36</td>
<td>1.29</td>
<td>2.54</td>
<td>.93</td>
<td>3.36</td>
<td>1.29</td>
</tr>
<tr>
<td>sound 10</td>
<td>2.54</td>
<td>1.57</td>
<td>2.73</td>
<td>1.42</td>
<td>3.64</td>
<td>.50</td>
<td>3.73</td>
<td>.47</td>
</tr>
<tr>
<td>sound 11</td>
<td>1.45</td>
<td>1.51</td>
<td>2.18</td>
<td>1.60</td>
<td>2.54</td>
<td>1.21</td>
<td>2.72</td>
<td>1.35</td>
</tr>
<tr>
<td>sound 12</td>
<td>3.00</td>
<td>1.26</td>
<td>2.73</td>
<td>1.35</td>
<td>3.73</td>
<td>.47</td>
<td>3.54</td>
<td>.69</td>
</tr>
</tbody>
</table>

Table 5.4: Mean Number of Correct Responses for Sounds for Each Phase
Table 5.5 shows the results of the simple effects calculation for each sound, between the phases. It can be seen from Table 5.5 that sounds 8, 10, 11 and 12 were shown to be significantly different across the phases. Therefore, post hoc tests (Tukey hsd) were conducted on sounds 8, 10, 11 and 12 in order to elicit which phase of the experiment had an effect on performance.

<table>
<thead>
<tr>
<th>sound</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>not significant</td>
</tr>
<tr>
<td>2</td>
<td>not significant</td>
</tr>
<tr>
<td>3</td>
<td>not significant</td>
</tr>
<tr>
<td>4</td>
<td>not significant</td>
</tr>
<tr>
<td>5</td>
<td>not significant</td>
</tr>
<tr>
<td>6</td>
<td>not significant</td>
</tr>
<tr>
<td>7</td>
<td>not significant</td>
</tr>
<tr>
<td>8</td>
<td>Sig p=&lt;0.01 (F=3)</td>
</tr>
<tr>
<td>9</td>
<td>not significant</td>
</tr>
<tr>
<td>10</td>
<td>Sig p=&lt;0.01 (F=4.6)</td>
</tr>
<tr>
<td>11</td>
<td>Sig p=&lt;0.01 (F=5.4)</td>
</tr>
<tr>
<td>12</td>
<td>Sig p=&lt;0.05 (F=2.33)</td>
</tr>
</tbody>
</table>

Table 5.5: Simple Main Effects of Phase

[F(11,44)=2.08  p=<0.05]  
[F(11,44)=2.80  p=<0.01]  

The follow up analysis (Tukey hsd) showed that the mean number of correct responses to sound 8 in Phase 2 (mental arithmetic) as significantly lower than the mean number of correct responses both in Phase 3 and Phase 4. No other significant differences between the phases were found (Tukey ps=<0.05).

The follow-up analysis for sound 10 showed that there were significantly fewer correct responses for sound 10 in Phase 2 (mental arithmetic) than in Phase 3 and Phase 4 (ps=<0.01).

The follow-up analysis performed on sound 11 showed that there were significantly fewer correct responses to sound 11 during Phase 2 (mental arithmetic) than in Phase 3 and in Phase 4 (ps=<0.01). No other significant differences were found.
The follow-up analysis showed that sound 12 produced fewer correct responses when the Phase 2 (observation task) was being undertaken than in Phase 3 and Phase 4 (ps=<0.05).

The significant differences for the correct responses to the sounds is summarised in Table 5.6.
<table>
<thead>
<tr>
<th>Sound</th>
<th>Mental arithmetic</th>
<th>Observation Task</th>
<th>Phase 3</th>
<th>Phase 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>not significant</td>
<td>not significant</td>
<td>not significant</td>
<td>not significant</td>
</tr>
<tr>
<td>2</td>
<td>not significant</td>
<td>not significant</td>
<td>not significant</td>
<td>not significant</td>
</tr>
<tr>
<td>3</td>
<td>not significant</td>
<td>not significant</td>
<td>not significant</td>
<td>not significant</td>
</tr>
<tr>
<td>4</td>
<td>not significant</td>
<td>not significant</td>
<td>not significant</td>
<td>not significant</td>
</tr>
<tr>
<td>5</td>
<td>not significant</td>
<td>not significant</td>
<td>not significant</td>
<td>not significant</td>
</tr>
<tr>
<td>6</td>
<td>not significant</td>
<td>not significant</td>
<td>not significant</td>
<td>not significant</td>
</tr>
<tr>
<td>7</td>
<td>not significant</td>
<td>not significant</td>
<td>not significant</td>
<td>not significant</td>
</tr>
<tr>
<td>8</td>
<td>significantly fewer correct responses than Phases 3 and 4 (p=&lt;0.05)</td>
<td>no significant difference</td>
<td>significantly more correct responses than Phase 2 (mental arithmetic) (p=&lt;0.05)</td>
<td>significantly more correct responses than Phase 2 (mental arithmetic) (p=&lt;0.05)</td>
</tr>
<tr>
<td>9</td>
<td>not significant</td>
<td>not significant</td>
<td>not significant</td>
<td>not significant</td>
</tr>
<tr>
<td>10</td>
<td>significantly fewer correct responses than Phases 3 and 4 (p=&lt;0.01)</td>
<td>significantly fewer correct responses than Phase 4 (p=&lt;0.01) and significantly fewer correct responses for Phase 3 (p=&lt;0.05)</td>
<td>significantly more correct responses than Phase 2 (mental arithmetic) (p=&lt;0.01)</td>
<td>significantly more correct responses than Phase 2 (mental arithmetic) (p=&lt;0.01)</td>
</tr>
<tr>
<td>11</td>
<td>significantly fewer correct responses than Phases 3 and 4 (p=&lt;0.01)</td>
<td>no significant difference</td>
<td>significantly more correct responses than Phase 2 (mental arithmetic) (p=&lt;0.01)</td>
<td>significantly more correct responses than Phase 2 (mental arithmetic) (p=&lt;0.01)</td>
</tr>
<tr>
<td>12</td>
<td>no significant difference</td>
<td>significantly fewer correct responses than Phase 3 (p=&lt;0.01) and Phase 4 (p=&lt;0.05)</td>
<td>significantly more correct responses than Phase 2 (observation) (p=&lt;0.01)</td>
<td>significantly more correct responses than Phase 2 (observation) (p=&lt;0.05)</td>
</tr>
</tbody>
</table>

Table 5.6: Significant Differences for Correct Responses to Sounds Across Each Phase of Experiment 4
The second analysis performed on the interaction examined the number of correct responses to the sounds within each individual phase of Experiment 4. The simple effects are shown in Table 5.7. The results showed that there were significant differences between the number of correct responses for the sounds in Phase 2 (mental arithmetic task and the observation task) and Phase 3 (mental arithmetic task, observation task and identification of the sounds). There were however, no significant differences between the number of correct responses for each sound within phase 4 of the Experiment. To elicit which sounds within Phase 2 and 3 produced significantly different correct responses, a post hoc test (Tukey hsd) was conducted on the data. The significant differences for the correct responses to the sounds within each Phase are summarised in Table 5.8.

<table>
<thead>
<tr>
<th>Mental Arithmetic (Phase 2)</th>
<th>Sig at p=&lt;0.01 (F=8.03)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observation (Phase 2)</td>
<td>Sig at p=&lt;0.01 (F=4.62)</td>
</tr>
<tr>
<td>Phase 3</td>
<td>Sig at p=&lt;0.05 (F=2.18)</td>
</tr>
<tr>
<td>Phase 4</td>
<td>N/S (F=1.74)</td>
</tr>
</tbody>
</table>

Table 5.7: Simple Effects of Sounds for Each Phase

[F(11,132)=2.47 (p<0.01)]
[F(11,132)=1.91 (p<0.05)]
<table>
<thead>
<tr>
<th>Phase</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 2 (Mental Arithmetic)</td>
<td>N/S</td>
<td>N/S</td>
<td>N/S</td>
<td>N/S</td>
<td>N/S</td>
<td>significantly more correct responses than sound 10 and 11 (ps=&lt;0.05)</td>
<td>N/S</td>
<td>N/S</td>
<td>N/S</td>
<td>N/S</td>
<td>N/S</td>
<td>N/S</td>
</tr>
<tr>
<td>Phase 2 (Observation)</td>
<td>N/S</td>
<td>N/S</td>
<td>N/S</td>
<td>N/S</td>
<td>N/S</td>
<td>significantly more correct responses than sounds 8, 10, and 12 (ps=&lt;0.05)</td>
<td>N/S</td>
<td>N/S</td>
<td>N/S</td>
<td>N/S</td>
<td>N/S</td>
<td>N/S</td>
</tr>
<tr>
<td>Phase 3</td>
<td>N/S</td>
<td>N/S</td>
<td>N/S</td>
<td>N/S</td>
<td>N/S</td>
<td>N/S</td>
<td>N/S</td>
<td>N/S</td>
<td>N/S</td>
<td>N/S</td>
<td>N/S</td>
<td>N/S</td>
</tr>
<tr>
<td>Phase 4</td>
<td>No significant difference between the correct responses for any of the sounds</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.8: Significant Differences for Correct Responses to Sounds Within Each Phase of Experiment 4
The follow up analysis showed that in Phase 2 (mental arithmetic task) sound 11 produced significantly fewer correct responses than every other sound (ps=<0.05). Sound 6 was shown to have a significantly higher correct response rate than sound 10, sound 11 and sound 8 (ps=<0.05).

Post hoc tests (Tukey hsd) on each of the correct responses to sounds in Phase 2 (observation task) showed that sound 11 produced fewer correct responses than sounds 1, 3, 4, 5, 6, 7, and sound 9 (ps=<0.01). Sound 6 produced significantly more correct responses than sounds 8, 10, and 12 (ps=<0.05).

In Phase 3 (mental arithmetic task, observation task and identification of the sounds) the follow up analysis for the correct responses to the sounds showed that sound 11 had significantly fewer correct responses than sounds 6 and 12 (ps=<0.01) and sounds 2, 4, and 10 (ps=<0.05).

No significant differences were found for Phase 4 (ps=>0.05).

5.3.2.2 Main Effects

There was a significant main effect for Phase (F(3,30)=4.0602, p=<0.05). The means of the correct responses are shown in Table 5.9. There was also a significant main effect for Sounds (F(11,110)=3.267 p=<0.01). The means are shown in Table 5.10.

<table>
<thead>
<tr>
<th>Phase</th>
<th>Mean Number of Correct Responses for Sounds</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mental arithmetic</td>
<td>3.098</td>
<td>1.210</td>
</tr>
<tr>
<td>Observation</td>
<td>3.220</td>
<td>1.187</td>
</tr>
<tr>
<td>Phase 3</td>
<td>3.455</td>
<td>.903</td>
</tr>
<tr>
<td>Phase 4</td>
<td>3.402</td>
<td>.987</td>
</tr>
</tbody>
</table>

Table 5.9: Mean Number of Correct Responses for Phase

A post hoc test, (Tukey hsd) was performed on the correct responses for each phase and it was shown that Phase 2 (mental arithmetic) produced significantly fewer correct responses.
(Tukey p=<0.05) than Phase 3 (mental arithmetic task, observation task and identifying the warnings). No other significant differences between the phases were found.

<table>
<thead>
<tr>
<th>Sounds</th>
<th>Mean Number of Correct Responses for Sounds</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>sound 1</td>
<td>3.386</td>
<td>1.02</td>
</tr>
<tr>
<td>sound 2</td>
<td>3.341</td>
<td>.94</td>
</tr>
<tr>
<td>sound 3</td>
<td>3.500</td>
<td>.79</td>
</tr>
<tr>
<td>sound 4</td>
<td>3.500</td>
<td>.90</td>
</tr>
<tr>
<td>sound 5</td>
<td>3.455</td>
<td>.97</td>
</tr>
<tr>
<td>sound 6</td>
<td>3.750</td>
<td>.53</td>
</tr>
<tr>
<td>sound 7</td>
<td>3.545</td>
<td>.66</td>
</tr>
<tr>
<td>sound 8</td>
<td>3.023</td>
<td>1.30</td>
</tr>
<tr>
<td>sound 9</td>
<td>3.386</td>
<td>1.17</td>
</tr>
<tr>
<td>sound 10</td>
<td>3.159</td>
<td>1.20</td>
</tr>
<tr>
<td>sound 11</td>
<td>2.227</td>
<td>1.46</td>
</tr>
<tr>
<td>sound 12</td>
<td>3.250</td>
<td>1.06</td>
</tr>
</tbody>
</table>

Table 5.10: Overall Mean Number of Correct Responses for Sounds (all phases)

A post hoc test (Tukey hsd) was performed on the sounds and showed that overall across the phases of Experiment 4 sound 11 produced significantly fewer correct responses than sounds 3, 4, 5, 6 and 7 (Tukey ps=<0.01). Sound 11 also produced fewer correct responses than sounds 1, 2, 9, and 12 (Tukey ps=<0.05). Although no other significant differences between the sounds were found, it can be seen in Figure 5.11 that overall sound 6 appeared to produce the highest number of correct responses while sound 11 seemed to produce the least number of correct responses.
5.3.3 Latency

5.3.3.1 Introduction

As discussed in the Introduction to this chapter, the program written for the Archimedes 4000 computer recorded the amount of time (latency) the participants took to identify the sounds. The data was recorded in milliseconds. If the participants failed to respond there was a 'timed-out' response of 90 seconds and consequently there were large variations in the data. The data was therefore transformed using a logarithmic transformation. The logarithmic transformation is used when the standard deviation is proportional to the mean and also when the data is markedly positively skewed (Howell 1987). The two-way Anova (phase by sound) is summarised in Table 5.11. As repeated measures were used the figure for Huynh-Feldt (H-F) Epsilon correction was calculated to give more conservative degrees of freedom (Howell 1987) and the correction is shown in Table 5.12.
### Table 5.11: Summary Table for Two-Way Anova (Phase by Sound)

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-Value</th>
<th>P-Value</th>
<th>H-F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>10</td>
<td>5.2027</td>
<td>0.5202</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase</td>
<td>3</td>
<td>0.6810</td>
<td>0.2270</td>
<td>4.3644</td>
<td>0.0115</td>
<td>0.0261</td>
</tr>
<tr>
<td>Phase*subject error</td>
<td>30</td>
<td>1.5604</td>
<td>0.0520</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sounds</td>
<td>11</td>
<td>2.4089</td>
<td>0.2189</td>
<td>1.7332</td>
<td>0.0752</td>
<td>0.1602</td>
</tr>
<tr>
<td>Sounds*subject error</td>
<td>110</td>
<td>13.8984</td>
<td>0.1263</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase*sounds</td>
<td>33</td>
<td>1.2600</td>
<td>0.0381</td>
<td>1.0191</td>
<td>0.4428</td>
<td>0.4359</td>
</tr>
<tr>
<td>Phase<em>sounds</em>subject error</td>
<td>330</td>
<td>12.3644</td>
<td>0.0374</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 5.12: Adjustment of H-F Degrees of Freedom

<table>
<thead>
<tr>
<th>Source</th>
<th>H-F Epsilon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment</td>
<td>0.6756611</td>
</tr>
<tr>
<td>Sounds</td>
<td>0.3712027</td>
</tr>
<tr>
<td>Experiment*sounds</td>
<td>0.3641634</td>
</tr>
</tbody>
</table>
The summary Anova Table shown in Table 5.1 demonstrates a significant main effect for Phase \[ F(3,30) = 4.3644 \text{ p}<0.05 \] (H-F correction \( p<0.05 \)). The means for the phases are shown in Table 5.13. Although there was no significant effect for sounds, \[ F(11,110) = 1.7332 \text{ p}>0.05 \] (H-F correction = \( p>0.05 \)) overall sound 6 appeared to produce the shortest latency (mean 3.21) and sound 11 seemed to have the longest latency (3.46). There was no significant interaction \[ F(33,330) = 1.0191 \text{ p}>0.05 \] (H-F correction \( p>0.05 \)).

<table>
<thead>
<tr>
<th>Phase</th>
<th>Mean of Response Latency (m-seconds)</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 2 (Mental Arithmetic)</td>
<td>3.344</td>
<td>.267</td>
</tr>
<tr>
<td>Phase 2 (Observation)</td>
<td>3.273</td>
<td>.263</td>
</tr>
<tr>
<td>Phase 3</td>
<td>3.249</td>
<td>.224</td>
</tr>
<tr>
<td>Phase 4</td>
<td>3.309</td>
<td>.299</td>
</tr>
</tbody>
</table>

Table 5.13: Mean Number of Response Latency for Each Phase of the Experiment

A post hoc test (Tukey hsd) was performed on the main effect of Phase to elicit in which phase there was a significant different in the response time to the sounds. The results showed that in Phase 2 (mental arithmetic task mean 3.344) produced significantly longer latencies than Phase 3 (mean 3.249) \( p<0.05 \). There were no other significant differences between the means of the phases \( ps>0.05 \). The means are shown in Table 5.13.
Figure 5.12: Latency of Response to the Sounds in Each Phase
Although there was no significant interaction between the phases and the latency of response to the sounds, the interaction is shown in Figure 5.12 and it can be observed that there was a trend towards sound 6 having the shortest latency in three of the Phases; Phase 2, (observation) (mean 3.139), Phase 3 (mean 3.164) and Phase 4 (mean 3.199), while sound 11 had a longer latency in the Phase 2 (mental arithmetic) (mean 3.556), Phase 3 (3.423) and Phase 4 (3.522).

5.3.4 Correct Responses for Secondary Tasks.

Overall, the pattern of results for the correct responses to the combination of the tasks follows a pattern similar to that which would be predicted by the literature regarding dual-task studies. Highest scores for the tasks were achieved when the tasks were undertaken individually, together with identifying the sounds. The overall results for the correct responses are illustrated in Figure 5.13 for the mental arithmetic task in each phase and in Figure 5.14 for the results of the observation task in each phase.

When the observation task and the mental arithmetic task were undertaken concurrently in Phase 3, together with the sound identification task, the number of correct responses decreased in comparison to when each task was undertaken separately (together with the identification of the sounds) as predicted in the dual-task literature. In the final phase (Phase 4) the number of correct responses for the observation task increased, while the number of correct responses for the mental arithmetic task decreased further. However, it is necessary to compare the secondary task performance with the primary task performance and therefore the statistical comparisons will be performed on the data for the correct responses to sounds, the correct responses for the mental arithmetic task and the correct responses for the observation task.
Figure 5.13: Observation Scores (Late, Early and Correct Responses for Each Phase)
Figure 5.14: Correct Scores for Mental Arithmetic Task for Each Phase
5.3.5 Comparison of Primary and Secondary Task Performance

The graphs illustrated in Figures 5.15 to 5.25 depict individual subject performance on the primary task of identifying the sounds and the secondary tasks of mental arithmetic, observation and talking. The figures used have been corrected for guessing (Dennis, Personal Communication). The graphs are included to demonstrate the different strategies and responses by the participants and to show that the overall pattern of correct responses is by no means consistent for each subject during each phase of the experiment.

The graphs also illustrate the predicted decrement in the performance on the secondary tasks from the dual-task and workload literature. The graphs show the correct responses for the sounds and for the observation task in percentages on the left axis and the raw number of correct responses for the mental arithmetic task on the right.
Figure 5.15: Correct Responses Subject 1 (All Phases, All Tasks)
Figure 5.16: Correct Responses Subject 2 (All Phases, All Tasks)
Figure 5.17: Correct Responses Subject 3 (All Phases, All Tasks)
Figure 5.18: Correct Responses Subject 4 (All Phases, All Tasks)
Figure 5.19: Correct Responses Subject 5 (All Phases, All Tasks)
Correct Response for Sounds

Observation

Mental Arithmetic

Figure 5.20: Correct Responses Subject 6 (All Phases, All Tasks)
Figure 5.21: Correct Responses Subject 7 (All Phases, All Tasks)
Figure 5.22: Correct Responses Subject 8 (All Phases, All Tasks)
Figure 5.23: Correct Responses Subject 9 (All Phases, All Tasks)
Figure 5.24: Correct Responses Subject 10 (All Phases, All Tasks)
Figure 5.25: Correct Responses Subject 11 (All Phases, All Tasks)
In order to examine whether the trends illustrated in Figures 5.15 to 5.25 were statistically significant a Friedman test was performed separately on the data for each task. The Friedman test was used to determine whether there was a significant decrement in the performance with respect to the participants' number of correct responses overall, as the phases increased in difficulty, or, for example, whether one particular task showed a greater decrement than the others.

Each subject's score for each task, for each phase of the experiment, was ranked. Because there were ties among the ranks the Friedman statistic to correct for ties was used (Siegel and Castellan 1988).

First, the ranks of the correct responses for the sounds in each phase of the experiment are shown in Table 5.14. The highest rank, (5), being allocated to the highest score and the lowest rank (1) allocated to the lowest score. The result of the Friedman calculation (corrected for ties) was significant ($F_r=9.6$, $df=4$, $p= <0.05$) The result suggests that the conditions were significantly different from each other. However, follow up analysis showed that any difference between the means was only approaching significance. The lack of a significant difference at follow-up was possibly due to the large number of tied ranks in the data. The $F$ statistic is increased by using the formula that corrects for tied ranks. Therefore a greater number of ties results in a greater increase in the $F$-statistic. Therefore the inflated statistic was significant but the differences between the individual sums of ranks were not.
<table>
<thead>
<tr>
<th>Subject</th>
<th>Learning</th>
<th>Phase 2 (mental arithmetic)</th>
<th>Phase 2 (observation)</th>
<th>Phase 3</th>
<th>Phase 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>2.5</td>
<td>4.5</td>
<td>2.5</td>
<td>4.5</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>2.5</td>
<td>1</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>3.5</td>
<td>5</td>
<td>3.5</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>4.5</td>
<td>1</td>
<td>4.5</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>2.5</td>
<td>3.5</td>
<td>5</td>
<td>3.5</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>3.5</td>
<td>1</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>3.5</td>
<td>2.5</td>
<td>4</td>
<td>3.5</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>4.5</td>
<td>3</td>
<td>4.5</td>
<td>4.5</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>4.5</td>
<td>1</td>
<td>4.5</td>
<td>4.5</td>
</tr>
<tr>
<td>Sum of Ranks</td>
<td>31.5</td>
<td>25</td>
<td>29</td>
<td>41</td>
<td>39.5</td>
</tr>
</tbody>
</table>

Table 5.14: Ranks of Correct Responses to Sounds
The ranks of the number of correct responses for the mental arithmetic calculations are shown in Table 5.15 (3 being designated to the highest number of correct responses and 1 being designated to the lowest number of correct responses).

The Friedman calculation (corrected for ties) showed that there was a significant difference between the conditions ($F_r=18$ df=2 $p=<0.01$) Although the Friedman statistic indicates that there is a difference between conditions is does not indicate where the difference is. Therefore the critical value is calculated to determine which conditions were significantly different (Siegel and Castellan 1988).

Critical $z$ value for comparisons between the conditions was calculated as 11.2 and the difference between the means must be greater than or equal ($\geq$) to this figure.
The results showed that the number of correct responses for the mental arithmetic calculations in Phase 2, (in which the participants calculated the answers to the mental arithmetic task together with the identification of the sounds) was not significantly different to Phase 3 (mental arithmetic and observation task together and the identification of the sounds). However, there were significantly higher correct answers (p=<0.05) in Phase 2 (mental arithmetic calculations and identification of the sounds) to the number of correct answers in Phase 4 (mental arithmetic, observation task, talking and identifying the sounds). The number of correct responses to the mental arithmetic calculations were also significantly higher in Phase 3 in comparison to Phase 4. The results are summarised in Table 5.16.

<table>
<thead>
<tr>
<th>Comparison between Conditions</th>
<th>≥ critical value of 11.2</th>
<th>Significant (p=&lt;0.05)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mental Arithmetic compared with Phase 3</td>
<td>7.5</td>
<td>Not significant</td>
</tr>
<tr>
<td>Mental Arithmetic compared with Phase 4</td>
<td>19.5</td>
<td>Significantly different</td>
</tr>
<tr>
<td>Phase 3 compared with Phase 4</td>
<td>12</td>
<td>Significantly different</td>
</tr>
</tbody>
</table>

Table 5.16: Summary Table for Comparisons of Correct Responses for Mental Arithmetic Task

The ranks of the correct responses for the observation task are shown in Table 5.17 (3 being designated to the highest number of correct responses and 1 being designated to the lowest number of correct responses).
Table 5.17: The Ranks for the Number of Correct Responses (Observation Task)

The Friedman calculation showed that there was a significant difference between the conditions ($F_r = 16.5 \text{ df} = 2 \ p < 0.01$). The critical value was then calculated to determine which conditions were significantly different (Siegel and Castellan 1988).

Critical $z$ value for comparisons between the conditions was calculated to be 11.2. The difference between the means must be greater than or equal ($\geq$) to this figure.

The results showed that the number of correct responses for the observation task in Phase 2, (observation task together with the identification of the sounds), was significantly different to both Phase 3 (mental arithmetic and observation task together and the identification of the sounds) and Phase 4 (mental arithmetic, observation task, talking and identifying the sounds). There was no significant difference between the correct responses in Phase 3 and Phase 4. The results are summarised in Table 5.18.
<table>
<thead>
<tr>
<th>Comparison of conditions</th>
<th>$\geq$ critical value of 11.2</th>
<th>Significant (p&lt;0.05)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observation compared with Phase 3</td>
<td>$33-16=17$</td>
<td>significantly different</td>
</tr>
<tr>
<td>Observation compared with Phase 4</td>
<td>$33-17=16$</td>
<td>significantly different</td>
</tr>
<tr>
<td>Phase 3 compared with Phase 4</td>
<td>$16-17=-1$</td>
<td>Not significantly different</td>
</tr>
</tbody>
</table>

Table 5.18: Summary Table for Comparisons of Correct Responses for Observation Task
5.3.6 Subjective Ratings

When each subject had completed the experiment they were asked to rate on a scale of zero to 100 how difficult they perceived each phase to be. The participants ratings are shown in Table 5.19.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Phase 2 (mental arithmetic)</th>
<th>Phase 2 (observation)</th>
<th>Phase 3</th>
<th>Phase 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50</td>
<td>30</td>
<td>90</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>50</td>
<td>100</td>
<td>100</td>
<td>90</td>
</tr>
<tr>
<td>3</td>
<td>50</td>
<td>40</td>
<td>79</td>
<td>99</td>
</tr>
<tr>
<td>4</td>
<td>60</td>
<td>60</td>
<td>80</td>
<td>92</td>
</tr>
<tr>
<td>5</td>
<td>40</td>
<td>20</td>
<td>60</td>
<td>80</td>
</tr>
<tr>
<td>6</td>
<td>50</td>
<td>15</td>
<td>75</td>
<td>95</td>
</tr>
<tr>
<td>7</td>
<td>60</td>
<td>30</td>
<td>60</td>
<td>90</td>
</tr>
<tr>
<td>8</td>
<td>45</td>
<td>25</td>
<td>95</td>
<td>95</td>
</tr>
<tr>
<td>9</td>
<td>25</td>
<td>10</td>
<td>70</td>
<td>99</td>
</tr>
<tr>
<td>10</td>
<td>30</td>
<td>30</td>
<td>60</td>
<td>90</td>
</tr>
<tr>
<td>11</td>
<td>15</td>
<td>15</td>
<td>55</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 5.19: Participants Rating for Perceived Difficulty of Each Phase

The a priori prediction was that each phase of the experiment would increase in difficulty and therefore a Page test for ordered alternatives was performed on the data (Siegel and Castellan 1988). The Page trend test examines the hypothesis that the groups will be ordered in a specific sequence. Therefore, the results should show that the participants found phase 4 the most difficult and Phase 2, the mental arithmetic task and the observation task performed individually, the least difficult. The ratings were ranked and the ranks are shown in Table 5.20.
The results of the Page Trend Test showed that there was a significant difference ($L=321.5$ $p=<0.001$) between each phase of Experiment 4. The result show that the participants perceived each phase as increasing in difficulty, as predicted a priori.

5.3.7 Significant Confusions Between Sounds for Each Phase.

The significant confusions ($p=<0.01$) were calculated using the same method of analysis described in the first set of Experiments (see Chapter 2). Multinomial distribution analysis was calculated on the errors, taking into account any response bias that may have occurred. The significant confusions are shown in Table 5.21.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Phase 2 (mental arithmetic)</th>
<th>Phase 2 (observation)</th>
<th>Phase 3</th>
<th>Phase 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3.5</td>
<td>3.5</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>1.5</td>
<td>1.5</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>2.5</td>
<td>1</td>
<td>2.5</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>1</td>
<td>3.5</td>
<td>3.5</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>1.5</td>
<td>1.5</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>11</td>
<td>1.5</td>
<td>1.5</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>15</td>
<td>33.5</td>
<td>41.5</td>
</tr>
</tbody>
</table>

Table 5.20: Ranks of Subjective Ratings
Table 5.21: Significant Confusions Between Warnings (p=<0.01)

<table>
<thead>
<tr>
<th>sound played:</th>
<th>identified as in Phase 2 (mental arithmetic)</th>
<th>identified as in Phase 2 (observation)</th>
<th>identified as in Phase 3</th>
<th>identified as in Phase 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1,1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>8,11</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>8</td>
<td>1,8</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In the four Phases of Experiment 4 involving the tasks, or combinations of tasks when identifying the set of twelve sounds, there are again certain sounds that are confused (for example, sound 1 was called sound 7; and sound 8 was called sound 11). In the set of Experiments reported in Chapter 2, the confusion between sound 1 and sound 7 was not an enduring confusion however, during Phases 2, 3 and 4 of the current study the two sounds are significantly confused. There were no significant non-confusions between the sounds in any of the phases of Experiment 4.
<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>*</td>
<td>0.0604</td>
<td>0.9456</td>
<td>0.9630</td>
<td>0.9986</td>
<td>0.3238</td>
<td>0.0062</td>
<td>0.8786</td>
<td>0.9456</td>
<td>0.8262</td>
<td>0.2542</td>
</tr>
<tr>
<td>2</td>
<td>0.2100</td>
<td>*</td>
<td>0.8584</td>
<td>0.8108</td>
<td>0.8428</td>
<td>0.8892</td>
<td>0.2028</td>
<td>0.2108</td>
<td>0.8584</td>
<td>0.0982</td>
<td>0.0152</td>
</tr>
<tr>
<td>3</td>
<td>0.9150</td>
<td>0.2008</td>
<td>*</td>
<td>0.8068</td>
<td>0.8386</td>
<td>0.8850</td>
<td>0.9000</td>
<td>0.2086</td>
<td>0.8542</td>
<td>0.3272</td>
<td>0.7906</td>
</tr>
<tr>
<td>4</td>
<td>0.9108</td>
<td>0.0014</td>
<td>0.8500</td>
<td>*</td>
<td>0.8344</td>
<td>0.8808</td>
<td>0.8958</td>
<td>0.9108</td>
<td>0.8500</td>
<td>0.6510</td>
<td>0.0146</td>
</tr>
<tr>
<td>5</td>
<td>0.1476</td>
<td>0.7836</td>
<td>0.7414</td>
<td>0.6982</td>
<td>*</td>
<td>0.0112</td>
<td>0.7836</td>
<td>0.7972</td>
<td>0.7414</td>
<td>0.0672</td>
<td>0.6836</td>
</tr>
<tr>
<td>6</td>
<td>0.3702</td>
<td>0.3628</td>
<td>0.3402</td>
<td>0.3176</td>
<td>0.0144</td>
<td>*</td>
<td>0.0174</td>
<td>0.3702</td>
<td>0.3402</td>
<td>0.2484</td>
<td>0.3100</td>
</tr>
<tr>
<td>7</td>
<td>0.2108</td>
<td>0.9044</td>
<td>0.8584</td>
<td>0.8108</td>
<td>0.8426</td>
<td>0.8926</td>
<td>*</td>
<td>0.9194</td>
<td>0.0202</td>
<td>0.6578</td>
<td>0.0140</td>
</tr>
<tr>
<td>8</td>
<td>0.2740</td>
<td>0.4896</td>
<td>0.5390</td>
<td>0.5932</td>
<td>0.5566</td>
<td>0.5056</td>
<td>0.2604</td>
<td>*</td>
<td>0.2220</td>
<td>0.7870</td>
<td>0.0001</td>
</tr>
<tr>
<td>9</td>
<td>0.0006</td>
<td>0.7848</td>
<td>0.7428</td>
<td>0.6994</td>
<td>0.7284</td>
<td>0.7710</td>
<td>0.7848</td>
<td>0.1256</td>
<td>*</td>
<td>0.5626</td>
<td>0.6848</td>
</tr>
<tr>
<td>10</td>
<td>0.8212</td>
<td>0.4660</td>
<td>0.5146</td>
<td>0.5682</td>
<td>0.7088</td>
<td>0.4816</td>
<td>0.4660</td>
<td>0.0001</td>
<td>0.5146</td>
<td>*</td>
<td>0.0050</td>
</tr>
<tr>
<td>11</td>
<td>0.0780</td>
<td>0.7724</td>
<td>0.2640</td>
<td>0.9340</td>
<td>0.8678</td>
<td>0.2392</td>
<td>0.2276</td>
<td>0.0028</td>
<td>0.2004</td>
<td>0.4734</td>
<td>*</td>
</tr>
<tr>
<td>12</td>
<td>0.7184</td>
<td>0.7352</td>
<td>0.0010</td>
<td>0.3812</td>
<td>0.8052</td>
<td>0.7522</td>
<td>0.7352</td>
<td>0.7184</td>
<td>0.7872</td>
<td>0.2492</td>
<td>0.0668</td>
</tr>
</tbody>
</table>

Table 5.22: Experiment 4 Phase 2 (Mental Arithmetic) Significant Confusions (p<=0.01) (shown in bold borders)
Confusions Approaching Significance (p<=0.05) (shown in double borders)
<table>
<thead>
<tr>
<th>Sound Played</th>
<th>Sound Identified As:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>*</td>
</tr>
<tr>
<td>2</td>
<td>0.0090</td>
</tr>
<tr>
<td>3</td>
<td>0.1563</td>
</tr>
<tr>
<td>4</td>
<td>0.5620</td>
</tr>
<tr>
<td>5</td>
<td>0.8372</td>
</tr>
<tr>
<td>6</td>
<td>0.0208</td>
</tr>
<tr>
<td>7</td>
<td>0.0586</td>
</tr>
<tr>
<td>8</td>
<td>0.5216</td>
</tr>
<tr>
<td>9</td>
<td>0.0068</td>
</tr>
<tr>
<td>10</td>
<td>0.4960</td>
</tr>
<tr>
<td>11</td>
<td>0.0022</td>
</tr>
<tr>
<td>12</td>
<td>0.6148</td>
</tr>
</tbody>
</table>

Table 5.23: Experiment 4 Phase 2 (Observation) Significant Confusions (p=<0.01) (shown in bold borders) Confusions Approaching Significance (p=<0.05) (shown in double borders)
Table 5.24: Experiment 4 Phase 3 Significant Confusions (p=<0.01) (shown in bold borders) Confusions Approaching Significance (p=<0.05) (shown in double borders)
### Table 5.25: Experiment 4 Phase 4 Significant Confusions (p=<0.01) (shown in bold borders) Confusions Approaching Significance (p=<0.05) (shown in double borders)
5.3.8 Comparison of Overall Performance, All Phases.

The overall performance for all tasks in each of the phases in Experiment 4 is summarised in Table 5.26 together with the latency and the subjective ratings of the participants for the perceived difficulty of each phase. The significant confusions (p=<0.01) for each phase are also shown.

Ranks are shown in parenthesis 3 (or 4) for highest number of correct responses and 1 for the lowest number of correct responses. The ranks for the latency are ordered as 1 for the shortest latency and 4 for the longest latency. The ranks for the subjective ratings are shown as 4 for the most difficult and 1 for the least difficult phase.

The highest number of correct responses for the identification of the sounds occurred in the following order;

i) Phase 3 (mean 3.455),
ii) Phase 4 (mean 3.402)
iii) Phase 2 (observation task mean 3.220)
iv) Phase 2 (mental arithmetic task mean 3.098).

The mental arithmetic task produced the highest number of correct response when the task was undertaken concurrently with the identification of the sounds. The least number of correct responses for the mental arithmetic task occurred in Phase 4 (mental arithmetic, observation, talking and identification of the sounds).

The secondary task of observation produced the highest number of correct responses when the task was undertaken together with the identification of the sounds, the least number of correct responses occurred in Phase 3 (mental arithmetic, observation and identification of the sounds).

The shortest response time to identify the sounds occurred as follows;
i) Phase 4 (mean latency 3.249ms).
ii) Phase 2 (observation task, mean latency 3.273ms)
iii) Phase 3 (mean latency 3.309ms)
iv) Phase 2 (mental arithmetic, mean latency 3.344ms).

The subjective ratings, as already discussed, followed the pattern predicted, Phase 2 (observation) being perceived as the easiest Phase, second, Phase 2 (mental arithmetic), third Phase 3 and most difficult Phase 4.
<table>
<thead>
<tr>
<th></th>
<th>Phase 2</th>
<th>Phase 2</th>
<th>Phase 3</th>
<th>Phase 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(mental</td>
<td>(observation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct Response</td>
<td>arithmetic task)</td>
<td>task)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sounds</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct Response</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>means Sounds</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct Response</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mental arithmetic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct Response</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>observation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Latency milli seconds</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subjective</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ratings for difficulty</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Significant Confusions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Sounds</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.098 (1)</td>
<td>3.220 (2)</td>
<td>3.455 (4)</td>
<td>3.402 (3)</td>
</tr>
<tr>
<td></td>
<td>351 (3)</td>
<td>-</td>
<td>(2)</td>
<td>(1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>228 (3)</td>
<td>116 (1)</td>
<td>128 (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(late 32 early 4)</td>
<td>(late 141 early 7)</td>
<td>(late 124 early 12)</td>
</tr>
<tr>
<td></td>
<td>3.344 (4)</td>
<td>3.273 (2)</td>
<td>3.309 (3)</td>
<td>3.249 (1)</td>
</tr>
<tr>
<td></td>
<td>(2)</td>
<td>(1)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td></td>
<td>1 called 7, 2 called 1 and 4, 3 called 12, 4 called 2, 5 called 10, 6 called 5, 8 called 11, 11 called 8, 12 called 3</td>
<td>1 called 7, 2 called 1 and 4, 3 called 12, 4 called 2, 5 called 10, 6 called 5, 8 called 11, 11 called 8, 12 called 3</td>
<td>1 called 7, 2 called 1 and 4, 3 called 12, 4 called 2, 5 called 10, 6 called 5, 8 called 11, 11 called 8, 12 called 3</td>
<td>1 called 7, 2 called 1, 4 called 2, 8 called 11, 12 called 3.</td>
</tr>
</tbody>
</table>

Table 5.26: Summary Table of Comparison of Correct Responses to Tasks, Latency, Subjective Ratings and Confusions Between Sounds for Each Phase. (The ratings are shown in parenthesis).
5.3.9 Comparison Between Experiment 1 and Experiment 4 (Learning Phase)

The learning Phases of Experiment 1 and Experiment 4 are discussed in the following section as the same sounds and the same procedure was used in both Experiments.

The pattern of responses made during the learning phases of both Experiment 1 and Experiment 4 in which the same sounds (Natural Names) and the same procedure were used were shown to be similar. The total number of correct responses, made from a possible total of 144, is shown in Table 5.27. A correlation was conducted on the correct responses to sounds for both Experiment 1 and Experiment 4 and was found to be significant ($r=0.883$ df=10 $p<0.01$). This indicates that the number of correct responses for each sound in both experiments followed a similar pattern.

<table>
<thead>
<tr>
<th>Sound</th>
<th>Experiment 1 (Learning)</th>
<th>Experiment 4 (Learning)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>123</td>
<td>114</td>
</tr>
<tr>
<td>2</td>
<td>111</td>
<td>111</td>
</tr>
<tr>
<td>3</td>
<td>129</td>
<td>127</td>
</tr>
<tr>
<td>4</td>
<td>111</td>
<td>112</td>
</tr>
<tr>
<td>5</td>
<td>109</td>
<td>105</td>
</tr>
<tr>
<td>6</td>
<td>120</td>
<td>123</td>
</tr>
<tr>
<td>7</td>
<td>118</td>
<td>114</td>
</tr>
<tr>
<td>8</td>
<td>90</td>
<td>101</td>
</tr>
<tr>
<td>9</td>
<td>106</td>
<td>103</td>
</tr>
<tr>
<td>10</td>
<td>129</td>
<td>122</td>
</tr>
<tr>
<td>11</td>
<td>101</td>
<td>93</td>
</tr>
<tr>
<td>12</td>
<td>121</td>
<td>118</td>
</tr>
</tbody>
</table>

Table 5.27: Number of Correct Responses for the Learning Phases of Experiment 1 and Experiment 4 (Natural Names Condition).
The significant confusions (p<0.01) in Experiment 4 (Learning) were considerable fewer than in Experiment 1 (Learning). The confusions for both experiments are shown in Table 5.28. There are certain sounds that were confused in both experiments, sound 3 was identified as sound 12; sound 8 was identified as sound 11; and sound 12 was identified as sound 3. In the learning phase of Experiment 4 there were, however, a larger number of confusions than in the learning phase of Experiment 1 approaching significance. The significant confusions and those approaching significance in the learning phase of Experiment 4 are shown in Table 5.29.

<table>
<thead>
<tr>
<th>sound played:</th>
<th>Identified as (Experiment 1 Learning Phase)</th>
<th>Identified as (Experiment 4 Learning Phase)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7, 11</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>2, 8</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>3, 11</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 5.28: Comparison Between Significant Confusions Experiment 1 and Experiment 4 (Learning Phase) (p<0.01)
Sound Identified As:

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>*</td>
<td>0.8878</td>
<td>0.7936</td>
<td>0.8354</td>
<td>0.9982</td>
<td>0.0782</td>
<td>0.0132</td>
<td>0.3778</td>
<td>0.2332</td>
<td>0.2772</td>
<td>0.5404</td>
</tr>
<tr>
<td>2</td>
<td>0.2676</td>
<td>*</td>
<td>0.6438</td>
<td>0.0322</td>
<td>0.8490</td>
<td>0.7496</td>
<td>0.7192</td>
<td>0.4962</td>
<td>0.0422</td>
<td>0.8384</td>
<td>0.7368</td>
</tr>
<tr>
<td>3</td>
<td>0.7454</td>
<td>0.7210</td>
<td>*</td>
<td>0.6856</td>
<td>0.7966</td>
<td>0.6574</td>
<td>0.5140</td>
<td>0.3816</td>
<td>0.5846</td>
<td>0.6302</td>
<td>0.7210</td>
</tr>
<tr>
<td>4</td>
<td>0.4376</td>
<td>0.0158</td>
<td>0.3606</td>
<td>*</td>
<td>0.2894</td>
<td>0.3606</td>
<td>0.9086</td>
<td>0.7120</td>
<td>0.4444</td>
<td>0.0386</td>
<td>0.1058</td>
</tr>
<tr>
<td>5</td>
<td>0.7370</td>
<td>0.1866</td>
<td>0.1506</td>
<td>0.8362</td>
<td>*</td>
<td>0.0140</td>
<td>0.7968</td>
<td>0.0648</td>
<td>0.0594</td>
<td>0.0632</td>
<td>0.1866</td>
</tr>
<tr>
<td>6</td>
<td>0.7454</td>
<td>0.1286</td>
<td>0.5670</td>
<td>0.6856</td>
<td>0.0158</td>
<td>*</td>
<td>0.7090</td>
<td>0.3816</td>
<td>0.1660</td>
<td>0.1718</td>
<td>0.7210</td>
</tr>
<tr>
<td>7</td>
<td>0.2998</td>
<td>0.7404</td>
<td>0.8464</td>
<td>0.6392</td>
<td>0.8024</td>
<td>0.5988</td>
<td>*</td>
<td>0.5474</td>
<td>0.0128</td>
<td>0.8938</td>
<td>0.3702</td>
</tr>
<tr>
<td>8</td>
<td>0.0324</td>
<td>0.8586</td>
<td>0.0192</td>
<td>0.3934</td>
<td>0.5730</td>
<td>0.0192</td>
<td>0.0274</td>
<td>*</td>
<td>0.2752</td>
<td>0.9272</td>
<td>0.0001</td>
</tr>
<tr>
<td>9</td>
<td>0.1380</td>
<td>0.0476</td>
<td>0.1296</td>
<td>0.4532</td>
<td>0.2078</td>
<td>0.9598</td>
<td>0.5818</td>
<td>0.2528</td>
<td>*</td>
<td>0.5368</td>
<td>0.1544</td>
</tr>
<tr>
<td>10</td>
<td>0.5670</td>
<td>0.7074</td>
<td>0.7876</td>
<td>0.2616</td>
<td>0.6184</td>
<td>0.7876</td>
<td>0.8702</td>
<td>0.5514</td>
<td>0.2978</td>
<td>*</td>
<td>0.7074</td>
</tr>
<tr>
<td>11</td>
<td>0.4116</td>
<td>0.0398</td>
<td>0.0660</td>
<td>0.3656</td>
<td>0.1188</td>
<td>0.7354</td>
<td>0.8308</td>
<td>0.0558</td>
<td>0.0210</td>
<td>0.6854</td>
<td>*</td>
</tr>
<tr>
<td>12</td>
<td>0.4770</td>
<td>0.4544</td>
<td>0.0001</td>
<td>0.4224</td>
<td>0.5252</td>
<td>0.3974</td>
<td>0.8480</td>
<td>0.5850</td>
<td>0.9422</td>
<td>0.9576</td>
<td>0.832</td>
</tr>
</tbody>
</table>

Table 5.29: Experiment 4 Phase 1 (Learning) Significant Confusions (p=<0.01) (shown in bold borders) Confusions Approaching Significance (p=<0.05) (shown in double borders)
5.4 Discussion

5.4.1 Primary and Secondary Task Performance

One of the aims of the current study was to establish whether participants could correctly identify the twelve sounds whilst concurrently undertaking a task or a combination of tasks. The multiple resources literature was used as a framework, but the tasks used also reflected the types of tasks undertaken by staff in the I.C.U. environment. An attempt was made in the current study to select tasks that were representative of what happens in the 'real' world, as a criticism of models used in some studies of selective attention is that the task situations used are rarely encountered in 'normal' interactions in the environment (e.g. Tipper et al. 1992). The tasks used however, are arguably not as well validated as those tasks used in more theoretically based studies. For example, some secondary tasks were originally designed to test theories of human attention and memory, whilst other secondary tasks were designed to reflect the 'real-world' environment. These tasks are largely less well validated but have greater ecological credibility. Many of the latter tasks reflect the vast amount of research undertaken in the aircraft industry. For example, tracking tasks feature predominantly amongst the workload literature and are similar to actual tasks that a pilot would undertake in the cockpit.

At a theoretical level in the dual-task and workload literature it has been well documented that some combinations of tasks can be undertaken more successfully than other combinations of tasks. One of the important considerations in deciding a particular combination of tasks is whether resources involved in the tasks are shared or separate.

One of the assumptions of the multiple resources model is that if a secondary task is used that emphasises central processing functions, then this will interfere with a primary task that also utilises central processing, as opposed to a primary task that requires, for example, perceptual processing. In the current
experiment the tasks used were superficially different from one another in many characteristics and confusion or co-operation between the tasks used was not observed to occur during the study. However, by attempting to represent tasks undertaken in the I.C.U. environment rather than to deliberately engage or disengage specific cognitive or perceptual processes for reasons of theory alone, the distinctions between the underlying resources used in the primary and secondary tasks was obscured to some extent.

Thus, the underlying resources used in the tasks in the current study overlapped to some degree. The mental arithmetic task, the identification of the sounds, and the talking task were all presumed to use verbal and central processing codes. The response modalities between the tasks described above differed. Manual responses were used in the mental arithmetic task and the identification of the sounds, while a vocal response was used in the talking task. The observation task was presumed to use different resources from those used by the above tasks, requiring resources from the visual and spatial modalities, and perceptual processing codes. The response mode for the observation task was a simple manual response.

The problem of resource similarity between the tasks used in Experiment 4 may have been avoided if the tasks had been selected according to the secondary task literature (e.g. Hart and Wickens 1990). There have been many studies in which secondary tasks have been used that have been shown to utilise different resources to those required by the primary task. For example, Casali and Wierwille (1984) manipulated perceptual load by varying presentation rate and complexity of hazard warnings, using time estimation and tapping tasks as secondary tasks, which, as Hart and Wickens (1990) state, impose central processing demands (as opposed to perceptual demands) and manual response resources, therefore requiring different resources to those required by the primary task.
The tasks used in Experiment 4 were not tasks specifically designed for a secondary task paradigm. As previously discussed the selection of a secondary task is an important consideration and is dependent on the context of the experimental situation (Ogden et al. 1979). There have been several reviews of different secondary tasks used to investigate operator workload (e.g. Ogden et al. 1979, Eggemeier and Wilson 1991) which describe a wide variety of secondary tasks used in combination with many different primary tasks. The secondary tasks are used to evaluate many different problems ranging from driver's performance in heavy traffic to pilot's workload and the assessment of stress. Ogden et al. (1979) state that the most frequently used tasks are choice reaction time tasks, mental arithmetic, tracking, and monitoring. These have been discussed in greater detail in the introduction to this chapter.

On reviewing the literature to determine the types of tasks used as secondary tasks, the information is rather vague. For example, Eggemeier and Wilson (1991) describe the mental arithmetic tasks used in a variety of studies. The types of calculations used in the various studies included adding together two pairs of digit numbers presented aurally, an auditory subtraction task, a series of digits presented aurally with the participant required to add three to the final digit and report the total and as a final example an experiment in which participants were required to retain a series of numbers in memory and to add a further number to each digit when requested to do so by the experimenter. The major difference between these examples and the mental arithmetic task undertaken in the current study is that the calculations were presented aurally and required a vocal response. However, in the current study the mode of presentation for the primary task was auditory and it therefore seemed appropriate to present the mental arithmetic task visually, requiring a written response. A vocal response might have been used, but this would then have interfered with the talking task also used.
In Experiment 4 one of the objective was to investigate the effect on the primary task (identifying the sounds) of secondary tasks that could represent aspects of tasks undertaken on the I.C.U., utilising the concept of embedded tasks as described by Hart and Wickens (1990). This was considered important during the final phases of Experiment 4 for two reasons. First, because during the pilot study one participant declined to continue with the experiment as she found it too difficult. Second, in some experimental situations the secondary tasks are seen as unimportant or uninteresting, and are subsequently ignored or dropped by the operator or by participants in the laboratory. One solution to this problem is to use embedded tasks which are designed to appear as if they are part of, or important to, the primary task. It has been demonstrated that this strategy improves the likelihood that all the tasks will be performed.

Overall, there was very little decrement in the performance of the primary task, the correct identification of the sounds. The only significant difference was between the mental arithmetic task of Phase 2 in which there were fewer correct responses to the sounds than in mental arithmetic task in Phase 3, and no other significant differences between the phases were found for the recognition of the alarm sounds. There were, however, considerable decrements in the performance in the secondary tasks. This deterioration in performance will be discussed in detail later.

Practice is an important consideration in dual task and secondary task literature. In novel, complex tasks resources may be involved in learning, co-ordinating and executing the various components of the task (Wickens 1991). Thus, another criticism of Experiment 4 is that as all the participants undertook the phases in the same order it is possible that there was a learning effect. The improvement in performance could therefore be explained by the fact that, initially, participants found it difficult to manipulate the features of the various tasks, but with each successive phase found the tasks easier to combine. Although Experiment 4 was presented in the same
order to each participant because of the predicted increasing difficulty of the phases, which was confirmed by the subjective ratings of the participants, a further experiment could be carried out in order to investigate the effect of randomising the order of the phases.

A further criticism of Experiment 4 is that the same pages of calculations were used in each experiment so again it is possible that there was a learning effect. However, if different calculations had been used for each experiment it would have been difficult to assess whether the participants' performance had altered due to the specific combination of tasks, or because of the effects of using a different set of calculations. A problem with the mental arithmetic task was that there were only two sheets of calculations available from the Human Assessment Laboratory (HAL) and although the sheets were alternated during Phases 2 and 3, during Phase 4 the participants would receive one of the two sheets again. Once they had received the sheet they always began at the beginning so they were exposed to the same calculations twice and it is possible that the participants might have become familiar with the calculations, and a learning effect ensued.

A suggestion to improve the mental arithmetic task in Experiment 4 would be to have had more pages of calculations thereby ensuring that the participants always began with different calculations during each phase.

The highest number of correct responses for the identification of the sounds was shown to occur in Phase 3. The second highest number of correct responses for the identification of the sounds was in Phase 4, although Phase 4 was rated by the participants as being the most difficult phase. With the introduction of the talking task in Phase 4, there were fewer correct responses to the mental arithmetic task, although performance of the observation task increased slightly in comparison to Phase 3. The participants found it less demanding to observe the F's and E's and talk than to undertake the mental arithmetic task and
talk. This result would appear in part to agree with the multiple resources literature, in that some tasks are easier to perform together than other combinations of tasks. However, performance on the observation task only improved slightly, even though different resources were supposedly being utilised.

Although the observation task was predicted to produce the largest decrement in performance as literature regarding vigilance over even a short length of time predicts poor performance, the mental arithmetic task actually showed the greatest decrement over the phases. Participants also rated the mental arithmetic task as being more difficult than the observation task.

During Phase 2, the mental arithmetic task alone (together with the identification of the sounds), the participants reported they were anxious to complete the calculations accurately and to also identify the sounds correctly. The participants also reported that if they were in the middle of a calculation they found the alarm intrusive and wanted to continue with the calculation rather than stop and identify the alarm. This would appear to add support to Damos' claim (1991) that task priority should be stated by the experimenter, as participants would undertake the task they found most salient in the absence of such direction.

The least number of correct responses for the identification of the sounds occurred in Phase 2 (mental arithmetic) which also had the longest latency recorded, and was rated by the participants as being less difficult than both Phase 3 and Phase 4. Although it is possible there was a learning effect, another reason that could be suggested is that during both stages of Phase 2 the participants felt under less time pressure hence produced longer latencies, and therefore less stress (as discussed by Derrick 1988) than in the final two conditions in which they had more tasks to complete in a similar time period.

Wickens (1991) suggests that operators may process tasks simultaneously if, for example, they have to undertake two five
minute tasks in seven minutes. It could be suggested that this happens when operators become very experienced in certain tasks. In the laboratory environment during the study it is possible that

the participants altered their strategy when confronted by an increase in task demands, by 'speeding up' their responses in order to complete the tasks.

In the current study participants appeared to use the strategy of accelerating their responses for the primary task in Phase 3 when the secondary task load increased, but there was no evidence of a speed-accuracy trade-off as the number of correct responses to the primary task increased in Phase 3, and there was the corresponding predicted decrease in the correct responses to the secondary tasks.

It could therefore be suggested that the participants did change their strategy as suggested for example by Eggemeier and Wilson (1991) during the final two phases of the experiment. As Damos (1991) states, undertaking two or more tasks concurrently is intrinsically different to performing one task alone. However, these results need to be viewed with some caution as during the final two phases the scenario was introduced which included the concept of embedded tasks as discussed, for example, by Eggemeier and Wilson (1991). During the pilot study, for this study, when the secondary tasks were introduced, some participants perceived the overall task as being too difficult and abandoned undertaking the secondary tasks. Therefore as suggested by Hart and Wickens (1990) and discussed previously, the secondary tasks were introduced as embedded tasks, and the strategy appeared to work as the secondary tasks were regarded by the participants as important to the overall task. The introduction of the scenario may have accounted to some degree for the improved performance in identifying the sounds and the shorter latencies. With hindsight it would have been much more satisfactory to have introduced the scenario during Phase 2, at the beginning of the study.
The response modalities used in the current study were quite similar across the different tasks, although participants could have used their left hand to respond to the identification of the sounds and their right hand to respond to the observation task. Had the touch screen been compatible with the sounds program the participants could have used their left hand to just touch on the screen which may have been easier than manipulating the mouse. The participants also stopped writing the answers to the mental arithmetic task in order either to use the mouse to click on the sound they thought correct, or to press the space bar for the observation task. As many of the response modalities in the current study were similar, e.g. manual responses using a hand, future studies could consider different response modes such as foot or vocal responses.

The multiple resources model proposes that verbal processing used different resources to those used in the processing of non-verbal material. In the current study participants differed in the processing strategies used. As for example, Weinstein (1987) argues in some tasks participants could use a verbal or a spatial strategy. For example, in the identification of the sounds both spatial and verbal mnemonics were used to recall the names of the sounds. For example, some participants remembered the sound as increasing or decreasing in pitch, or as going up and down which could be described as spatial mnemonics. Other participants used the word sound of the name to fit with the auditory sound, using verbal processes, and another strategy used as a mnemonics was visual imagery. Some participants used a combination of strategies for different sounds.

5.4.2 Correct Identification of Sounds

In the current study sound 6 had the highest number of correct responses and the shortest latency overall, while sound 11 had the least number of correct responses and the longest latency. Therefore sound 6 was quickly identified and had the least number of incorrect responses in comparison to the other sounds.
in the experiment. Sound 11, as in the first set of experiments, was frequently incorrectly identified and took participants longer to identify than the other sounds.

Sound 6 is a distinctive sound and some participants used 'Beethoven's Fifth' as a mnemonic strategy to identify the sound. However, overall if a different mnemonic strategy was used by participants, sound 6 was still quickly identified. It is possible that whilst participants were undertaking the tasks sound 6 was relatively distinctive amongst the other sounds, although in Phase 3 sound 6 was significantly confused with sound 5, a similar melodious sound, but not as distinctive as sound 6.

Sound 11, as discussed in Chapter 2, appeared to be difficult for participants to identify, with a temporal pattern that was frequently confused with sound 8, as in the first set of experiments. Sound 8 was also frequently incorrectly identified and consistently confused with sound 11.

Sounds 8, 10 and 11 produced fewer correct responses in the mental arithmetic phase than in both Phase 3 and Phase 4. This could be explained as task trade off as in Phase 2, the mental arithmetic task, participants performed well on the secondary task. In Phase 3, performance on the primary task of identifying the sounds improved whilst the secondary task of mental arithmetic showed a considerable decrement.

5.4.3 Significant Confusions

A further aim of the current study was to establish whether confusions between sounds changed or increased as a result of the additional tasks. During the learning phase of Experiment 4 there were considerably fewer significant confusions between the sounds than had been observed in the learning phase of Experiment 1, although there were a substantially higher number of confusions approaching significance in Experiment 4 than in Experiment 1.
The 'natural' sounds had been chosen for use in Experiment 4 as they had been shown in the first set of experiments to have been the easiest to learn. Although the procedure used in both experiments was identical, the participants in Experiment 4 did not confuse the sounds as much as the participants in Experiment 1. A possible reason for this is that they were just better at learning the sounds. However, the confusions became more apparent during the phases when the tasks were introduced. As previously discussed, the identification of auditory sounds does not usually occur in isolation, and the focus of interest in the current study was the confusions that occurred when the tasks were introduced.

A confusion that occurred consistently during Phases 2, 3 and 4, (but not during Phase 1), of Experiment 4 was the asymmetrical confusion between sound 1 and sound 7. However, sound 7 was not significantly confused with any other sound during the study. It could be suggested that whilst undertaking the concurrent tasks the participants perception of the sounds altered. As discussed above in the correct responses sound 6 was confused with sound 5 although they are very different sounds when heard one after another. Sounds 6, 5, 1 and 7 are all melodious. Sound 1 is a two-tone pattern, whilst sound 7 has two pulses, both sounds having short intervals between either the tone or the pulse. By introducing the tasks the participants could, for example, have just recognised the sounds as being 'melodious' whereas without the tasks the more detailed features of the sounds were recognised.

There is an issue of the generalisation of the alarm-related results to be addressed both in Experiment 4, and the first set of studies, Experiments 1, 2 and 3 which are reported in Chapter 2. When experiments take place over a short period of time, the task of stimulus identification is rather more of a relative discrimination task than would be expected to occur in the working environment, where sounds generally need to be identifiable on an absolute basis. When a single alarm occurs in
practice, there may be no other alarm with which to compare it for identification, and so it may be less identifiable as a result.

It is to be expected that the total number of alarms which could be recognised under normal conditions would be lower than that which can be achieved under laboratory conditions. For example, Weiss and Keushner (1984) demonstrated that a set of twelve warnings could be readily discriminated in a laboratory study, where comparisons between sounds can easily be made. As Sanders and McCormick (1993) suggest, if absolute identification between sounds is required, then a maximum of five warnings may be appropriate.

The ease with which the transition from relative to absolute identification is made will, however, depend on the cues used to recognise the warning and these cues will in part depend on the range of sounds used in the set of sounds. As previously discussed, pitch information decays over time (Deutsch 1978) so one would expect longer time intervals between alarms differing only by pitch (such as continuos tones, for example) to lower people's ability to distinguish between two or more such sounds. As it is, the results of the experiments reported here show that even with short time intervals between the successive playings of such sounds they are hard to differentiate. On the other hand, if there was only one such sound in the complete set of warnings, and as a result it was remembered in terms of a verbal cue such as 'continuous', then it may be readily discriminated both in relative and absolute terms.

Sounds too which can be readily labelled may be more resistant to forgetting during the relative/absolute transition, but further work needs to be undertaken on this topic. For example it is possible that the sounds investigated by Weiss and Kershner (1984) labelled 'wow' 'yelp' and 'ping' might still be recognised in an absolute identification task because of their acoustic imagery and labelling qualities. Other sounds (such as 'low warble', 'warble', and 'high warble' and 'beeps', 'long beeps' and 'short beeps' may suffer more in working conditions even though
they could be discriminated under laboratory conditions. However, the results presented in this thesis suggest that differentiation on the basis of different on/off cycles may not be that good even in laboratory conditions.

In the warning set tested in the experiments reported in Chapter 2 and in this chapter, it is likely that sounds which are easy to label, for example sound 6 labelled as 'Beethoven's Fifth' by many participants, will still be identifiable under absolute identification conditions, and that sounds which are in some way unique may still be recognisable under naturalistic working conditions. However, it is necessary to undertake further research in this area, particularly in relation to the verbal labelling of sounds.

5.4.4 Subjective Measures.

Once the experiments had been completed the participants were asked to rate, on a scale of 1 to 100, which phase of the experiment they had found the most difficult. This was a unidimensional scale which only measured one aspect of how the participants perceived the phases overall.

Derrick (1988) 'culled' (sic) a rating scale from existing workload rating scales and used the scales that appeared the most appropriate for the tasks used in his study. However, Derrick does not comment as to whether or not the culled scales were validated.

It would have been interesting in the current study to have been able to use a multi-dimensional scale to obtain more detailed information regarding the subjective views of the participants. For example, the stress levels produced by the cognitive demands of the task combinations would have been interesting to measure. However, although multidimensional scales were considered, for example Situational Awareness Rating Technique (SART e.g. Fracker 1988, Selcon et al. 1991) which has been developed as a subjective evaluation technique in aerospace...
system design evaluation, a problem with the majority of the subjective rating scales is that they have been designed with the aircraft industry, and specifically pilots, in mind which may not be appropriate under these circumstances.

The task characteristic literature suggests that the more complexity, less feedback and greater time pressure should combine to increase perceptions of workload (Derrick 1988). Derrick found that these measures do not combine in a simple way, in fact they demonstrate differential effects on the overall perception of the workload. The participants in the current study were given no feedback for any of the tasks during Phases 2, 3 and 4. Therefore, if there was confusion between sounds these would have endured throughout the entire experiment. However, mere exposure to the sounds could contribute to some of the learning.

Feedback from participants at the completion of Experiment 4 regarding the difficulty of the phases supported Wickens and Liu’s (1988) suggestion that in certain situations operators actually try harder to complete a task although at increased effort to the operator, which is unobserved. As Derrick (1988) reports, using tasks alone to predict workload may lead to erroneous results and the inclusion of subjective measures is therefore important.

Participants trying harder would explain the improvement in performance in Phase 3 and 4 for the correct identification of sounds, and the shorter latencies but a higher rating regarding the difficulty of the phases.

Some of the participants stated that during the combination of tasks, as the tasks increased in difficulty, they gave up on either the vigilance task or the mental arithmetic task, and many of the participants commented that had the study been in a real clinical environment, their patient would have died!
Whilst studies have suggested that if a task is practised alone and becomes automated then two tasks can be performed without any decrement (e.g. Schneider and Fisk 1982), in the current study the participants did not receive enough practise on any of the tasks for them to become automated. It could be suggested that during the current study the participants timeshared the tasks. However, staff in the clinical area of the I.C.U. may develop a very efficient time sharing strategy and aspects of the tasks may then become automated. Furthermore, they may develop a true skill in timesharing, for example, knowing when to sample from the displays instead of continually scanning the monitors.

5.4.5 Future Research

It has been demonstrated that in single task experiments by giving specific instructions to the participants, performance has improved, for example Vidulich and Wickens (1984) instructed participants to try harder and Yeh and Wickens (1988) gave more stringent performance criteria. Several studies have looked at motivational incentives, which again have improved performance.

Future studies could examine the effect on the identification of the sounds as the primary task, as opposed to the secondary task, increases in difficulty (e.g. Navon and Gopher 1979). For example, the 'controlled' sounds and the 'neutral names' used in the first set of experiments could be used in future experiments. The relative priority of the tasks used in the current study could also be manipulated. For example, the primary task could be the mental arithmetic task with the identification of sounds as a secondary task.

As discussed above, tasks from the secondary task literature could be used which incorporate different resources from those resources required by the primary task. For example, a time estimation task (e.g. Hart 1986), in which participants are asked to assess retrospectively how much time has passed, could be
used. The premise of a time estimation task is that during conditions of high workload, an interval of time is generally underestimated. Time estimation techniques are described by Hart (1986) as being an effective measure of operator workload.

A similar task that could also be considered is an interval production task, in which a participant produces a series of regular intervals by performing a motor response at a specific rate.

A further alternative could involve the participants being instructed to maximise performance on all tasks. There exists a large number of studies that have provided instructions to participants to vary the allocation of effort between two tasks in different proportions (e.g. Vidulich 1988 Wickens et al. 1983). In general the results showed that the performance of the two tasks trade off reciprocally, that is resources are withdrawn from one task and are redeployed to improve the performance of the secondary task.

Future research could also use physiological measures such as heart rate variability or event related potentials in the laboratory environment. As the tasks were perceived by the participants to be very demanding it could be predicted intuitively that the increases in task demands would produce corresponding physiological changes.

Not all the possible combinations of the secondary tasks were undertaken in this study. For example, talking together with identification of the auditory warnings, the observation task plus talking and identification of auditory warnings or the mental arithmetic task talking and identification of auditory warnings could also be combined in different experimental conditions. Furthermore, as already discussed, to prevent the possibility of an order effect occurring, future experiments could also vary the order of Phases 2, 3 and 4.

In conclusion, although the secondary tasks used to represent tasks undertaken on the I.C.U. did not entirely fit the criteria of
using different resources to those used by the primary task, the results indicate that when the primary task was combined with two or more of the secondary tasks performance could not maintained. This suggests that there could be considerable overload for staff working in the average I.C.U.
Chapter 6: Discussion

6.1 Overview

One of the main aims of the current study was to investigate confusions that occur between auditory warnings currently in use at the I.C.U. Derriford Hospital. As Kerr (1985) argues, problems regarding auditory warnings cannot be addressed in isolation and a second aim therefore, was to examine whether the introduction of tasks affected the performance on the primary task, the identification of the auditory sounds. To ascertain the types of activities undertaken by the staff when auditory alarms are activated, two observational studies were undertaken, one using a video camera and the second, direct observation. The findings from the observation studies were then used to develop a range of tasks that represented aspects of actual tasks undertaken in the I.C.U. environment.

The multiple task literature proposes that tasks using different resources can be performed together more successfully than combinations of tasks using different resources. Therefore, in the final experiment tasks were developed that were representative of tasks undertaken in the I.C.U. environment, and that were assumed to use different resources. Although the tasks appeared superficially different from one another, there was some degree of overlap in the tasks used in the final study. For example, the mental arithmetic task, the identification of the sounds and the talking task were all assumed to use central resources, verbal processing resources, and a manual response. However, the observation task used resources that were assumed to be different, which included perceptual processing, spatial resources, and a manual response. The results and implications of the studies will be discussed in greater detail in the following sections.
6.2 General Discussion of Results

It has been demonstrated in the current experiments that the participants learned the set of sounds very quickly, many of the participants learning the sounds in less than 1 hour. The learning was retained for over one week with almost perfect performance when the retention of the sounds was tested in the return phase. These are similar findings to Patterson and Milroy (1980) who examined how quickly participants could learn and identify a set of ten aircraft warnings and found that after having learnt the sounds during the first session, the participants then returned one week later and were asked to re-identify the same set of sounds. Their results showed that during the return session the participants achieved almost perfect performance.

Overall, there was no one sound that was the easiest to learn and that produced the highest number of correct responses across all experiments. In Experiments 1 and 2, sound 3 produced the highest number of correct responses, and in Experiment 3, sound 10 produced the highest number of correct responses. When the tasks were introduced in Experiment 4, sound 6 produced the highest number of correct responses overall.

The sound that consistently produced the least number of correct responses throughout the majority of the experiments was sound 11. In Experiment 3 sound 2 produced the least number of correct responses.

These results indicate that the overall properties of a sound may be important factors when participants are identifying sounds, with some properties being more salient than others. For example, sound 6 sounded very much like Beethoven's Fifth, a distinctive, well-known sound. The findings would also appear to indicate that the verbal labels that individuals attach to arbitrary sounds are important mnemonic strategies when identifying such sounds.
Sound 3 was a distinctive, continuous sound of 1000Hz, and in Experiments 1 and 2, this produced the highest number of correct responses. It could be suggested that the overall properties of the sound, a continuous tone, was relatively easy for participants to identify, but the sound was also consistently confused with sound 12 which was again a continuous sound, but with a higher pitch of 2350 Hz. The two sounds, 3 and 12, when heard one after another, could be easily discriminated, when they were heard at random, during the experiments, the participants consistently confused them. Moreover, the time intervals between the sounds would be much greater in practice than they were during the experiments. The confusion between sounds 3 and 12 demonstrate that, as Deutsch (1978) showed, individuals find it difficult to discriminate between warnings on the basis of pitch alone as memory for pitch decays over time. Furthermore, sounds 3 and 12 had no temporal characteristics to distinguish between them.

Overall, there was very little decrement in the performance of the primary task, the correct identification of the sounds. The highest number of correct responses for the identification of the sounds occurred in Phase 3 while the least number of correct responses occurred in Phase 2.

Participants found sound 11 the most difficult sound to remember. The sound consisted of a repeated tone of 2200 Hz, lasting 0.35 seconds with an equal sized gap of 0.35 seconds between the tones. Sound 11 was consistently confused with sound 8, which was a similar sound of a repeated tone lasting 0.5 seconds at 1760Hz, with a 1-second pause between the tones. Even though there was a difference in the timing, sound 11 was much faster with approximately 15 beats in a 10 second period while sound 8 was slower with 6 beats in a 10 second period, (sound 8 being about 2.5 times slower than sound 11). The fact that the confusion was asymmetrical suggests that participants could identify the faster tempo when sound 11 was heard, but when they heard sound 8, they could not distinguish between the tempi. Patterson and Milroy (1980) also identified consistent confusions and non-confusions
and concluded that a similarity of temporal characteristics, in particular repetition rate, can cause confusion.

Most of the confusions that occurred in the present study were asymmetric, for example sound 4 was frequently named as sound 2 but not vice versa (2 was not named as 4 in the same way). As discussed in Chapter 2, Dawson and Harshman (1986) found that asymmetries that occur in visual perception, for example, they found that the letter 'Q' was more frequently mistaken for the letter 'O', than the letter 'O' was for the letter 'Q'. The alphabetic confusion research may provide a framework for addressing the issue of asymmetrical auditory confusions found in the current study.

In the learning phase of Experiments 2 and 3 there were also significant non-confusions that occurred. A non-confusion was determined by the fact that for a particular pair of sounds there was no error, not even as a guess. For example in Experiment 2, a non-confusions was sound 6 which was never called sound 10. Non-confusions were not present with the 'natural' sounds used in Experiments 1 and 4. This indicates that although some of the cues that may be used in identifying a sound were removed (such as the medical name, for example) the participants recognised that sound 6 was definitely not sound 10 and if they did make a guess at the response it was not for that sound.

The implications of the findings regarding confusions between auditory sounds could be of relevance to the British Standards Institute (BSI), The Comite European de Normalisation (CEN) and the International Standards Organisation (ISO) who are currently drafting standards on auditory alarms for medical equipment. For example, sound 3 was readily identified by participants in Experiments 1 and 2. This indicates that although a continuous tone may not be acoustically the most suitable sound to use, it is recognised specifically because it is a continuous tone. However, when a second continuous tone is introduced this then causes confusion between the two sounds. Similarly, if two 'melodious' sounds are used the overall labelling of the two sounds could result in confusions occurring, as was shown in the
current study by the confusions between for example, sound 5 and sound 6. It could therefore be suggested that only one continuous sound should be used in an environment where there are several different auditory alarms. Similarly, one melodious sound, one sound with an on-off rhythm and so on. This suggestion may help prevent confusions occurring in environments such as the I.C.U.

The mnemonic strategies used by participants to identify the sounds were similar across all the experiments. Participants used a variety of mnemonic strategies in order to identify the auditory sounds. Some participants used imagery, for example, humidifier was remembered by one participant as a hot sunny day with people running around, (the image perhaps being suggested by the word 'humid'). Another sound was labelled by a participant as the 'Spiderman theme.' As Baddeley (1990) demonstrated, visual imagery had a powerful effect on learning word lists and found that words which could be imagined (e.g. bullet and grey) were learnt easier than abstract words (e.g. gratitude and infinite).

In the first set of experiments, the more neutral names used in Experiment 3 had an effect on the mnemonic strategy developed by participants. The same kind of images were not evoked by the neutral names, for example, participants in Experiment 3 appeared to find it difficult to conjure an image for names like Uniform and Alpha. Sound 10 produced the highest number of correct responses in Experiment 3 and three participants identified the sound, which was called 'Sierra', by imagining the Sierra car, which is not an immediately obvious choice of mnemonic. Sound 2 produced the least number of correct responses in Experiment 3 only, the name for the sound was 'Bravo', which perhaps participants found difficult to develop a mnemonic strategy for.

Some participants used a verbal mnemonic strategy by fitting the name to the sound. For example, Aircall is a two-tone sound that fitted into the two syllables of the name. Other
participants used a spatial strategy to identify sounds by remembering specific features of the sound, for example an increase or reduction in pitch. As Baddeley (1990) states, imagery may have both visual and spatial dimensions and in the experiments reported participants appeared to use a combination of strategies. The multiple resources model proposes that verbal processing used different resources to those used in the processing of non-verbal material. In the current study participants appeared to differ in the processing strategies used to identify the sounds, as Weinstein (1987) suggests, in certain tasks participants could use either a verbal or a spatial strategy.

The acceptance and development of a more succinct set of auditory warnings than those in use at the present time, that are less confusing and more appropriate in terms of urgency is often delayed because of lack of clinical testing. Developing a protocol to test an improved set of warnings within an I.C.U. environment would seem a priority before new warnings are introduced. There is a need for appropriate testing and validation of any new warnings designed in the hospital environment. There is also a need to develop protocols within the hospital environment. A starting point may be to simulate an I.C.U., with alarms occurring at various locations and participants having to undertake various tasks whilst responding to the alarms.

6.2.1 Latency

During Experiment 4 the participant's response time when identifying the sounds was recorded. The results showed that the mental arithmetic task (Phase 2) produced a longer response than the other phases. Phase 3 produced the shortest response latencies.

In the working environment time pressure is regarded as a major source of mental workload (e.g. Moray et al. 1991). There are few environments in which there is unlimited time in which to complete tasks. Hart and Staveland (1988) amongst others have
found that time pressure influences increased perceptions of subjective workload. The least number of correct responses for the identification of the sounds also occurred in Phase 2 (mental arithmetic task) and was rated by the participants as being less difficult than both Phase 3 and Phase 4. Although it is possible there was a learning effect, another reason that could be suggested is that during both stages of Phase 2 the participants felt under less time pressure, hence the longer latencies, and therefore less stress than in the final two conditions in which they had more tasks to complete, in a similar time period.

Although the results for the individual sounds were not significant overall sound 6 appeared to produce the shortest latency and sound 11 seemed to have the longest latency. The absence of a significant effect for the individual sounds could be explained by the fact that the response latencies recorded on the Archimedes were overall fairly similar.

6.2.2 Performance on Secondary Tasks

In the current research the primary task of identifying the auditory warnings was emphasised as the priority which, as Damos (1991) points out is important, otherwise participants would decide on their own priorities. The primary task was identified, unlike the Gaba and Lee (1990) study, in which only the performance on the secondary tasks was measured. In the current study performance on both the primary and secondary tasks was recorded.

In the dual-task and workload literature it has been well documented that some combinations of tasks can be undertaken more successfully than other combinations of tasks. One of the important considerations in deciding on a particular combination of tasks is whether resources involved in the tasks are shared or separate. Another important assumption of the multiple resources model is if a secondary task is used that emphasises central processing functions, then this will interfere with a
primary task that also utilises central processing, as opposed to a primary task that requires perceptual processing.

In the current experiment the tasks used were superficially different from one another in many characteristics and confusion or co-operation between the tasks used was not observed to occur during the study. However, according to the multiple resources literature, the underlying resources used in the tasks in the current study overlapped to some degree. The mental arithmetic task, the identification of the sounds, and the talking task were all presumed to use verbal processing codes, central processing and manual responses were used in the mental arithmetic task and the identification of the sounds, and a vocal response was used in the talking task. The observation task, however, was presumed to use different resources, requiring resources from the visual and spatial modalities, perceptual processing and a simple immediate manual response.

Overall, there was very little decrement in the performance of the primary task, the correct identification of the sounds. The only significant difference was between the mental arithmetic task of Phase 2 in which there were fewer correct responses to the sounds than in mental arithmetic task in Phase 3, and no other significant differences between the phases were found for the recognition of the sounds.

There were however, considerable decrements in the performance in the secondary tasks. When undertaking the mental arithmetic task (Phase 2) participants produced the highest number of correct response when the task was undertaken concurrently with the identification of the sounds and the least number of correct responses occurred in Phase 4 (mental arithmetic, observation, talking and identification of the sounds).

When undertaking the observation task, participants produced the highest number of correct responses when the task was undertaken together with the identification of the sounds, and
the least number of correct responses occurred in Phase 3 (mental arithmetic, observation and identification of the sounds).

The participants appeared to find it less demanding to observe the F's and E's and talk, than to undertake the mental arithmetic task and talk. This result would appear, at least in part, to agree with the multiple resources literature, in that some tasks are easier to perform together than other combinations of tasks. However, performance on the observation task only improved slightly, even though different resources were supposedly being utilised.

Tasks may interfere with each other if they are too difficult to undertake concurrently. The tasks used in the current study were considered to be relatively simple tasks, similar to the types of tasks used in secondary task paradigms. Yet when the tasks were combined, performance deteriorated considerably. Kahneman (1973) proposed that tasks will interfere with each other when the amount of attention necessary to complete each task exceed the capacity of the operator, and this appears to have been the case in the current study. As Craig (1991) reports, in some situations a 'difficult' task would still be manageable because it does not exceed capacity, and that decrements may occur only when there is an already demanding primary task and performance will subsequently suffer from any increase on that load.

Although the observation task was predicted to have the largest decrement in performance as literature regarding vigilance over even a short length of time predicts poor performance, the mental arithmetic task showed the greatest decrement over the phases. Participants also rated the mental arithmetic task as being more difficult than the observation task.
6.3 Strategies and Workload

The direct observation study demonstrated that staff working in the I.C.U. often undertake multiple tasks concurrently. It is possible that when tasks are undertaken in the laboratory different strategies are used to when similar tasks are performed in the working environment. As Eggemeier et al. (1991) point out laboratory experiments are conducted for a relatively short period and participants do not receive the training or the opportunity to develop efficient or effective timesharing skills which staff on the I.C.U. may have developed with practice when undertaking concurrent tasks.

In the laboratory environment during the study it appeared that participants altered their strategy when confronted by an increase in task demands, by 'speeding up' their responses in order to complete the tasks. The strategies used by participants during the final experiments appeared to be very rapid switching between the tasks. This was particularly apparent during Phase 3 when the secondary task load increased. However, there was no evidence of a speed-accuracy trade-off as the number of correct responses for the primary task increased and the shortest latency occurred in Phase 3.

As the phases were undertaken by the participants in the same order there could have been a learning effect. The participants could have developed initial time sharing strategies and as they progressed through each phase they became more familiar with the tasks. An important consideration is that in novel, complex tasks resources may be involved in learning, co-ordinating and executing the various components of the task (Wickens 1991). However, the introduction of the scenario in Phases 3 and 4 could also have had some effect on performance, the improvement identifying the sounds and the shorter latencies. As Hart and Wickens (1990) discuss, it is beneficial if the secondary tasks are perceived by participants as being important to the primary task otherwise the secondary tasks could be ignored all together.
Eggemeier et al. (1991) also state that often in a laboratory setting instructions to participants emphasise similarity of performance, for example to perform all tasks as well as possible. In the current experiments the emphasis was on the performance of the primary task. However, in the I.C.U. environment when multiple tasks are being undertaken staff determine their own priorities. It is argued by some (e.g. Hart and Staveland 1988, Eggemeier et al. 1991) that strategies used by operators in multiple task environments modify the workload and subsequent resource expenditure.

In multiple task environments it is postulated that some tasks can be undertaken by rapid timesharing between tasks while other tasks need to be performed concurrently. The multiple resources model also predicts that tasks that use different modalities can, in general, be performed with less decrement than tasks that share resources.

In the current experiment it was not possible for participants to perform some of the tasks concurrently because of physical constraints. If, for example during the observation task the participant was observing the screen for the letter 'E' to appear and an auditory warning occurred, the participant would have to look away from the observation task presented on one computer in order to identify the sound on the other computer and click on the name. Similarly, if a participants was undertaking the mental arithmetic task and an alarm sounded they had to stop the calculations in order to look at the Archimedes in order to identify the sound. It was physically impossibility for individuals to look in two directions at once. Likewise, the response modalities were similar, participants had to stop writing the answers to the mental arithmetic calculations in order to either manipulate the mouse of the Archimedes computer, or to press the space bar on the B.B.C. computer. Other studies (e.g. Damos 1991) have suggested using different response modes such as a foot pedal or a verbal response.
The strategies operators use in multiple task environments depend on the training and practice of individuals. As Lintern and Wickens (1991) suggest it may be advantageous to develop general timesharing skills or strategies in multi-task environments. Eggemeier et al. (1991) suggest that the way an operator processes information may be influenced by training and its impact is particularly obvious in multi-task situations when it is often necessary to timeshare various tasks.

Eggemeier et al. (1991) further suggest that training and improved strategies of information processing by an operator can lead to aspects of a task becoming automated, which develops with extended practice. Because automatic processing is associated with less resources being used, adequate training in timesharing skills could result in more efficiency and reduce the levels of workload in multi-task environments. However, Moray et al. (1991) concede that determining the best sequence in which to perform a series of tasks under time constraint is a complex matter. In the working environment operators may have very little time to plan any strategy.

In complex, multi-task environments operators use cues in order to predict certain events, for example when intervention is necessary in a particular system (e.g. Cuqlock-Knopp et al. 1991). In the I.C.U. environment experienced staff may also use cues to predict certain events. For example, if it is known that the drug being administered via a syringe pump is due to finish at 3 pm and an alarm occurs at 2.58 pm the staff nurse responsible for that patient may use these cues to assess which piece of equipment is alarming and the reason why the alarm has been activated.

Cuqlock-Knopp et al. (1991) also suggests that one of the most straightforward ways to reduce the demands of a task is to design the task so that the number of mental operations required for its completion are small. They further suggest that if information was presented in a holistic way the operator would be able to combine several separate dimensions of information
into one image. This in turn would reduce the effort or attention required to perceive the information displayed and result in greater processing efficiency. This is applicable to the I.C.U. environment as computerised systems are gradually being introduced (e.g. Green et al. 1991)

One of the criticisms of the multiple resources model is that it fails to take into account aspects such as environmental stress (e.g. Gopher and Donchin 1986). It would seem that an important consideration when examining the mental and physical components of workload, is the stressors specific to the environment being investigated. The following section examines some of the stressors that are specifically related to the I.C.U. and hospital environment which may need to be considered in further studies regarding workload on the I.C.U.

6.4 Factors Relevant to the I.C.U. and Hospital Environment

There are many studies that have investigated stress in the I.C.U., (e.g. Fletcher 1987). Hay and Oken (1972) reported that staff in I.C.U. who are involved with seriously ill patients and continual contact with death often experience grief, anxiety, guilt, exhaustion, over commitment and over stimulation. The consequences of these psychological demands often results in staff withdrawing from the patient by, in many instances, relating more to the machines and equipment.

Increases in medical technology in the I.C.U. also cause stress amongst staff. McCarthy (1985) found that inadequate training particularly with respect to the monitors and equipment caused stress, often no formal training was provided with the nurses picking up pieces of information from their colleagues, who had in turn been told about the equipment by someone else. McCarthy also found that nurses reported often feeling like extensions to the machines rather than highly trained specialists. If a more formal system of training was implemented, the learning of auditory warnings could be incorporated quite easily as staff could at least be given an introduction to important alarms and
possibly even learn the sounds in a very short time, as the current study has demonstrated.

Other stressors are characterised by the very nature of hospital work in which employees work around-the-clock. Nursing staff usually work a set shift of 8 or 12 hours and are then relieved by another shift while the overall work of caring for patients continues. Price (1984) reports that 12 hour shifts have a negative impact on patient care and an adverse effect on the quality of the nurses' time and behaviour. However, Fields and Loveridge (1988) compared nurses who worked 8 or 12 hour shifts to investigate if those nurses working the 12 hour shifts in an I.C.U. suffered from increased fatigue and impaired critical thinking ability. The study found no difference between the two shifts, showing that while all the nurses were more fatigued at the end of their shift than at the beginning, their critical thinking abilities tended to be higher at the end of the shift.

However, many studies have demonstrated that operators who continuously perform cognitive tasks for extended periods do demonstrate performance decrements and become less efficient at detecting visual or auditory signals (e.g. Stroh 1971, Mackie 1977, Warm 1984, Angus and Heslegrave 1985). It is also suggested that vigilance decrement can occur as early as 20-35 minutes are the initial onset of a task. To remedy this problem Warm (1984) recommends that short monitoring periods should be undertaken, in reality this does not seem plausible if staff are working 12 hour shifts. Conversely, other studies have argued that the efficiency in detecting critical signals depends on the level and complexity of the signal used in the task (e.g. Thackray and Touchstone 1989). It would therefore seem important to introduce into environments such as the I.C.U. auditory alarms that are not complex, are easy to learn and remember, and informative about the situation they are signalling.

Sleep loss increases the onset of decrements in cognitive performance, particularly on attention demanding vigilance tasks.
during sustained work (Haslam and Abraham 1987; Haslam 1982; Martin et al. 1986). Night workers experience disruption of circadian rhythms (Tepas and Monk 1987, Monk and Embrey 1981). Many hospital staff are required to rotate periodically from day to night duty or work permanent night duty and follow a conventional social pattern during their days off which consequently disrupts circadian rhythms.

Night work presents other problems such as a decrease in alertness as the body temperature decreases to its lowest level from 03.00 - 06.00 hours (Minors and Waterhouse 1985, Monk et al. 1985). In general, workplace performance is poorest during the early morning hours (Folkard and Monk 1979). These factors could have implications for workload measures used to assess how staff perform on various tasks. If staff are fatigued because they are on night duty and their circadian rhythms are in disarray, it could be suggested that performance on certain tasks may be impaired. These factors may also affect the subjective ratings of staff, for example, a tasks undertaken in the early hours of the morning may be perceived as being more demanding than the same task undertaken during the day.

The type of work in I.C.U. is continuous and often is not differentiated by day or night. It could be suggested that a decrement in vigilance is more likely to occur during the early hours of the morning, particularly if the member of staff is tired. Subsequently, if a patient suddenly deteriorated between 03.00 - 06.00 hours, when alertness is at its lowest, if auditory warning sounds are very similar there is the potential for confusion between the sounds and a possible critical incident could then occur.

6.5 Subjective Ratings

If future studies were undertaken in the I.C.U. environment it would be appropriate to obtaining the subjective opinions of staff regarding the mental effort involved in performing the various tasks. A task that may objectively be rated as a
superficially 'easy task' (e.g. talking to visitors), may in fact be rated subjectively, as a highly stressful task, requiring considerable mental effort by the individual. The rating may depend on the subject matter of the conversation. For example, the conversation may involve a discussion with a distressed relative, who can often appear very angry and even aggressive to staff, and dealing with such a situation may require considerable mental resources and effort.

Subjective ratings are useful in determining the perceived mental workload of the operator but only if they are properly validated (e.g. Hill et al. 1992). There have been several subjective rating scales developed particularly in the aviation industry, for example the Cooper-Harper Scale (Reid and Nygren 1988) was developed to investigate the handling of aircraft. More recently the Situational Awareness Rating Technique (SART e.g. Fracker 1988, Selcon et al. 1991) has been developed as a subjective evaluation technique in aerospace system design evaluation. One of the problems with subjective rating scales is that they have been designed with the aircraft industry and specifically pilots in mind.

Another problem with implementing subjective rating scales in an environment such as I.C.U. is that staff may be too busy to respond to a subjective rating scale. This may result in a delayed report of the perceived mental effort experienced during a particular task. Delayed responses are problematic since the individual has to recall the situation and information may subsequently be forgotten or inaccurate.

6.6 Future Research

If future research programmes were developed an area of research that could be developed further is the investigation of the mnemonic strategies individuals use to identify abstract sounds both in the laboratory setting and in the I.C.U. environment. There has been considerable research regarding environmental sounds (e.g. Ballas 1993) but there appears to be
limited research regarding the strategies used to identify abstract sounds.

In future laboratory studies the effect on the identification of the sounds as the primary task increases in difficulty (as opposed to the secondary tasks in the current study) could be examined. For example, the 'controlled' sounds or the 'neutral names' used in the first set of experiments could be used in future experiments as these sounds were shown to be more difficult than the 'natural' sounds for participants to identify.

The relative priority of the tasks used in the current study could also be manipulated. For example, the primary task could be the mental arithmetic task with the identification of sounds as a secondary task. Another alternative could involve the participants being instructed to maximise performance on all tasks.

Tasks could be used that incorporate different aspects of tasks identified in the observation study, or that emphasise a specific type of task for example, observation. The molar categories used in the current study were very broad and drugs category for example, could be structured as a molecular category. Conversely, tasks that are entirely different and unrelated to the I.C.U. environment could be used in conjunction with a series of tasks such as those used in the current study. For example, a time estimation task (e.g. Hart 1986), in which participants are asked to assess retrospectively how much time has passed could be used. The premise of a time estimation task is that during conditions of high workload, an interval of time is generally underestimated. Time estimation techniques are described by Hart (1986) as being an effective measure of operator workload. A similar task that could also be considered is an interval production task, in which a participant produces a series of regular intervals by performing a motor response at a specific rate, for example, tapping in intervals of 10 seconds. The assumption is that the busier an operator is the less attention or resources will be available to judge the time. It could be
hypothesised that if participants were undertaking an interval production task, the time intervals would be longer for sounds that participants found difficult to identify because more resources were being used to distinguish that sound.

Future research could develop protocols to examine mental workload in the I.C.U. and hospital environment. This could include identifying a primary task of interest, which could be varied, for example a feature of the observation task could be the primary task or the different procedures involved in the drugs category. The method of measuring the mental workload involved would have to be considered with care. As Casali and Wierwille (1983) point out if tasks are intrusive when the primary task is being performed when an aircraft is in flight the results could be hazardous and this is also applicable to the I.C.U. environment.

As discussed previously, a starting point could be a simulated I.C.U. perhaps using an unused ward or side room in a hospital or I.C.U. If this could be achieved then staff who work in environments such as the I.C.U. could participate in studies with ease. Future studies could include evaluating performance on various tasks during different shifts and also at the beginning of a shift and comparing performance at the end of a shift.

Workload in the hospital environment at present appears to be a pseudonym for 'manpower requirements' with little, if any, mention of the mental effort involved in tasks. The tasks are very patient oriented with little regard for aspects such as whether there are increased cognitive demands as a result of the increases in technology on the I.C.U. Merely observing an individual perform a task cannot identify the attentional demands imposed by the task being undertaken. Workload literature regarding the hospital environment takes little account of the vast amount of research that has been conducted in other environments where there have been similar increases in technology. Some of the findings are not applicable to I.C.U. but other suggestions may improve the working environment for
the staff. For example, Carswell and Wickens (1987) found that graphical representation of information resulted in improved performance as opposed to numerical presentation of data. In the I.C.U. the majority of information is displayed numerically on medical equipment.

6.7 Conclusion

In conclusion, it is well established that auditory warnings in environments such as I.C.U. are too numerous and many staff cannot identify warnings when they are heard out of context (Momtahan et al. 1993). Many studies have examined how humans process information and factors that contribute to effective and efficient processing of that information and the implications are important for the design of equipment for use in I.C.U. and other multi-task environments.

The research described in this study has elicited that untrained participants can learn a number of arbitrary sounds and retain the information for a period of over one week. Some of the sounds were demonstrated to be easier for participants to learn than other sounds. Certain features of sounds used in the current study were shown to cause consistent confusions throughout the experiments. For example, a continuous tone (sound 3) was correctly identified and in one experiment produced the highest number of correct responses. However, it was also consistently confused with another continuous tone (sound 12). Many confusions were asymmetrical suggesting that aspects of a certain sound were more salient to participants than features in a similar sound.

The detailed observation study elicited a variety of tasks that were undertaken when the auditory warnings were activated on the I.C.U. Aspects of these tasks were then used to developed a set of tasks, using the multiple resources model as a framework, to use in the return phases of the final experiment.
As the dual task literature predicted when two or more tasks were undertaken concurrently there is a decrement in performance on the secondary task. Other studies have suggested that depending on the type of task, performance on both tasks can be as good as when each task is performed alone. However, that was not demonstrated in the current study.

The majority of the literature concerning mental workload in aviation describes the importance of overload on the operator and the consequences if this should occur (e.g. Casali and Weirwille 1983). This concern is also important for the I.C.U. environment. Understanding workload in an environment such as the I.C.U. would be beneficial from the perspective of efficiency and safety regarding the care of patients and also in respect to the well-being of staff. However, as Jex (1988) states operator overload is difficult to predict as there are many variables to consider such as motivation, training, skill and fatigue that greatly complicate the picture. Measuring mental workload is complex and fraught with methodological problems. For example, as Markou (1991) argues workload prediction and measurement techniques are often validated by correlating them with each other. A high correlation indicates that the measures are measuring the same thing it does necessarily mean that what they are measuring is workload.
REFERENCES


Barker, R.G. and Wright, H.E. (1955) *Midwest and it's Children* Evanston Ill Row Peterson


Dennis, I. Department of Psychology, University of Plymouth. Personal Communication


Fields, W.L. and Loveridge, C. (1988) Critical Thinking and Fatigue How Do Nurses on 8 and 12 Hour Shifts Compare? Nursing Economics. 6, pp 189-191


Frame, R.E. (1979) Interobserver Agreement as a Function of the Number of Behaviours Recorded Simultaneously. The Psychological Record. 29, pp 287-296.


Green, C.A. Personal Communication


365


Kerr, J.H. and Hayes, B. An Alarming Situation In The Intensive Therapy Unit. Intensive Therapy Medicine 9, pp 103-104.


367


372


Stanton, N. (1992) Alarms in a Coronary Care Unit Health and Safty Unit and Applied Psychology; Aston University, Birmingham.


376


