APPROPRIATE AUTOMATION OF RAIL SIGNALLING SYSTEMS:

A HUMAN FACTORS STUDY

Nora Balfe

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy

July 2010

ABSTRACT

This thesis examines the effect of automation in the rail signalling environment. The level of automation in a system can be described as ranging along a continuum from manual control to fully autonomous automation and development of appropriate automation for a system is likely to enhance overall system performance. Network Rail, the company which owns, operates, and maintains the rail infrastructure in the UK, envisions increasing levels of automation in future rail systems, but prior to this research, little structured evaluation of current automation had been undertaken.

The research performed for this thesis set out to examine the impact of automation on rail signalling. A rail automation model was developed to illustrate the levels of automation present in different generations of signalling system. The research focussed on one system in particular, the Automatic Routing System (ARS). The ARS has been present in modern signalling centres since the late 1980s. It uses timetable information to set routes for trains arriving on its area of control and incorporates complex algorithms to resolve conflicts between trains.

Multiple methods were used to investigate current signalling automation. An understanding of the signalling domain underpinned the research, and a model was developed to illustrate the type and level of automation present in different generations of current signalling systems. Structured observations were employed to investigate differences in activity between individual signallers. As a part of this study, a relationship was found between observed intervention levels and some of the trust dimensions identified from the literature. A video archive analysis gave initial insight into some of the issues signallers had with automation, and semi-structured interviews carried out with signallers at their workstations built on these themes. The interviews investigated four areas; signallers' opinions of ARS, system performance issues, knowledge of ARS, and interaction with ARS. Data were gathered on a wide variety of individual issues, for example on different monitoring strategies employed, interaction preferences, signallers' understanding of the system and their ability to predict it. Data on specific issues with ARS also emerged from the interviews, for example the impact of poor programming and planning data, and the poor competence of the system, particularly during disruption. An experiment was performed to investigate the differences between different levels of automation under both normal and disrupted running. The experiment gathered quantitative data on the effect of different levels of automation on workload and performance in addition to eye tracking data which were used to gain insight into signaller monitoring strategies. The results indicate that ARS does reduce workload and increase performance, and it does so in spite of deficiencies in terms of feedback to the signaller. This lack of feedback makes it difficult for the signaller to understand and predict the automation and, hence, creates difficulties for the operator. In addition, the methods for controlling ARS are limited and it can be difficult for the signallers to work cooperatively with the system.

Principles of good automation were identified from the literature and recommendations based on these and the findings of the research were developed for future signalling automation systems. These highlighted the importance of improving feedback from ARS and the ability of the signaller to direct the system. It is anticipated that these improvements would allow the signaller and the automation to work more closely together in order to maximise overall system performance. The principles of automation are intended as a generic guidance tool and their application is not confined to rail signalling. There may also be wider implications from the research such as the influence of operators’ ability to understand and predict automation in automation use, and the existence of different types of monitoring behaviour.

ACKNOWLEDGEMENTS

My thanks first go to John Wilson and Sarah Sharples for their excellent supervision and support throughout the research, and for complementing each other in the advice they gave me throughout. I would also like to thank Theresa Clarke for her guidance and support, and for the funding!

Thanks to all those who assisted in the research, in particular Bob Muffett and John Robinson, without whom signalling would still remain a confusing tangle of symbols, lines, colours and rules. In addition, thanks to all the participants in the studies, particularly the signallers who gave up their time to explain signalling and answer my questions. Thanks also for the endless cups of tea offered, which unfortunately I don't drink.

Many thanks to Team Ergo, both for the professional and not so professional support offered. In particular the other students whose willingness to discuss their experiences and collective trauma kept me motivated and sane – Alex, Tom, Pedro and Yassi. Also those in the noisy bay at Network Rail: Fiona, Justin, and Philippa. Thanks for the lunch and chocolate.

Thanks to my family and friends for helping me have a life outside my PhD. Fiona deserves special mention for being kind enough to listen to my complaints and still find time to proof read for me. And finally, my thanks to Simon, particularly for letting me bore the pants off him with railway talk when no-one else would listen.

LIST OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

GLOSSARY

AB	Absolute Block – a method of signalling control
ARS	Automatic Route Setting
CBI	Computer Based Interlocking – a signalling safety system
CCF	Control Centre of the Future – a system giving information on train delays
CCTV	Closed Circuit Television
CSR	Cab Secure Radio – a radio system between signallers and train drivers
CTRL	Channel Tunnel Rail Link – railway linking the UK to Europe
ECRO	Electrical Control Room Operator – a job role responsible for controlling electrical power to the railway
DOS	Disk operating system
EPSRC	Engineering and Physical Sciences Research Council
FCFS	First Come First Serve – a method of signalling control
FOC	Freight Operating Company
GP Screen	General Purpose screen – an ARS display giving information on alarms and train paths
ICT	Information and Communication Technology
IECC	Integrated Electronic Control Centre – a type of signalling system
IWS	Integrated Workload Scale – a signalling workload measurement tool
LOA	Level of Automation
LOM	Local Operations Manager – a job role responsible for managing signal boxes
MOM	Mobile Operations Manager – a job role responsible for trackside management of incidents on the railway
NX	Entry-Exit Panel – a type of signalling system
OCS	One Common Switch – a type of signalling system
ODEC	Operational Demand Evaluation Checklist – a signalling workload measurement tool
OLE	Overhead Line Equipment – a method of delivering electrical power to trains
ORR	Office of Rail Regulation
PICOP	Person In Charge Of Possession – a job role responsible for managing engineering and maintenance work on the railway
PPM	Public Performance Measure – a railway performance measure
RRI	Route Relay Interlocking – a signalling safety system
RSSB	Railway Safety and Standards Board
SA	Situation Awareness
SAGAT	Situation Awareness Global Assessment Technique
SART	Situation Awareness Rating Technique
SD	Standard Deviation
SME	Subject Matter Expert
SPAD	Signal Passed At Danger
SPAM	Situation Present Assessment Measure
SPT	Signal Post Telephones
SSI	Solid State Interlocking – a signalling safety system
STP	Special Timing Pattern – a control mechanism within ARS
SWAT	Subjective Workload Assessment Technique
T12	Form of protection for track workers
T2	Form of protection for track workers
T3	Possession of the line by track workers
TC	Track Circuit
TCB	Track Circuit Block – a method of signalling control
TD	Train Describer – unique identifier for trains
TOC	Train Operating Company
TOPS	Total Operating Processing System – a system giving information on train movements
TORR	Train Operated Route Release – a signalling support system
TRB	Train Record Book – a manual recording system

TRUST	Train Running System on TOPS – a system giving information on train movements and delays
TSDB	Timetable Services Database – a system holding the railway timetable
TTO	Timetable Order - a method of signalling control
TTP	Timetable Processor – a method of changing train timetables in ARS
UK	United Kingdom
VDU	Visual Display Unit
WON	Weekly Operating Notice – a publication giving details of planned railway operations

CHAPTER 1: INTRODUCTION

1.1 Chapter Overview

This chapter introduces the research completed for this thesis and the context in which it was conducted; that is, rail signalling. The background to the work is described, the aims and objectives and the overall research approach are all outlined, and an overview of the thesis structure is given. In addition, the domain of rail signalling is briefly introduced and the specific system under investigation, the Automatic Routing System (ARS), is described.

1.2 Background

This research examined automation in rail signalling systems. The most frequently cited definition of automation is when a machine (usually a computer) assumes a task that is otherwise usually performed by a human operator (Parasuraman & Riley, 1997). It may be introduced into a system for a variety of reasons, but frequently cited reasons for use of automation include achieving tasks more efficiently and reliably than human operators. Cited benefits include reduction of operator workload and error coupled with a reduction in labour costs. These benefits make automation very attractive to businesses wishing to increase efficiency while reducing costs. However, automation can lack the flexibility of human operators in the face of novel situations and thus difficulties can be encountered when the designers attempt to replace human problem solving abilities with automation. Thus, humans are likely to remain vital to system performance for many years (Parasuraman & Wickens, 2008).

Automation has been present in rail signalling systems for many years. At a basic level the interlocking systems which ensure that signallers do not set conflicting routes for trains can be regarded as an early form of automated decision support. These have been in place since the 1800s. However, this work primarily concerns the most advanced form of automation currently in use on the UK network, namely, Automatic Route Setting, or ARS.

Automatic Route Setting was first introduced circa 1989 and is now used in 11 signalling centres across the UK. Fundamentally, it works by using the programmed timetable for the train service to set the appropriate routes for trains arriving in its area of control. This successfully eliminates the mundane work of route setting for the signaller. However, ARS also attempts to deal with conflicts between trains; that is, when two trains arrive at a section of railway at the same time; the most common reason for this would be late running of one or more of the trains. To make effective decisions on which train to route first requires expert knowledge of both static and dynamic properties of a given situation. Static properties would include the train service pattern and the infrastructure layout. Dynamic properties could include the relative delay and speeds of the trains involved and the state of the infrastructure at that time. Although ARS attempts to account for these variables it is not always successful and the signallers then step in to take over. Despite the presence of ARS for almost 20 years, little formal research has been conducted on the effectiveness of the system and the interactions between the human operator and the automated system. Without human factors research it is difficult to improve upon the current system. The investigation of how signallers go about working with ARS under these kinds of circumstances forms the basis of this research.

Network Rail is currently involved in a major programme of work to upgrade the UK rail network and in the future the company has plans to centralise its control facilities. At present the intention is to go from approximately 800 signal boxes to 15, with a consequent reduction in the workforce. Automation is expected to play an integral part in these plans. In order to facilitate the design of the new automated systems which will support the centralised control it is important to understand the use of and issues with the current system. This future use of automation has been an important driver for this work and an outcome of the research will be recommendations for new systems drawn from the research conducted on the present system.

1.3 Aims and Objectives

The aim of this research is to study a real world automated system (ARS) in order to understand the impact of automation on the human operators who work alongside it and identify how automation can be implemented to best support overall system performance.

Three objectives have been defined to support this aim:

1. To develop a theoretical framework within which to research and implement human centred automation in rail signalling.
2. To study current use of automation within rail signalling and understand the effect automation has on the signalling task, including:
    a. How signallers monitor the system;
    b. How signallers interact with the system;
    c. Signallers' understanding of how the system works;
    d. Overall system performance.
3. To develop recommendations for development and implementation of automation in future rail signalling systems.

These objectives were addressed through the development of a conceptual framework and a research framework. The conceptual framework describes how the concepts investigated in this research relate and how they contributed to the development of recommendations, while the research framework illustrates the methods used throughout the research.

1.4 Research Approach

Figure 1-1 illustrates the conceptual approach taken to address the objectives. Theory and knowledge were identified in the literature and provided a basis to support the investigation of the design and use of automation. The review of the literature also identified best practice which supported the development of guidance. The research undertaken examined the impact of the design of automation on the operator, in terms of trust and workload and the performance and behaviour of operators using automation, with key themes of monitoring and situation awareness.

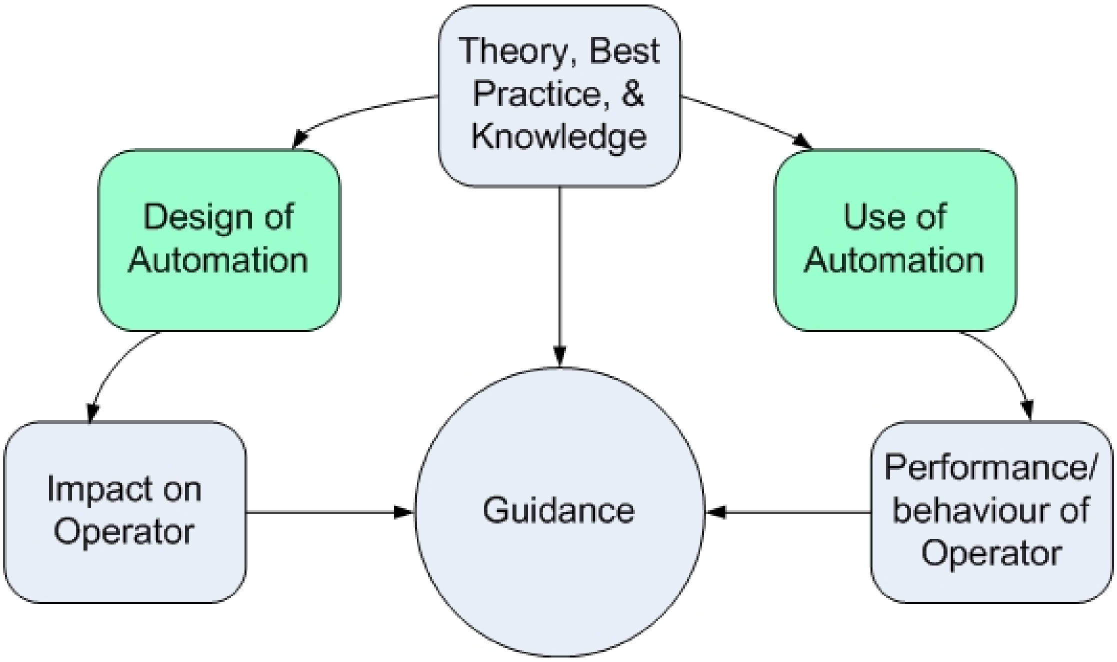


Figure 1-1: Conceptual Framework

This research was jointly funded by Network Rail and the Engineering and Physical Sciences Research Council (EPSRC). As such, the researcher was given the opportunity to work within Network Rail for the duration of the research and spent three years working full time in the company's headquarters in London. During this time she was fully integrated within Network Rail's Ergonomics National Specialist Team. This immersion provided many opportunities in terms of participation in meetings and projects aligned with the research area as well as facilitating direct access to people and environments for this research. Working from within Network Rail facilitated access to personnel and work sites which otherwise would have been very difficult to arrange. This allowed the development of a comprehensive understanding of the work tasks and environment associated with rail signalling.

Three research approaches underpinned the work; real world research, mixed method research and grounded theory. Figure 1-2 illustrates the research approach taken. Real world research is that in which the problem being investigated is set in the real world in contrast to a more controlled laboratory setting (Robson, 2002). This research was undertaken in the rail environment examining a real world automated system and the people who operate it. The complexity, messiness, and inability to control the environment associated with the real world approach were apparent throughout the research. Hence, the flexible design associated with the real world approach was also utilised, and this incorporated a mixed method approach allowing the use of multiple methods to investigate problems (Hignett & Wilson, 2004).

Hypotheses were not generated at the beginning of the research, but were formed by an iterative process throughout the course of research in the manner of Grounded Theory (Pidgeon, 1996).

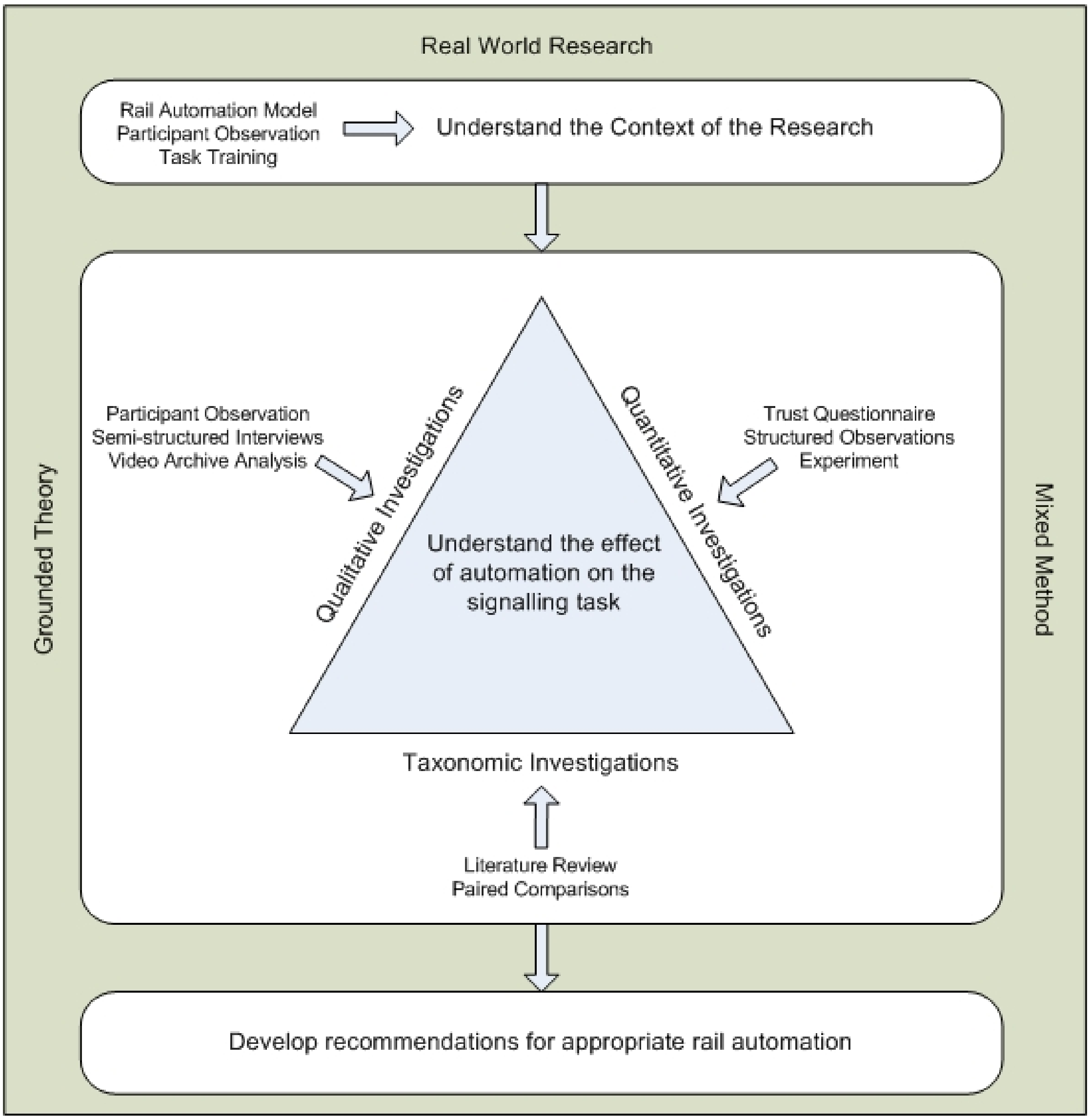


Figure 1-2: Research Framework

The starting point was to understand the context of the research, specifically to develop a firm appreciation of the nature of the signalling task. This is known as 'boot-strapping' in the cognitive task analysis domain (Chipman, Schraagen, & Shalin, 2000). Without this it would have proved impossible to plan studies to investigate aspects of the signalling task potentially affected by automation and to analyse the data gathered in these studies. Knowledge of signalling was achieved initially through participant observation, including direct field observations of signallers in a variety of signal boxes. "Field observations support a discovery process, they serve to draw attention to significant phenomena and suggest new ideas, the validity and generality of which can then be evaluated through additional studies” (Vicente et al., 2001. p. 835). In total, 89 signal box visits were undertaken throughout the course of the research (Appendix A) and these facilitated a general understanding of the whole system and where and how automation fits within it to develop over time in the mind of the researcher. Placement within Network Rail also made possible the use of the participant observation approach (Hammersley & Atkinson, 1995). All members of the team and other colleagues encountered in the organisation were aware of the researcher’s role. Although not classic and fully structured participant observation much of the direction of the research and interpretation of findings has been influenced by situations and conversations encountered on a daily basis, whether within the Ergonomics team, engineering teams, operations teams, or front line signalling sites. However, much of this was informal and is presented within the thesis as a supporting method which allowed greater confidence in the validity of the data gathered using more formal methods and to better interpret the findings. Attendance at signalling school also provided some specific task training which consolidated knowledge. The work undertaken to understand the context of the research was drawn together in the development of a rail automation model which describes the automation present in three generations of signalling systems.

Once a preliminary understanding of the research context was established a three prong approach was taken to the research. The qualitative investigations into ARS included semi-structured interviews, questionnaire data, a video archive analysis and the overall participant observation approach which provided insight and allowed deeper interpretation of the findings from the other methods used. Structured observations, questionnaire data and an experiment were the methods which contributed quantitative data to the research. All the research methods used are summarised in Appendix B. The taxonomic investigation into good automation was achieved through reviews of the literature the principles of automation were generated from this and validated using the paired comparisons technique (Sinclair, 2005). Each of the individual methods used in these investigations will be discussed in detail in the relevant chapter. The final goal was to develop recommendations for appropriate automation and this was achieved by drawing together the findings from all the research undertaken.

1.5 Signalling Environment

The research was conducted entirely within the domain of rail signalling and it is important to understand this context. This section gives a brief overview of signalling and ARS. Railways require signalling systems to operate safety and efficiently. The primary aim of signalling is to ensure separation between trains but signalling systems also control the points movements required to set routes for trains (i.e. they are also responsible for ensuring trains get to their destination). Separation between trains is most often ensured by allowing only one train into a section at a time. Entrance to a section is controlled by signals and the presence of a train in that section is detected either manually or through sensors such as track circuits or axle counters. There are three main forms of signalling system in operation on the UK railway today (Figure 1-3); lever frames are the oldest, dating from the 1800s, and use levers attached to signals and points to set routes for trains. Entry Exit (NX) panels were introduced in the 1950s and allowed the signaller to operate the points and signals through button presses, with the physical labour of moving them being undertaken automatically. The most recent form of signalling system is the Integrated Electronic Control Centre (IECC), developed in the 1980s. This operates like an NX panel, but uses Visual Display Units (VDU) and incorporates an automated route setting system (ARS). The main focus of this research is ARS, but comparisons will be drawn with these other generations of signalling system.

The signaller’s job becomes more complex when he/she is faced with competing demands for track occupancy, for example, two trains want to travel over the same piece of infrastructure at the same time. This situation is known as a conflict, and the signaller must ‘regulate’ the train service to resolve it. Discussions with rail operations subject matter experts (SME) determined that regulation may be defined as:

> The planning and implementation of train paths over the available infrastructure in order to optimise the train service, mitigate the effects of disruption, and support recovery from disruption.

Signallers working on lever frames or NX panels are responsible for making these decisions, but in IECC signal boxes ARS is capable of making decisions between trains.

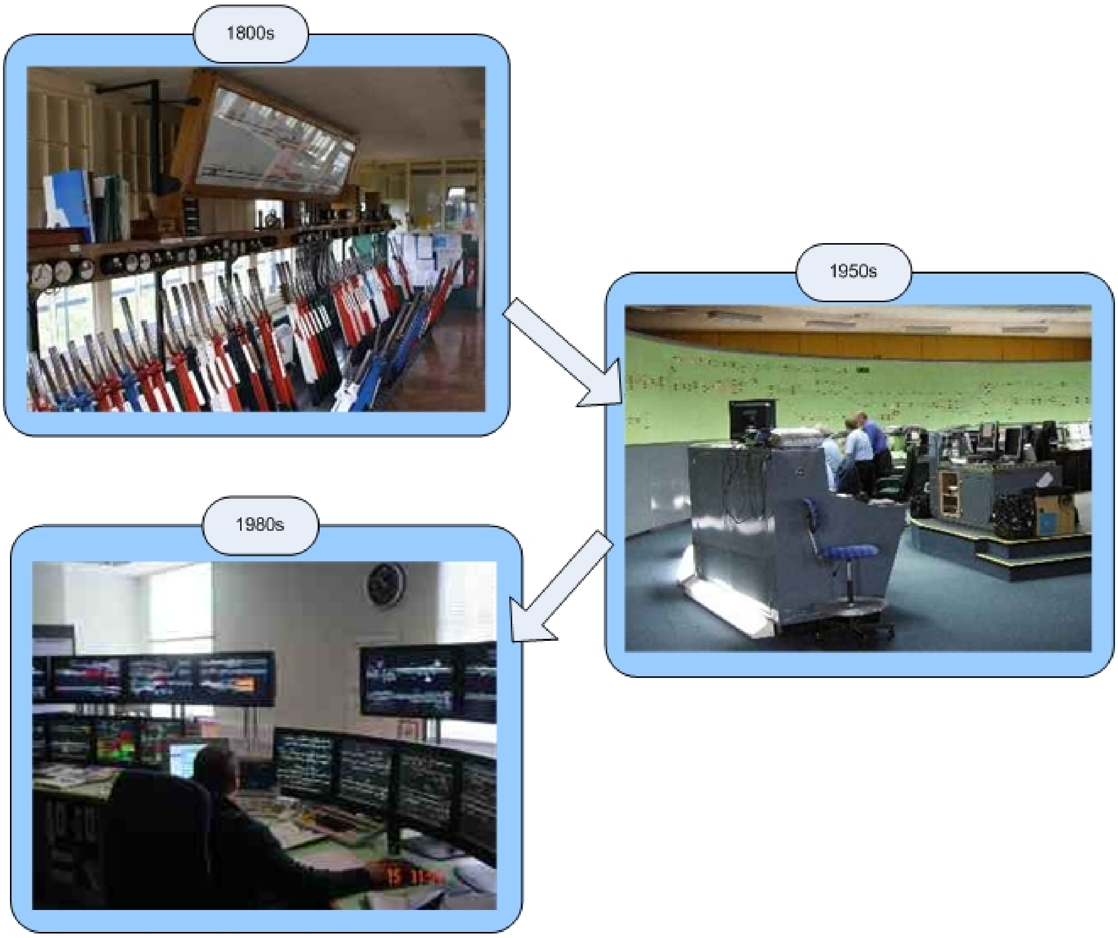


Figure 1-3: Examples of Signalling Systems

Automatic Route Setting (ARS) has been in place since the late 1980s, and was first introduced in Liverpool Street IECC. DeltaRail, who now develop ARS, stated that (DeltaRail, 2008):

> "ARS optimally routes trains using timetable data, current train position and an internal representation of the rail network. It can handle severely disrupted service patterns and assist the signaller in the event of train or infrastructure failures."

ARS has access to the central timetable services database (TSDB) and each day downloads the timetable for all the trains in the area it controls. It then uses codes from the timetable to determine the route and timings for each train. As each train enters the control area ARS automatically sets the route ahead of the train. Algorithms are also incorporated in ARS to compare trains on the workstation to decide which to route first. Less advanced forms of route setting automation could either route strictly according to the timetable or operate on a first come first serve basis, but ARS attempts to regulate the train service to attain optimal performance.

It does this by holding a list of trains currently in the area and the routes they require. All these trains are compared to identify which trains potentially conflict. There are three ways a train may conflict; they may travel over the same section of track in the same direction, travel over the same section of track in opposite directions, or travel over lines which cross. If two conflicting trains require the same route at the same time, ARS uses a set of parameters such as train priority, current train delay, and predictive forward movements from the timetable to determine the weighted delay for each train in this situation. It then sets a route for the train which it calculates causes the least delay.

The signaller has no insight into this process. The signalling screens only display when routes have been set by ARS and although there is an ability to query ARS through the general purpose (GP) screen, this information is not always informative to the signaller, particularly if they do not fully understand the processes ARS uses to make its decisions. It is impossible for ARS to give information on what it is planning to do as it does not make decisions in advance but immediately implements decisions made.

Signallers can use reminder devices to constrain ARS. Reminder devices were traditionally used in mechanical systems to physically prevent signallers pulling a lever to set a route, usually because of some form of blockage on that route. In IECC systems, reminder devices placed over signals prevent ARS setting a route to or from that signal. The reminder also serves the traditional function of reminding the signaller not to set that route. Although intended as a safety device, reminders are frequently used by IECC signallers to control ARS as it is a direct and easy way to inhibit route setting.

Not all trains are in ARS; this is most likely to be because there is no timetable or an incomplete timetable for them in the database. Trains which are not in ARS are shown in pink and must be controlled manually. Signallers may also choose to take trains out of ARS. This allows signallers to maintain control over that train as it must be routed manually, although it can be put back in to ARS if the signaller wishes.

Further information on rail signalling, including descriptions of roles and systems and more detail on ARS processes, can be found in Appendix C. The reader may find it useful to refer to this section for background information.

1.6 Thesis Structure

The remainder of this thesis is presented in seven chapters:

- Chapter 2 – review of the human factors literature with respect to automation. This chapter covers the benefits and issues typically encountered with automation. Key human factors themes are discussed and principles of automation are drawn from the literature.
- Chapter 3: Rail Automation Model – This chapter describes the development of a rail automation model which illustrates the variation in levels of automation in different generations of UK signalling systems.
- Chapter 4: Structured Observations of IECC Signallers. This chapter presents the method, results and discussion of observation studies carried out in four IECC signal boxes.
- Chapter 5: Signaller Interviews. This chapter presents the method, results and discussion of semi-structured interviews with signallers. The results from analysis of pre-existing videos of interviews with signallers in IECCs are also presented and discussed in this chapter.
- Chapter 6: Level of Automation Experiment. This chapter presents the method, results and discussion of a simulator experiment designed to examine the differences in workload, performance, monitoring, and signaller activity between three different levels of automation.
- Chapter 7: General Discussion. The ARS system is discussed in the light of the findings and the methods employed are evaluated. Recommendations for future automated signalling systems are given.
- Chapter 8: Conclusions. The work is concluded in Chapter 8. The recommendations are summarised and the impact of this research is stated. Recommendations for future research are also outlined.

1.7 Chapter Summary

This chapter has introduced the domain of rail signalling and described the system under investigation (ARS). The aims and objectives of the research were outlined and the approach to the research described. Finally, the structure of the thesis was outlined. The next chapter will introduce automation and present the taxonomic investigations, discussing the relevant human factors literature and producing 12 principles of automation.

CHAPTER 2: AUTOMATION AND HUMAN FACTORS

2.1 Chapter Overview

This chapter introduces automation of control systems and gives some background on the benefits and issues that have been associated with the introduction of automation. The existing research on key human factors themes with regard to automation is discussed, including trust, situation awareness (SA), workload, monitoring, and mental models. The different approaches to the design of automated systems are discussed and, finally, 12 principles of automation are drawn from the literature.

2.2 Introduction

The first stage in this research was a taxonomic investigation of automation. This was achieved through a review of the relevant literature and the definition of 12 principles of automation from that review. These principles were validated subsequently through presentation at conferences and a paired comparisons exercise with human factors professionals (Appendix D). The position of this research in the research framework is illustrated in Figure 2-1.

The research in the field of human factors of automated systems is presented in this chapter. An overview of automation is first provided, with discussion of the benefits and issues with introducing automation into a system. Frameworks are then presented to describe how the level of automation may vary within a system. Such frameworks are often used to investigate the impact of automation on human factors concerns such as trust, SA, and workload (e.g. Endsley & Kiris, 1995; Kaber, Perry, Segall, McClernon, & Prinzel, 2006). These concerns are discussed in the third section of this chapter in addition to the impact of automation on operator monitoring. Approaches to the design of automated systems are presented before the existing research is drawn together in the form of the 12 principles of automation.

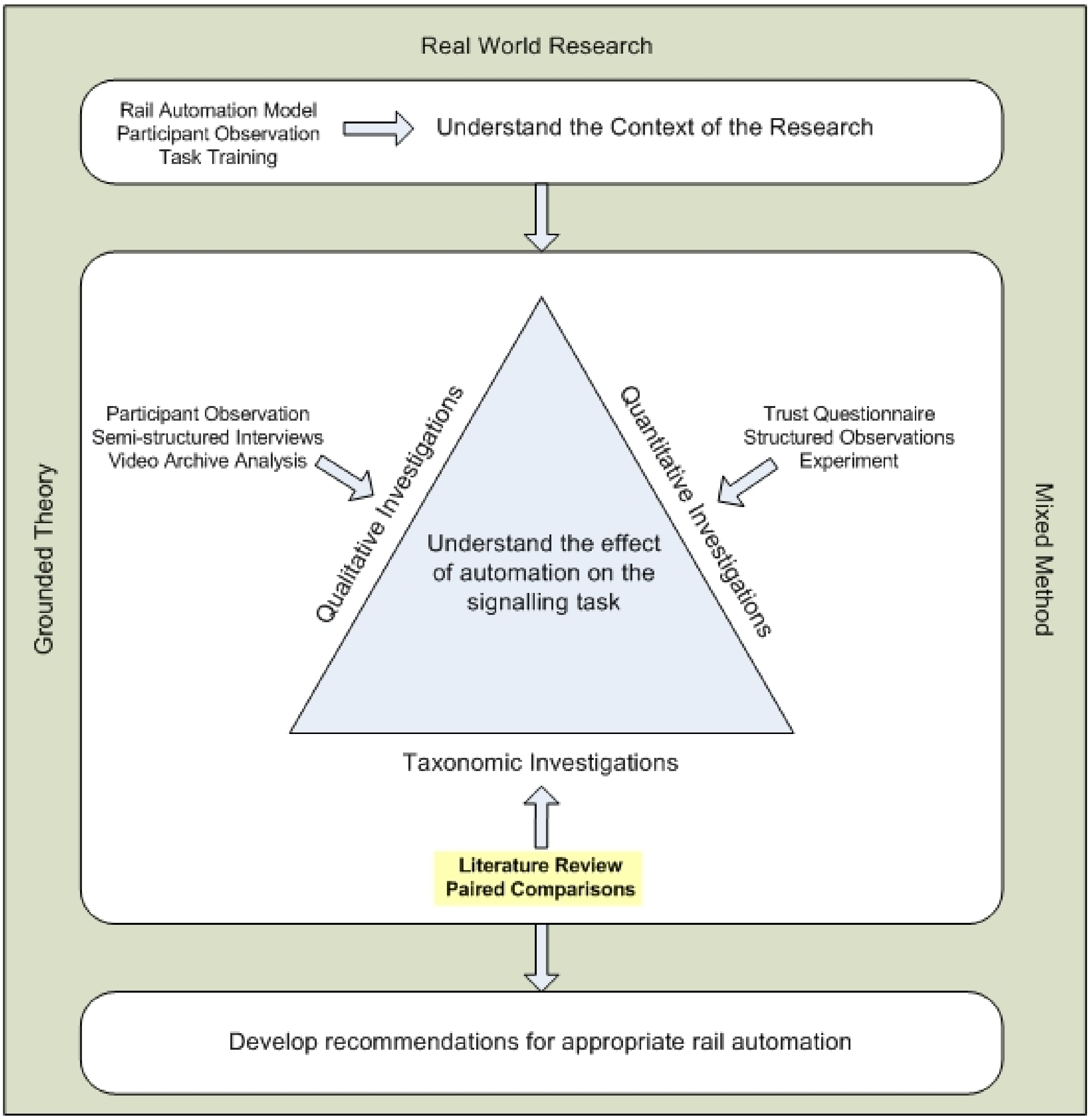


Figure 2-1: Position of the Literature Review in the Research Framework

2.3 Overview of Automation

Automation is developed and introduced to replace tasks previously performed by human operators (Parasuraman & Riley, 1997). However, even highly automated systems need humans for supervision, adjustment, maintenance, expansion and improvement and Parasuraman and Wickens (2008) suggest that humans are likely to remain vital to system performance for some time. Automation has many strengths, including precision, speed, lack of emotion and distraction but it also has weaknesses, not least that it lacks the flexibility which humans possess that allows us to adapt to novel or unexpected situations.

Technology is the driving force behind automation, and as technology continues to increase in power and reduce in size and cost, it is likely to drive automation even further (Wiener & Curry, 1980). The potential for automation to operate systems more economically has also added to the lure of automation. This push of technology coupled with the pull of potential efficiency gains has made automation of complex systems increasingly common. However, wholly automated systems (i.e. systems with no human operator) in complex industries are rare and are generally confined to closed systems, such as industrial manufacturing. Most automated systems have at least one human operator to monitor their performance. The interaction between the human and automation creates a number of human factors issues. This chapter discusses the concept of automation, the benefits and issues, and the associated human factors issues.

There are a number of perceived benefits of automation; these include a reduction in human error, a saving on labour costs and a reduction in human workload (Bainbridge, 1983; Dekker, 2004; Hollnagel, 2001). Automation certainly contains the potential to bring about a reduction in human error, labour costs and workload but these benefits are not always realised when an automated system is introduced. Human error may be reduced in the task performance; however, machines are manufactured, programmed and maintained by humans and an error may occur at any one of these stages which does not become manifest until the operational stage (Wiener & Curry, 1980). Thus, human error may still occur in highly automated systems. In addition, these errors may be hidden and lie unknown within the system and so they have the potential for severe consequences (Wickens, 1992). It may be more accurate to say that automation can reduce the human variability associated with task performance, rather than human error.

In respect of labour costs, it must be noted that automation does not usually replace the human operator in totality; usually a new role is created for a human supervisor or operator. In addition, there are new job roles associated with the design, manufacture, programming and maintenance of the automated machine and these roles may be more skilled, higher paid jobs than those the automation is designed to replace. Hence, the saving in labour costs may not be as high as is sometimes perceived. Well designed automation may lead to a reduced workload; however, it is often the case that while a reduction in physical workload is achieved, there is a potential increase in mental workload for the operator. Automation may also lead to peaks and troughs in workload (Woods, 1996) if it reduces workload during periods when workload was already low but becomes a burden during higher workload phases.

Introduction of automation is often based more on these perceived benefits rather than whether it is appropriate. Wickens (1992) listed three circumstances where it is appropriate to introduce automation; automation which is employed to perform a function that is beyond the capabilities of a human operator, for example performing complex calculations at high speed or highly precise measurement; automation which performs functions at which human operators are poor, for example monitoring a system for a single failure event; and automation which provides assistance to human performance, for example augmenting information on display systems.

Aside from the benefits there are also many problems in which automation of a system may result; Bainbridge (1983) highlighted a number of such problems in an important discussion on the ironies of automation. She suggested that the introduction of automation is often ironic as it replaces tasks humans perform reasonably well or easily and leaves the operator to perform tasks which were too difficult to automate. For example, operators may be required to take over from the automation under unusual or failure conditions. This is the time when the tasks are most difficult, but when automated support is often lowest. Other examples of issues with automated systems include low reliability automated systems or those with a tendency to err as this may induce low operator trust which results in low usage of the automation (Sheridan, 1999). Operators who are not actively involved in the control may suffer from out of loop unfamiliarity (Wickens, 1992) or loss of SA (Endsley, 1996), and this can become a major problem if they are required to take over from the automation, especially during emergency circumstances. Another problem may be the loss of skills on the part of the operator as they are no longer required to use them regularly (Bainbridge, 1983); again, this may be an issue during system failures when the operator is required to take over from the automation quickly and effectively.

Automation therefore may be implemented with varying degrees of success, and introducing it into a system requires careful analysis, planning, and testing to ensure maximum benefit is achieved. The level of automation in a system is a key factor in determining the benefits and issues which may arise from the introduction of that automation.

2.4 Levels of Automation

The level of automation employed in an automated system may vary along a continuum from no automation at all through to fully autonomous operation. The literature contains a number of models describing the level of automation (Billings, 1991; Endsley & Kiris, 1995; Parasuraman, Sheridan, & Wickens, 2000). These models all range from complete manual control through to autonomous automation but use different scales to describe the intermediate levels of automation. Manual operation is usually included in such models as a base level. Figure 2-2 shows a model developed by Billings.

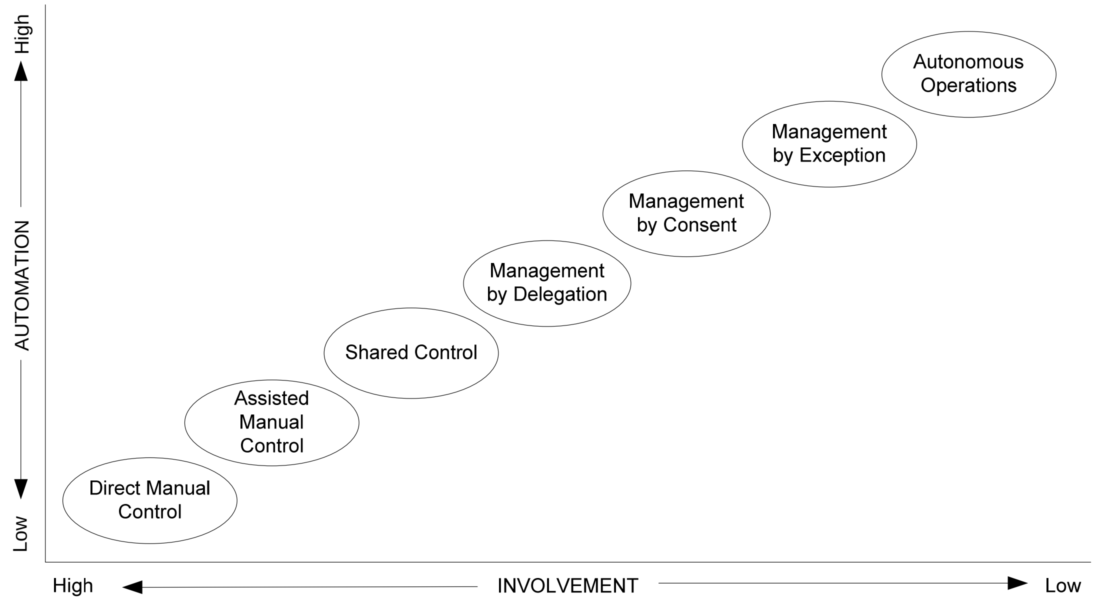


Figure 2-2: Levels of Automation (Billings, 1991).

During the lower three levels in this model the operator still has a degree of manual control (i.e. he/she is still responsible for physically completing tasks). As automation moves into the next three levels the operator takes a more managerial role, either instructing the automation to perform specific tasks (management by delegation), giving consent for automatically identified options to be executed (management by consent) or vetoing options chosen by the automation (management by exception). Autonomous operation has no operator involvement in normal operations. The highest levels, management by exception and autonomous operation, exclude the operator from the decision making process and are therefore undesirable as the operator cannot work cooperatively with the automation and is ill prepared to assume control if it fails (Hollnagel, 2001). Therefore, in the higher levels of automation, management by delegation or consent may be preferable because they keep the operator in the loop and this view is supported by research which found higher levels of performance for management by consent levels of automation as compared to higher and lower levels (Ruff, Narayanan, & Draper, 2002).

Endsley and Kiris (1995) developed a similar scale with only five points (Figure 2-3), although other 10 point scales have been developed by Endsley and Kaber (1999). This is very similar to the model developed by Billings (1991) with the lower levels of automation having the human operator make decisions and the higher levels moving the decision making increasingly towards the system.

		Roles	
Level of Automation		Human	System
None	1	Decide, Act	-----
Decision Support	2	Decide, Act	Suggest
Consensual AI	3	Concur	Decide, Act
Monitored AI	4	Veto	Decide, Act
Full Automation	5	-----	Decide, Act

Figure 2-3: Levels of Automation (Endsley & Kiris, 1995)

Although different levels of automation have been developed by different researchers, they all use the same approach of creating a scale on which to rank the level of automation (Moray, Inagaki, & Itoh, 2000). The scales vary in wording and graduation but most are complementary and employ one scale to describe the automation, often focussing on where the responsibility for decision making lies. The exception to this is a model for types and levels of automation developed by Parasuraman et al. (2000) and based on original work by Sheridan and Verplank (1978).

This model used a four stage model of human information processing to describe different levels of automation (Parasuraman et al., 2000). Human information processing describes the human perception and analysis of information to reach and implement decisions. Wickens and Carswell (1997) presented the model shown in Figure 2-4 as a typical representation of human information processing. Information is received through the senses and cognitively acted upon. A response is selected and executed and the effect is perceived and fed back through the model (Wickens, Gordon, & Liu, 1998).

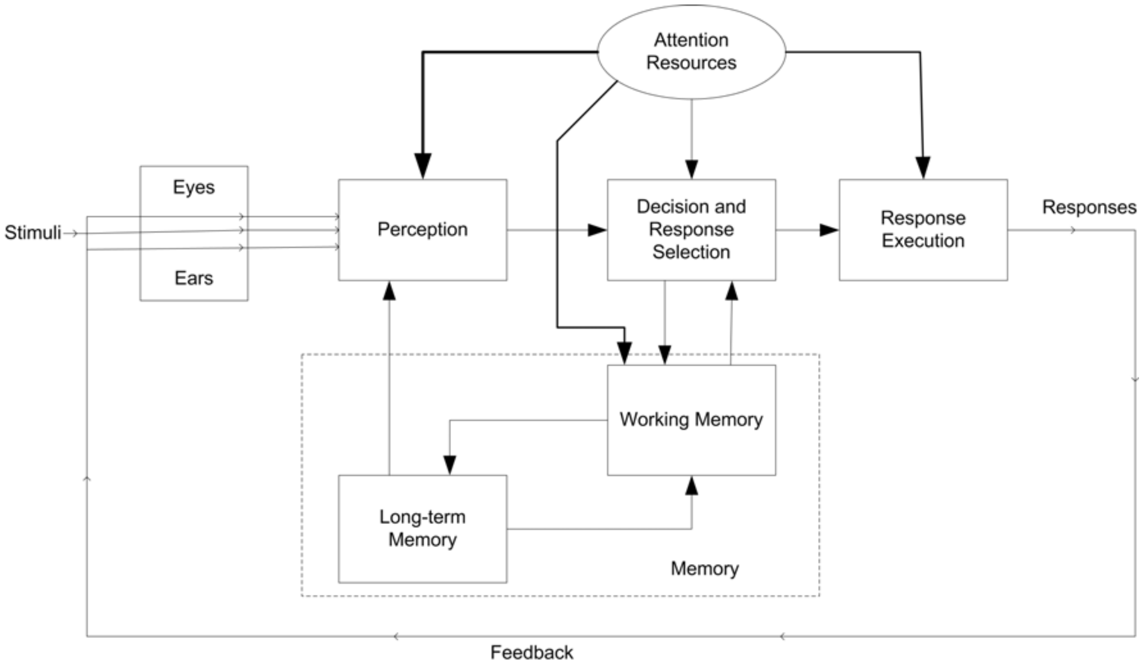


Figure 2-4: Human Information Processing (Wickens & Carswell, 1997, p. 91)

Parasuraman et al. (2000) simplified this model for their work on automation into the four stage model shown in Figure 2-5. Interestingly, human information processing models were developed using the metaphor of the digital computer (Wickens & Carswell, 1997), so by applying it to automation, Parasuraman et al. are reapplying this metaphor to its origins.

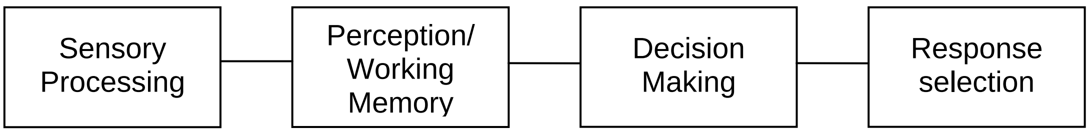


Figure 2-5: Human Information Processing (Parasuraman et al., 2000)

The four stages of the model are information acquisition (sensory processing), information analysis (perception/working memory), decision and action selection (decision making), and action implementation (response selection). Information acquisition refers to the sensing of data and includes positioning and orienting of sensory receptors, sensory processing, initial pre-processing of data prior to full perception, and selective attention. Information analysis involves conscious perception and manipulation of processed and retrieved information in working memory. Decision and action selection involves choosing from the decision alternatives. The final stage, action implementation, refers to the execution of the action choice and typically replaces the human hand or voice. Each of these functional dimensions is assigned a level of automation, for example on a scale of 1-8. A potential scale is shown in Figure 2-6 (Sheridan, 1998).

LOW	1.	The computer offers no assistance; the human must do it all
	2.	The computer suggests alternative ways to do the task
	3.	The computer selects one way to do the task, and
	4.	…executes that suggestion if the human approves, or
	5.	…allows the human a restricted time to veto before automatic execution, or
	6.	…executes automatically, then necessarily informs the human, or
	7.	…executes automatically, then informs the human only if asked
HIGH	8.	The computer selects, executes, and ignores the human.

Figure 2-6: Levels of Automation (Sheridan, 1998)

Once levels have been assigned a graph can then be produced for an automated system, showing how the automation varies through each stage (Sheridan, 1998). In Figure 2-7, three examples of different systems are illustrated. The circles represent a potential voting system for an organisation; the acquisition of information is manual, although email could be used and so this is assigned to Level 2. The results are analysed by the computer and the winner decided automatically. Power is transferred to the winner with the aid of the computer for passing on information. The black squares represent advice on a new air traffic control system; information acquisition and analysis are recommended to be highly automated (e.g. radar, weather information, etc.). Decision making is recommended to be manual and since implementation is in the hands of the pilots, autopilots are likely to be involved. The open squares represent a typical robotic manufacturing task. A computer system acquires all the data and performs the analysis, and the results are available for a human supervisor to check. The analysis results are passed on to a decision algorithm and the results of this are displayed to the operator. The decision is implemented by a robot in a fully automatic fashion (Sheridan).

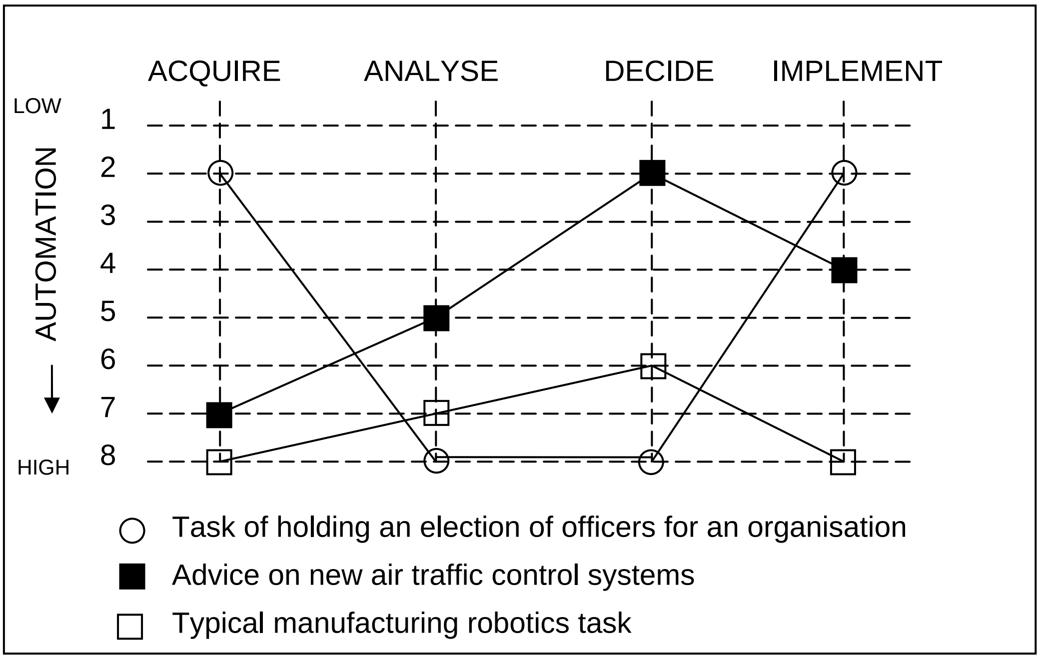


Figure 2-7: Graphical Representation of Levels of Automation (Sheridan, 1998)

The advantage of this model, as described in these examples, is the ability to differentiate between types of automation within a system. This means that individual automation systems can be modelled more accurately. For example, an automated system may have high levels of information acquisition and action implementation automation but leave the operator to analyse the information and make decisions. It is not possible to represent this situation on any of the other models found in the literature.

Classifying levels of automation using models such as those described provides a framework to support the design of automation (Parasuraman, Sheridan, & Wickens, 2008) but it also usefully provides a structure within which to research optimal levels of automation, and this approach is commonly found in empirical research on automation (Endsley & Kaber, 1999; Kaber et al., 2006). Some of the issues which have received specific attention in the research are discussed in the following sections.

2.5 Human Factors Concerns

The introduction of automation, and the level/design of that automation, creates new issues for the human operator. The impact of automation on four such issues is discussed in this section: trust, SA, workload and monitoring.

2.5.1 Trust in Automation

Trust has been identified as a potentially important construct by researchers who theorise that low levels of trust in automation may influence operators' usage (Muir, 1987; Sheridan, 1999). Muir (1994) stated that if it was not possible to build automated systems which are trustworthy, then we could not build automated systems at all. Several studies have found a correlation between trust levels and use of automated systems (de-Vries, Midden, & Bouwhuis, 2003; Lewandowsky, Mundy, & Tan, 2000; Moray et al., 2000; Muir & Moray, 1989). Operators only use automation to the extent that they trust it; if operators distrust automation they will reject it, preferring to perform the task manually.

There is much discussion in the literature regarding a definition of trust, but the work of both Barber (1983) and Rempel, Holmes, and Zanna (1985) form the most common basis for a definition. These were developed to represent interpersonal trust but have been commonly used to define human-automation trust (Madhavan & Wiegmann, 2007). Both are three stage definitions, and can be regarded as overlapping somewhat. The first stage involves the creation of an accurate mental model which allows the operator to understand and predict the behaviour of the system (Muir, 1987). This implies that trust is dependent upon understanding of the system (Lee, 1991). The second stage concerns the ability of the system to correctly perform its tasks and can be regarded as the most important for human-automation trust (Muir). This might also be called reliability or competence and refers to the performance of the system (Lee). Barber identified three types of technical competence that one human might expect from another: expert knowledge, technical facility, and everyday routine performance. These three factors roughly correspond to Rasmussen's (1983) taxonomy of knowledge, rule, and skill based behaviour. Automation may be capable of carrying out only one of these three factors but still be able to perform its individual task satisfactorily. The final dimension can be labelled faith, and becomes important when the automation is more competent than the human operator. The operator is therefore unable to evaluate the automation and must rely on an assessment of the automation's responsibility. These three stages may be sequential. Operators initially trust a system if they find it to be predictable; once predictability has been established, they find the system to be dependable. Faith requires belief in the referent beyond that for which there is direct evidence. Faith in automation may be based upon the evidence gathered during the predictability and dependability stages but also upon belief that the machine can cope with certain events, even though those exact events may not have been encountered before (Rempel et al.).

Both the above definitions suggest that trust is a multi-dimensional concept and there are many factors which can influence an operator's trust in an automated system. Research, much of it using a pasteurisation plant simulation, has consistently shown that automation reliability is closely related to operator trust (Lee & Moray, 1994; Muir & Moray, 1989; Wiegmann, Rich, & Zhang, 2001). In fact there are two facets to reliability, an automated system may be reliable in the sense that it does not suffer mechanical failure, but it must also be reliable in the sense of making correct decisions consistently or performing its function well. This second facet can be labelled 'competence' for clarity (Madsen & Gregor, 2000; Muir & Moray, 1996; Parasuraman et al., 2000). System competence has been found to be the greatest predictor of the operator's overall trust (Muir & Moray, 1989) and operator trust may be affected differently by different levels of system incompetence. Small errors, even those which do not affect performance, may greatly reduce trust while operators have been found to become increasingly less sensitive to larger errors (Lee & Moray, 1992; Muir & Moray). Automation must therefore be extremely reliable if high levels of trust and usage are to be achieved.

Such research highlights the importance of highly reliable and competent automation; however, operators may perceive even unreliable automation to be better than manual operation. Riley (1996) suggested that operators' trust in, or decision to rely on, automation is strongly influenced by the operators' self confidence. If an operator has more confidence in his/her own abilities than in the automation then they are more likely to perform the task manually, and research using the pasteurisation plant simulation has confirmed this relationship (Lee & Moray, 1992, 1994). Operators used the automated system when conditions became such that they could not manage the system manually (e.g. during faults). Despite the low reliability, the automation became useful to the operator (Sheridan, 1999). The type of automation error and the consequences of that error also influence usage (Jiang et al., 2004); for example, if the automation makes an incorrect decision which causes further problems for an already overloaded operator they are more likely to discontinue using the automation. The interplay between competence, usefulness and self confidence may be quite complex, but to ensure automation is useful and utilised it is clear that the first requirement is reliability, both in the sense of repeated consistent functioning and competent decision making.

Safety critical systems are likely to be highly reliable and competent, and operators highly trained and confident in their abilities, and in these cases other factors may influence trust. Feedback from the automation becomes particularly important as automation becomes more complex and possibly even exceeds operator competence. Operators require explicit and appropriate feedback on its intentions in order to develop appropriate expectations (Sarter, Woods, & Billings, 1997; Sheridan, 1999). Good feedback may even counter the loss of trust in automation with low reliability and increase automation use. Research has shown that if operators are given an explanation as to why the automation might err then trust and usage levels can be maintained (Beck, Dzindolet, & Pierce, 2007; Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003). Simpler automation systems may not require advanced levels of feedback as the operator may be capable of understanding and predicting the automation without such prompts. It is the ability to develop an accurate mental model which the operator can use to understand and accurately predict future behaviour of the system which facilitates trust (Madsen & Gregor, 2000; Muir & Moray, 1996; Sheridan, 1999).

The investigation of mental models would be a study in itself (Bristol, 2005); a short overview of mental models is given here due to their relevance to trust in automation. "A mental model is an individual's cognitive representation of how a system operates. Mental models enable an individual to describe, explain, and make predictions about system operations" (Scerbo, 1996, p. 54). The purpose of mental models is illustrated in Figure 2-8.

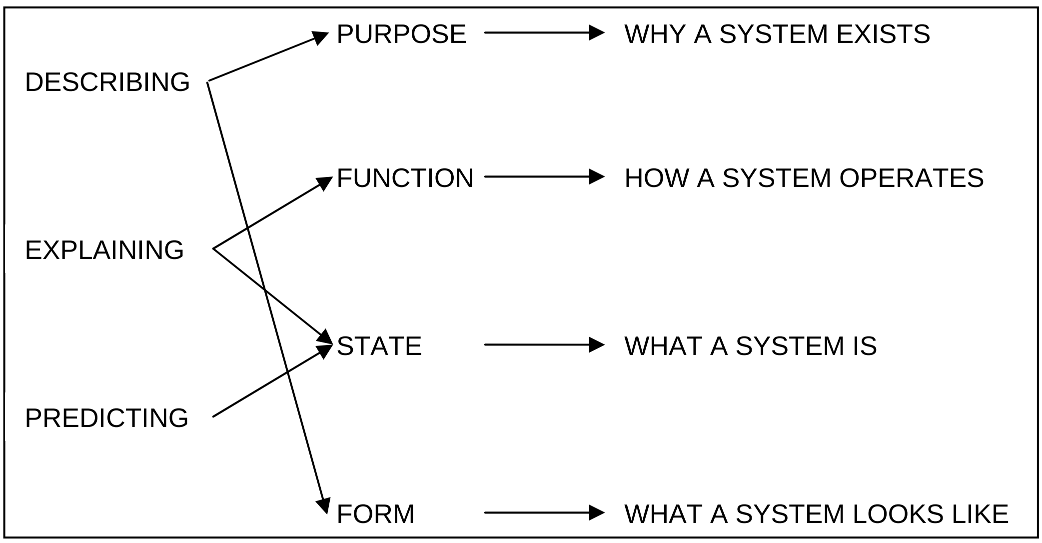


Figure 2-8: Purposes of Mental Models (Rouse & Morris, 1986, p. 351)

Wickens (1992) stated that successful performance in control rooms depends on a good mental model of the system allowing operators to anticipate future system states, formulate plans, and troubleshoot effectively and poor or inaccurate mental models have been associated with incidents and accidents (Sheridan & Parasuraman, 2006). Operators who possess an accurate mental model can make correct judgements on when an automated system can be relied on and when it should not be relied on. This is referred to as ‘trust calibration’.

Calibration of trust refers to the correspondence between a person’s trust in the automation and the automation capabilities (Lee & See, 2004; Madhavan & Wiegmann, 2007). If trust is miscalibrated the result is inappropriate reliance on the automation, either overtrust or undertrust. For a system to work optimally, the operator’s level of trust in the automation must be correctly calibrated (i.e. it should match the actual capabilities of the automation). These capabilities may vary in different circumstances; for example, automation may be competent in one set of circumstances but not in another. Operators should be able to recognise when automation can be relied upon and when it cannot. However, trust is not always uniform between different operators. Pre-existing factors such as experience with technology and familiarity may influence operator trust (Sheridan, 1999) meaning that individuals may have different trust levels for the same automated system. Merritt and Ilgen (2008) found that individual differences did affect perceptions of automation competence and hence influenced trust. Interestingly, they also found that individuals who had higher expectations of and a propensity to trust automation had the largest negative impact on trust when the automation failed. This suggests that correct calibration, and optimal automation usage, for less than perfectly reliable automation may more likely be achieved by individuals who are not predisposed to trust the automation.

As trust is a multi-dimensional concept and is dependent on circumstances, features of the automation, and individual differences it is difficult to measure. There is no direct objective measurement of trust and so measurement tends to depend on subjective ratings on the dimensions believed to influence trust, including reliability, competence, understandability, faith, personal attachment, and deception (Atoyan, Duquet, & Robert, 2006; Bisantz & Seong, 2001; Jian, Bisantz, & Drury, 2000; Madsen & Gregor, 2000).

The key findings on trust are summarised in Table 2-1. Experimental studies have provided evidence of the tendency for operators not to use automation if they distrust it, although this result is tempered by an operator's level of self confidence in handling the system without the aid of the automation. The effect of reliability or error rates on trust and subsequent usage of trust has been conclusively proven. Evidence also exists to support the notion of competence as a key dimension in trust and there is some support for the idea that understanding automation can improve the rating of trust. However, it is important to note that some aspects of trust are not influenced by the system itself but are related to the individual and outside the control of system designers.

Key Finding	*Author*
There is a correlation between trust in and usage of automation.	Muir & Moray, 1989; de-Vries, Midden, & Bowhuis, 2003
High reliability and competence are fundamental requirements for trust in automation.	Muir & Moray, 1989; Wiegmann, Rich & Zhang, 2001
Operator self-confidence and the usefulness of the automation also influence usage.	Lee & Moray, 1992; 1994
For complex systems, explicit feedback is required to develop trust.	Sarter et al., 1997; Sheridan, 1999; Dzindolet et al., 2003
Trust must be well calibrated to ensure optimal use of automation.	Lee & See, 2004; Madhavan & Wiegmann, 2007
Accurate mental models are important to ensure correct calibration of trust.	Sheridan & Parasuraman, 2006
Individual differences influence trust.	Merritt & Ilgen, 2008

Table 2-1: Summary of Key Research on Trust

2.5.2 Situation Awareness and Automation

Situation awareness can be thought of as a person's real time mental model of the world around them and is central to effective decision making and control in dynamic systems. The most common definition of SA is "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future" (Endsley, 1996, p. 164). These three elements of SA are typically referred to as levels with each level building on the previous and representing a more advanced state of SA. The three levels are: the perception of critical factors in the environment (SA level 1), understanding what these factors mean, particularly in relation to the achievement of goals (SA level 2), and being able to predict what will happen to the system in the near future (SA level 3).

Automation holds the potential to affect an operator's SA by reducing their direct involvement in the system and hence their perception of the environment may be reduced, either through complacency or through reduced quality in the feedback from the automation. Complacency may be a particular risk for highly reliable systems in which the operator rarely, if ever, is required to intervene (Endsley, 1996). Over longer periods, use of automation may also degrade the operators' capability to understand what changes in the environment mean, affecting the higher levels of SA. The complexity of some automated systems can also make the higher levels of SA harder to achieve as the systems become less understandable and have a resulting impact on the operator's ability to predict future system state. Although operators may not require a high level of SA if the automation is performing well, it may prove critical following a failure. In automated systems, this potential loss of SA is often referred to as out-of-the-loop unfamiliarity. In contrast, automation may also hold the potential to improve operator SA if it is designed and implemented in a manner which optimises operator workload, gives good feedback, and keeps the operator involved in the system (Billings, 1991).

In order to understand the impact of automation on SA it is necessary to measure operator SA. There are three principal methods of measuring SA: freeze probe measures, real-time probe techniques, and self-rating scales. Situation Awareness Global Assessment Technique (SAGAT; Endsley, 2000) is the most common freeze probe technique (Salmon, Stanton, Walker, & Green, 2006) and has the advantage of being an objective technique. However, generally it can only be used in conjunction with a simulator as it involves freezing the simulation at random intervals, blanking the display, and questioning the operator on their current perceptions of the situation. These answers are then compared to the actual situation to determine the participants' accuracy (Endsley, 1996). The requirement to use a simulator means that the use of SAGAT is limited to experimental settings and cannot be used in real world settings. Realism is further reduced by the requirement to freeze the simulation to administer the tool. Real-time probe measures such as Situation Present Assessment Measure (SPAM; Durso & Dattel, 2004) alleviate the necessity to freeze the simulation by incorporating the probes into the simulation. The measurement of SA is based on the time taken for the operator to respond to the probes. Another advantage of this method is that measurement of SA using this tool is not affected by memory decay, as could be the case with SAGAT (Wickens, 2008). It is also possible that, with care, this measure could be applied in a real world setting and so is not confined to simulators.

Both probe techniques represent an intrusion for the operator which the researcher may wish to avoid. In these cases, self-rating scales are commonly used and Situation Awareness Rating Technique (SART) is the most common of these. This tool was originally developed to assess aircrew SA (Taylor, 1990, in Salmon et al., 2006) and asks the participant to rate themselves on 10 dimensions post trial. Three key dimensions from SART of attentional demand, attentional supply, and understanding have since been identified and form the 3D-SART tool (Jones, 2000). Although SART is easier and less intrusive to administer than probe measures, it can only collect data on the whole trial so SA cannot be measured for specific circumstances. Less commonly, SA may also be measured using observer rating scales. Observer rating scales typically involve an SME observing an operator and rating their SA (Salmon et al., 2006). However, for this method to be effective there must be visible cues for the SME to observe. This is likely to be a particular issue with automated systems as operators' physical interactions, and hence visible cues, are reduced. The final point to make on measurement of SA is common to all measures, and that is the requirement to first understand what SA consists of. Different systems and tasks have different SA requirements and so the first step in measuring SA in a new context must be a thorough understanding of the elements necessary to build and maintain SA in that context (Endsley & Rodgers, 1994).

Research on SA and automation aims to answer the question of whether operators are less aware of changes in the environment when those changes are under the control of another agent and so experimental studies in the area have typically focussed on how SA varies with level of automation (Durso & Sethumadhavan, 2008). The levels of automation investigated range from full manual control, through intermediate levels where the operator is responsible for decision making, to fully autonomous automation. Few effects of high levels of automation have been found on Level 1 SA, although one experiment found that Level 1 SA improved when information acquisition was automated but decision making was manual (Kaber et al., 2006). This suggests that the use of information is more important for SA than the gathering of that information.

Two studies in two different domains have shown Level 2 SA to be significantly increased in intermediate levels of automation as compared to full automation (Endsley & Kiris, 1995; Kaber & Endsley, 2004). These results suggest that keeping the operator involved in the decision making process produces increased SA and although this increase in SA was not always accompanied by a corresponding increase in system performance, it did allow participants to perform better following automation failures, a result also found by Kaber, Onal, and Endsley (2000). However, SA is not always highest at intermediate levels of automation, as illustrated by Endsley and Kaber (1999). Their experiment found SA was increased during the highest levels of automation. This experiment differed in that participants were novices who had only a very short training time on the system. It seems likely that these participants struggled to control the system manually and so when their resources were freed during the higher levels of automation they were able to pay more attention to the system and hence improve SA. This study highlights an important point; SA does not only suffer when operators are taken out of the loop by automation, but also when operators are overloaded.

All the empirical research found in the literature was based in artificial simulated environments with no empirical data coming from real world research. However, accident data from the aviation domain may give some insight into the effect of automation on SA. There are a number of accident investigations which have identified lack of SA, or out-of-the-loop unfamiliarity, as a causal factor, for example, an accident in New York in 1984 in which the pilot was unaware of the airspeed as this was under automatic control at the time of the accident (Wickens, 1992). Such accidents suggest that lack of SA due to automation does have a real effect; however, it is also possible that such accidents were due to other more critical factors such as pilot distraction or overload which resulted in low SA (Dekker, 2004). Nevertheless, it can be concluded that automation systems must be designed to support operator SA under a variety of conditions to ensure safe performance.

Table 2-2 summarises the key findings of the literature review on SA. It is clear that SA can be affected by the level of automation in a system. That effect may be positive or negative, depending on the tasks that must be achieved, which are automated, and how they are automated. Automation which is designed to keep the operator in the loop while eliminating menial tasks to reduce workload is likely to create optimal SA.

Key Finding	*Author*
Level 1 SA is higher during automated operation of information acquisition, suggesting that the use of information is more important for SA than gathering the information.	Kaber et al., 2006
Level 2 SA is higher during intermediate levels of automation.	Endsley & Kiris 1995; Endsley & Kaber, 1999
Level 2 SA may be improved by automation during high workload conditions.	Endsley & Kaber, 1999
Performance during automation failures is better with higher SA.	Kaber et al., 2000
Well designed automation has the potential to improve operator SA.	Billings, 1991
SA is also affected by high workload conditions.	Endsley & Kaber, 1999

Table 2-2: Summary of Key Research on Situation Awareness

2.5.3 Workload and Automation

Automation is commonly introduced into systems on the basis that it will reduce operator workload and hence facilitate more efficient operation. Workload is a difficult concept to define but refers to the load or demand imposed on the human operator (Wickens & Dixon, 2007). These demands may be physical or cognitive in nature. The simple four stage human information processing model (Parasuraman et al., 2000) can be used to separate physical demands from cognitive demands, with information acquisition and action implementation generating predominantly physical demands for the operator, and information analysis and decision and action selection generating predominantly cognitive demands. In many ways, mechanisation and automation have reduced or eliminating much of the physical demand on humans, but the introduction of technology into society may have increased daily cognitive demands, for example by driving a car or interacting with a computer (Megaw, 2005). Similarly with automation of control systems, the information acquisition and action implementation phases of systems are typically easier to automate so operators have a reduced physical workload but their cognitive load may not be reduced. However, the physical demands can sometimes be more obvious and system designers may seize this reduction as an opportunity to increase the control area, increasing the cognitive workload for the operator as he is required to assimilate greater amounts of information to maintain SA (Macdonald, 1999). It is important to note that automation does hold the potential to assist the operator during high workload conditions, and automated systems are most successful when they achieve this (Dixon & Wickens, 2006).

Workload also varies between normal and disrupted, or failure, conditions. It may be acceptable during normal operations, but if the automation fails, or conditions occur which are outside its capabilities, workload may rise to levels at which operators cannot maintain satisfactory performance. The capacity for automation to reduce workload during normal conditions, when workload may already have been acceptable, but to fail to assist during high workload conditions is known as 'clumsy automation' (Wiener, 1989; Woods, 1996). With highly reliable systems it may be the case that failures are so rare that any performance decrement during a failure is an acceptable risk and so workload during failures is not a design consideration. However, workload which is too high is only one side of the issue, performance may also be affected if operators are underloaded and highly reliable automation which leaves the operator with too little to do may risk this (Young & Stanton, 2002). Operators who are underloaded may become bored and suffer lower SA as they become distracted. In order to understand the overall effect of automation on workload it is necessary to choose appropriate techniques with which to measure it.

There are four main categories of workload measurement; performance based measurement, physiological, analytical, and subjective measures (Megaw, 2005). Performance based measures may assess the performance of the participant at the task under review (primary measure) or their performance on a secondary task (secondary task measure). A secondary task may be useful if it is difficult to measure the success of a primary task. It aims to establish the spare capacity of the participant after completing the primary task and uses the amount of spare capacity as a measure of primary workload. Physiological measures such as cardiac activity, brain activity, galvanic skin response, eye function, and hormonal analysis have the advantage of being objective, but they may only apply to one dimension of workload and can be difficult and expensive to administer, particularly in a real world setting (Megaw). There is also variability in the results and no physiological measure has yet been accepted as standard (Sheridan, 2002). Analytical measures use system models based on information processing and resource theories to estimate workload but they require significant time and effort to develop and use (Pickup, 2006). Subjective measures are the most frequently used method of assessing workload. The National Aeronautics and Space Administration Task Load Index (NASA-TLX; Hart & Staveland, 1988) is a common subjective measure of workload and provides a reliable measure of overall workload on a scale of 0 to 100. Operators are asked to rate their workload on a number of pre-determined scales including mental demand, temporal demand, physical demand, performance, effort, and frustration (Harris, Hancock, Arthur, & Caird, 1995). Alternatives include the Cooper-Harper and Bedford scales developed to assess the controllability of new aircraft, and the Subjective Workload Assessment Technique (SWAT) which uses three three-point scales to rate subjective workload (Megaw).

Using these measurement techniques, experimental studies have shown that automation has the potential to reduce workload in domains such as telerobotics and aviation (Endsley & Kaber, 1999; Harris et al., 1995; Kaber et al., 2000; Kantowitz, 1994). A study by Kaber et al. (2006) determined that there is a larger workload reduction when information acquisition and action implementation are automated as compared to information analysis and decision making. Although workload is generally reduced by automation, some studies have suggested that a requirement to continuously monitor automation actually increases workload (Kaber & Endsley, 2004; Warm, Dember, & Hancock, 1996). The effort associated with remaining vigilant is high (Warm et al.) and workload may also be increased by the necessity to extract data regarding analysis and decisions made by the automation. This would explain why automation is more successful at reducing workload in the information acquisition and analysis functional dimensions. Operators' cognitive workload may be just as high, or even higher, during automated decision making as they make their own decisions as a basis for comparison with the automation. Data on the effect of automation on workload in the real world come from analyses of the Aviation Safety Reporting System (Renyard, Billings, Cheaney, & Hardy, 1986) in which incidents involving automation and workload were found to be more common on advanced automation glass cockpit flight decks than less advanced aircraft (Kantowitz, 1994). The incidents also tended to be more severe. This suggests that automated systems hold the potential for workload to spiral upward to the point where operators can no longer cope. In contrast, manual systems are more easily paced by the operator. For real world automated systems, it may be more important to ensure operators can perform safety and efficiently under all circumstances than at low workload under some.

These findings suggest that careful consideration of cognitive workload needs to be taken when automation is introduced, particularly with regard to abnormal operations or situations. Maintaining balanced and manageable level of operator workload during different phases of control is likely to result in an optimal level of performance. Table 2-3 summarises the key findings on workload and automation.

Key Finding	*Author*
Automation can reduce workload during normal operations.	Kantowitz, 1994; Endsley & Kaber, 1999; Harris et al., 1995; Kaber et al., 2000
Automation of information acquisition and action implementation have a greater effect on workload.	Kaber et al., 2006
Monitoring of automation may increase workload.	Kaber & Endsley, 2004; Warm et al., 1996
Automation may increase workload during incidents.	Kantowitz, 1994

Table 2-3: Summary of Key Research on Workload

2.5.4 Monitoring of Automation

A major effect of automation is the transition of the operators' role from one of monitoring which is integrated with control to one of monitoring for automation failures. This difference is often referred to in terms of 'active control' and 'passive monitoring' (Endsley & Kiris, 1995; Liu, Fuld, & Wickens, 1993; Metzger & Parasuraman, 2001). The suggestion is that manual control is an active process whereas monitoring of an automated system is a more passive role for the human operator. This may not be a positive change as humans are not well equipped for the monitoring task due to working memory limitations (Shorrock & Straeter, 2006; Wiener, 1985). Ironically, automation is often applied because the system designers believe it can do the job better than the operator, but the operator is still required to monitor that it is working effectively (Bainbridge, 1983). This monitoring work may be more mentally demanding than manual control (Parasuraman, Mouloua, & Molloy, 1996), especially considering the number of different components it may be necessary to monitor, many of which may operate faster and in a more complex manner than humans are capable of (Shorrock & Straeter). Furthermore, automation may remove the operator from the 'loop' potentially hindering effective monitoring and in the event of a problem requiring intervention, operators who are only involved in monitoring may have slower reactions as they gather the information necessary to understand the situation (Endsley, 1996).

Endsley and Kiris (1995) implied that 'passive monitoring' is associated with lower cognitive processing, suggesting that when operators are monitoring they are passively rather than actively processing information. There is evidence to suggest that passive processing of information is inferior to active processing (Cowan, 1988) and Metzger and Parasuraman found that expert air traffic controllers in a laboratory experiment took longer to find conflicts when they only had to detect them, as compared to detect and resolve conflicts. However, it may be the case that in real world operations operators do actively process system information while monitoring. If the information the operator is receiving is meaningful and they make judgements, including predicting future system states and deciding whether to intervene, based on it then some level of active processing would be required. In support of this, the subjective workload associated with monitoring has been shown to be quite high (Warm et al., 1996).

Despite the concerns regarding monitoring, there is relatively little empirical research dealing with the subject of monitoring, and in particular human monitoring behaviour in relation to automation (Liu et al., 1993). The majority of the literature relating to monitoring is in the fields of vigilance and complacency, and assumes that the desired level of monitoring is a constant but the operators' ability to maintain this varies over time.

Vigilance refers to the ability of an individual to remain alert over a period of time and is typically measured by the number of missed signals. Operators are required to maintain vigilance when working with control systems to ensure they can control them effectively, but this may be harder to achieve when the operator's main role is monitoring as there is little to keep him/her actively involved. Vigilance decrement is a phenomenon first simulated by Mackworth in 1950 (Parasuraman, 1987), and describes the situation where an operator begins to miss vital cues after a sustained period of attention. It has been repeatedly demonstrated in laboratory settings (Parasuraman; Warm et al., 1996) but most of these studies used simple sensory tasks that were not representative of the complexity of monitoring dynamic real systems (Moray, 2003). In addition, the tasks required the participants to monitor for infrequent signals that carried little significance or meaning to the individual further increasing the difficulty of remaining vigilant, and hence these findings may not be applicable to the real world where operators would be more likely to attach both meaning and importance to signals (Parasuraman, Molloy, Mouloua, & Hilburn, 1996; Wiener, 1987). One experiment addressed these concerns by using the pasteurisation plant microworld, previously used to examine the role of trust in automation use, to examine the vigilance decrement in a complex realistic task (Moray & Haudegond, 1998). No vigilance decrement was found suggesting that the vigilance decrement does not exist for some kinds of dynamic tasks. Therefore, although the vigilance decrement has been clearly demonstrated in laboratory experiments, it is less clear whether this effect carries over into the real world. Complex, dynamic, real world systems may keep operators sufficiently involved and engaged for a significant reduction in vigilance to be avoided. It seems likely that only highly autonomous automated systems pose a vigilance problem to operators.

Complacency also concerns unnoticed cues by operators, but in this case it is due to the operator having come to rely upon the automation and failing to monitor it, possibly due to a false sense of security or because they prioritise other tasks (Endsley & Kiris, 1995; Parasuraman, 2000; Sarter et al., 1997). Although the outcomes of vigilance decrement and complacency are the same, they differ in that complacency is a failure to sample correctly when monitoring whereas vigilance decrement is a difficulty in remaining focussed on the monitoring task (Moray, 2003). Research demonstrated that the time taken to detect failures rises when automation of one system in a multi task system is constantly reliable as compared to variable reliability automation of that system (Parasuraman, Molloy, & Singh, 1993). The effect was eliminated when participants had only to monitor one system. However, Moray (2003) argued that complacency is not concerned with the detection of signals but rather with attention and so a more appropriate measure would be to calculate the optimal sampling frame for a given system and assess operator monitoring against this. Operators who monitor a system more frequently than required are likely to distrust that system, while sampling less frequently than required would indicate complacency. For highly reliable systems the optimal sampling rate would be low, but this means that if a signal was to occur immediately after the operator has sampled it may not be noticed for a considerable period even though the operator is displaying optimal sampling behaviour (Moray & Inagaki, 2000). In this way, experimental studies can suggest a complacency effect where it is not actually present. Further research using the same experimental set-up as Parasuraman et al. demonstrated that participants' mean time between fixations rose significantly for constant reliability automation (Bagheri & Jamieson, 2004) suggesting that the operators did adjust their sampling in response to the increased reliability of the automation. Muir (1987) also found a relationship between trust levels and how often operators monitor automation. This would seem to be a reasonable response to automation which appears reliable and operators adjust their trust and sampling frequency accordingly. If all signals were to be responded to immediately, even those which occur very infrequently, it would require huge monitoring resources and in the normal course of events, operators would likely be accused of mistrust in the automation. As the reliability and competence of automated systems increases, the requirement to monitor should decrease as a result of the operator's well calibrated trust.

Despite the research focus on vigilance and complacency, studies have also been undertaken to examine how operators' go about monitoring complex real world systems (Sarter, Mumaw, & Wickens, 2007; Vicente, Mumaw, & Roth, 1998; Vicente, Roth, & Mumaw, 2001, 2004). Vicente et al. (2004) developed a model to describe operator monitoring in the nuclear domain. A simplified version of this model is shown in Figure 2-9.

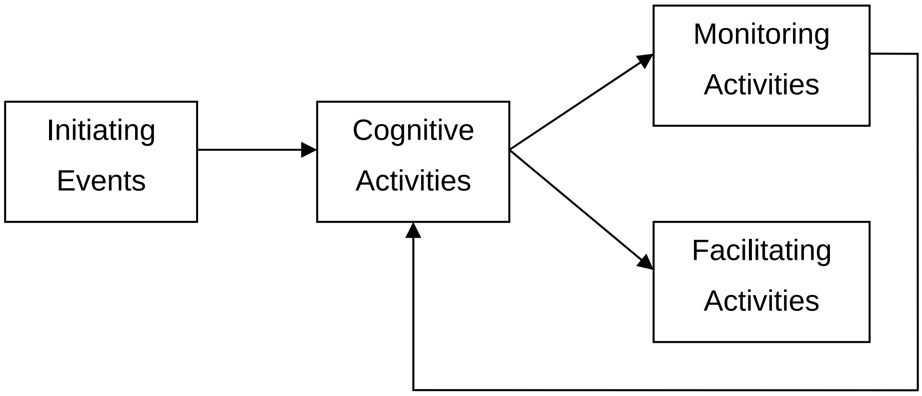


Figure 2-9: Operator Monitoring Model (Vicente et al., 2004)

The model suggested that monitoring is triggered, or initiated, by scheduled tasks or activities, policies, or alarms. The operator then engages in cognitive activities, for example evaluating the input, identifying the relevant data, finding the data, and developing a monitoring plan. Facilitating activities may be undertaken by the operator, for example configuring an interface, but these may not always be necessary. Monitoring activities in the nuclear domain included monitoring indications, conducting a field tour, monitoring alarm screens, communicating with other operators, or reviewing logs. As this model was developed exclusively within the nuclear domain it is not certain whether it can be generalised to other domains. However, it seems likely that monitoring of control systems in other domains would follow a similar model, although the specific tasks and activities undertaken in each stage may vary. Within the model, operators develop strategies to facilitate effective monitoring (Vicente et al., 1998). These strategies involve using knowledge gained through experience to anticipate events and to schedule work around monitoring activity. Such strategies are dependent on the feedback from automation systems and the operators' mental model of the system and if these are poor the strategies may be ineffective (Sarter et al., 2007). In these cases, operators may rely more on raw information and neglect monitoring of automated systems, with a resultant loss in the benefit of automation.

Eye tracking equipment may be used to gather information on the specifics of how operators monitor their systems. Methods for precise tracking of eye fixations have been around for over 100 years (Jacob & Karn, 2003) and although the technology continues to improve becoming more accurate, easy to use, and reliable, eye tracking systems continue to be invasive and/or uncomfortable to use. Remote eye tracking systems are more comfortable for the participant but they must be careful not be move out of the range of the equipment. Head mounted eye trackers allow more movement, but are physically uncomfortable and become less precise the more the participant moves. The extraction and analysis of eye tracking data can also be difficult and labour intensive (Jacob & Karn). Such limitations have meant that the volume of research is not as large as might be expected given how long eye tracking has existed. However, eye tracking has been used to study operators' use of systems in control environments (e.g. Anders, 2001; Ottati, Hickox, & Richter, 1999) results of these studies may be used to help design better interfaces or to improve training programmes (Dishart & Land, 1998; Ottati et al.). Such studies provide important and useful data, but are very domain specific. Unfortunately, it is impossible to generalise results on the importance of a particular display in one domain to a domain which incorporates a completely different set of displays. Within the rail domain, research has been undertaken to study train driver visual strategies (Luke, Brook-Carter, Parkes, Grimes, & Mills, 2006) but no research was found on rail signalling.

The key findings from the research on monitoring are summarised in Table 2-4. Overall, the concerns regarding monitoring of automation appear to have little empirical basis, particularly with regard to vigilance and complacency. The reduction of monitoring of reliable automated systems which is sometimes labelled complacency seems more likely to be a reasonable prioritisation on the part of the operator, and part of an overall strategy which operators develop to help them manage tasks and activities effectively. Although eye tracking may be used to investigate the mechanics of monitoring, no published research was found in the rail signalling domain.

Key Finding	***Author***
The subjective workload associated with monitoring may be high.	Warm et al., 1996
‘Passive monitoring’ may reduce awareness as compared with ‘active control’.	Endsley & Kiris, 1995; Metzger & Parasuraman, 2001
Although a reliable laboratory result, no evidence of a vigilance decrement has been found in real world systems.	Moray & Haudegond, 1998
Alleged complacency may be due to the calibration levels of trust.	Moray & Inagaki, 2000; Bagheri & Jamieson, 2004
Operators develop strategies to monitor automation effectively.	Vicente et al., 2004

Table 2-4: Summary of Findings on Monitoring

2.6 Design of Automated Systems

Whether the concerns discussed in the previous section become real issues in new automated systems is typically decided at the system design stage and may depend on the design strategy employed (Waterson, Older Gray, & Clegg, 2002). Automation design strategies determine where the line of control is between the human operator and the automation. Initial automation strategies were to automate everything technically possible to achieve widely anticipated benefits (Dekker, 2004), and this led to the “left-over” principle. Under the ‘left-over’ principle, the operators tasks are determined by whatever proved too difficult to automate (Hollnagel, 2001). This system therefore can place a heavy burden upon the operator, particularly if the left over elements are diverse and complex. Unfortunately, given the limitations of current technology and the flexibility of humans, this can often be the case.

Another approach, the substitution or compensatory principle (Hollnagel, 2001), uses MABA-MABA lists (Men-Are-Better-At, Machines-Are-Better-At) to allocate functions between the operator and the automation. This approach was first developed in the 1950s by Fitts (Fitts, 1951), and is still referred to as Fitts’ List. It only takes into account isolated capabilities, and so does not consider the full complexity and demands of a given situation and how this would best be controlled. Added to this is the fact that introducing automation into a system creates new functions and tends to transform a task often in unanticipated ways (Dekker & Woods, 2004). So while function allocation may have some uses in guiding allocation of functions (Parasuraman et al., 2008), it is also important to consider the role of automation and the human in the context of the task as a whole.

More recently, a strategy has emerged called the complementary principle. The purpose of this strategy is to sustain and strengthen human ability to perform efficiently and involves cooperation and coordination between the automation and the operator (Dekker & Woods, 2004). The main concern is not the momentary level of efficiency, “but rather the ability of the system to sustain acceptable performance under a variety of conditions” (Hollnagel, 2001). This situation is also described as the ‘cyborg metaphor’ in which the operator and the automation act together and the removal of one would render the system useless (Lee, 2008).

Within rail signalling, research has been undertaken to guide the development of new automated signalling systems in Sweden (Hellstrom, Frej, Gideon, & Sandblad, 1997). Analysis of the requirements for an autonomous automation system suggested that it is potentially impossible to fully automate Swedish train control (Kvist, Hellstrom, Sandblad, & Bystrom, 2002). A description of the Swedish rail network (Kauppi, 2006) suggests that it is considerably less complex than the UK network, with a large proportion (more than 80%) of the network consisting of single lines rather than the more complex double or multiple track layout commonly found in the UK. If rail signalling cannot be fully automated it suggests that a cooperative human-machine system is required to control railways (Kauppi). Automated systems developed to support this requirement will need to address the concerns outlined in this chapter. The final section of this chapter describes 12 principles of automation drawn from the literature which may be used to help guide the design of automated systems.

2.7 Principles of Automation

In order to address the concerns outlined in the previous section and to assist in the design of cooperative automated systems, principles of ‘good automation’ were drawn from the human factors literature. Key pieces of literature were reviewed and any recommendations or guidance on the development of automated systems were noted; 12 key principles were then defined from these data. Application of these principles to the design of automated systems should help ensure optimal levels of trust, SA, monitoring, and workload.

2.7.1 Reliable

The automation should function consistently.

The adverse effects of reliability on operator's trust in and use of automation have been repeatedly demonstrated (Dzindolet et al., 2003; Wiegmann et al., 2001). Different failure rates have different effects on the operator. A difference has been drawn here between reliability and competence but this does not always seem to be the case in the literature. Reliability here has been taken to mean the repeated consistent functioning of an automated device (Sheridan, 1999), but its ability to do the job correctly is understood to be competence.

2.7.2 Competent

The automation should perform tasks correctly given the information that is input.

Muir and Moray (1996) identified competence as a key dimension in development of trust in automation and suggested that designers of automated systems should consider whether automation will be able to carry out a function effectively as any weaknesses will reduce the likelihood that operators will use the automation. Control failures that are as a result of programming (i.e. where the automation acted as designed but the result was not desirable) would be classified as incompetent automation rather than unreliable. This in contrast to much of the literature in the area where automation competence is frequently referred to as reliability (e.g. Moray et al., 2000; Riley, 1994).

2.7.3 Visible

All decision relevant information for a given situation should be available to the operator.

This may be taken as supporting the first stage of SA. It differs from the next principle (observability) in that it refers to the provision of information regarding the system being controlled, while observability relates to the provision of information concerning the automation's decisions and actions. Dekker (2004) warned of automation making information 'invisible', hiding events which may be of interest to the operator and Billings (1991) and Endsley (1996) both recommended that operators always have basic information on system parameters being monitored available in a clear and easily interpretable format as this allows them to remain involved and aware of the system.

2.7.4 Observable

Automation should provide effective and immediate feedback to the operator allowing him/her to maintain awareness of system state.

Observable automation can be achieved through good quality feedback from the automation. Parasuraman and Riley (1997) stated the importance of providing feedback regarding automation state, actions, and intentions to enable the operator to monitor and intervene effectively. Woods (1997) identified that systems that provide only weak feedback on their activities are more likely to surprise the operator and recommended that observability of automation activities be improved. Sarter and Woods (1997, p. 554) defined observability as “the ability of available feedback to actively support operators in monitoring and staying ahead of system activities and transitions”. They stated that it involves more than just availability of the data but also the cognitive work involved in extracting it. The method of feeding back information also affects observability. Sheridan and Parasuraman (2006) suggested that automation which follows good etiquette (e.g. patient and non-interruptive) in feeding back information supports better system observability. Limited observability is likely to impact on the operators’ ability to understand the automation and develop a correct mental model, and may restrict the use of automation (Norman, 1990).

2.7.5 Understandable

Decisions made by the automation should be understandable to the operator given the current state of the system and environment.

Automation which can be easily understood by the operator enables them to predict and work cooperatively with the automation. Parasuraman and Riley (1997, p. 248) argued that “better operator knowledge of how the automation works results in more appropriate use of automation”. This understanding forms the basis of the development of a mental model. Development of an accurate mental model allows the operator to predict future actions of the automation (Sheridan, 1999). Woods (1997) stated that automation surprises are more likely in situations where operators have gaps in their mental models of how the automation works in different situations. Hopkin and Wise (1996) argued that successful human-machine relationships require the operator to understand the criteria taken into account by the computer, but it is not necessary for understanding to extend as far as how the individual algorithms work (Lenior, Janssen, Neerincx, & Schreibers, 2006). Comprehensible and predictable automation should be developed, even at the cost of reduced flexibility or power of the automation (Billings & Woods, 1994).

2.7.6 Directable

The operator should be able to direct the automation easily and efficiently.

Woods (1997) recommended that users be given the ability to direct the automation as a tool in achieving goals, and Dekker (2004) that the human operator be allowed to assume a strategic role in directing the automation. This ability to direct the automation helps achieve a more cooperative system, with the resulting benefits including improved SA, mental workload, and overall system performance (Miller et al., 2005). Without the ability to influence and direct the automation the recommendations on observable and understandable automation are useless as the operator is essentially powerless (Christoffersen & Woods, 2001).

2.7.7 Robust

The automation should be able to perform under a variety of conditions, not just normal operating conditions.

Sheridan (1999) termed the ability of automation to cope with a variety of conditions ‘robustness’. This should be both in terms of the operating envelope within which the automation is capable and avoiding ‘clumsiness’; that is, automation which is of most assistance during normal operations but becomes less helpful, or even a burden, during abnormal operations (Billings, 1997). Billings (1991) suggested that automation should be of most help during times of highest workload and somewhat less help during times of lowest workload. In practice this may be hard to achieve, but it is desirable that automation is designed to be helpful, and not a burden, during times of high workload.

2.7.8 Accountable

The operator should be responsible for overall performance and therefore in charge of the automation

Billings (1991) suggested that responsibility is an important concept in human-automation relationships. Sarter et al. (1997) stated that computers cannot be expected to be responsible beings and so it follows that the operator must be provided with the means to control the system. Automation which is autonomous and lacking in accountability is more likely to surprise the human operator, making it difficult for him/her to maintain effective control of the system (Woods, 1997). Aircraft pilots have expressed a strong preference for management by consent automation, but when workload demands were high they preferred a management by exception system (Olsen & Sarter, 1999). Research has indicated that operators who are not fully conscious of their role in ensuring high performance are less likely to intervene (Mosier, Skitka, & Korte, 1994). Ensuring that the automation is accountable to the operator enables the operator to take responsibility for the overall system performance. Of course there are some circumstances where it may be possible to completely remove the human from the control loop, usually in areas which have very definite data and characteristics. Train driving may be an example of this, however, even in these cases a human override is deemed necessary (Sheridan, 1999).

2.7.9 Error Resistant

The automation should make it difficult for the operator to make an error.

Billings (1991, p. 78) suggested that error resistance may be achieved by “clear, uncomplicated displays and simple, intuitive procedures”. He also suggested that error resistance could be achieved through confining the operator’s potential actions to an envelope protected by the automation, although to ensure accountability a human override should be incorporated. Rasmussen and Vicente (1989) argued that reliable human-machine systems can be developed by designing interfaces which minimise the potential for error and support recovery from errors. They describe a number of principles for reducing operator error which fall under the category of making systems error resistant. These include consistent mapping of cues for action and symbols of process function; tools provided to the operator to experiment and test hypotheses for use in unanticipated situations; provision of appropriate information for monitoring purposes; and development and maintenance of mental models. Norman (1983) suggested that error resistance can be maximised through improved feedback, functional organisation of screen displays, command languages or menu headings which are distinct from one another, minimising the ease with which actions that have serious implications or are not reversible can be performed, and consistency of the system structure and commands.

2.7.10 Error Tolerant

The automation should have the ability to mitigate the effects of an operator error.

Error tolerance in a system can be increased by the monitoring of other stakeholders in the system (Billings, 1991); automation can provide this support by giving clear warnings when unsafe actions are attempted. Rasmussen and Vicente (1989) also described a number of guidelines for systems to cope with operator error that could be classified as methods to enhance error tolerance in a system. These include making the limits of acceptable performance clear to the operator before the effects disappear or become irreversible and providing feedback on the effects of actions to counter any delay between operator action and observable effect. Norman (1983) suggested that actions should be reversible whenever possible.

2.7.11 Proactive Control

The system should support the operator in predicting and controlling ahead rather than controlling reactively.

Dekker (2004) suggested that automation should support the operator in reasoning in advance and knowing what to expect, allowing them to develop a plan in advance. Endsley (1996) stated that supporting SA by allowing operators to keep up with changing system parameters and understanding the effect of these allows operators to proactively optimise system performance and prevent future problems. Proactive control can be enabled by ensuring that the automation is predictable. Sandblad, Andersson, Bystrom, and Kauppi (2002) recommended a proactively controlled system for railway control which allows the operators to monitor the development of the system over time and prevent disturbances.

2.7.12 Skill Degradation

The automation should incorporate a method to guard against operator skill degradation.

Bainbridge (1983) suggested that skill degradation is a likely but undesirable trait of automation. As the reliability of the automation increases the opportunity for the operator to practice manual control is reduced and the effect may be to reduce the operator's skill in understanding and controlling the underlying system (Dekker, 2004). The result may be an operator who is required to take over when the automation reaches its limits, but who is no longer skilled enough to do so adequately (Endsley & Kiris, 1995; Hoc, 2000). It may be difficult to guard against skill degradation in some highly automated systems but in these cases high fidelity simulators may be provided to help operators maintain their skills.

2.7.13 Summary

The 12 principles presented in this section were drawn from the literature and validated with human factors professionals (Appendix D). They are intended as a guide for the design of automated systems, but may also be used to structure an evaluation of existing systems.

2.8 Conclusions

Automation has the potential to add benefit to control systems through the reduction of workload and increase in performance (Sheridan, 1999). However, the introduction of automation faces many challenges, including ensuring the correct level of automation, to ensure that the benefits are achieved and system performance is optimised. Table 2-5 describes findings for each of the key themes with respect to automation. The table illustrates that automation has an effect on SA, workload, and monitoring of automation and these are influenced by the operators' trust and mental models of the automation.

Theme	*Key Finding*
Trust	• There is a correlation between trust in and usage of automation. • High reliability and competence are fundamental requirements for trust in automation. • Operator self confidence and the usefulness of the automation also influence usage. • For complex systems, explicit feedback is required to develop trust. • Trust must be well calibrated to ensure optimal use of automation. • Accurate mental models are important to ensure correct calibration of trust. • Individual differences influence trust.
Situation Awareness	• Level 1 SA is higher during automated operation of information acquisition, suggesting that the use of information is more important for SA than gathering the information. • Level 2 SA is higher during intermediate levels of automation. • Level 2 SA may be improved by automation during high workload conditions. • Performance during automation failures is better with higher SA. • Well designed automation has the potential to improve operator SA • SA is also affected by high workload conditions.
Workload	• Automation can reduce workload during normal operations. • Automation of information acquisition and action implementation have a greater effect on workload. • Monitoring of automation may increase workload. • Automation may increase workload during incidents.
Monitoring	• The subjective workload associated with monitoring may be high • 'Passive monitoring' may reduce awareness as compared with 'active control'. • Although a reliable laboratory result, no evidence of a vigilance decrement has been found in real world systems. • Alleged complacency may be due to the calibration levels of trust. • Operators develop strategies to monitor automation effectively.

Table 2-5: Summary of Key Findings

To tackle these concerns, 12 principles of automation were drawn from the literature and validated with human factors professionals. These principles are summarised in Table 2-6.

Principle	*Description*
Reliable	The automation should function consistently.
Competent	The automation should perform tasks correctly given the information that is input.
Visible	All decision relevant information for a given situation should be available to the operator.
Observable	Automation should provide effective and immediate feedback to the operator allowing him/her to maintain awareness of system state.
Understandable	Decisions made by the automation should be understandable to the operator given the current state of the system and environment.
Directable	The operator should be able to direct the automation easily and efficiently.
Robust	The automation should be able to perform under a variety of conditions, not just normal operating conditions.
Accountable	The operator should be responsible for overall performance and therefore in charge of the automation.
Error Resistant	The automation should make it difficult for the operator to make an error.
Error Tolerant	The automation should have the ability to mitigate the effects of an operator error.
Proactive Control	The system should support the operator in predicting and controlling ahead rather than controlling reactively.
Skill Degradation	The automation should incorporate a method to guard against operator skill degradation.

Table 2-6: Summary of Principles of Automation

2.9 Chapter Summary

This chapter has described the benefits and issues associated with automation. The effect of automation on human factors themes of trust, SA, workload, and monitoring were discussed and principles were developed from the literature to help minimise issues arising from these concerns. Different modelling techniques for describing the level of automation were also discussed and the advantages of the model presented by Parasuraman et al. (2000) were highlighted. This model will be used in the following chapter as the basis for a rail automation model describing the types and levels of automation present in different generations of UK signalling systems.

CHAPTER 3: RAIL AUTOMATION MODEL

3.1 Chapter Overview

A structured framework was developed for the description of types and levels of automation in rail signalling. The framework took as a starting point the model for types and levels of automation proposed by Parasuraman et al. (2000), although scales have been modified to make them more applicable to rail signalling. Data were gathered to support the development of these scales and three different types of signalling system were plotted using the final scales. The final model describes the levels of automation present in the UK railway, and can be used to identify areas where the level of automation may be inappropriate.

3.2 Introduction

Automation in rail signalling varies hugely depending on the type of signalling system employed. Three types of signal box are predominant in UK rail signalling; lever frame, NX panel and IECC and these three systems have been analysed to support the development of the rail automation model. Lever frame signal boxes were introduced in the 1800s and are still in widespread use today across the rail network. Routes are set for trains by physically pulling large, and sometimes weighty (Muffett, 2007), levers which are directly connected to the trackside equipment. These boxes are very limited in terms of the support provided to the operator. Installation of NX panels was undertaken in the 1950s and represented a huge leap forward in assisting the operator with the physical workload associated with moving points and signals. The signaller simply presses buttons on the panel and the physical movement of trackside equipment is achieved automatically. The most recent form of signalling system are IECCs and principles similar to NX panels are used for manual signalling, but also introduced decision making automation capable of setting routes automatically (ARS). See Appendix C for more information on these signalling systems.

The description of the levels of automation present in different generations of rail signalling systems was identified as a starting point for this research. This was a key part of the first research step identified in the Research Approach Framework (Figure

3-1), understanding the context of the research. It helped generate knowledge of general signalling principles which supported the research overall, but also identified which aspects of automation had advanced, and which had not. As such, it provided a basis on which to continue with the more targeted research.

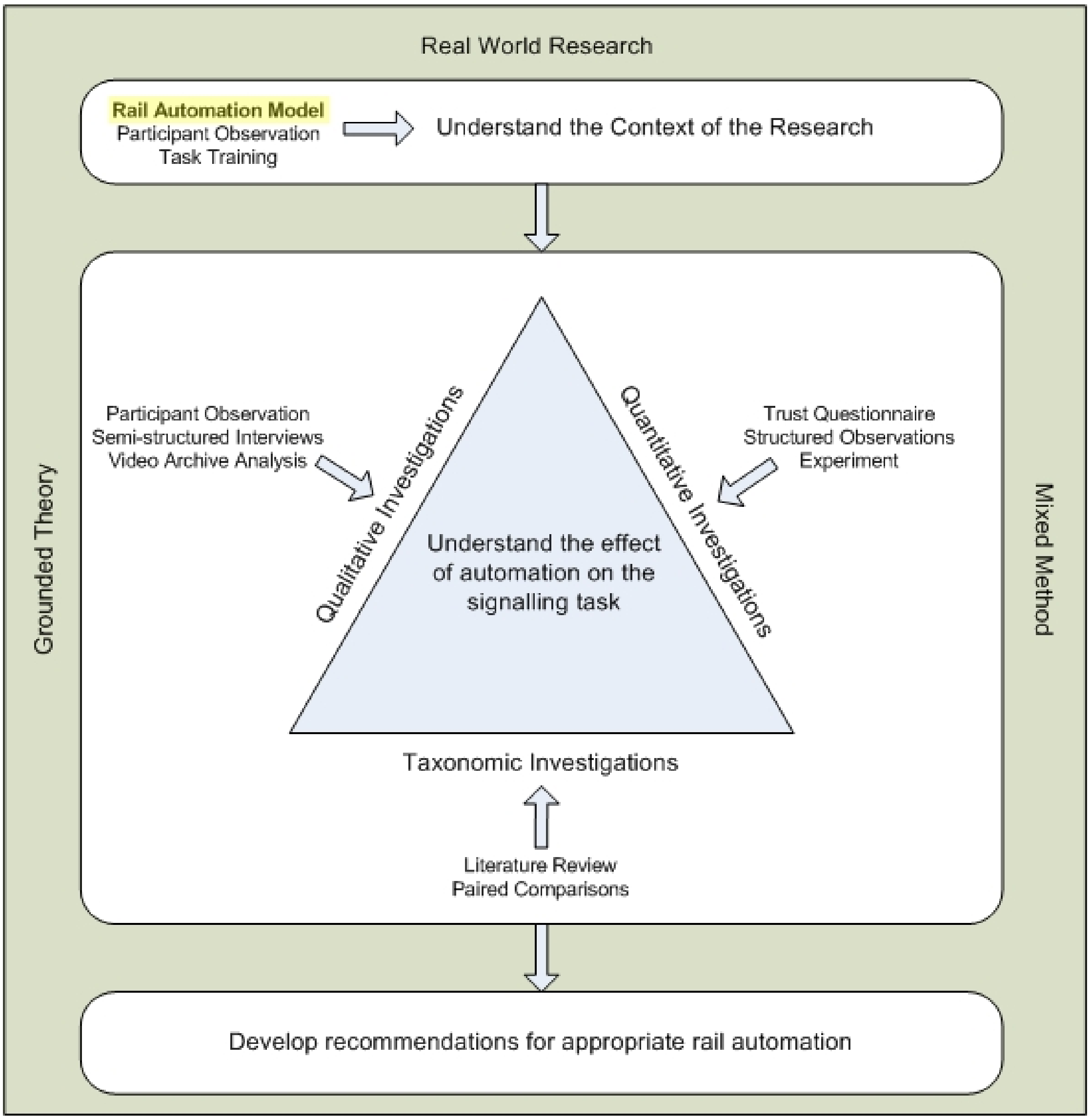


Figure 3-1: Position of Rail Automation Model in the Research Framework

Models of levels of automation have typically been used to investigate the impact of different levels of automation on key issues such as SA (e.g. Endsley & Kiris, 1995; Kaber et al., 2000; Kaber et al., 2006) and workload (e.g. Kaber & Endsley, 2004; Kantowitz, 1994). The levels of automation identified in the models can be used as independent variables in experimental designs, allowing the effect of automation to be described on a continuum. Models of levels of automation may also be used to support the design of automated systems by providing a framework on which to base decisions on allocation of function (Parasuraman et al., 2000). The rail automation model was originally developed to illustrate the differing levels of automation in different generations of signalling systems, which may be a new application of this type of model. The model may also be used in one of its more traditional roles; to identify areas where the level of automation may be inappropriate and hence potentially guide the development of new rail automation systems.

There are a number of models detailed in the literature (e.g. Billings, 1991; Endsley & Kiris, 1995) but the model chosen as the basis of this research is the model for types and levels of automation described by Parasuraman et al. (2000). The ability of this model to discriminate between four functional dimensions of Information Acquisition, Information Analysis, Decision and Action Selection, and Action Implementation means it is a more powerful method of analysis. Simply describing automated systems along one continuum does not give an appreciation of the variability of automation which may be present within systems.

Parasuraman et al. (2000) detailed their interpretation of how automation may vary in each functional dimension. For information acquisition, a low level of automation was suggested which simply helps gather the information; a mid-level is when the automation organises the information in some form, perhaps forming priorities; and a high level is where the automation filters the information so that a full set of raw data is not provided to the operator. Low levels of information analysis automation may involve the use of algorithms to extrapolate incoming data over time or predict, and a higher level may involve integration of input variables into a single value. Automation may assist the operator with decision making, for example by using conditional logic. Parasuraman et al. proposed that decision automation level increases as the automation narrows the decision alternatives. Automation of the final stage, action implementation, may be the easiest of types of automation to understand, with the level being defined by how much of the physical activity is replaced by automation.

The rail automation model developed levels for each functional dimension on the basis of site visits and uses these levels to plot the three generations of signalling systems included in the model.

3.3 Method

3.3.1 Participants

The participants in this study were based in signal boxes which were visited to gather data on the requirements and capabilities for each of the four functional dimensions. In total nine signal boxes were visited to gather data for this study: three lever frame, three NX and three IECC. The number of signallers observed for the study was approximately 20. The data from these visits formed the basis of the rail automation model, although it was supplemented and further developed following less structured visits to signal boxes throughout the course of the research.

3.3.2 Apparatus

The data were recorded manually; no apparatus was used.

3.3.3 Procedure

Visits to signal boxes were arranged through SMEs and visits were undertaken with SMEs. The following questions, based on the model from the literature (Parasuraman et al., 2000), were asked at each signal box:

- What information is required?
- How is this information acquired?
- What analysis is performed upon the information?
- How is this analysis performed?
- What decisions are made on the basis of the information?
- How are these decisions made?
- What actions must be implemented?
- How are the actions implemented?

The results for each question were noted and tabulated. The table was used, in conjunction with levels of automation found in the literature (Parasuraman et al., 2000; Sheridan, 1999; Sheridan & Verplank, 1978) to generate new levels for each functional dimension. The levels of automation applied to each functional domain of each generation of signalling equipment were validated by SMEs, and the levels were also re-evaluated at stages throughout the period of research to ensure they were still applicable in light of the information emerging from the research.

3.4 Results

The results from the site visits were collated and the data collected are summarised in Table 3-1. Automated elements are highlighted in green. Information that is required to be gathered for signalling purposes includes:

- Train movements – train entering area and track occupation;
- Train information – class, destination, timetable, and delay;
- Infrastructure – position of points, signal aspects, routes set, infrastructure failures, and planned restriction of infrastructure.

Analysis of this information during normal running was simply to determine which trains required a route to be set, which had priority, and which signals and points would need to be operated. The decision selection phase involved choosing which train to set a route for. The actions required once a routing decision has been made are:

- Set points;
- Clear signal aspects;
- Cancel route after train has passed;
- Communicate with relevant parties;
- Complete paperwork.

			Lever Frame	*NX Panel*	*IECC (ARS)*
Acquire	Train Movements	Train entering area	Block bell[1]	Panel (via TD[2])	VDU (via TD)
		Train location in area	Block instrument[3]/visual	Panel (via track circuits)	VDU (via TC[4])
		Train leaving area	Visual	Panel	VDU
	Train Information	Class	Block bell	Panel (via TD)	VDU (via TD)
		Destination	Timetable	Panel (via TD)	VDU (via TD)
		Delay	Timetable	Timetable/TRUST/CCF[5]	Timetable/TRUST/CCF
		Special/Additional Trains	Telephone	Telephone	Telephone
	Infrastructure	Control area	Track diagram/visual	Panel	VDU
		Position of points	Lever position	Point position switch	VDU
		Signal Aspects	Lever position	Panel	VDU
		Route set	Lever position	Panel	VDU
Analyse	Analysis of acquired information and planning of regulating decisions		Manual	Manual	Automatic (ARS)
Decide	Decide actions to be taken to regulate		Manual	Manual	Automatic (ARS)
Action	Set Route	Set points	Lever	Button	ARS
		Clear signals	Lever	Button	ARS
	Communicate	Train movements	Block bell	TD	TD
	Paperwork	Record train movements	Manual	TD	TD
	Cancel Route	Signals back to danger	Lever	TORR[6]	TORR
		Set points to normal	Lever	TORR	TORR

Table 3-1: Requirements for Normal Train Routing

[1] Block bells are a form of telegraphic communication used between lever frame boxes.
[2] Train describers (TD) are 4 digit alphanumeric identifiers for individual trains. They are displayed on the panel/workstation.
[3] Block instruments are manually controlled devices showing the condition of the line between signal boxes, i.e. clear or occupied.
[4] Track circuits (TC) are electric circuits running through the rails which detect the presence of a train.
[5] Train Running System (TRUST) and Control Centre of the Future (CCF) provide information on train timetables and delays.
[6] Train Operated Route Release (TORR) automatically releases routes following the passage of trains.

The scales developed for each of the functional dimensions are shown in Table 3-2.

Information Acquisition		
1	None	Human gathers all information without assistance from computer or technology, using senses for dynamic information and paper based sources for static information
2	Low	Human gathers all information but with assistance from IT (telephone/fax/email/CCF/TRUST)
3	Med	Information acquisition is shared between the automation and the human
4	High	Computer and technology provide the majority of the information to the human
5	Full	Computer gathers all information without any assistance from human
Information Analysis		
1	None	Human analyses all information.
2	Low	Computer analyses information as it is received and detects conflicts only as they occur.
3	Med	Computer gives a future prediction based on basic information for the short term (e.g. current trains on the workstation).
4	High	Computer gives a future prediction based on fuller information (e.g. trains arriving in future, infrastructure state, current situation on other workstations), and highlights potential problems/conflicts over a longer period of time.
5	Full	Computer gives a long term future prediction using all relevant data (e.g. up to date information on train speeds, infrastructure state etc.).
Decision and Action Selection		
1	None	Human makes all decisions, without any support.
2	Low	Computer provides decision support to the human to help ensure decision is not unsafe.
3	Med	Computer performs basic decision making (e.g. first come first serve, run trains to timetable) and leaves perturbed modes to the human.
4	High	Computer performs mid-level decision making (e.g. apply set rules to delayed trains) and has basic plans for implementation during perturbed operations.
5	Full	Computer makes all decisions under all circumstances using complex algorithms to determine the optimal decision (e.g. based on a high level prediction of the future state, optimal conflict resolution) and provides flexible plans for disrupted operations.
Action Implementation		
1	None	Human implements all actions and communications.
2	Low	Computer augments human’s physical labour (e.g. hydraulic assistance on lever).
3	Med	Computer implements physical actions but human is required to perform communications (possibly with assistance from ICT).
4	High	Computer implements physical actions and basic communications but human is required to perform complex or unusual communications.
5	Full	Computer implements all actions and communications.

Table 3-2: Levels of Automation in the Rail Automation Model

3.5 Discussion

New levels have been generated as a part of this work, as the original levels were found not to adequately describe the levels found in the rail environment. It may be that although the original levels were developed to be generic (Parasuraman et al., 2000) they are not sufficiently powerful to be applied to any system. Sheridan (1998) used a similarly generic scale to describe levels of automation and plot the graph shown in Chapter 2; however, his interpretation of the scale for plotting the graph was quite liberal. It is also worth noting that the scales used by Sheridan have varied over time (Parasuraman et al., Sheridan, Sheridan & Verplank, 1978). This may indicate a difficulty in applying one scale to different systems. Such a criticism would not be limited to this model, as other examples were found of researchers changing levels used to describe automation between different studies (e.g. Endsley & Kaber, 1999; Endsley & Kiris, 1995). A further criticism of the scales developed by Parasuraman et al. is that they combined the functional dimensions, creating one scale for information acquisition and analysis and a second for decision making and action implementation. The advantage of their model is the division into the four functional dimensions, and combining these within two scales compromises some of the power of this approach. This model uses a separate scale for each of the functional dimensions and five levels of automation have been defined within each; (1) none, (2) low, (3) medium, (4) high, and (5) full.

3.5.1 Information Acquisition

The levels for information acquisition proposed by Parasuraman et al. (2000) included some form of analysis at the higher levels to determine the most relevant information and what information can be discarded. Pure information acquisition would more properly refer only to the sensing of relevant data. In this model, the level of automation depends entirely upon the level of assistance the automation/computer/technology gives the operator in gathering the required information. Information acquisition has moved beyond pure manual conditions in all signal boxes as they are required to be equipped with some Information and Communications Technology (ICT). Even the most basic signal box has a telephone and fax machine and the majority have a computer providing CCF and/or TRUST; however, in lever frame boxes, the primary method of information acquisition is manual, either visually or through the block instruments. Therefore, Lever Frame boxes have been assigned to Level 2. In NX Panels and IECCs, the basic information of train position has been automated, via track circuits, but the signaller is still required to gather large quantities of information manually, including paper based timetables and information on incidents which comes in via the telephone. Therefore these have both been assigned to Level 3.

3.5.2 Information Analysis

An alternate approach to information acquisition is taken in this model to that of Parasuraman et al. (2000) whereby the levels of automation are based upon how far into the future an analysis is performed and the degree of accuracy of the prediction. In respect of rail signalling, any analysis performed is in support of regulating decisions which may need to be made, and in particular in determining where there may be conflicts in the future. Only where there is a route setting agent present (i.e. ARS) is any analysis currently automated and even then, conflicts are only detected and dealt with as they occur. Therefore, Lever Frames and NX Panels are both assigned to Level 1 and IECCs to Level 2.

3.5.3 Decision and Action Selection

As well as increasing the decision making power of the automation, the rail automation model proposes that decision automation can also be in the form of supporting the correct decision process, in rail signalling this is performed by the interlocking. The higher levels of decision automation in this model involve the ability of the automation to decide the route of trains and the levels increase with the complexity with which the automation can cope.

The main decision required in signalling is which route to set for each train, and when to set it. Mechanical interlocking has been in use since the 1800s to support the signaller’s decision on route setting. The interlocking ensures that the route set is not unsafe. Basic route setting agents are capable of routing trains either according to first come first serve or running strictly to timetable. More complex systems, such as the ARS in operation in IECCs, use a set of rules at each junction to determine which route to set. Full automation could be envisaged as setting routes according to the output of a prediction and conflict resolution tool which ensures that the route set is optimal and does not impact negatively on other trains across the network. Lever Frame boxes and NX panels have been assigned to Level 2, as they have the interlocking systems to support decision choices. The more advanced IECCs are assigned to Level 4.

3.5.4 Action Implementation

Lever Frame boxes are largely manually operated. There is a possibility, where heavy levers exist, to employ a hydraulic actuator to assist in pulling the lever but this is not often implemented on the railway. This level has been included because it is viable solution and in the wider context of automation it is one that is often used. Therefore, Lever Frames have been assigned to Level 1. On NX Panels the signaller is required only to push buttons to select a route and the system then ensures that all signals and points are changed accordingly. This is therefore Level 3 and IECCs are similar with regard to action implementation and have also been assigned to Level 3. The rail industry requires a large number of communications as compared to some other industries with a high level of automation. This is probably due to the complexity of the rail industry which is reflected in the number of different people with whom a signaller must communicate, for example train drivers, control staff, station staff, level crossing staff, members of the public, delay attribution clerks, and other signallers. At present there is relatively little support from automation in these communications.

3.5.5 Rail Automation Graph

On the basis of these scales the following graph was produced plotting each of the three generations of signalling technology, Lever frame, NX panel and IECC (Figure 3-2). The case of IECC operating with ARS switched off has not been plotted separately as in this case the IECC essentially reverts to the levels of NX panel.

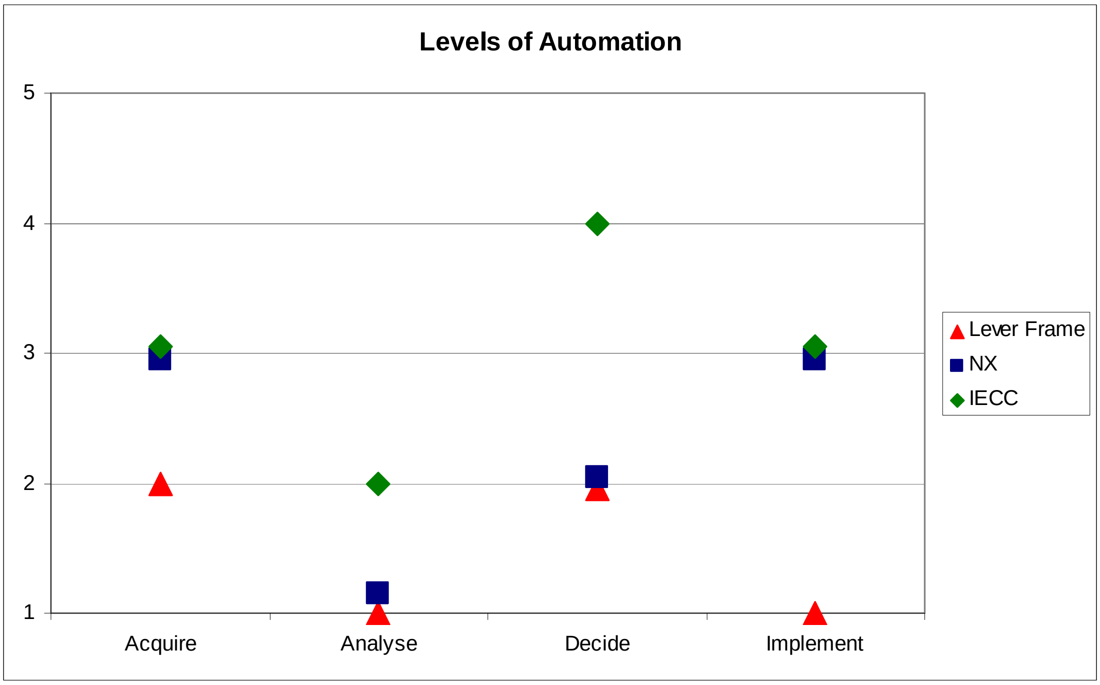


Figure 3-2: Model of Levels of Rail Signalling Automation

The model indicates that the automation of Information Acquisition and Action Implementation have not been increased by the introduction of IECC. As these are areas in which automation can be of great assistance to the operator, for example by supporting SA (Kaber et al., 2006), there may be potential to increase automation in these areas. It is also clear that little automatic analysis of information is achieved, and this may be another area where there is scope to increase the level of automation to support the operator and improve overall system performance. Currently the operator must integrate information from many different sources to obtain a complete picture of the situation in his/her area of control. Good signallers also use the information at their disposal to generate predictions of future states which allow them to step in early. There is scope for automation to support operators in these early interventions, both by assisting in identifying where action needs to be taken and by predicting the impact of any changes made. Decision and action selection showed the largest increase in automation with the introduction of IECC. Such a high level of automation is likely to require high competency levels and good feedback from the system to ensure the operator trusts and uses it (Dzindolet et al., 2003; Muir & Moray, 1989). If these are not provided this may not be an appropriate level of automation for this functional dimension, particularly without the supporting automatic analysis of information. Action implementation was not greatly increased in the move from NX to IECC technology, and this is primarily due to the lack of progress on automation of communications. Automation of the transfer of information between railway control staff may be another area in which there is scope to increase levels of automation to support more efficient performance.

3.6 Chapter Summary

The model presented provides clarity on the different types of signalling systems and the levels of automation present in each. It also illustrates those areas where there is scope for expansion, and this is most evident in the information analysis functional dimension. Both information acquisition and action implementation may also hold potential for further automation in order to support signalling. However, the levels of decision and action selection may not be appropriate if supporting analysis and feedback are not provided.

The development of this model facilitated understanding of the context of the research but did not provide any data on the use of, or issues with, the main system under investigation, ARS. An observation study was undertaken in order to begin the quantitative investigation of ARS by examining the levels of usage. This study will be presented in the following chapter.

CHAPTER 4: STRUCTURED OBSERVATIONS OF IECC SIGNALLERS

4.1 Chapter Overview

An observation study of signallers using ARS at four IECCs was conducted in order to establish and record the behaviours signallers exhibit while using automation and to gather some initial data on the factors that influence automation usage. The results showed clear differences in signaller activity, even when circumstances on the workstations were very similar. The study was also designed to investigate whether a relationship exists between signallers' trust in ARS and their observed interactions with it. A questionnaire was administered to support this aim and significant differences were found between groups exhibiting different levels of intervention and quiet time.

4.2 Introduction

The rail automation model described how automation levels increased in IECC but did not provide any information on the impact of that automation on the human operators. This study developed an observation method to gather empirical data on signaller behaviour in IECCs while using, or choosing not to use, ARS. Using this method, the study provided data regarding the magnitude of the differences in five activities between signallers and between IECCs. Anecdotal reports and opinions of signallers' use or non use of the automation are common within Network Rail but this was the first empirical study of system usage. A questionnaire was also administered to gather data on individual operators' levels of trust in ARS and the results from this questionnaire were related to the observed activities. As illustrated in Figure 4-1, these were two of the three quantitative methods used to understand the effect of automation on the signalling task.

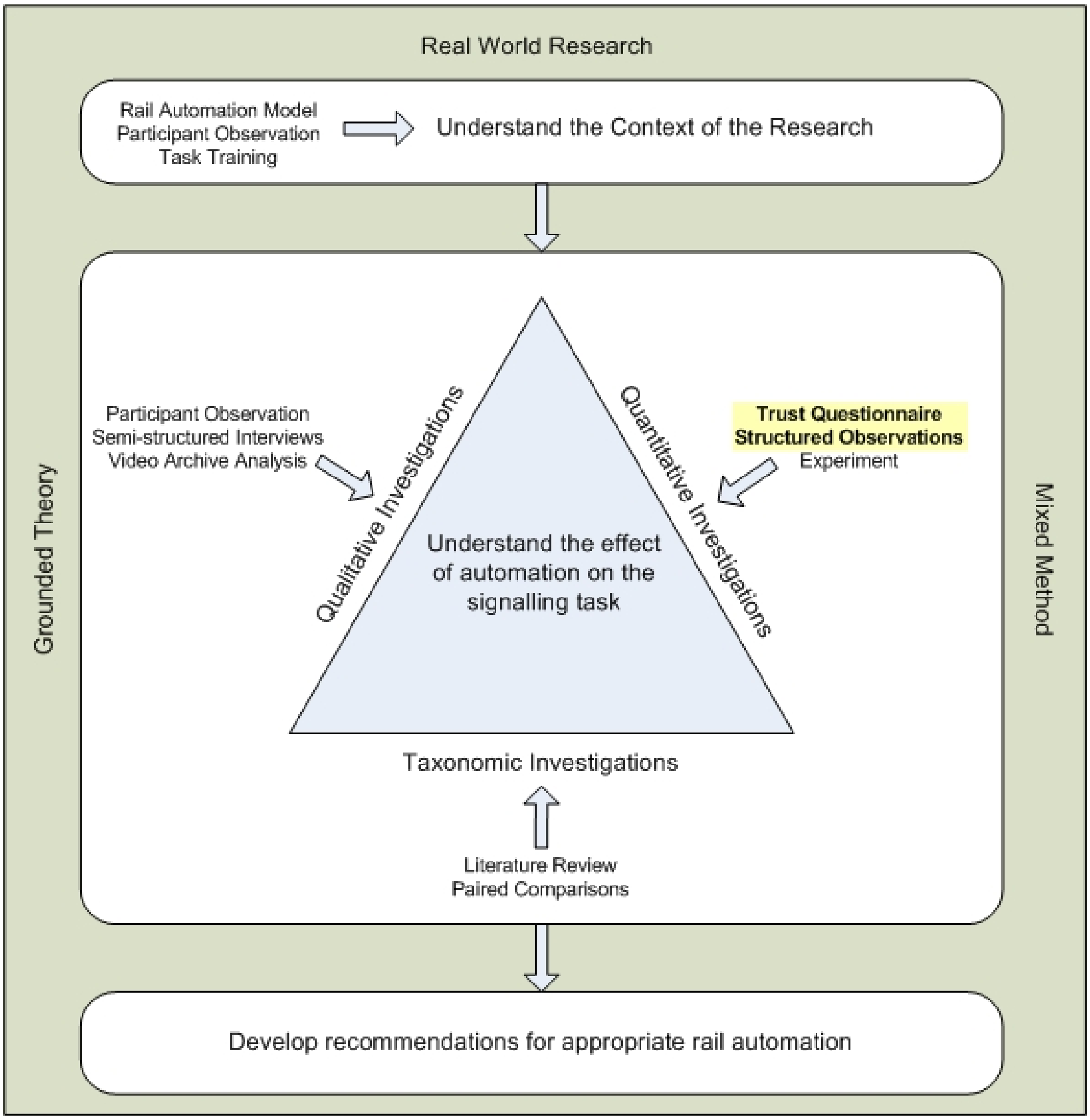


Figure 4-1: Position of the Observation Study and Trust Questionnaire in the Research Framework

Specifically, the aim of the study was to develop and apply a methodology for observing signallers at their workstations. This methodology was applied in pursuit of the following objectives:

1. To determine what proportion of signallers' time is spent monitoring, controlling, planning, communicating or not actively involved in signalling during normal operations.
2. To determine whether different signallers have different strategies in how they use ARS and to establish whether attitudes and strategies within signal boxes are more similar than attitudes between signal boxes.
3. To establish whether trust is related to how often the signaller intervenes and their level of monitoring.

Observation studies have previously been used to gather information in support of an ethnographic approach to systems design (e.g. Bentley et al., 1992). Observations have been undertaken in the rail domain to evaluate work systems and processes (Kauppi, 2006; Roth, Malsch, & Multer, 2001). Reinach (2006) created a framework to describe train dispatcher (signaller) activity and created six high level categories within this: actuating controls, issuing directives, granting permissions, carrying out communications, record keeping, and reviewing reference materials. This framework was intended for use in measuring signaller workload, although measurement of workload through observable tasks may not give a full indication of actual workload as it is not necessarily representative of cognitive workload (Pickup, 2006). However, it may be appropriate to use such a framework as part of a suite of workload tools. Reinach noted that frameworks such as this provide valuable information on number and diversity of tasks involved in signalling and may be used to create models of safety and performance. Lenior (1993) developed four categories for an observation supporting an analysis of cognitive processes of signallers in The Netherlands. The categories were route setting, train movements, telephone communications, and communications with colleagues. This framework differs from Reinach's framework in that it is less exhaustive; all activities of the signaller would not be captured by Lenior's framework. The categories relating to control activities are also more focussed on the outcomes or purposes of the activity. These differences are representative of the variation in research focus; Reinach was primarily interested in describing signaller activity while Lenior was conducting an investigation into cognitive processes. It is important to tailor an observation framework to the issue under investigation. The signalling systems involved in the research undertaken using these frameworks did not include automation, and so the frameworks developed are more applicable to the development of the method in this study than the results.

Trust has been identified as a key issue in the use of automation; operators will not use automation if they do not trust it. Reliability of the automation is known to be a fundamental requirement in the development of trust (Wiegmann et al., 2001), but as signalling systems are safety critical, the automation is required to be highly reliable. It was expected therefore that other dimensions in the development of trust may emerge, some of which may not have as strong an empirical basis in the research. These included feedback (Sarter et al., 1997), understandability and predictability (Sheridan, 2002), and faith (Muir, 1987). Competence of automation is also fundamental to trust development (Muir & Moray, 1989).

The following key dimensions were identified from the literature to be included in the measurement of trust:

- Reliability – in terms of both mechanical reliability and consistent functioning over time (Madsen & Gregor, 2000; Muir & Moray, 1996; Sheridan, 1999).
- Robustness – the ability to function under a variety of different circumstances (Sheridan, 1999; Woods, 1996).
- Understandability – the ability to understand what the automation is doing, why it is doing that and how it is doing it (Madsen & Gregor, 2000; Sheridan, 1999).
- Competence – the perceived ability of the automation to perform its tasks (Madsen & Gregor, 2000; Muir & Moray, 1996).
- Explication of intention – the ability of the automation to explicitly give feedback on its intended actions (Norman, 1990; Sheridan, 1999).
- Dependability – the extent to which the automation can be counted on to do its job (Muir & Moray, 1996; Rempel et al., 1985).
- Personal Attachment – the extent to which operators like to use the automation (Madsen & Gregor, 2000).
- Predictability – the ability of the operator to predict the actions of the automation (Muir, 1994; Rempel et al., 1985).
- Faith – the extent of belief that the automation will be able to cope with future system states which it may not have yet encountered (Madsen & Gregor, 2000; Muir, 1994; Rempel et al., 1985).

Previous research has linked trust to automation usage (de-Vries et al., 2003; Muir, 1987), but these studies have examined the use of automation as either all or nothing, on or off. This is possibly due to the nature of the automated systems used in the experiments which did not allow participants to simply intervene to force a decision; however, this approach is possible with the ARS system. This study therefore aimed to investigate the link between trust and the level of intervention in the system. Another strength of this study is that it examines expert users of a real world system in the live environment as opposed to simulations.

4.3 Method

4.3.1 Participants

Opportunity sampling was used for this study. Signallers were not specifically chosen for the study; the decision was purely based on who was working on the workstations of interest on the days the researcher was available, and whether they were willing to take part in the study. On three occasions the same signaller was observed on both workstations, but these occurred by chance. Therefore the total number of participants was 21. All participants were male and had at least five years IECC signalling experience.

4.3.2 Apparatus

A questionnaire was administered to gather data on signallers' trust in the automation (Appendix E). Statements for each of the key dimensions identified in the literature were taken from previously validated questionnaires (Bisantz & Seong, 2001; Jian et al., 2000; Madsen & Gregor, 2000; Muir & Moray, 1996). The questions were slightly modified to suit the signalling environment. The 19 statements on the questionnaire are shown below.

1. ARS is always available for use (Mechanical Reliability)
2. ARS is capable of performing under a variety of different circumstances (Robustness)
3. It is easy to understand what ARS does (Understandability 1)
4. ARS is capable of signalling trains as competently as a signaller (Competence 1)
5. ARS gives explicit information on its intended actions (Explication of intention)
6. I can count on ARS to do its job (Dependability)
7. I have a personal preference for using ARS (Personal Attachment)
8. I can predict what ARS will do from moment to moment (Predictability 1)
9. If ARS makes a routing decision which I am uncertain about I have confidence that ARS is correct (Faith 1)
10. I understand how ARS works (Understandability 2)
11. ARS performs well under normal running conditions (Competence 2)
12. ARS is very unpredictable, I never know what it is going to do (Predictability 2)
13. I can rely on ARS to function as it is supposed to (Reliability 2)
14. Even if I have no reason to expect that ARS will be able to deal with a situation, I still feel certain that it will (Faith 2)
15. I understand why ARS makes the decisions it does (Understandability 3)

16. ARS performs well under disturbed conditions (Competence 3)
17. ARS is very consistent (Predictability 3)
18. ARS will always make the same routing decision under the same circumstances (Reliability 2)
19. I trust ARS

4.3.3 Design

Eight workstations in four signal boxes were included in the study. Workstations A in each signal box were comparable in terms of workload and the type of demands placed on the signallers. Workstations B were also chosen to be comparable. The first three signal boxes chosen for the study, York, Liverpool Street and Ashford, were picked on the basis of the reported usage of ARS in each. Usage is reportedly high in Liverpool Street, low in Ashford, and variable in York. Tyneside was added to the study at a later date to gather additional data and was chosen on the basis of having comparable traffic levels and complexity to the other workstations in the study. Figure 4-2 describes the workstations involved in the study.

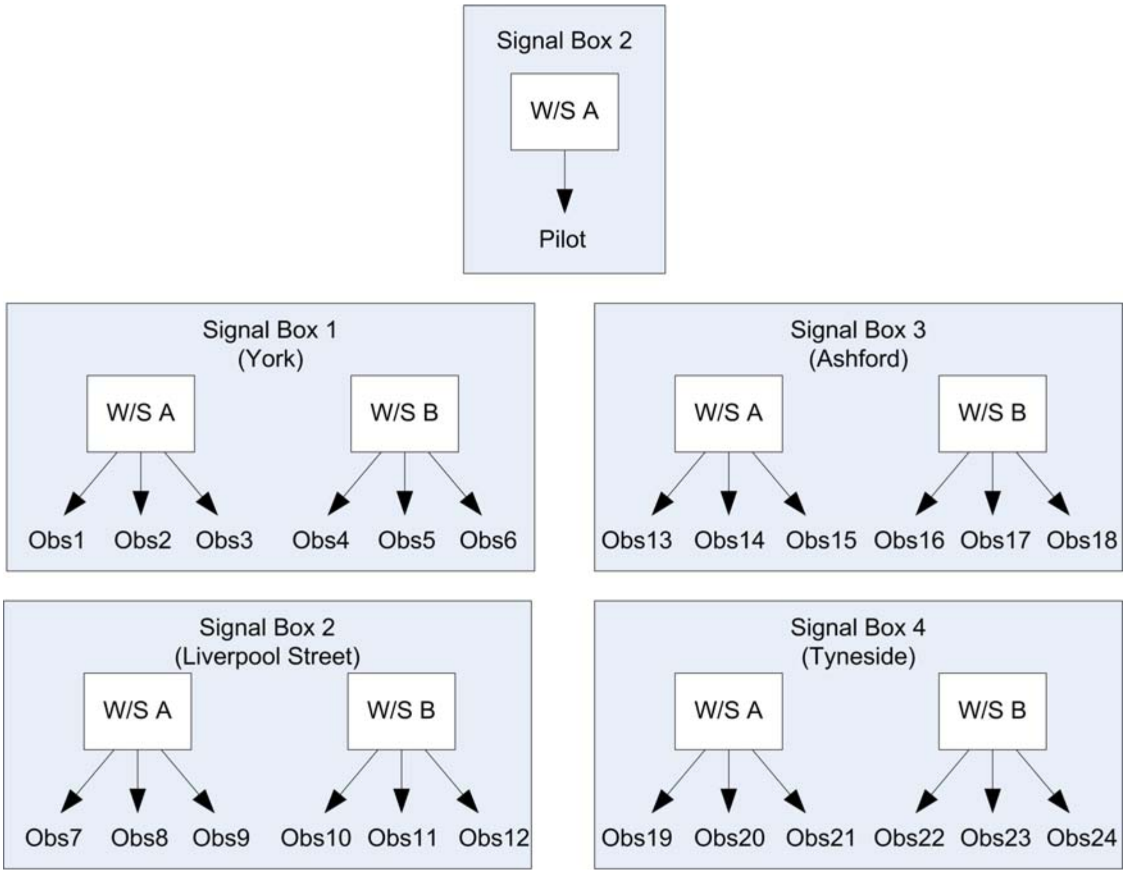


Figure 4-2: Study Design

Complexity of workstations was measured using Operational Demand Evaluation Checklist (ODEC) scores and verified by Subject Matter Experts (SME). The ODEC tool was developed to measure the demand placed upon the signaller due to the infrastructure on a particular workstation (Pickup & Wilson, 2007). The tool measures quantifiable aspects of the workstation such as number of signals, number of level crossings, and speeds of trains and then ranks each entity as high, medium, or low. Visits were undertaken and data collected to complete ODECs for all workstations in the four IECCs and, in order for the study to be comparable, the ODEC scores were matched as closely as possible for the workstations chosen for the study so the workstations were as similar as possible (Appendix F). In addition, SMEs were consulted to ensure that the specific demands of the chosen workstations were comparable, for example, the four workstations in Group A all have high traffic levels through station areas while all four in Group B have a depot. Although every effort was made to make the workstations in the study as comparable as possible there are no two areas on the railway which precisely match and this variability must be accepted as a limitation of the study.

York South workstation controls the area around York station. It is a relatively complex workstation controlling over a thousand trains a day. Leeds East workstation controls the area around Leeds station and is similarly complex. Shenfield and Ilford workstations control portions of the railway leading in towards Liverpool Street station. They have similar train service levels to the York workstations and although neither controls a large station, they both have several smaller stations. The number of trains is much lower on North Kent and Ashford 4 workstations in Ashford IECC but they were the highest scoring workstations in Ashford IECC. The Darlington workstation in Tyneside was similarly less complex than other workstations in the study. Newcastle workstation controls the area around Newcastle station and has similar traffic levels to York.

It is a limitation of the study that the predicted workload demand from ODEC was estimated to be less on the Ashford and one of the Tyneside workstations. However, SMEs consulted agreed that they would be broadly comparable for the purposes of this study. Unfortunately there was a further unexpected limitation on Ashford 4 workstation as the observations were carried out too early to catch the peak time. In general, the peak hours are between 16:00 and 19:00, and for this reason the observation time was chosen to be 16:30-18:00, but the peak is later at Ashford as it takes over an hour for the peak trains from London to reach Ashford.

The North Kent 1 workstation is not actually an IECC workstation but, aside from a minor difference in that the GP screen is not always displayed, the automation runs identically to an IECC workstation. The difference is in terms of the underlying structure of the signalling system, but the ARS system is identical. Most signallers do not use ARS on Ashford 4, but unfortunately due to opportunity sampling, two of the three signallers observed used ARS on this workstation even though this is not the norm.

Figure 4-3: Newcastle Workstation

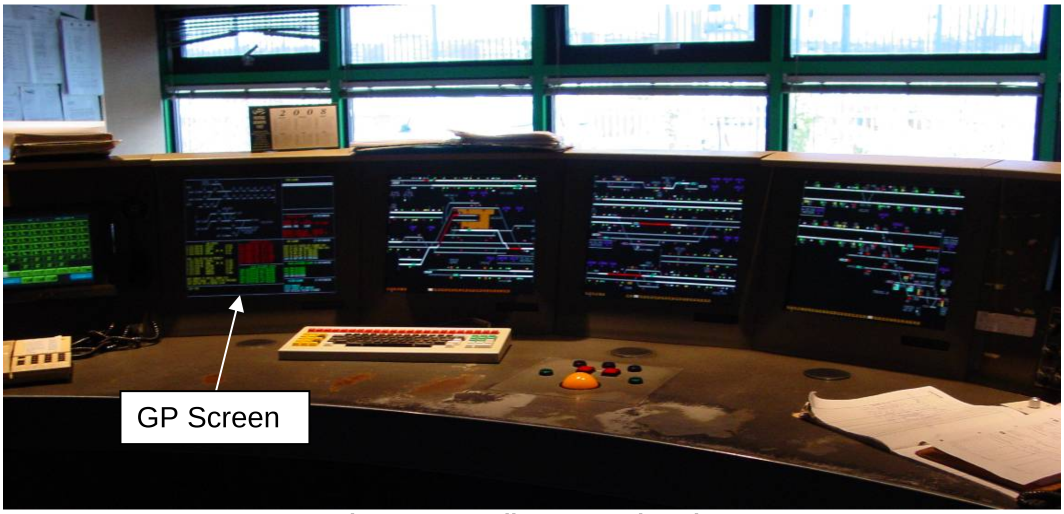


Figure 4-4: Darlington Workstation

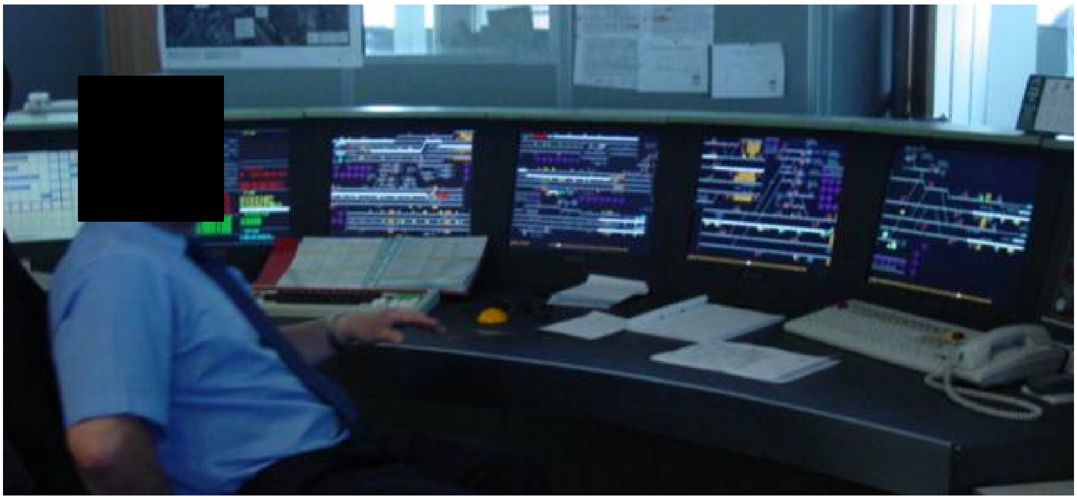


Figure 4-5: Ilford Workstation

Figure 4-6: Shenfield Workstation

4.3.3.1 <u>Coding Scheme for Data Collection</u>

A coding scheme was developed to support manual real time observations in the field. Five basic codes were used:

- Monitoring
- Intervening
- Planning
- Communicating
- Quiet Time

A sixth supplementary code (Closed Circuit Television; CCTV) was added for one of the signal boxes included in the study.

It is important to note that monitoring was coded when it was the only activity the signaller was engaged in. Within these five categories, additional sub-categories were coded where possible. The full coding system including all sub-categories is described below.

Monitoring

MA: Active monitoring.
Monitoring was coded as active if the signaller was sitting up while monitoring.

MP: Passive monitoring.
Monitoring was coded as passive if the signaller was sitting back while monitoring.

Intervention

TB: Trackerball.
Trackerball usage was only noted if the signaller used the button.

K: Keyboard.
This was the keyboard attached to the ARS system only. Other planning systems on the signaller's workstation also have keyboards but use of these was classified as planning behaviour.

Planning

PS: Simplifier
The simplifier is a printed simplified version of the timetable which tells the signaller what time each train should be at different points on the workstation.

PT: TRUST.
TRUST is a system that the signaller can interrogate for information on specific trains. It provides them with the scheduled route and current delay for individual trains (Appendix C).

PC: CCF.
CCF is a map based system showing the running of trains. Trains are colour coded reflecting their delay and the timetable can also be shown for each train with a prediction of future delay (Appendix C).

PP: Paperwork.
Paperwork such as completing the TRB[7] was classified under planning behaviour in the absence of a more appropriate category. Instances of paperwork activities were very rare during the study period.

---

[7] TRB (Train Register Book) is a log book kept on some workstations to log any incidents or occurrences (Appendix C).

R: Reading ARS Output.

ARS can be interrogated for information on the scheduled routes of specific trains and to discover which train ARS is giving priority to at a junction. Therefore, reading of ARS output on the GP screen[8] was classified as planning behaviour although in some instances the signaller may have actually been cancelling irrelevant alarms. Again, instances of this were sufficiently rare not to have had a significant impact on the data.

Communications

T: Telephone.

Any telephone calls were classified under this heading.

CS: Voice communications.

Signallers frequently communicate with the signaller on an adjacent workstation or the shift manager. Only information which was relevant to the immediate signalling situation was thus classified. Conversations regarding, for example, situations which occurred in the past were coded as ‘Quiet Time’ as they would not have been relevant to the signalling at that time.

CI: Intercom communications.

Some IECCs use intercoms to communicate with Control. Effectively this replaces telephone communications with Control[9].

Quiet Time

Q: Quiet time.

This included any time when the signaller was involved in an activity not directly related to signalling. Conversations with other signallers or staff, conversations with the researcher, reading newspapers or magazines were all examples of activities classed as quiet time.

QA: Signaller away from workstation.

Signallers occasionally took time away from the workstation, for a variety of reasons but most commonly to make a cup of tea. If the workstation was left unattended this activity was classed as quiet time away from the workstation.

---

[8] GP Screen – General Purpose Screen. The GP screen gives details on alarms, trains approaching the control area, and responses to signaller queries. See Appendix C for a description of the GP screen.

[9] Railway control staff take a more strategic view of the railway. It may be necessary for signallers to co-ordinate with them to implement changing plans.

Closed Circuit Television

Only one of the sites (Tyneside) had CCTV screens on the workstations. These screens are used to monitor and operate level crossings. When the signaller was involved in either monitoring or operating these, CCTV was coded.

4.3.4 Pilot

The study was piloted in Liverpool Street on Shenfield workstation. One of the aims of the pilot study was to establish whether it would be necessary to video the signaller. This would have provided additional data but some signallers were known to be uncomfortable with being videoed. It was established in the pilot study that video would not add to the study as the main missing information was what was happening on the screens and why the signaller was making the observed interventions. A static video camera would not have captured these data adequately as the data were spread across different systems and screens and so it was decided not to use video.

The pilot study was also intended to test the coding scheme to ensure that the codes were exhaustive and that it was possible to easily differentiate between the behaviours. The basic coding scheme worked well during the pilot but a major finding was the existence of different levels of monitoring behaviour. The signaller's position was observed to change substantially during the monitoring task and during the pilot study five different levels were identified. The highest level had the signaller sitting up, watching the screens intently with his hand on the trackerball and was very common when the signaller was waiting for the right moment to intervene. In the next level the signaller again was sitting up, with his hand still on the trackerball but scanning the screens rather than watching one spot intently. This monitoring was common when the signaller felt it was likely he may have to intervene but he had not yet decided where. The next level was similar but the signaller did not have his hand on the trackerball. This was inferred to be pure information gathering monitoring behaviour. It was often seen when the signaller was preparing to leave the desk, had just returned, or after an intervention. The next type of monitoring behaviour identified was passive monitoring, the signaller was sitting back but it was clear from his movements and posture that he was watching what was happening on the screens. The final type of monitoring behaviour seen during the pilot was complete passive monitoring. Often the signaller put his hands behind his head and it was not possible to tell if he was even focussing on the screens. Although these five levels were clearly observed during the pilot study, to attempt to code these for all observed signallers subsequently would have greatly complicated the observer's task. Therefore a decision was made to only differentiate between active and passive monitoring in subsequent observations. Pan, Gillies, and Slater (2008) stated that body movement is an easily observable indicator of a person's state. Cues in the posture and behaviour of signallers were used to infer whether they were engaging in active or passive monitoring. The results of the pilot allowed the procedure for the observations to be finalised.

4.3.5 Procedure

The observations were carried out at the same time of day on each workstation. The observation time was 16:30-18:00. The researcher arrived at the signal box at approximately 16:00 and approached the signaller on the workstation of interest. Usually the signaller had been made aware of the study in advance to ensure they were happy with being observed, but in some cases this was not possible and a brief outline of the study was required before proceeding. The study was then explained in more detail and the signaller was given a consent form to read through and sign (Appendix G). Following this the researcher asked about the current state of the area under their control and whether there were any particular problems. Any instances of disruption or late running trains were noted. Once everything was explained satisfactorily the signaller was instructed to ignore the researcher as much as possible and to act as if she was not there.

The observations commenced as close to 16:20 as possible. This was earlier than the scheduled time of 16:30, but there were a number of occasions where it was not possible to complete the observation period or where the signaller was replaced by the relief signaller[10] for a few minutes, and so starting early gave an extra 10min of data to fill any gaps. The data were coded every 5s using the coding system outlined earlier. An Excel spreadsheet bound into a book was used to record the data (Appendix H).

[10] Relief signallers are often used in bigger signal boxes such as IECCs. They rotate around each workstation during a shift allowing the signaller on duty to take a break.

At the end of the observation period, participants were given the questionnaire and asked to rate their agreement with each statement on a five point Likert scale. They were also asked to give a short debrief on any unusual occurrences on the workstation during the observation. Finally, they were asked what their most common reason for intervening during the observation was, and which intervention method they favoured.

4.4 Results

4.4.1 Overall Results

Figure 4-7 describes the results of the observations. Each bar on the graph describes the distribution of activity for one signaller and each group of three bars describes the three observations for each workstation in the study. There are clear differences in activity between signallers. The same signaller was observed on Obs 3 and Obs 4, and these two graphs are remarkably similar. Obs 8 and Obs 11 also show the same signaller, but these graphs are different. In this case there was disruption during Obs 11 which contributed to the difference in the graphs. Finally, the same signaller was observed for Obs 23 and Obs 24 but these graphs show a big difference, particularly in intervention. It is not possible to account for this difference as the workstation was reportedly running smoothly on both days. The signaller on Obs 17 was the only one to choose not to use ARS.

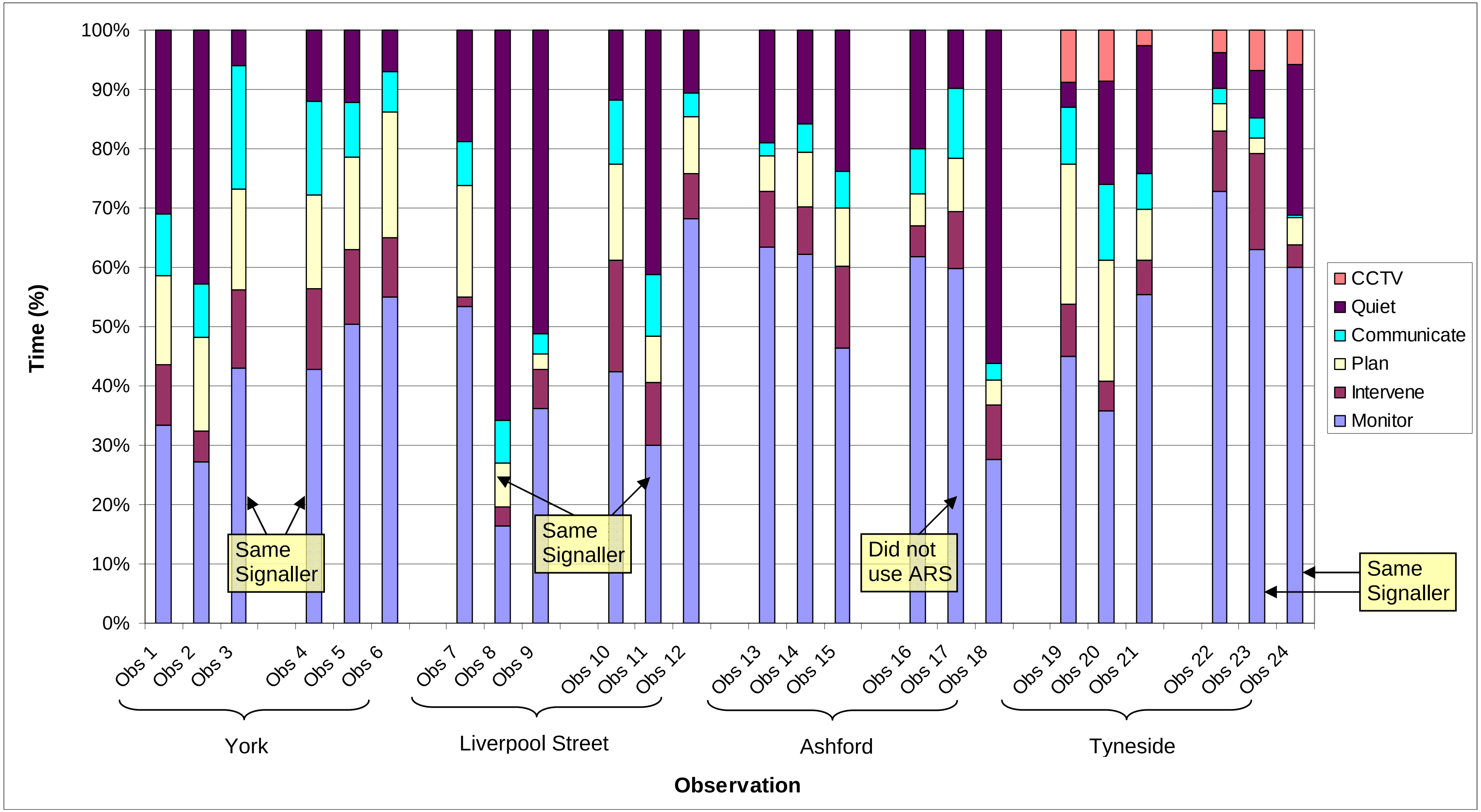


Figure 4-7: Observation Data

	Monitoring	*Intervention*	*Planning*	*Comms*	*Quiet Time*	*CCTV*
York A	34% (8%)	9% (4%)	16% (1%)	13% (7%)	27% (19%)	n/a
York B	49% (6%)	12% (2%)	17% (3%)	11% (5%)	10% (3%)	n/a
Liv. St. A	35% (19%)	4% (3%)	10% (8%)	6% (2%)	45% (24%)	n/a
Liv St. B	47% (19%)	13% (6%)	11% (4%)	8% (4%)	21% (17%)	n/a
Ashford A	57% (10%)	11% (3%)	8% (2%)	4% (2%)	20% (4%)	n/a
Ashford B	50% (19%)	8% (3%)	6% (3%)	8% (5%)	29% (24%)	n/a
Tyneside A	46% (10%)	6% (2%)	17% (7%)	10% (4%)	14% (9%)	7% (3%)
Tyneside B	65% (7%)	10% (6%)	4% (1%)	2% (2%)	13% (10%)	6% (2%)

Table 4-1: Average Percentage Occupancy and Standard Deviation per Workstation

Table 4-1 describes the average time dedicated to each behaviour for each workstation in the study. The standard deviation (SD) is also given and the high values for these illustrate the variability of the data on workstations.

The circumstances on the workstations during the observations were recorded (Appendix I). Out of the 24 observations, 13 had entirely smooth running with no problems whatsoever, eight had minor problems which the signallers stated had little or no effect on their work, and three had more major problems which had a slight effect on their work; late running following earlier failure of the overhead line equipment (OLE)[11], a track circuit[12] which was not operating correctly, and some major congestion due to trespassers further along the railway. Figure 4-8 shows the results of the observations ordered by these three groupings. This graph clearly shows that the three observations with some disruption did not have the highest monitoring or intervention levels. It is likely that the disruption would have had some effect on the observed behaviour of the signaller, but that effect was not large enough for these observations to be prominent.

[11] OLE supplies power to trains through electric wires strung above the railway.

[12] Track circuits use an electric current through the rails to detect the presence of a train (Appendix C).

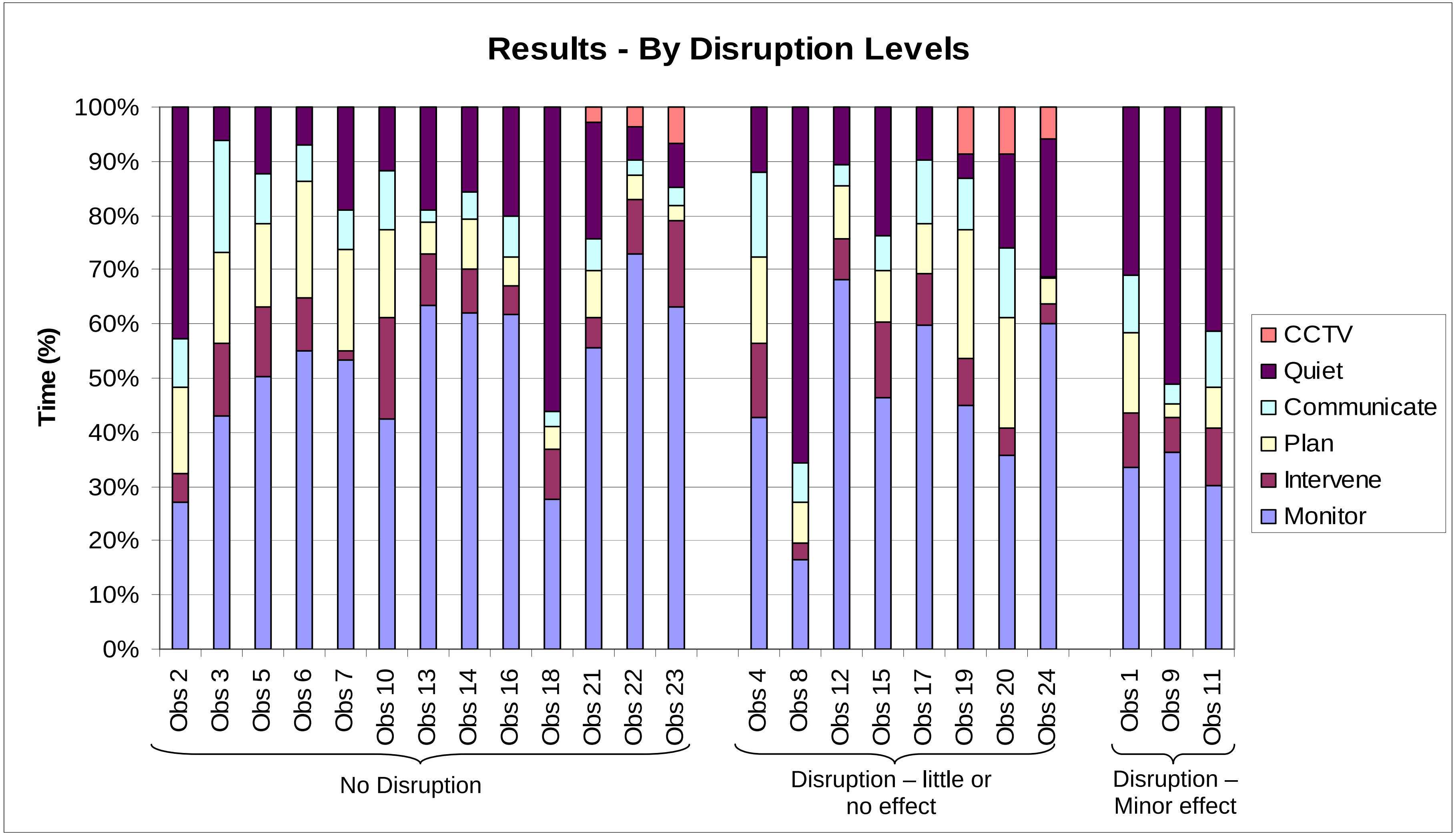


Figure 4-8: Results by Disruption Levels

4.4.2 Monitoring

The mean percentage of time spent monitoring was 48%, a maximum of 73% and a minimum of 16% of the total 90min observation. Two types of monitoring behaviour were identified as a result of the pilot study and were coded during the remainder of the studies; active monitoring and passive monitoring. Passive monitoring was typically carried out for longer periods of time than active monitoring; the average length of time spent passively monitoring was 27s and the average length of time spent actively monitoring was 13s. The proportion of passive monitoring was higher than that of active monitoring, with means of 27% and 21% of total time respectively.

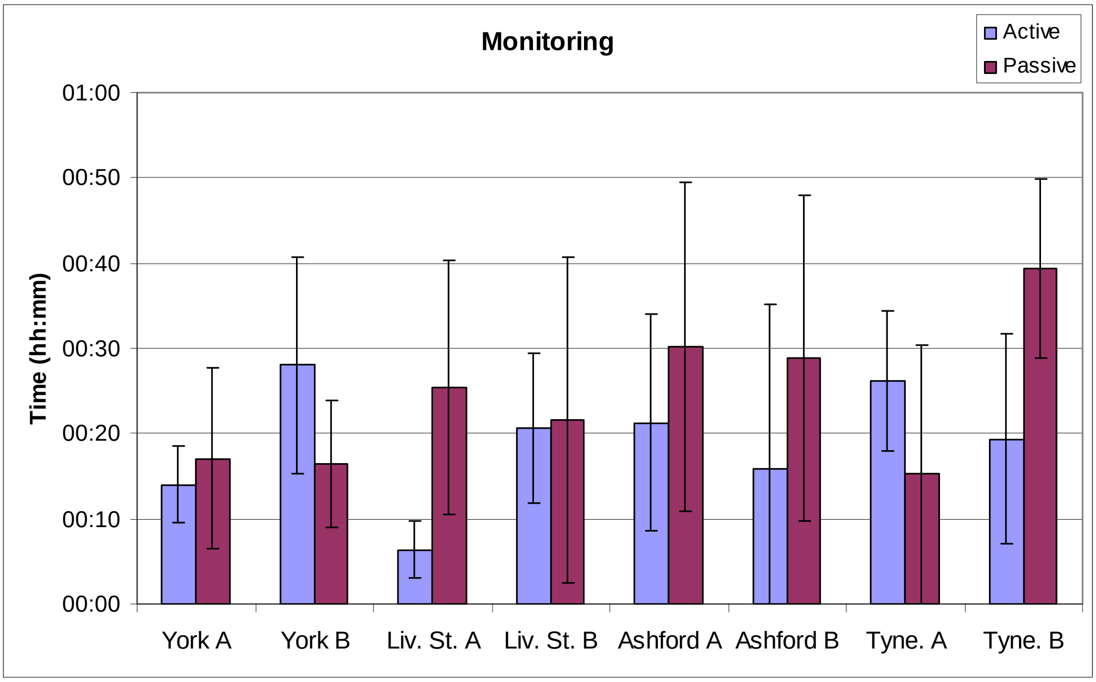


Figure 4-9: Monitoring Results

Figure 4-9 shows the average level of active and passive monitoring observed on each workstation in the study. The SD is also shown and the high SD values suggest that the individual rather than the workstation drives the monitoring level. As the observations were carried out at the same time of day the traffic encountered should have been very nearly identical. Although some of the workstations do show comparable monitoring levels (e.g. Obs 5 and Obs 6, see Figure 4-7), in the light of the other data gathered it is likely that this is a coincidence.

4.4.3 Intervention

The average percentage time spent intervening over the course of an observation was 9%, the maximum was 19% and the minimum 1%. Two types of intervention were coded, use of the trackerball and use of the keyboard. The trackerball allows the signaller to set routes and other directive activities. These activities can also be achieved through the keyboard, but the keyboard may also be used to query ARS or to look up timetable information. As can be seen from the Figure 4-10, use of the trackerball was considerably higher than use of keyboard. The average time for an intervention with trackerball (8s) was only slightly longer than keyboard interventions (6s). Overall, use of the keyboard was very low as compared to use of the trackerball, but use was highest in York (Obs 1 - 6).

Figure 4-10 shows the mean and SD of trackerball and keyboard use for each workstation in the study. Similar to monitoring, intervention levels differed greatly between individuals, as can be seen by the high SD for trackerball use in Figure 4-10. Since the three observations for each workstation were conducted at the same time of day, the train running pattern should have been almost identical and thus, barring any incidents or infrastructure problems, the workload and tasks encountered should have been very similar. An increase could be seen on workstations which experienced incidents during the course of the observation but even then these were not the highest observed levels of intervention.

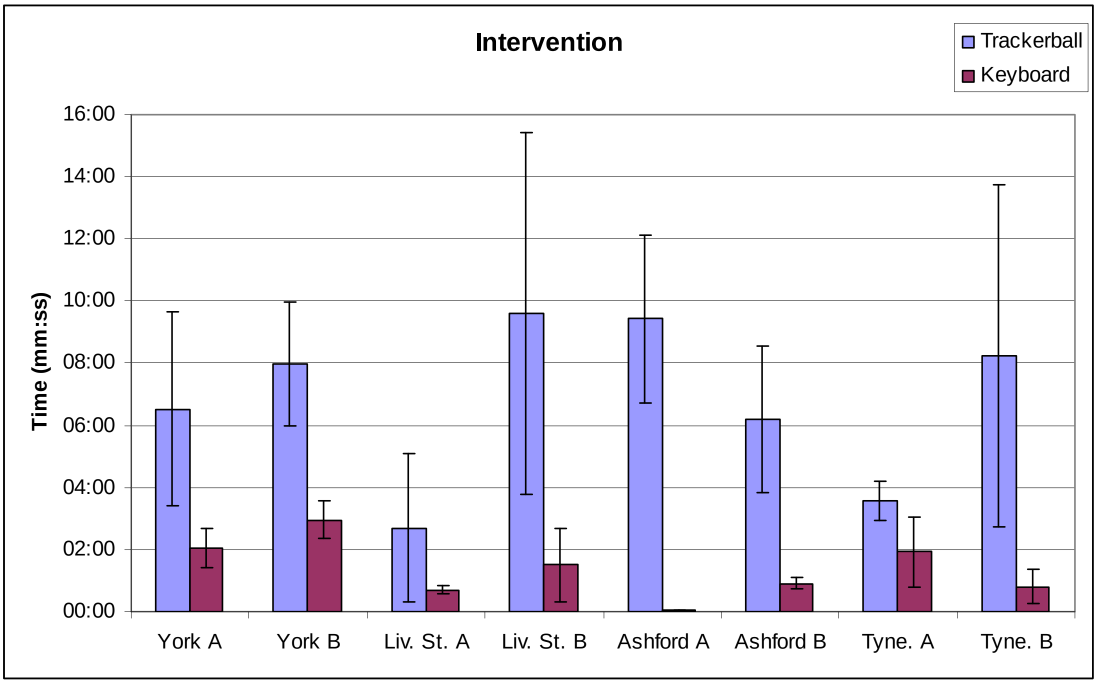


Figure 4-10: Intervention Results

4.4.4 Planning

The average percentage of total time spent occupied with planning activities was 11%, with a maximum of 23% and a minimum of 3%. The distribution of planning activities varied greatly across different IECCs, as shown in Figure 4-11. Use of the simplifier was highest in York, use of CCF was highest in Liverpool Street and both Ashford workstations and Darlington showed much lower CCF and TRUST use than the others. Use of TRUST was highest on the York A workstation in York, and why this was greatly reduced on the York B workstation is not known. Paperwork and reading ARS output occupied very little of the signallers' time overall. The average length of time for an individual planning activity was 9s for use of the simplifier, 20s for TRUST, 17s for CCF, 22s for paperwork and 5s for reading ARS output.

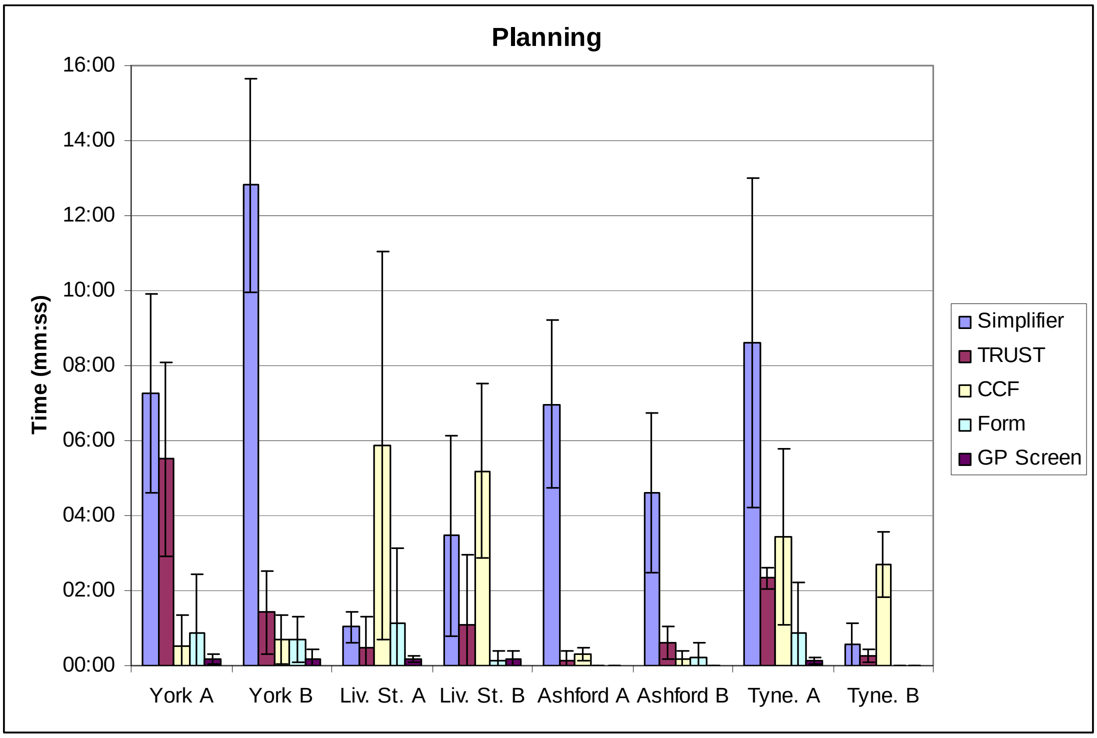


Figure 4-11: Planning Results

Planning is the only area where a difference between the IECCs can be seen. In particular the use of TRUST and CCF varied greatly between boxes. All four IECCs had both systems available to them but in Liverpool Street the TRUST terminal was shared between two workstations and so required the signaller to move away from his workstation slightly, whereas in York and Ashford both CCF and TRUST were available on the workstation. The use of TRUST was highest in York, perhaps because the signallers there have had access to TRUST for longer than CCF and so are more used to using it. The use of CCF was highest in Liverpool Street, probably because it is available on the workstation and TRUST is not. Signallers in Ashford made little use of either system; the reason for this is not known but may reflect the lower complexity of the workstation. Tyneside showed a big difference in the use of planning tools between the two workstations. Tyneside B had the lowest use of planning tools in the whole study and this may be due to the lower complexity of that workstation. Tyneside A, however, had the highest use of planning tools. The observed use of planning tools is summarised in Table 4-2

	Simplifier	*CCF*	*TRUST*
York	High use – printed off daily and used to record passage of trains	Low use – some non use	Varied use between workstations
Liverpool Street	Low use – reference only	High use – most utilised planning tool	Low use – some non use
Ashford	Medium use – reference only	Low use – some non use	Low use – some non use
Tyneside	Varied use between workstations. Reference only	Medium use	Varied use between workstations

Table 4-2: Summary of Planning Characteristics

4.4.5 Communications

The average percentage of time overall spent on communications was 8%, with a maximum of 21% and a minimum of 0.4%. The average time of an individual telephone call was 28s, whereas conversations with other signallers on adjacent workstations were 9s on average and the average use of the intercom was 30s. Only signallers in York had an intercom which allowed them to talk to Control, and this was used much more on the York South workstation (Obs 1-3), probably because of the presence of York Station on that workstation, requiring more coordination with Control. The majority of communications were carried out over the telephone. Conversations between signallers on adjacent workstations varied across the IECCs but were highest in York.

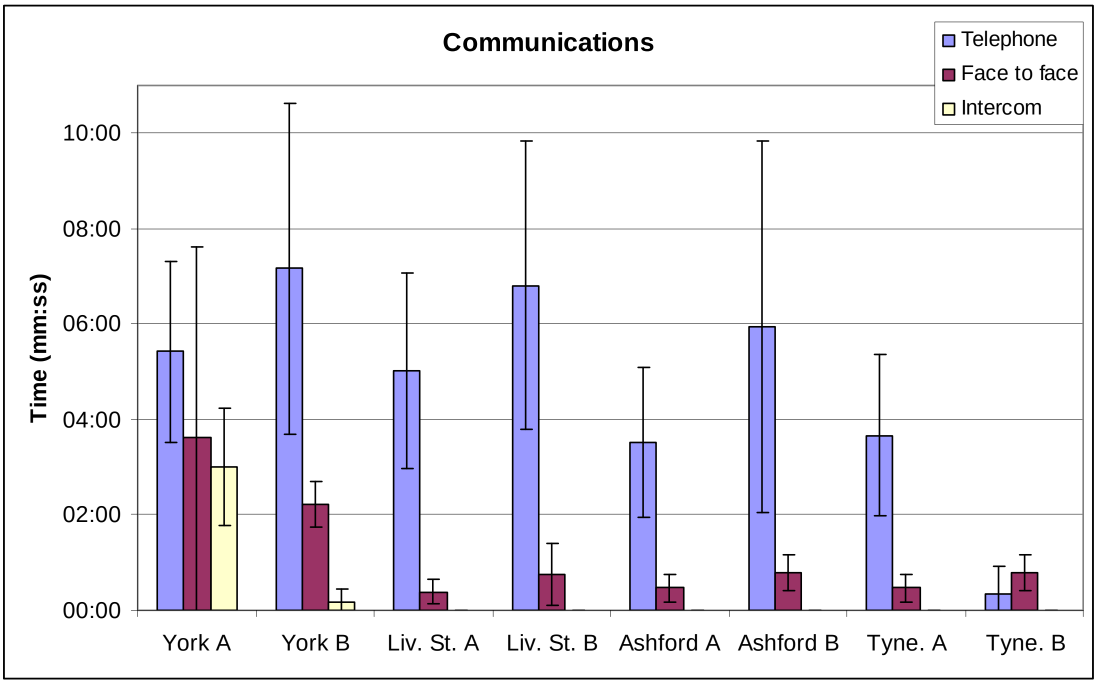


Figure 4-12: Communications Results

4.4.6 Quiet Time

The average percentage of overall quiet time during the observations was 22%, with a maximum of 66% and a minimum of 4%. Two types of quiet time were coded, time spent at the workstation not actively involved in signalling, and time spent away from the workstation. Individual quiet periods at the workstation lasted for an average of 20s, compared to 1min for quiet periods away from the workstation. The longest quiet period spent at the desk without monitoring or engaging in any other signalling activity was 4min 35s, but this was an unusually long time, the next highest time was 2min 55s. As can be seen from Figure 4-13, quiet time at the workstation was considerably more common than quiet time away from the workstation. The longest time spent away from the workstation by any of the observed signallers was 2min 55s.

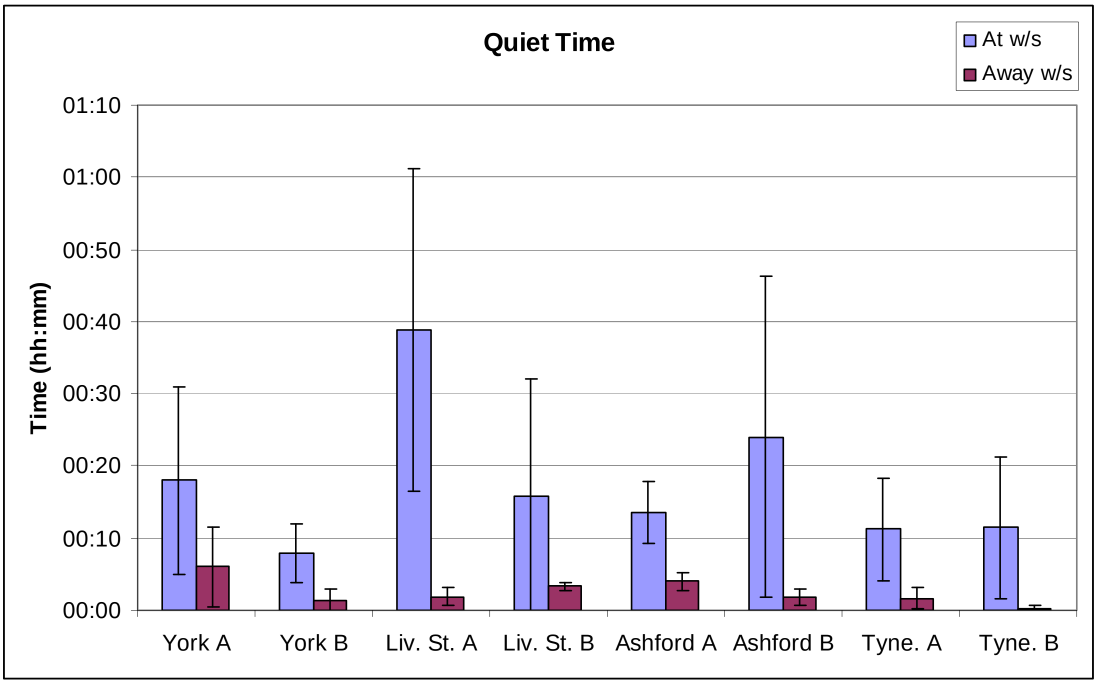


Figure 4-13: Quiet Time Results

4.4.7 Closed Circuit Television Operation

Closed Circuit Television screens required to operate level crossings were only present on the workstations in Tyneside IECC. Signallers were required to lower the barriers for each train and confirm that the crossing was clear for trains to pass. This occupied a reasonable chunk of the signallers' time on these workstations, between 3% and 9% of the total observation time.

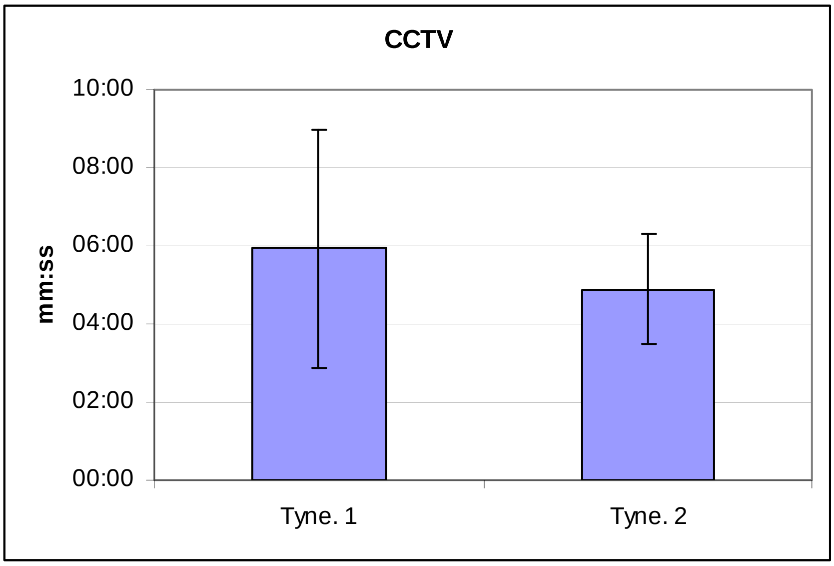


Figure 4-14: CCTV Results

4.4.8 Trust

The results of the questionnaires examining each signaller's perceived trust in the automation are described in this section. Although there were 24 observations, three signallers were observed twice but only completed the questionnaire after one observation. Therefore there were 21 questionnaire respondents. The results of the questionnaire are shown in Table 4-3 and the means for each question are graphed in Figure 4-15.

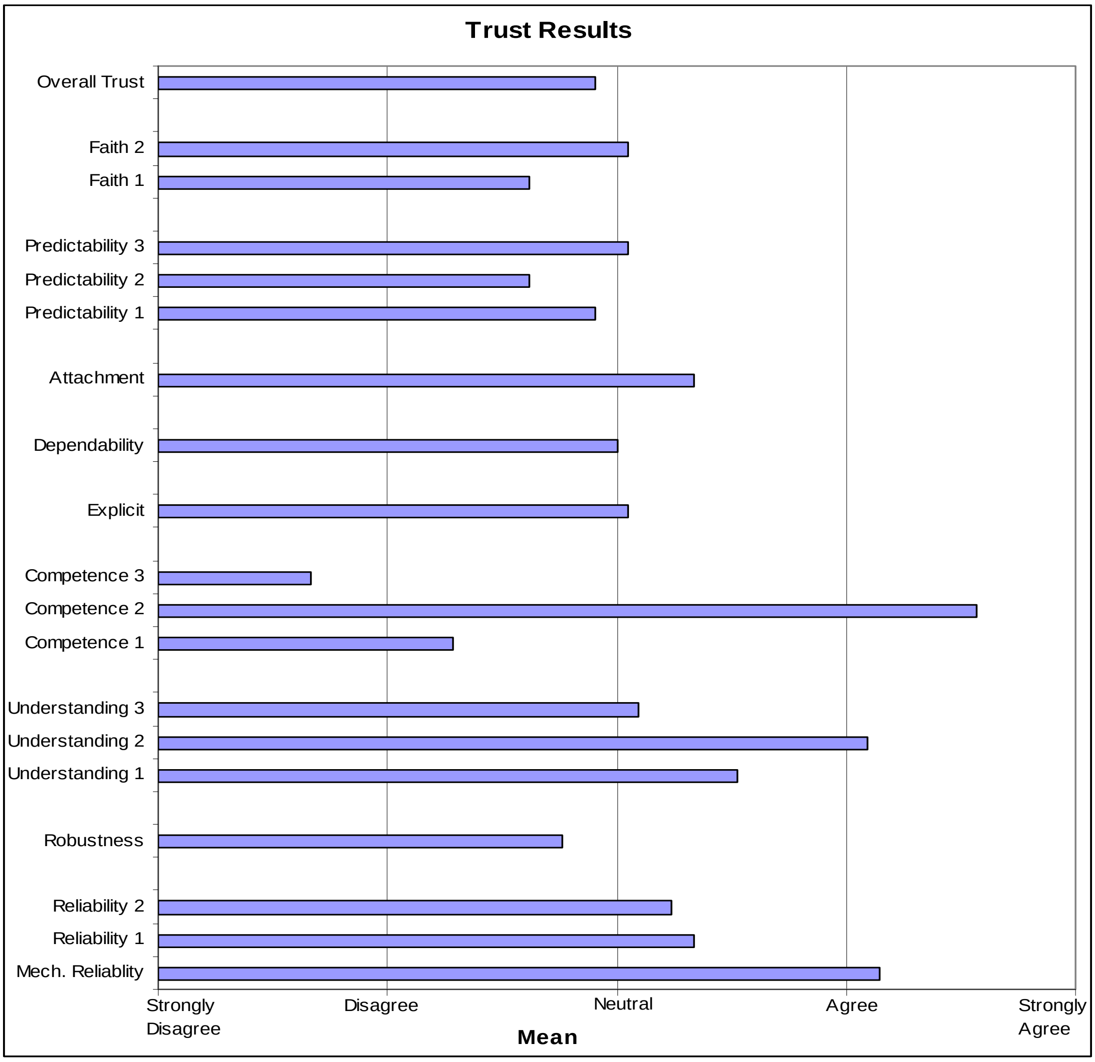


Figure 4-15: Trust Questionnaire Results

Question	*Name*	*Mean*	*SD*
ARS is always available for use	Mech. Reliability	4.14	1.11
I can rely on ARS to function as it is supposed to	Reliability 1	3.33	0.73
ARS will always make the same routing decision under the same circumstances	Reliability 2	3.23	1.78
ARS is capable of performing under a variety of different circumstances	Robustness	2.76	0.89
It is easy to understand what ARS does	Understanding 1	3.52	1.08
I understand how ARS works	Understanding 2	4.10	1.00
I understand why ARS makes the decisions it does	Understanding 3	3.10	1.09
ARS is capable of signalling trains as competently as a signaller	Competence 1	2.29	1.10
ARS performs well under normal running conditions	Competence 2	4.57	0.51
ARS performs well under disturbed conditions	Competence 3	1.67	1.02
ARS gives explicit information on its intended actions	Explicit	3.05	0.92
I can count on ARS to do its job	Dependability	3.00	0.71
I have a personal preference for using ARS	Attachment	3.33	1.11
I can predict what ARS will do from moment to moment	Predictability 1	2.90	1.34
ARS is very unpredictable, I never know what it is going to do	Predictability 2	2.62	0.92
ARS is very consistent	Predictability 3	3.05	1.07
If ARS makes a routing decision which I am uncertain about I have confidence that ARS is correct	Faith 1	2.62	1.12
Even if I have no reason to expect that ARS will be able to deal with a situation, I still feel certain that it will	Faith 2	3.05	0.76
I trust ARS	Overall Trust	2.90	1.00

Table 4-3: Trust Questionnaire Results

In order to analyse further the results of the questionnaires, the signallers observed were divided into groups of high, medium and low in terms of monitoring, intervention, and quiet time. Planning and communications were not analysed as levels for these were likely to be affected by factors outside of the signallers' direct control. Each observation was compared to the other two observations on the same workstation to determine the groupings. Two observations (Obs 1 and Obs 9) were excluded from this part of the study on the basis that there was significant disruption on these workstations during the observations and this may have affected the observed levels of each activity. These exclusions were in addition to Obs 3, Obs 11 and Obs 23 which were omitted as the signaller in each of these had already been observed, and therefore had previously completed the trust questionnaire. The sample size was therefore 19. As data gathered using Likert scales can be regarded as pseudo-interval data (Tabachnick & Fidell, 2007), t-tests were run between the high and low groups in each category to test for significant differences between the two groups.

No differences were found in terms of monitoring, but a number of differences were found in terms of intervention:

- Explication of intention – “ARS gives explicit information on its intended actions”, $t(11)=2.385$, $p<.05$. Low interveners were more likely to agree with this statement.
- Understandability 2 – “I understand how ARS works”, $t(11)=2.851$, $p<.05$. Low interveners were more likely to agree with this statement.
- Predictability 2 – “ARS is very unpredictable; I never know what it is going to do”, $t(11)=-2.337$, $p<.05$. Low interveners were less likely to agree with this statement.
- Reliability 1 – “I can rely on ARS to function as it is supposed to”, $t(11)=2.434$, $p<.05$. Low interveners were more likely to agree with this statement.
- Faith 2 – “Even if I have no reason to expect that ARS will be able to deal with a situation, I still feel certain that it will”, $t(10)=2.373$, $p<.05$. Low interveners were more likely to agree with this statement.
- Understandability 3 – “I understand why ARS makes the decisions it does”, $t(11)=2.782$, $p<.05$. Low interveners were more likely to agree with this statement.
- Overall trust – “I trust ARS”, $t(11)=2.478$, $p<.05$. Low interveners were more likely to agree with this statement.

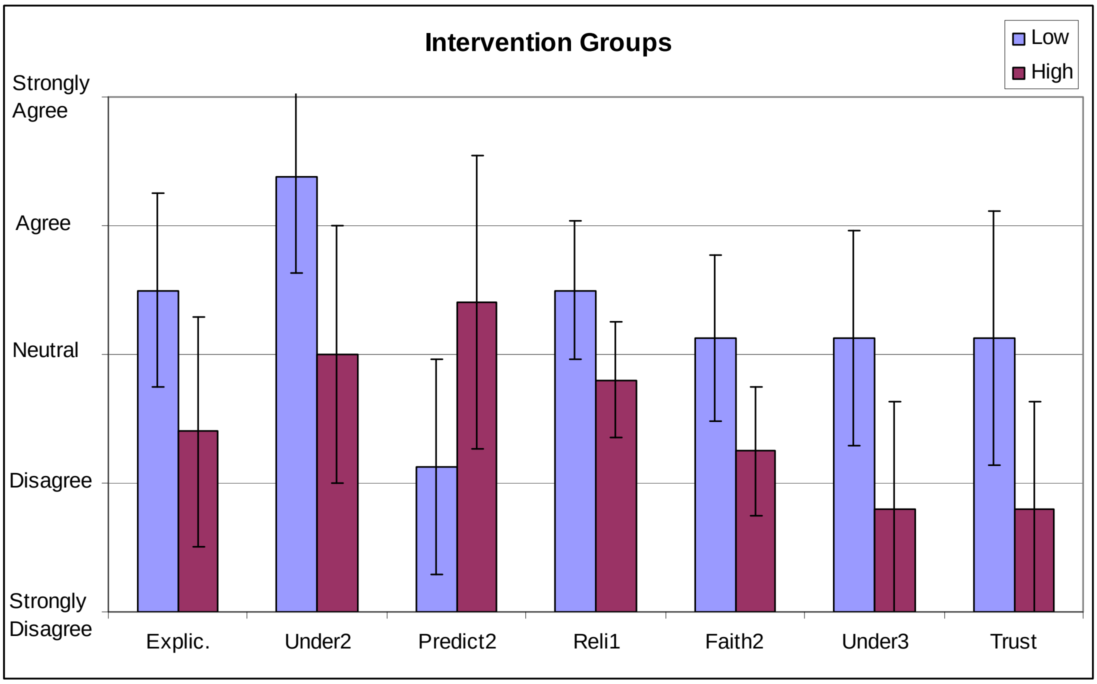


Figure 4-16: Significant Differences for Intervention Groups

A difference was also found between the groups for the overall understandability dimension ($t(11)=2.571$, $p<.05$) with low interveners rating their understanding of the automation higher.

Two significant differences were also found between signallers engaging in high and low levels of quiet time:

- Understandability 1 – “It is easy to understand what ARS does”, $t(12)=-2.178$, $p<.05$. Signallers displaying high levels of quiet time were more likely to agree with this statement.
- Faith 1 – “If ARS makes a routing decision which I am uncertain about I have confidence that ARS is correct”, $t(12)=-2.756$, $p<.05$. Signallers displaying high levels of quiet time were more likely to agree with this statement.

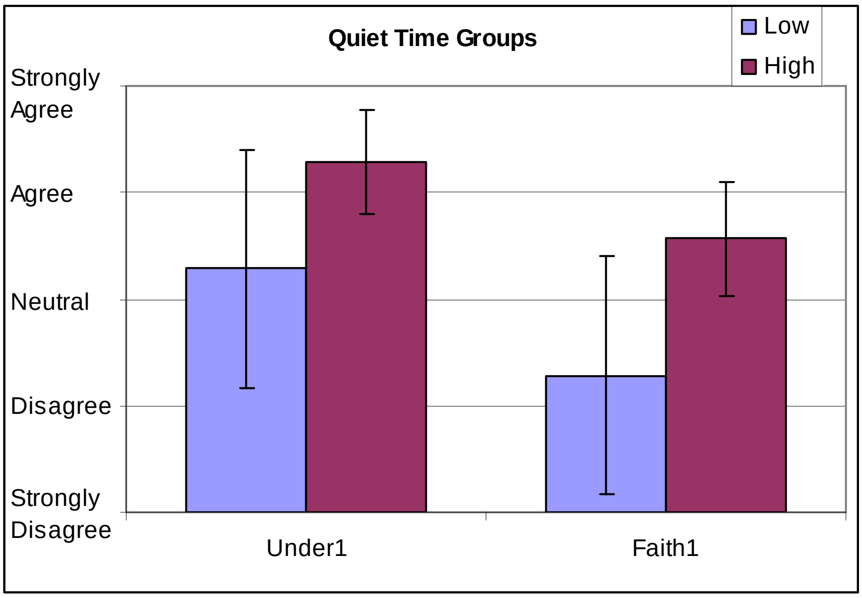


Figure 4-17: Significant Differences for Quiet Time Groups

4.4.9 Inter-Observer Reliability

Inter-observer reliability is the extent to which the results of two or more observers of the same situation agree (Robson, 2002). Cohen's Kappa (*K*) can be used to determine the level of inter-observer reliability. Cohen's Kappa uses a proportion of agreement (P_o) between observers (i.e. the proportion of occasions when the observers used the same code for the same time interval) and a proportion of chance (P_c) (i.e. the probability of both observers using the same code at the same time) to assess the level of inter-observer reliability. The formula for Cohen's Kappa is:

$$K = \frac{P_o - P_c}{1 - P_c}$$

Robson (2002) gave the following rules of thumb for interpreting the results:

K = 0.40 – 0.60: fair inter-observer agreement
K = 0.60 – 0.75: good inter-observer agreement
K > 0.75: excellent inter-observer agreement

Four of the observations were undertaken by two different observers to determine inter-observer reliability for the method. There was a crucial difference between the observers as the first observer had considerably more knowledge of the signalling task than the second. Figure 4-18 shows the comparison of results between the observers.

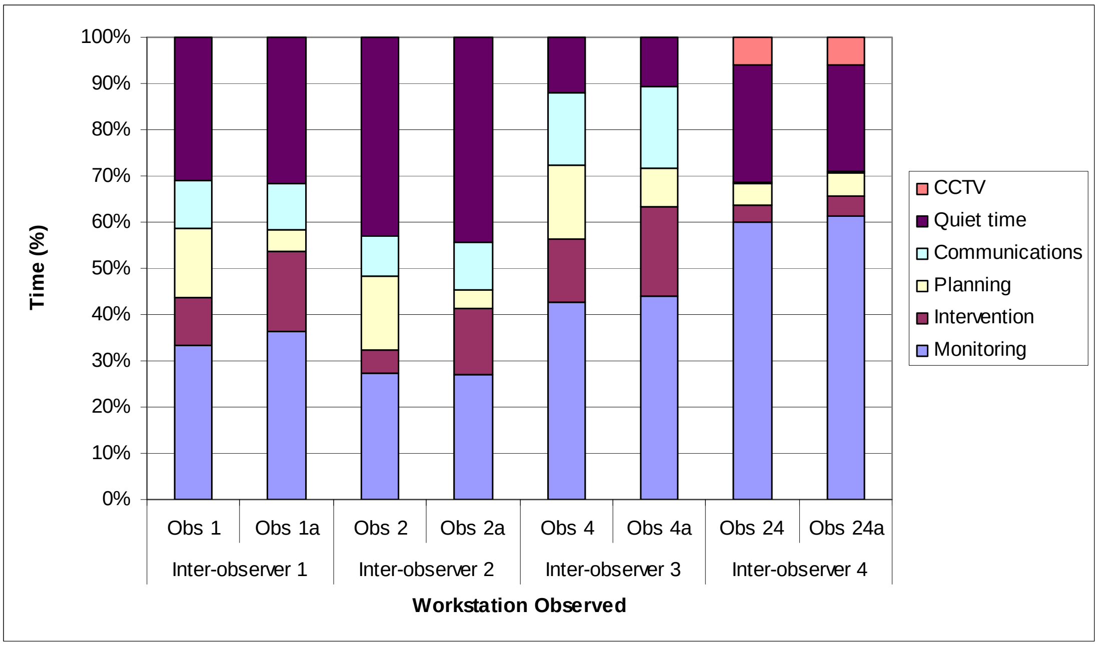


Figure 4-18: Inter-Observer Comparison

The proportion of agreement for each 5s block and Cohen's Kappa for each of the inter-observer reliability studies is shown in Table 4-4.

	Proportion of Agreement	*Cohen's Kappa*
Inter-observer 1	67%	55%
Inter-observer 2	66%	52%
Inter-observer 3	65%	52%
Inter-observer 4	80%	65%

Table 4-4: Inter-observer Statistics

4.5 Discussion

4.5.1 Monitoring

Monitoring behaviour showed considerable variation between observed signallers despite the observations on individual workstations experiencing similar conditions. It seems likely therefore that monitoring levels under normal running are driven by the individual rather than the workstation.

Two types of monitoring were identified and coded during the study; active monitoring and passive monitoring. No reference to different states or levels of monitoring was found in the literature on automation. However, a similar concept arose in the discussion on differences between ‘active control’ and ‘passive monitoring’ (Endsley & Kiris, 1995) where it was suggested that automation induced passive processing of information which was inferior to active. Both active and passive monitoring were frequently engaged in throughout the study but active monitoring was more common between interventions while passive monitoring was associated more with quiet time. It seems likely that rather than signallers working with automation constantly suffering from inferior information processing as is suggested in the literature, they actually actively process information when they believe decisions may be required and engage in a more relaxed form of monitoring (passive monitoring) when they feel the demands of the workstation are lesser. Although it was not possible to determine how much attention the signaller paid during passive monitoring, there are frequent examples of interventions following a period of passive monitoring so it can be concluded that information is still being processed. Cowan (1988) suggests that attention can be automatically triggered even during passive information processing states. If this is the case, the use of different levels of monitoring behaviour may be a very effective strategy for reducing workload associated with monitoring, which has been shown to be high (Warm et al., 1996), while still maintaining awareness of the system. There are probably more types of monitoring behaviour to be observed in other signallers, indeed one signaller in the study monitored the system through CCF (the map-based planning tool), but as monitoring was not the singular focus of this study the number of coded monitoring behaviours was reduced to two.

High levels of monitoring (either active or passive) were associated with low levels of quiet time, and low monitoring was associated with high levels of quiet time. This would seem to suggest that monitoring and quiet time are interchangeable and signallers who can find a ‘distraction’ (i.e. someone to talk to in most cases) will use the time otherwise used for monitoring purposes. There is probably a lower threshold of monitoring below which signallers would feel uncomfortable, but further research would be required to identify what this might be. However, establishing that boundary would contribute towards understanding the necessary levels of operator awareness of the system. The association of active monitoring with interventions and passive monitoring with quiet time provides some validation of the decision to differentiate between the two types of monitoring. More research is required to look at monitoring behaviours alone to identify what triggers each one and the quantity, quality and type of information gathered at each level.

It is important to note that all observed signallers engaged in routine monitoring behaviour and the longest observed period when they were away from the workstation and could not monitor it was just under 3min. It would appear that monitoring is a critical ongoing activity for the signallers which they are not willing to neglect. This suggests that they place a high priority on maintaining awareness of their control area, even when ARS is running all trains.

4.5.2 Intervention

The results show a high degree of variation in intervention levels between signallers. Unfortunately it is not known why some observations had higher intervention levels than others. It is clear that it is not totally due to particular circumstances on the workstation, as the events experienced should have been very similar, but whether the difference in the times is due to some signallers intervening in more circumstances than others or to some signallers using more efficient strategies is unclear. For example, perhaps some signallers intervened in advance and were thus able to deal with problems quite quickly; perhaps even making some interventions to prevent a situation developing, or some interventions may be more efficient than others (e.g. STP[13] as compared to manual route setting). Therefore, like monitoring levels, intervention levels appear to be driven by the individual rather than the workstation and research supports the theory that individual differences may account for these differences (Merritt & Ilgen, 2008).

The trackerball was used far more often than the keyboard, probably because it represents a more direct method to interface with the system, in a similar manner to the common preference for mouse to keyboard commands in PC usage (Krisler & Alterman, 2008). Use of the keyboard was highest in York IECC but as this study recorded only intervention levels rather than the purpose of the interventions, the

[13] STP – Special Timing Pattern. STPs are pre-programmed timetables for common routes across the workstations which can be applied to trains which have no entry in the timetable.

reason for this difference is not known. Further research could investigate common interventions, how they are achieved and the length of time required for each.

Following the observations, signallers were asked for the most common reasons for interventions during the observation period. Common reasons were to regulate due to late running (most often achieved with reminders), to facilitate permissive working[14], to ensure routes for trains were set in the correct order, and to route trains out of a depot. Some signallers also reported that they intervened to set manual routes because they did not trust ARS to do it. Setting manual routes using the trackerball and applying reminders to control ARS were the most common forms of intervention. This reflects the ease of use of these methods as compared to others. It was not possible to pursue this information further as part of this study, but these themes were picked up in the subsequent interviews which will be presented in Chapter 5.

4.5.3 Planning

The use of four different planning tools was coded. It was impossible to tell when the signallers were looking up information for immediate use, for background information or simply as something to do. There was an initial assumption that signallers would only look up information that applied to their signalling goals, even if it was not strictly necessary. However, a signaller in one observation reported that he was looking up the running times of the Eurostar. He was not involved in any aspect of signalling the Eurostar trains and only enquired about them to pass the time. Other signallers were observed to query specific trains which they or their colleagues intended travelling home on.

Signallers in York had generally higher use of the simplifier; this was because they have a habit in York of going through the simplifier and crossing off trains as they pass through the workstation. This is not a requirement but every signaller observed in York did this, while no other signaller involved in the study did. Other signal boxes use the simplifier as a reference and do not mark it, in fact in other signal boxes the simplifier is kept in plastic sheets whereas York signallers print off a new copy each day. The advantage of striking trains off the simplifier is not clear, but one advantage of having a disposable copy was that any alterations to train running could be made

[14] Permissive working refers to two trains occupying the same section of track, usually a platform, at the same time; ARS is restricted from setting these routes.

on the simplifier, whereas signallers in other IECCs had to make a note of changes separately.

Reading ARS output (i.e. GP Screen) and filling in forms or TRB were very infrequent activities and accounted for very little of the signaller's time. The North Kent 1 workstation in Ashford is not an IECC workstation and it does not actually have a GP screen so the signallers have to exit a signalling screen and call up a separate screen to interrogate the ARS. This activity was not observed at all on this workstation suggesting that having the ability to interrogate ARS is 'nice to have' but signallers are not willing to go out of their way to use it, possibly because the quality of the information it supplies is poor.

4.5.4 Communications

Telephone communications are not within the direct control of the signaller; they are very much influenced by the events on the workstation at a particular time. Communications are not potentially influenced by the automation in the way that some of the other activities are and for both these reasons a detailed discussion is not required. However, increased voice communications between adjacent signallers indicates a perceived need to pass each other information on train running. This need should not exist, or should be greatly reduced, if the automation is running well. It appears that a bigger factor in the levels of voice communications is the control room layout; voice communications were greatly increased in York as compared to the other IECCs and the most obvious explanation for this is the relative openness of York IECC. Tyneside B (Darlington - Obs 22-24) had the lowest communications, probably reflecting the lower complexity of that area.

4.5.5 Quiet Time

Activities that were coded as quiet time included chatting to fellow signallers, looking at mobile phones, drawing up rosters, checking email, staring at fingernails, reading (letters, books, newspapers, magazines), tidying the workstation, searching in bags or drawers, doodling, eating, or simply looking around the control centre. Some signallers also found it impossible to ignore the researcher and time spent talking to the researcher was coded as quiet time. The researcher was careful not to initiate any conversations with the signallers and so any conversations were initiated by the signaller at times when he clearly felt he had time to spare. Although it was preferable for the signaller not to speak to the researcher, the ongoing nature of this research meant that the researcher could not afford to alienate any potential participants. Two signallers in particular spent a good deal of time talking to the researcher, these were Obs 8 and Obs 18 (see Figure 4-7) and this is reflected in their high proportion of quiet time.

Quiet time was largely interchangeable with monitoring time; there are many instances in the data where the coding flicked between monitoring, particularly passive monitoring, and quiet time. This indicates that when signallers are monitoring the system it is not always strictly necessary and sometimes they may be doing it in the absence of anything else to do. Another indication that monitoring and quiet time are interchangeable is that signallers who had a low monitoring percentage had a high quiet time, and vice versa.

Signallers commonly left the workstation to have a tea, cigarette or toilet break. Sometimes a relief signaller took over for these, but particularly for tea breaks the workstations were frequently left unattended. The observed signaller rarely moved away from his workstation to have a chat with a signaller on another workstation but visits from other signallers on duty were frequent and gatherings of signallers around other workstations were noted during the observations. The act of being observed may have prevented the observed signaller from engaging in this behaviour when normally they would have spent considerably more time away from their own workstation.

The level of quiet time is probably quite dependent on the individual signaller's personality. More sociable and outgoing signallers probably spend more time chatting to other signallers both at their own workstation and away from it. More introverted signallers may spend the time monitoring, when perhaps it is not strictly necessary. Although signallers were requested to ignore the researcher, different signallers were comfortable with this to a greater or lesser degree. It is impossible to tell if they behaved exactly as they would have if the researcher had not been present but some signallers were capable of not speaking to the researcher at all. Some signallers spoke during quiet periods when they were obviously bored, and one or two signallers were not capable of ignoring the researcher at all and engaged in conversation throughout the observation.

4.5.6 Closed Circuit Television Operation

As CCTV was only present in Tyneside it is not possible to compare operation times in different IECCs. However, it is unlikely that there is any interaction between this task and ARS so the data are not particularly relevant to this study and were gathered for thoroughness only.

4.5.7 Trust

4.5.7.1 Reliability

The perceived mechanical reliability of ARS was considered high with most signallers reporting that it is always available for use. One of the signallers who disagreed with this statement was in Ashford IECC and the ARS had been unavailable on that workstation the previous day, but this is a very rare event and as there is redundancy in the ARS system it was only unavailable for a short period. The responses to this question were as expected, reflecting the very high mechanical reliability of the automation, as would be expected in a safety critical industry.

There was general agreement that ARS could be relied upon to function as it is supposed to, with only two signallers disagreeing with this statement. However, when asked whether they agreed that ARS will always make the same routing decision under the same circumstances there was a higher level of disagreement among the signallers (six compared to two). Although there was more agreement than disagreement with this statement, six signallers felt that ARS does not always make the same routing decision under the same circumstances. As with any computer, ARS follows rules to arrive at its decisions and so it will always make the same decision under the same circumstances. That some signallers disagreed with this statement indicates that they may use different criteria in their decision making and they probably do not fully understand the factors on which ARS bases its decisions. This is likely to be due to poor feedback and has strong implications for the predictability of the automation.

4.5.7.2 Robustness

Only one question examined robustness, whether ARS is capable of performing under a variety of circumstances. Just five signallers agreed with this statement reflecting ARS's inability to perform efficiently when the railway is disrupted. This was expected as ARS performs well when all trains are running to timetable and there are no incidents but this leaves a wide variety of circumstances when the signaller is required to step in. However the impact of this upon trust may not be very great, signallers simply calibrate their trust accordingly by developing a set of situations that they trust ARS with and they assume manual control or inhibit the automation for other situations.

4.5.7.3 Understandability

Three questions examined signallers' perceived ability to understand the automation. There was general agreement with the first statement 'It is easy to understand what ARS does' with only four signallers disagreeing. Signallers also agreed with the statement 'I understand how ARS works'. The final statement was 'I understand why ARS makes the decisions it does' and although overall the response to this question was positive, there was a higher negative response rate than in other questions on understandability. This indicates that signallers may have a lower understanding of why ARS makes certain decisions than of what it is doing and how it does it. It may be that the basic operation of ARS is well understood by signallers but the more complex area of ARS decision making and conflict resolution is less well understood. Understanding why the automation makes the decisions it does is fundamentally important for predicting and controlling the automation and lack of understanding is likely to impact strongly upon overall trust and use of automation (Lee, 1991). Of course, the questionnaire only examined perceived understanding; actual understanding may be different to, and perhaps lower than, perceived understanding.

4.5.7.4 Competence

Signallers universally agreed that ARS performs well under normal running conditions, and this reflects very positively on ARS's abilities during normal operations. There was a strongly negative response to the disturbed conditions question, which is not unexpected as ARS is widely reported to work less well under disturbed conditions. These two questions clearly illustrate the divide in ARS' abilities between normal and disturbed conditions.

The majority of the signallers disagreed with the statement comparing ARS competence to signaller competence (Q4). This is not surprising, both because ARS reportedly does not perform well under disturbed conditions, and because to agree would diminish their own standing and undermine their employment. In this light, it is surprising that there was not stronger disagreement with this statement. The statement could perhaps have been clearer with respect to what 'signalling trains' exactly referred to. However, overall the responses were negative reflecting the inability of ARS to signal trains in a variety of circumstances, but also reflecting the confidence signallers have in their own abilities.

As with robustness, the impact of the lack of competence of the automation in some situations may not impact overly on trust as long as the signallers are aware of the automations strengths and weaknesses and are able to take control when a situation arises that ARS cannot deal with (Lee & Moray, 1994; Madhavan & Wiegmann, 2007).

4.5.7.5 Explication of Intention

The results for this question examining feedback were noncommittal, with the majority of signallers opting for a neutral response. The question may not have been well phrased or the signallers may indeed have felt neutral regarding the feedback they get from ARS. This could be explained by a lack of imagination on how information could be fed back better.

ARS does not automatically give information on its intentions but it is possible to query ARS to determine which train it intends to route next at a junction. The feedback of such information could certainly be better supported and improving the feedback should have the result of increasing understanding and predictability, thereby improving trust and allowing the signallers to better control the automation (Dzindolet et al., 2003).

4.5.7.6 Dependability

Again the majority of responses to this question were neutral with an equal number of agreements and disagreements, but no signallers chose to score strong agreement or disagreement. In this case the phrasing of the question seems less likely to have been an issue and so it would appear that signallers were genuinely noncommittal regarding this question. It more likely reflects the discrepancy in ARS performance between normal and disturbed conditions and potentially some confusion about what

ARS's job is. In as far as it routes trains it can be counted on, but some signallers may expect more from it and resent how often they are required to step in. If signallers are expecting more from it than is actually delivered there is bound to be an impact on trust (Merritt & Ilgen, 2008).

4.5.7.7 Personal Attachment

Again this question had a high neutral response but there was a tendency in the results towards agreement. It would appear then that most of the signallers in the study were at least happy to use ARS, with two from York and one from Liverpool Street the only signallers who indicated otherwise. It is interesting to note that the only observed signaller who did not use ARS was in Ashford IECC but he did not disagree with this statement.

4.5.7.8 Predictability

The responses to the three questions on signallers' ability to predict ARS varied. The first (Q8) received a wide variety of responses with the most common being agreement, neutral and strong disagreement. No trend within the IECCs was apparent and so it seems that individual signallers' ability to predict ARS varies considerably. The high number of signallers who strongly disagreed with the statement does not reflect well on the predictability of the automation. The majority of signallers were neutral in their responses to the second question (Q12) but there was a skew in the responses towards disagreement. This suggests that signallers do find ARS to be somewhat predictable. The responses to the final question (Q17) were again largely neutral, but with a slight skew towards agreement. This again indicates that signallers find the system somewhat predictable. The combined results of the three questions suggest that signallers feel they can predict ARS most of the time but are not comfortable predicting its every action. Better feedback and understanding of the underlying logic should improve signallers' ability to predict the automation, but a display of the automation's future intentions would be the best solution.

4.5.7.9 Faith

Two questions investigated the signallers' faith in the automation, or belief that it will make the correct routing decisions. The interesting thing about the results for these questions was how positive the responses were. Although the overall response for the first was negative, six signallers reported that if ARS made a routing decision they were uncertain about they would have confidence that ARS was correct. These responses came from across all four IECCs. The responses to the second question (Q14) were predominantly neutral, perhaps indicating that signallers are not sure which situations ARS can deal with. Most likely in this situation, the signallers would not leave a train in ARS to see what happened but would take it out and signal it manually and thus they rarely test their knowledge of or faith in ARS's capabilities.

4.5.7.10 Overall Trust

Signallers replied to a general question on trust and again the responses were predominantly neutral but with three strong disagreements. This would appear to indicate that the overall trust in the system is low. It is an interesting situation because ARS runs on top of the interlocking, which the signallers do very much trust. Therefore, they can be sure that the interlocking will not allow ARS to make any grievous mistakes in terms of safety. So, while trust in the system may be low, this is likely to be in terms of efficiency and the safety risk is not very high. The impact of low trust is therefore not as strong as it might otherwise have been.

4.5.7.11 Differences between Groups

The results indicate that correlations can be found between the observed behaviours of signallers and their reported trust in ARS, particularly in relation to the amount of time spent intervening. Although the sample size was small, the direction of the differences between groups all indicate that lower trust results in higher intervention.

Differences were found in questions relating to feedback, understanding, predictability, reliability and faith. Apart from reliability and feedback (Dzindolet et al., 2003; Wiegmann et al., 2001), no other research is known to have found empirical evidence supporting the relationship of understanding, predictability, and faith to automation usage. The literature suggests a strong relationship between competence and automation usage (Muir & Moray, 1989), however this relationship was not found in this study. It is not known why this was. The responses to the questionnaires indicate that ARS is not a robust system (i.e. it is not competent during disruption). It may be that the perception of the competence of ARS is reasonably stable between signallers, but their ability to predict and understand it varies and it is this that differentiates their usage of the system.

4.5.8 Discussion of Method

The framework developed for this study was designed to classify all activities observed. This was achieved through the identification of five high level activities and in this way it differed from frameworks found in the literature which attempted to capture the purpose of the observed activities (Lenior, 1993; Reinach, 2006). However, some commonalities do exist, for example, communications were captured in all three frameworks and Reinach used a category of ‘actuating controls’ which is very similar to ‘intervention’.

The five main behaviours were comprehensive and easily distinguishable. However, the difficulty in determining the purpose behind interventions is a limitation of the method. Without expert knowledge of the signalling domain it may be difficult to interpret the events on the workstation, and even with expert knowledge it is not possible to fully understand the reasons behind each individual’s observed behaviours. This is compounded by the strict coding and unobtrusive nature of the method which prohibits asking the signallers for information on their behaviours and actions. Thus, although the method is reliable at recording observed behaviours, it is not sufficient to determine the reasons behind those behaviours. However, as with the framework developed by Reinach (2006), it may be used in conjunction with other methods to obtain a fuller picture of signaller workload. Alternatively, a less formal observation may give more insight into the reasons behind the observed behaviours, but it is also possible to use other methods, such as interviews, to gather these data.

The workstations involved in the study were carefully chosen to match the demands as closely as possible. However, it is impossible to precisely match workstations due to the variability of rail infrastructure and this limits the confidence which can be placed on conclusions drawn from comparisons of workstations. Finally, a high demand is placed on the researcher over the 90min period of observation due to the requirement to code the data at 5s intervals. Alternative methods of coding the data could be investigated to reduce this burden.

4.5.8.1 Inter-observer Reliability

The inter-observer reliability study undertaken demonstrated the reliability and validity of the method. Good inter-observer reliability is highly desirable as it demonstrates that the coding system developed is applied consistently.

The graph comparing the two observers (Figure 4-18) shows very similar results in terms of monitoring, communications and quiet time, but there is a discrepancy in terms of intervention and planning in the first three graphs (Inter-observer 1, 2 and 3). This can be explained by the difference in knowledge of the two observers. The more knowledgeable observer was aware of which input devices were attached to the signalling system and which were for interacting with the planning systems while the less experienced observer coded all use of input devices as intervention. Once this discrepancy is removed, as it was in the final inter-observer reliability study (Inter-observer 4), the inter-observer reliability appears very high.

The discrepancy in the first three observations is reflected in the statistics but nevertheless the statistics indicate reasonable inter-observer reliability. The final observation shows much improved results with good agreement between the observers. These statistics are lower than might be expected given the similarity of the graphs and one possible explanation for this is the frequency with which the data were coded. It was possible that there was a slight lag between the observers and this would have been picked up by the statistics which compare each 5s block of data, but the graphs would not have reflected this. Using the overall times for each of the five behaviours the percentage agreements between the observers were 78%, 76%, 83%, and 95% respectively.

4.5.8.2 Smoothness of Data

Graphs were developed to illustrate how the data changed over the period of the observations and to determine how smooth the data were over time (Appendix J). Although there are small changes visible over time in each observation, in general the data are remarkably smooth. The concern was that as the railway is a real time dynamic system the signallers' activities would be driven by occurrences on the railway and so their activities over time could differ greatly. This was not found to be the case, although small changes can be seen. This is an interesting result, as it reinforces the theory that individual signallers drive their own activities. It also suggests that shorter observations which would be less resource intensive both in the data collection and data analysis stages would yield valid results.

4.6 Conclusions

The main aim of this study was to develop a method to observe signallers at their work. The method developed used five basic activity or behaviour codes and resulted in graphs of the signallers' division of time across these five activities. Clear differences could be seen in different signallers using this method and it also showed good inter-observer reliability. The data were also shown to be quite stable over the 90min period, suggesting that shorter observations of perhaps an hour would be equally valid. Further research, perhaps using verbal protocols in addition to the observation method, would be required to identify the strategies underlying the graphs generated from the observations. The following chapters build on the data gathered during these observations and investigate the reasons behind signaller interventions and monitoring strategies.

The framework divided signaller activity into five observable behaviours. Although the data gathered using this method are very variable, excluding quiet time, monitoring was typically the predominant activity followed by planning. The variability of the data indicates that signallers have different strategies in their approach when working with ARS, although this method was not sufficiently powerful to determine what these strategies are and why they vary.

Two different forms of monitoring were clearly identifiable during the study. These were labelled active and passive as in the first the signallers appeared to be more involved in seeking information. Passive monitoring was usually characterised by removal of the signallers hand from the trackerball and the signaller sitting back in his chair, but maintaining his gaze towards the signalling screens. It is likely that the recognition of patterns of trains on the workstation suggests to the signallers that interventions are unlikely to be required for a few moments and they will take that opportunity to relax.

The trust questionnaire results show that the perceived reliability of ARS is high, as would be expected from a safety critical system. Although signallers found ARS competent in normal operations its competence was rated much lower during disruption. The responses on signallers' perceived understanding of ARS were quite positive but the responses on their ability to predict ARS were more varied and inclined to be negative. Overall, the trust in the system does not appear to be particularly high. A number of significant differences were found between groups of high and low interveners. This would suggest that the signallers' trust in the automation does have a noticeable effect on the strategy of use.

4.7 Chapter Summary

This chapter has presented a framework for structured observations of signallers and the results of 24 observations of signallers using ARS. Differences between signallers were found using the framework, and analysis of the trust questionnaire found that signallers who intervene more often report lower understanding, ability to predict, and faith in the automation. They also perceive the reliability and feedback from the automation to be lower. The study was limited by an inability to determine the reasons behind observed behaviours; hence interviews with signallers were undertaken to gain a richer picture of the impact of automation. These interviews will be presented in Chapter 5.

CHAPTER 5: SIGNALLER INTERVIEWS

5.1 Chapter Overview

Two qualitative methods were used to gain insight into the use and opinions of ARS. Both were essentially interview based, but one was an analysis of pre-existing video tapes of unstructured interviews with signallers at their workstation, and the other was semi-structured interviews undertaken for this research. The existing videos had not previously been analysed for the purpose of studying automation. Themes under the headings of opinion of ARS, system performance issues, knowledge of ARS, and interaction with ARS were identified from the data. These themes are discussed in this chapter. Key findings include signallers' descriptions of how they go about monitoring their workstations, preferred methods of interaction, and signallers' understanding of ARS.

5.2 Introduction

Qualitative data were collected in order to understand the knowledge, attitudes and opinions of signallers towards ARS. The previous observation study had provided data on the behaviour of signallers while using ARS but did not provide any information on the reasons behind the observed behaviours. This study aimed to address this gap and was an important step in understanding the effect automation has on the signalling task, one of the aims of this research. The people who use the system daily are best placed to provide information on its use and issues arising from it. Using interviews it was possible to elicit this information and to probe for additional information in areas such as signallers' understanding of and ability to predict the automation.

Figure 5-1 illustrates the methods used in the qualitative investigations detailed in this chapter. Two main methods were used to elicit information on ARS; a video interview archive analysis and semi-structured interviews with signallers. The overall participant observation approach to the research supplied additional information and context used to develop these methods and supplement and interpret the findings. The results of both studies have been collated in this chapter to give a rich description of the strengths and weaknesses of the current automation.

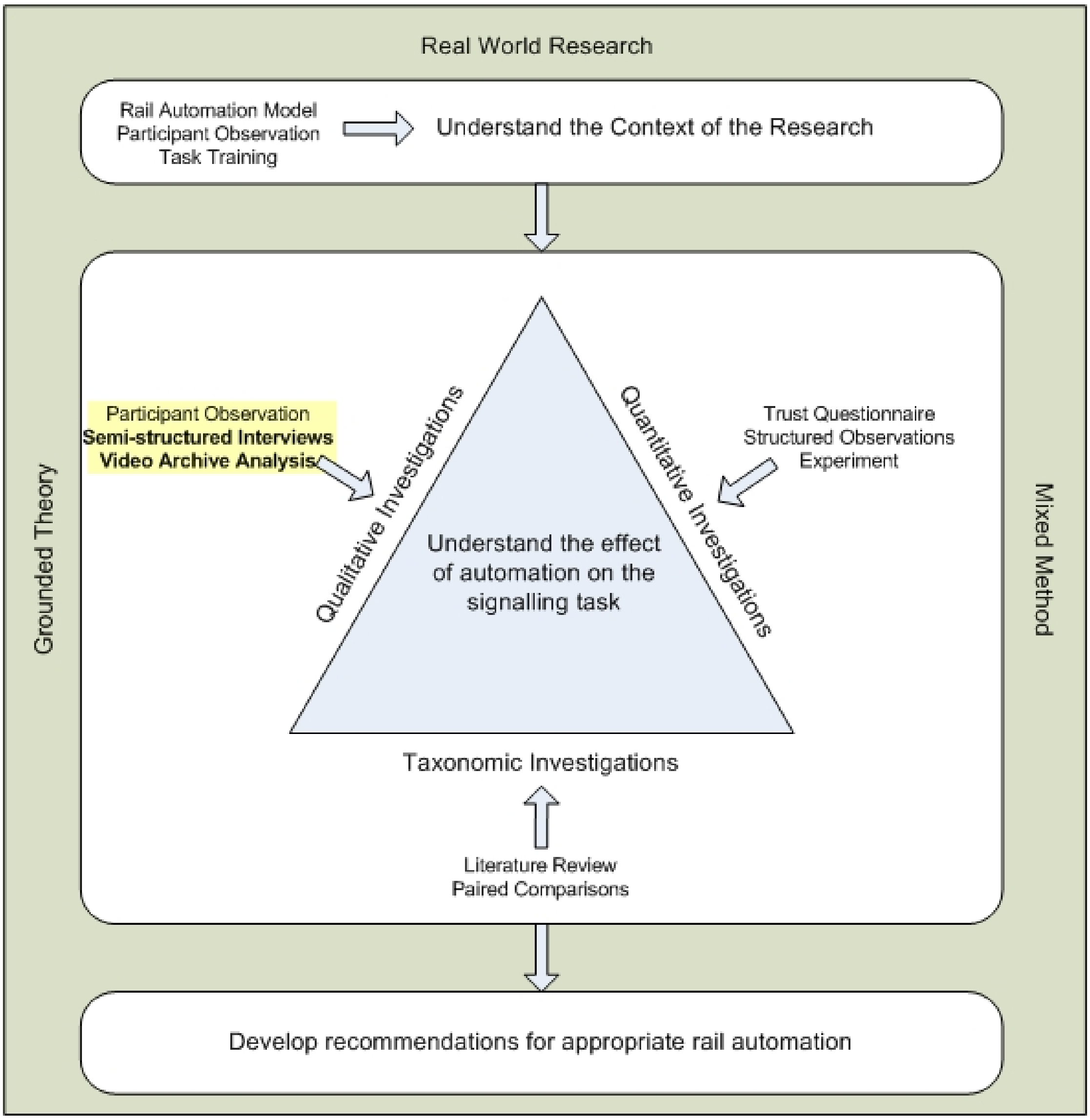


Figure 5-1: Position of Interviews, Video Archive Analysis, and Participant Observation in the Research Framework

The video archive analysis was based on eight previously recorded videos of interviews with signallers in an IECC which were made available to the researcher. These were undertaken by two researchers, Dr. Peter Timmer and Dr. Adam Stork, from University College London in 1999. The videos varied in length but were an average of approximately 3 hours long. Originally 10 videos existed however two had degraded over time and were unwatchable. In order to preserve the remaining eight they were transferred to DVD. The original research did not focus on ARS but rather was a more general investigation into signallers' strategies. As ARS was in use at the signalling centre where the research was undertaken it featured frequently in signallers' comments and discussions with the researcher and thus signallers' comments about ARS arose in a natural and unbiased manner throughout the interviews. The original researcher used an unstructured interview technique where the events on the workstation primarily drove the conversation. This meant that there was an element of verbal protocol in some of the data which emerged. Two video cameras were used in each interview; one trained on the signaller and one on the signalling screens. These were edited together to appear on one screen.

The semi-structured interviews were designed to build on the data extracted from the video archive analysis. Each interview was held at the signaller's workstation, allowing the interviewees to illustrate examples of issues and also to discuss anything that happened on the workstation during the interview. The structure of the interviews permitted any relevant themes to be pursued; hence the data collected were not constrained to pre-identified themes.

In addition to the themes built on from the video archive analysis, themes from the literature were included in the interviews. Automation competence was identified as a principle of good automation (Muir & Moray, 1996). The questions on competence in the trust questionnaire showed a negative response concerning disrupted operation and so these interviews aimed to probe the areas in which ARS's competence may be low in order to expand knowledge of its weaknesses. Similarly, other factors pertaining to trust, including understanding, feedback (visibility, observability, and querying in Figure 5-2), predictability, and expectations of automation (Merritt & Ilgen, 2008; Sheridan, 1999) were probed in order to more fully understand whether future automation could better support these.

Previous research in Sweden found that the automated tools provided to signallers are not predictable due to internal complexity and that they can surprise the operator by performing control actions which contradict the controller's plan (Kauppi, 2006). To avoid surprises, particularly during disruption, the controller is required to take control manually and inhibit the automation. This form of control has been labelled 'control by exception' (Sandblad et al., 2002). The Swedish researchers advocated a system which allows the operator to 'control by awareness' in which operators would be able to see the development of the system over time and prevent disturbances. Lenior et al. (2006) suggested that feedback from the automation may be particularly important in rail signalling operations as the system is not as predictable as other industrial processes due to imprecision in the information provided to the operator. The literature therefore suggests that feedback, understanding and predictability are important themes in rail signalling and may not be adequate in some signalling systems.

The ability to control, or instruct, automation becomes important in such a system. Kauppi, Wikstrom, Hellstrom, Sandblad, and Andersson (2005), in studying Swedish signalling, stated that the nature of train operations means that the initial timetable can quickly become obsolete during disruptions. It follows then that the operator should be able to easily modify the existing plan or formulate a new plan, but automated signalling systems do not always facilitate this (Kauppi, Wikstrom, Sandblad, & Andersson, 2006). Better directability has been achieved in the Netherlands where signallers control disrupted situations by changing the timetable, thus allowing the automation to route trains according to a new plan (Lenior et al., 2006). This is a powerful method of controlling the automation, but is difficult to achieve with ARS. On the basis of this research, methods used by signallers to interact with the automation was identified as a key theme for this study.

The different levels of monitoring found in the observation studies prompted an interest in the investigation of monitoring, and this was added to the interviews. Previous research has found that nuclear operators were able to describe how they monitor nuclear operations (Vicente et al., 2004). However, it is unlikely that the specifics of monitoring in one domain will be applicable in another domain. Even research within the rail signalling domain (Kauppi et al., 2005; Roth et al., 2001) is of very limited use regarding monitoring strategies as vastly different interfaces were examined. Automation is also expected to impact on workload (Kantowitz, 1994) and this study offered an opportunity to gather users' views on how ARS has affected their workload. As signallers have traditionally worked their way up through signalling grades by working in different signal boxes the comparison between NX panels and IECC was probed. Finally, accountability was identified as a principle of automation (Billings, 1991) and so the signallers' views on whether they were responsible for system performance were sought.

The data from both studies were combined during the analysis and so the results and discussion are presented together in this chapter. Figure 5-2 illustrates the themes extracted from both the visual interview archive analysis and the semi-structured interviews.

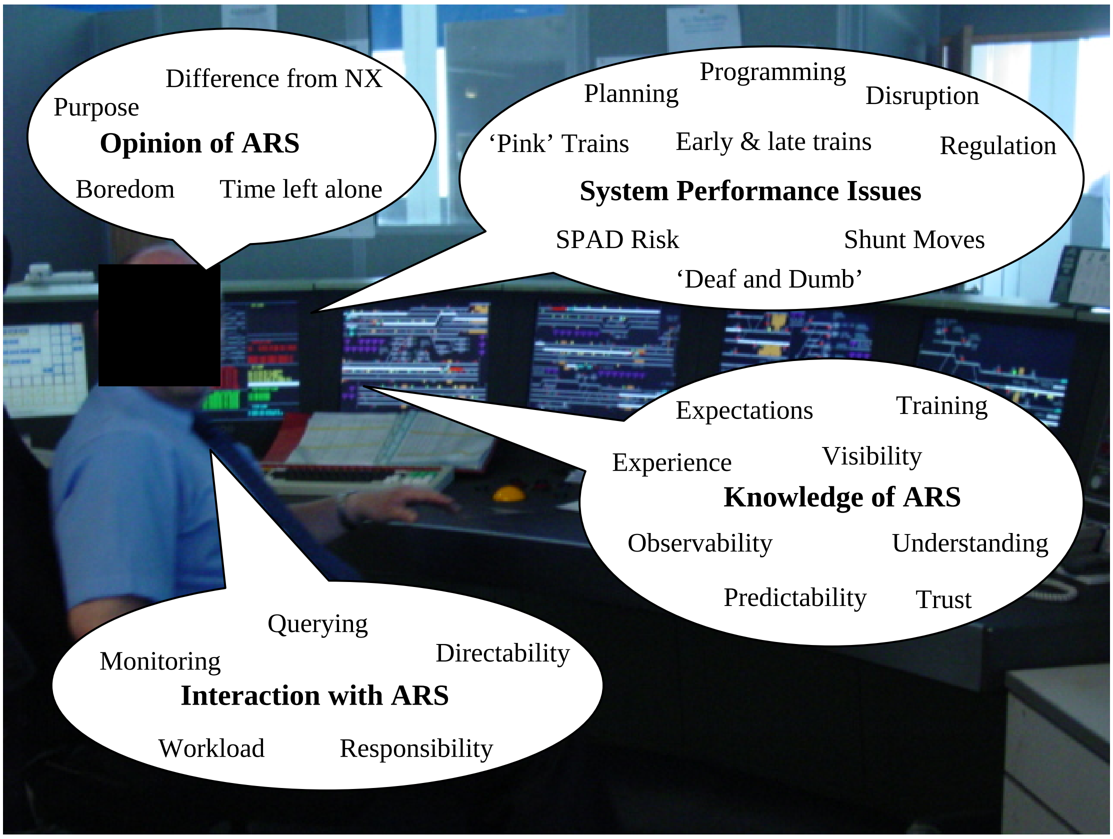


Figure 5-2: Qualitative Data Themes

Place names, including junction, station and depot names, and train destinations, have been changed to maintain anonymity.

5.3 Video Archive Analysis Method

5.3.1 Participants

No data concerning participants were included in the original report on the video analysis by Timmer and Stork (2000), so no details exist of age or experience of participants. From viewings, it can be said that all participants were male and were fully competent signallers at Liverpool Street IECC. A plan of observations was produced prior to the study and it seems likely that opportunity sampling was used.

5.3.2 Apparatus

The apparatus for this study consisted of the videos themselves and a DVD player.

5.3.3 Procedure

The analysis of the videos for this research involved watching each one and making detailed notes each time ARS was mentioned. It was not possible to use the video data to study signallers' behaviour towards and detailed interaction with ARS because the videos did not focus on this specific aspect of the signallers' task so critical data for such a study were missing from the videos. However, the interviews on the tapes did provide information on signallers' use and opinions of and interaction with ARS. Themes were drawn from the data using inductive thematic analysis and each theme was written up in a preliminary document. This was then used to help develop the questions for the semi-structured interviews.

5.4 Semi-structured Interview Method

5.4.1 Participants

In total 10 semi-structured interviews were undertaken. Two of these were with signalling SMEs who no longer work as signallers but together have more than 80 years experience working in the operational environment (Interviews 1 & 2); both have worked extensively with ARS and have managed IECC signal boxes. The remaining eight semi-structured interviews were with current signallers. All participants were male and had at least 5 years experience in IECC signalling.

5.4.2 Apparatus

A list of questions and probes was developed to support the interviews (Appendix K). Data collected were analysed using NVivo software.

5.4.3 Procedure

Interviews were arranged in advance with the Local Operations Manager (LOM) but individual signallers did not usually have advance warning of the interviews, especially in the bigger signal boxes, so sampling was opportunity based. Signallers from five boxes were interviewed; two London based signal boxes (three interviews), two boxes in Southern England (three interviews), and one Scottish box (two interviews).

Participants signed a consent form before the interview (Appendix L) and were given contact details of the researcher. The consent form explained the purpose of the interviews and assured participants that the data would be anonymous and not used for any other purpose. Participants were also given an opportunity to ask any questions they wished at this stage.

Interviews typically took between 40min and 1 hour and were digitally recorded. The opportunity to conduct the final two interviews occurred late on in the research period and were not recorded or transcribed. Handwritten notes were taken instead and these have been used to supplement the findings of the earlier interviews. A list of questions was used but the wording and order of the questions was flexible and probes or prompts were used to elicit more information on a particular topic. The interviews were typed up immediately after each interview and analysed using NVivo.

5.5 Results

Transcripts of the first six interviews were typed up as soon as possible following each interview and were added to the detailed notes from the video archive analysis. Both were then analysed using theory-led thematic qualitative analysis (Hayes, 2000). Each transcript was coded three times to ensure that all relevant data for each theme were picked up and the transcripts were analysed in NVivo. An example of a coded transcript page can be found in Appendix M. Card sorting was used to group the final themes together to better structure reporting of the findings.

Top level themes were defined to group together similar concepts for presentation of results. The four top level themes were:

- Opinion of ARS
- System Performance Issues
- Knowledge of ARS
- Interaction with ARS

Table 5-1 shows the sources and frequency for each of the themes which are discussed in the following sections. The table also indicates which themes have been particularly influenced or supplemented by the knowledge obtained through the participant observation approach.

	Video archive analysis	*Semi-structured Interviews*	*Freq. (comment)*	*Freq. (participant)*	*Part. Obs. Data*
Opinion of ARS					
General Opinion		X	29	10	X
Difference from NX	X	X	12	8	
Boredom		X	8	7	
Purpose	X	X	6	4	
Time left alone		X	6	6	
System Performance Issues					
Planning	X	X	19	11	
Programming	X	X	13	6	X
'Pink' Trains	X	X	5	4	
SPAD Risk		X	2	2	
Shunt Moves	X	X	3	3	
'Deaf and dumb'	X	X	4	3	
Early/late trains	X	X	5	4	
Disruption		X	11	5	X
Regulation	X	X	21	11	X
Knowledge of ARS					
Expectations		X	7	6	X
Training		X	5	5	X
Experience		X	2	2	
Visibility		X	4	4	
Observability		X	3	2	
Understanding		X	23	10	X
Predictability	X	X	53	12	
Trust		X	11	7	X
Interaction with ARS					
Monitoring					
Individual Trains	X	X	6	6	
Route Setting		X	6	5	
'Hot-spots'	X	X	12	8	
Overview	X	X	7	5	
Plain Line		X	5	5	
CCF		X	1	1	
Querying	X	X	20	12	
Directability					
Manual Control	X	X	12	9	
Reminders	X	X	20	12	
STPs	X	X	10	6	
Contingency Plans		X	1	1	X
Proactive Control	X	X	17	9	X
Key/Trackerball	X	X	4	3	
Workload	X	X	37	15	X
Responsibility	X	X	24	12	X

Table 5-1: Sources and Frequency for each Theme from both the Video Archive Analysis and Semi-structured Interviews

These themes are discussed extensively in the following discussion section. Quotations are used to illustrate the discussion and the interpretation of the findings.

5.6 Discussion

5.6.1 General Opinion of Automatic Route Setting

This section discusses the general attitude of signallers towards ARS and some of the more general effects it has had on the signallers and the signalling task. The signallers in the video archive analysis were not directly asked their opinion of ARS, and ARS did not arise in any conversation in such general terms, so the comments in this section are based on the 10 semi-structured interviews. Overall the interviewed signallers' opinions of ARS were positive, but usually with some reservations.

> *"Generally I think it's quite good. There are a number of weaknesses in the system" (Interview 5)*

> *"ARS is a wonderful tool, but that's all it is" (Interview 3)*

Only two signallers expressed reservations, one saying that he did like ARS but preferred to use it as a back up (Interview 8) and one preferred not to use it at all (Interview 6). Both these signallers were working workstations with lower traffic levels and less complexity than the other signallers interviewed, and this might explain why they felt able to handle the workload without ARS aid. The only unqualified support for ARS came from Interview 7.

> *"I like ARS. I think it is excellent." (Interview 7)*

The most interesting comment on ARS was from Interview 3.

> *"It's like working with a woman. It's like working with somebody you don't understand and they're working and yet you're supposed to be equal, and we're not." (Interview 3)*

This comment, although somewhat strangely expressed, reflects the difficulties signallers sometimes have working with ARS; ARS is extremely competent at routing on time trains when there are no restrictions of infrastructure (i.e. the entire railway is available for trains to run). This is ARS's 'bread and butter' and it achieves it successfully.

> *"The headcode is blue, the ARS is pulling off for it, all my sub-areas are on…so we shouldn't have any problems." (Video 3)*

In a general sense it seems signallers are happy to work with ARS and although there were a number of areas which may cause concern, and which will be discussed in the remainder of this chapter, only one interviewed signaller preferred to work without it. Much of what follows in this chapter deals largely with the failings of ARS, so it is worth pointing out that the system does work well under normal running and signallers do rely on it to a large extent. This is supported by signaller comments.

> *"I know that ARS will perform most of the time, 85-95% of the time it'll do what it's supposed to do." (Interview 1)*

> *"You do rely on ARS to do it." (Interview 7)*

The following sections are intended to support the development of new systems which address the weaknesses of ARS for which signallers must compensate, but it is important to acknowledge that the system also has strengths.

5.6.1.1 Difference from Panel Technology

ARS changes the work of the signaller significantly and this was evident in comments on the differences between working an IECC and NX panels[15]. These comments were focussed on two main areas. The first was the change from manual route setting to monitoring as the computer set routes (6 comments).

> *"It was very difficult for me to come off what I was used to and sit down and watch the computer doing the job for me. Because that's what you're doing, you're just sitting back and watching something, some piece of machinery doing the work for you." (Interview 3)*

> *"You are just sitting there watching a computer doing everything, whereas all the other signal boxes we've worked in, whether it be levers or NX panels, you're doing it all yourself." (Interview 8)*

> *"NX was a good system. What you did was a direct thing." (Interview 10)*

Secondly, the presence of ARS means that trains will not necessarily come to a stop if the signaller does nothing (4 comments). This is in contrast to NX panels where trains only have the authority to move if the signaller him/her self sets the route. Where ARS systems are employed the signaller must be prepared to step in quickly to stop ARS setting routes in any circumstances which require stopping trains.

> *"It is reverse thinking with this…in an NX box, if you had a problem you just didn't pull off the signal, you replaced what you needed to replace. With this, you have to prevent it from pulling off. Reverse thinking…it's just different; you've got to react quicker. (Interview 4)*

---

[15] NX – Entry-Exit Panels. This is a very common type of signal box which many signallers would have operated prior to working in an IECC (Appendix C).

Comments such as this highlight that although signallers are physically less involved in signalling when working with ARS, they must still maintain awareness of system state so that they can react quickly and step in if necessary. However, there is a negative tone to the first set of comments which suggests that signallers preferred direct involvement with the system and do not enjoy the change to a system with a higher proportion of monitoring.

5.6.1.2 Boredom

The change to a system with a much higher degree of monitoring and lower degree of intervention, at least under normal circumstances, means that there is increased potential for signallers to become bored. Four interviewees admitted that the monitoring elements could be boring.

> *"If it's all running 100%, I mean, even in the evening peak you just sit there and watch it, you could say it's boring, definitely." (Interview 7)*

> *"If everything's running on time and there's no decisions to make there's nothing to do." (Interview 9)*

Some signallers may cope with this by switching ARS off and working manually.

> *"Yeah, it just becomes boring. So again, there'll be days, especially at weekends when it's quieter when I might switch all the ARS off and just do it all manually." (Interview 8)*

However, two signallers claimed not to get bored with monitoring.

> *"I'd never find it boring as such, because I never know what it's going to do next." (Interview 3)*

> *"There's always something slightly different going on; keeps you interested." (Interview 10)*

It is not known why some signallers find the increased monitoring boring while others do not. Possible explanations include personal attributes of individual signallers, relative complexities of workstations, the level to which individual signallers think about situations, or a combination of all these factors and potentially others.

5.6.1.3 Purpose of Automatic Route Setting

Four signallers commented on their perception of the purpose of ARS. The first perceived purpose was as a tool to assist with manual route setting.

“ARS is supposed to assist us” (Interview 4)

“The amount of trains that you would have to manually route without ARS, ARS is a useful tool.” (Interview 5)

“This system was designed to take the mundane aspects of our job away from us.” (Video 4)

Comments such as these have a positive tone which suggests that signaller perceive the automation as there to help them. However, two signallers noted that the introduction of ARS has meant that signallers control a larger area than previously and seemed to perceive ARS as less of a benefit to them.

“ARS allows management to give us a bigger area.” (Interview 4)

“It means I can work a bigger panel.” (Interview 2)

Only one signaller mentioned ARS as an assistant during disruption.

“The most important function of ARS in my mind is that if anything goes wrong, the signaller can deal with whatever they need to deal with in that area to make the railway safe, and ARS hopefully should go on happily running trains in other areas.” (Interview 5)

The consensus from these few interview comments is that ARS is a tool to assist the signaller, primarily by relieving him/her of mundane train routing. However, there were some signallers who believed that ARS was introduced not to benefit the signaller but to enable him to control a larger area.

5.6.1.4 Time left alone

Signallers were asked how long they would be happy to leave their workstation, and ARS, alone for. In general the replies were between 2min and 5min, about the time taken to get a cup of tea.

“I could leave it happily for 2 minutes maybe. That’s how long it takes to make a cup of tea.” (Interview 3)

“A couple of minutes really. You don’t want to be any longer or you’re sort of panicking as to what it’s doing.” (Interview 8)

It was not possible to determine whether the signaller is unwilling to leave ARS or whether they do not want to be far away should an incident occur which requires their attention. It may be that there is an element of both in their unwillingness. However, it does highlight that signaller involvement is still an important and frequent element in the signalling system.

5.6.1.5 Summary

This section has discussed overall signaller opinions towards ARS, and these are largely positive. There are reservations towards certain aspects of the system and these will be explored in more detail in the following sections. Signallers do rely on ARS to relieve them of the more mundane aspects of the signalling task (e.g. repeatedly setting routes for trains). However, some regret was noticeable at the corresponding reduction in their direct involvement with signalling trains. Four signallers also admitted they sometimes become bored while monitoring the system, although others claimed never to find it boring. The literature suggests that if involvement is sufficiently low for operators to become bored there may be a risk of vigilance decrement (Parasuraman, 1987; Warm et al., 1996) but no suggestion was found that signallers in IECCs suffer from such an effect. This supports the view that vigilance decrement may not exist for real world systems (Moray & Haudegond, 1998). It appears that signallers do try to maintain a high degree of awareness of the system as indicated by their unwillingness to leave the system alone for any length of time.

5.6.2 System Performance Issues

A number of issues with ARS were raised during the course of the videos and interviews and these are discussed in this section. It is unlikely that these are a comprehensive list but are more likely to be the topics which concerned the interviewed signallers at that time. They do however give an insight into the types of problems encountered with the system.

5.6.2.1 Planning

The planning department is responsible for developing the timetable and inputting it into Timetable Services Database (TSDB) which ARS uses to run trains. Therefore, planning has a major impact on how well ARS runs.

> *“Something we’ve had a lot of problems with historically is the information provided from TSDB to ARS” (Interview 5)*

> *“It was always having all these codes missing so it was coming to the ARS computer here as basically useless information.” (Interview 8)*

Incomplete or inaccurate information (i.e. missing codes) means that ARS does not have all the data it needs to run a particular train correctly. Incomplete information means that the signaller must take responsibility for routing that train, as even one missing code will result in ARS not recognising the train. Inaccurate information can result in ARS setting the wrong route for a train resulting in delays to the service. In this way, planning can significantly affect ARS performance. However, the staff in the planning department are somewhat removed from the operational environment and may not always realise the impact they have.

> *"You're looking at a chain of people who sit in an office with very little experience of the running lines, I'm not being rude towards them, it's not a requirement of theirs, but we think perhaps it could be." (Interview 3)*

> *"Very few people in our train planning really understand what ARS does" (Interview 5)*

In addition to a lack of awareness of the importance of complete and accurate information to ARS, the planning team do not have a detailed understanding of the potential complexities of local infrastructure. The result is programmed train timetables which take no account of any local complexities and may make implementation of that timetable difficult, or impossible.

Problems with the timetable are rectified over time as major issues are fed back to the planning department, so the effect is most noticeable following a timetable change.

> *"Usually the programming is ok, until a new timetable comes in. Then you find a lot of problems with the extra programming of the service." (Interview 4)*

Engineering work is one particular area in which the planning department can have a strong impact on system performance. As engineering work restricts the infrastructure available for running trains, the planning department should plan trains around the works and input this new plan into the TSDB. In practice this happens quite rarely.

> *"When I cleared the signal for that move earlier with 4L82 and I cleared the signal across it went pink*[16] *because the train had actually been programmed to go straight through a T3 possession*[17] *instead of around it. Which is not only unhelpful but potentially dangerous as well." (Video 3)*

---

[16] Pink trains are those which ARS does not control.
[17] Possessions refer to the control of a portion of the railway by engineering staff during engineering works.

> *"Planning affects ARS…Engineering work over the weekend and during night not being pre-programmed into ARS" (Interview 2)*

Only in one site was replanning of traffic around engineering work frequent.

> *"Most of the planned engineering works, most of the stuff's in ARS. Someone somewhere is doing their work." (Interview 7)*

Signaller workload tends to increase during engineering work as they have a higher than usual volume of communications and paperwork associated with facilitating the engineering work. Manually routing trains which have not been reprogrammed represents an additional load for the signaller.

IECCs do have a timetable processor (TTP) which allows them to change the timetable for trains in their area in the short term. However, the interface for this system is disk operating system (DOS) based and staff have found it difficult to use.

> *"Right behind you is a machine called the timetable processor. Which I have taught myself how to use, and I can get into it and I can look at train data and adjust things where need be. The thing is, it's 1988 technology, very old, very slow, very user unfriendly. And it's also, to quote an old railway phrase, 'not really anyone's job to do that'." (Video 8)*

Very recently there has been an upgrade to the TTP which has made it much easier to use, but it is still 'not really anyone's job to do that' so full benefit from this system is unlikely to be achieved.

An unstructured interview with an experienced member of the train planning team revealed frustration at their end with the additional burden of work ARS places on the planners. Train plans for non ARS areas can, and do, omit codes as the signaller can easily compensate. As discussed above, this is not the case with ARS areas and a significant workload may be required to ensure that all codes are correctly entered. In addition, there is no formal feedback from the IECCs to the planning team so problems with the timetable go unresolved until they cause delays. At this point, delay is attributed to the originator and thus the planning team become aware of problems with their timetable.

The planning department is perhaps the department which has the most direct effect on ARS performance but being removed from the operational environment they may not realise just how important their inputs are. This can be quite frustrating for signallers who work with the system but have very little control over the inputs which dictate system performance. A structured feedback process from the IECCs to the planning team is one way in which ARS performance could be greatly improved.

5.6.2.2 Programming

Programming of ARS is another large area of influence on system performance. If the programmers get it right, or largely right, ARS performance will be much improved. Unfortunately the complexity of the railway means that a bespoke ARS must be designed and programmed for every workstation it is installed on. This programming involves drawing up complex tables of data for each section of the workstation in order to provide ARS with weighting for the factors it uses in its conflict resolution algorithms. The complexity of the infrastructure, and hence the data, means that it is almost impossible to ensure the algorithms calculate the optimal choice in every situation and the result is usually ‘quirks’ in the programming which signallers learn over time and must pre-empt.

> *“that’s a regular with empties…pull off into that depot, even though it knows that train is booked first, just don’t know why it does it.” (Interview 8)*

> *“It makes some very strange decisions at times, so I think the weighting factors could do with tweaking” (Interview 5)*

> *“there’s a couple of dodgy things you’ve got to watch out for, but you can override that and make use of, perhaps, reminders and lock up where you know there’s a dodgy.” (Interview 7)*

Signallers learn through experience where ARS programming is not optimal and they either monitor these areas carefully to ensure they can intervene in time or they use reminder devices to stop ARS making any moves in an area.

In addition to the existing programming issues, as the systems age the traffic patterns over a workstation may change and the programming at junctions becomes out of date.

> *“What you’ve got is an ARS system the logic of which, or the structure of which, has been designed 1988/89 thereabouts and nothing since. Things have come in since, like regulating policies and business priorities and things but it’s not in ARS” (Video 8)*

It is possible to change the programming, but it requires a data change which must be extensively tested and carries a large cost.

> *"The problem is to have any amendment to the system costs big money…we've made quite a few suggestions. In all fairness a lot of things have been ironed out but there have been over the years a lot of software problems and a lot of ARS problems that have had to go back and back and back to sort out" (Video 3)*

> *"At the moment I think it's too complicated a process, the hoops you have to jump through to get anything changed…and the money you have to pay, which is prohibitively expensive." (Interview 5)*

Signallers therefore learn to live with many of these problems as only those issues which are a safety concern would generate a business case for immediate change. Other issues are logged over time and if the designers can develop a fix this may be implemented during a periodic upgrade to each ARS system.

These programming issues are potentially the single most disruptive issue for signallers working with ARS and they stem from the system designers' attempts to make ARS as 'intelligent' as possible. However, the result is unreliable automation whose complexity is such that signallers find it difficult to predict. Problems with prediction will be discussed further in Section 5.6.3.7.

5.6.2.3 <u>'Pink' trains</u>

Pink trains are those which are not in ARS. This can be for a number of reasons:

- they were not put into the timetable database;
- key information is missing in the timetable database;
- they have been routed off their planned path;
- the signaller has removed them from ARS to control them manually;
- the ARS sub-area containing the train is switched off.

They are called pink trains because the headcode for any train which is not being routed by ARS appears pink on the signalling screens. Trains which are in ARS have a blue headcode (Figure 5-3).

Figure 5-3: Pink and Blue Train Headcodes

Freight trains are commonly pink.

> *"Unfortunately you get a lot of freight trains that aren't properly programmed in this area" (Video 6)*

As they are not in ARS, signallers are responsible for routing these trains manually or putting them into a special timing pattern (STP). These methods of interacting with ARS will be discussed further in Section 5.6.4.3.

> *"4Z53 in Appleby approach berth is not in timetable…so now as far as all us lot are concerned now that's a manual train, we've got to look after it manually. Not a problem." (Video 8)*

This is not a big area of concern, particularly as relatively few trains are commonly pink, and even may be beneficial during routine operations as they keep the signaller involved in the signalling task. In fact, one IECC has requested that freight trains be left out of ARS control as they often run outside their timetabled time and cause problems on the workstation. However, during disrupted working when the signaller is already busy these pink trains may represent an additional task which may easily be forgotten.

5.6.2.4 Signal Passed At Danger Risk

Although SPADs are primarily a risk for train drivers rather than signallers there are a couple of ways in which signallers can reduce the risk. First, the signaller must not change the signal aspect to red as a train approaches it, unless absolutely necessary. Secondly, the signaller can attempt to ensure that trains do not encounter red signals; that is, if the route ahead of the train is available then the signaller will try to ensure that the route is set for a train and no train approaches a red signal unnecessarily. One interviewed signaller complained that ARS frequently signals trains up to red signals before setting the rest of the route.

> *"ARS will signal trains up to red signals all the time… they're constantly on the back of signalmen about SPAD mitigation, sending trains up to red signals possibly tricking drivers into passing signals at red, and yet ARS does it as a matter of course." (Interview 7)*

This is likely to be a local problem where ARS has been programmed to keep options open in case a higher priority train comes along on a conflicting route. A similar problem was encountered in Liverpool Street.

> *"This is where ARS is quite infuriating when you watch it, because a normal signalman would pull off for that signal straight away…ARS won't prioritise it*

> *until the last minute in case another move comes along that's more important." (Video 6)*

This is a good example of the types of problems that can occur with ARS programming, and creates additional work for the signaller to monitor these areas where they feel ARS is not operating competently.

5.6.2.5 Shunt moves

Shunt moves involve moving a train a short distance, usually to change platform or attach to another train. As these are usually un-timetabled, ARS cannot set routes for these shunt moves.

> *"I'm going to clear the next signal because the equipment doesn't understand how to shunt the train. So when I clear this next signal 5F02 goes pink and the equipment says it's off planned path...It's saying you can't do it and I'm saying 'well, there's not a lot of other ways for it to be done'." (Video 3)*

> *"It's not programmed. I've shunted that manually." (Interview 7)*

In the case of one IECC some of the ARS sub-areas are permanently turned off due to the number of shunt moves required in the area.

> *"It can't be trusted….there are a number of shunt routes…that ARS can't physically clear the route" (Interview 5)*

Although these shunt moves impose only a small load on the signaller they are an illustration of one part of the signalling task which ARS cannot achieve. As with pink trains, these shunt moves do not pose a problem during normal operations but represent an additional demand which may serve to push an already increased workload higher during disrupted operations.

A method to 'trick' ARS has been developed in some areas where each possible shunt move (for example between a depot and each of the platforms in a station) has a unique train headcode associated with it and the relevant timetable is permanently programmed in to ARS. In this case the shunter can input the headcode for a required shunt move and ARS will be able to recognise and route that train. This is an example of a strategy devised by one signal box in conjunction with the system designers to improve system performance.

5.6.2.6 ‘Deaf and dumb’

The ARS gives priority to trains based on its calculations of relative delay. However, ARS has been labelled ‘deaf’ as it cannot be aware if the train it is prioritising cannot move, perhaps because of problems with that train or because the route ahead of that train is already occupied. A situation then develops where ARS will not route other trains with potentially conflicting routes until the route has been set for the priority train and hence several trains come to a standstill. This information is not fed back to the signaller by ARS (i.e. it is ‘dumb’) and so it is their responsibility to determine which train is causing the standstill and take it out of ARS, or take over manually.

> *“Let’s knock 5D11 out of the system”*
> *- And you’ve knocked 5D11 out because…?*
> *“It’s delaying other trains”*
> *- It’s stuck?*
> *“Yeah, it can’t go anywhere and other trains are waiting for it” (Video 7)*

Similarly with priorities for trains, ARS does not know when there are problems on the infrastructure and this can cause problems as it continues to route trains as though it has full access to the complete infrastructure.

> *“ARS doesn’t know when anything is wrong does it, at the moment, that’s what catches a lot of signallers out.” (Interview 7)*

Signallers are then required to keep on top of any situations and ensure that ARS does not set incorrect routes towards problems or for trains which have reported problems. To do this the signallers must maintain an awareness of the state of the infrastructure on their workstation at all times as well as constantly reviewing the trains in the area and predicting their progress and ARS’s intentions for them.

5.6.2.7 Early and late running trains

ARS works extremely well when trains are running on time and there are no restrictions of the infrastructure. It does not always work so well with late running trains.

> *“Usually it starts to mess up when you’ve got late running. It can’t make a correct decision, so you have to prompt it.” (Interview 4)*

Late running trains are likely to conflict with other trains and so the conflict resolution algorithms come into play. Statements such as that shown above by the signallers indicate that the algorithms are not producing correct answers as far as the signallers are concerned. Similarly it can cause problems with trains which are running early.

> *"It can't differentiate between early running trains and late trains. Where sets of empties come out of Kidderminster if they're coming up early then it'll just run them into the station and delay other trains, right time trains. So you have to constantly watch that." (Interview 4)*

Signallers must monitor for these kinds of problems and intervene to ensure the train service runs smoothly. Thus, a good signaller's awareness of the system must remain very high, even if he/she is not making many interventions.

5.6.2.8 <u>Disruption</u>

ARS does not deal with disruption well, and in many cases not at all.

> *"Usually you switch it off. It's got to go straight away" (Interview 9)*

> *"If things are disrupted it doesn't make logical decisions" (Interview 10)*

However, ARS was praised for its ability to keep one part of the workstation running while there were problems on another part. This allows the signallers to devote their attention to the disruption and leave ARS to run the rest of the workstation.

> *"The most important function of ARS in my mind is that if anything goes wrong the signaller can deal with whatever they need to deal with in that area to make the railway safe and ARS hopefully should go on happily running trains in other areas." (Interview 5)*

> *"I've got to admit; when you've got big trouble, particularly one end of the area…you do rely on ARS to do it." (Interview 7)*

> *"Your attention goes to that area so you've got to let ARS try to run the rest of the area" (Interview 9)*

However, not all signallers agreed that ARS worked well during disruption, particularly if there is a 'trickle down' effect from the disrupted area.

> *"ARS should be there to help you when things are degraded but you find it actually makes it worse, because you're actually switching it all off, like if you've got a big major incident somewhere, ok, I'll deal with that and Poppleton will deal with itself because ARS is on, it wouldn't, it would totally crucify you." (Interview 8)*

> *"It gets more difficult if there is a problem on one part of the workstation and late running to cope with as well." (Interview 9)*

It is likely that both the layout of workstations and the programming of ARS determine how much help it is likely to be during disruption. If ARS can assist the signaller during disruption it becomes a very valuable tool.

5.6.2.9 Regulation

One of the signallers' main goals is to 'regulate' trains. A definition of regulation is given in Chapter 5, section 5.3.6.1. When regulating, signallers aim to set routes for trains in order to minimise deviations from the timetable. The interviewed signallers all agreed that ARS was not capable of regulating as a signaller could.

> *"ARS does not regulate. It gives the impression of a regulating system and it is not a regulating system." (Interview 1)*

> *"If you want something to pull signals, it's fine, but when you want something to regulate, no, I wouldn't say it's the best regulating piece of kit in the world." (Interview 8)*

Signallers carefully evaluate the relative delay to trains and relative importance of different trains and use their experience to decide which train should be routed first in the event of a conflict. In doing this they take their whole control area into account and may even consider the consequences for other signallers down the line. Their decision does not always match that made by ARS.

> *"The later the train gets, the greater priority it gets. So ARS thinks 'I don't care if you're a Class 6 carrying a thousand tons or more, you're late, you go first'. And all these Class 1 trains are braking everywhere as this train goes across in front of them. It's not quite how it should be." (Video 8)*

> *"It can only do what it's programmed to do really. It's kind of one dimensional. It will only look at one train at a time, it can't think about three or four trains…it can't really see the bigger picture" (Interview 9)*

It seems that despite the complexity of its algorithms ARS does not take into account all the factors that signallers would in making regulating decisions, for example the relative speeds of trains and restrictions ahead of them.

> *"It's quite simplistic in the way it makes its decision, and it doesn't always consider factors that, for example it would be helpful if ARS was to have speed restriction information programmed into it, and that kind of thing…it's very black and white the way it's been designed to regulate." (Interview 5)*

Recently a new performance measure, Public Performance Measure (PPM), has been introduced under which trains must arrive at each destination less than 5min late to avoid incurring a penalty. This new performance measure was introduced after

ARS was developed and programmed so ARS does not attempt to take it into account.

> *"Obviously something else that is a weakness is ARS doesn't work to PPM in the same way that signallers would" (Interview 5)*

Although the ability to regulate is one of the selling points of ARS, the results from the interviews indicate that it is not one of its strengths, and compensating for this forms the basis of a large part of the signallers' job.

5.6.2.10 Summary

This section discussed some aspects of ARS capabilities and system performance. It is most competent when dealing with on time trains for which it has a complete timetable. However, a number of problems relating to automation competence were identified, including the more complex issues of dealing with disruption and regulating late trains. As with any automated system, ARS is only as good as the information provided to it, and the quality of the data input does affect the competency of ARS (Sheridan, 1996). Both the data entry by the planning department and the initial programming by the designers play a critical part in determining system performance but both these factors are largely outside the control of the signal box and signallers. It might be expected that these issues with competence would strongly affect signaller trust and hence usage of the system (Muir & Moray, 1989), but it appears that signallers' knowledge of the system performance issues allows them to calibrate their trust accordingly and judge when they can rely on ARS (Madhavan & Wiegmann, 2007).

The system performance issues discussed in this section influence the quality of the automation and the level at which it can operate. The issues illustrate the challenges signallers face in working with ARS. In general, signallers can cope with the areas in which ARS is not as competent as might be wished. They accept its limitations and are prepared to work around them to achieve the best results they can. However, the efforts ARS makes in regulating and disruption can cause problems for the signaller. It is clear that the automation is not sufficiently capable in these areas and perhaps should instead focus on supporting the signaller rather than attempting to make these decisions itself.

5.6.3 Knowledge of Automatic Route Setting

This section discusses the expectations, training and experience signallers have with ARS and their resulting understanding, ability to predict ARS, and trust in the system.

5.6.3.1 Expectations

Signallers and SMEs over the course of the study frequently spoke of the expectations they had of ARS before they started working with it and how it has failed to live up to these expectations. This is particularly noticeable among signallers who worked with ARS when it was first introduced.

> *"Basically we were led to believe that the signallers' job was to monitor ARS. We weren't here to actually signal trains, we were here to monitor it and they expected it to be that good…I did really think that I was coming up here for that…ARS just wasn't capable of that." (Interview 7)*

> *- What were you told it would do?*
> *"It would basically work the job for you" (Interview 2)*

> *"You think it's going to do everything" (Interview 9)*

The expectation was that ARS would be capable of running almost autonomously and the signaller's role would be to manage incidents when they occurred. In practice the signaller is involved in considerably more routine running of trains than was probably anticipated.

Only one signaller declared low expectations of the system.

> *"It was pretty early in my career down here that I was told it wasn't bought for regulating, it was just a signal puller, so I suppose that's part of the reason why I haven't trusted it since." (Interview 8)*

Signallers interviewed, both in the course of the semi-structured interviews and during informal box visits suggested that while ARS did not usually live up to their expectations, their expectations still influenced their use of and trust in it.

> *"I'm not really anti ARS, what I'm saying is, when we were taught to use it and it stems back from the way it was hailed, and it really is ingrained in your grey matter what it would do." (Interview 2)*

It seems likely that the difference between signallers' expectations and actual system performance has influenced attitudes towards the system. As with any system, it may take a considerable amount of time working with ARS to reverse any negative initial perceptions.

5.6.3.2 Training

There is no standardised training for ARS. New signallers attend a nine week signalling training course where they learn signalling principles and rules but ARS is not covered at all at this stage. They then go to the signal box in which they will be working and acquire 'local knowledge' allowing them to work each workstation in the box. This typically takes several months depending on the size of the box and complexity of the infrastructure. Signallers transferring to an IECC must learn to operate the workstations without ARS so there is limited opportunity for training on ARS. The focus is on learning the workstation, not on learning about ARS and how to work it.

> *"I learnt it without ARS. If I got to use it a bit more then perhaps I would use it." (Interview 6)*

Additionally, training in the box is not standardised and tends to be an experienced signaller showing the trainee the ropes. What training there is on ARS takes the form of informal lessons or knowledge passed on by the existing signallers to new recruits. Consequently, information passed on is haphazard and over time key pieces of information are likely to be lost.

> *"Signallers have never been fully briefed on what they're supposed to do in certain circumstances with ARS" (Interview 5)*

> *"If you're not programmed to work with it, you won't understand it. And therefore you don't know what it's going to do next." (Interview 3)*

Training is likely to be a key missing element which would help signallers to both understand and predict ARS (discussed in sections 5.6.3.6 and 5.6.3.7).

5.6.3.3 Experience

The lack of training on ARS means that much of signallers' knowledge of the system comes from their experience using it.

> *"Through the years we've just built up enough competence to work with it and understand it. You can't understand ARS in say 12 months, I've been at it several years and I still don't understand it." (Interview 3)*

Commenting on signallers who choose to turn ARS off, one interviewee said:

> *"I think they'd find the job a lot easier over there if they did turn it on, got more confident with it, you know. If you don't switch it on you're not going to get confident with it" (Interview 7)*

This learning through experience has been demonstrated recently at Edinburgh IECC where signallers at first preferred to work without ARS. Management staff insisted that they leave it on and gradually the signalling staff became more competent at using it and now seem to prefer to work with it. However, it should be noted that improvements in both the programming and timetable data entry are likely to have also contributed to this change of attitude.

5.6.3.4 Visibility

ARS is not particularly good at displaying and feeding back relevant information to the signaller. The only ARS pertinent information permanently displayed is the train colour, which indicates whether the train is in ARS or not; that is, blue if ARS is routing the train according to timetable, pink if it is not in ARS, and brown if the train is running to a contingency plan or STP (see Figure 5-3 on page 118).

One piece of information which would be very useful to the signaller is the status of the trains (i.e. whether they are running right time, early, or delayed). This information is currently available from CCF on most workstations but this information might be better displayed on the signalling screens.

> *"The biggest problem for me is that it does not show you that it actually is late running" (Interview 1)*

The use of different systems to present complementary information reflects the gradual evolution of signalling systems and the slow introduction of pieces of technology developed in isolation. Integration of all the existing systems could greatly assist the signaller by gathering all pertinent information into one place rather than fragmented across different systems.

5.6.3.5 Observability

Signallers are very keen to know what ARS is planning to do before it does it; however, this is impossible under the current programming as ARS only considers potential conflicts as they arise and it re-runs its calculations every 10s. Therefore if queried, it can tell you what it is planning to do now, but that might not be what actually happens when the route is set.

"It won't necessarily tell you the truth. That's probably the best way of looking at it." (Interview 3)

Providing the information on a constant basis, rather than the signaller having to query it for each train might be a better solution, allowing the signaller a constant window into ARS operations. This may have the benefit of increasing the signallers' awareness of ARS in real time as well as increasing their knowledge of how ARS works.

"I know you can call it up, i.e. if I want to find out what this train is going to do here, so it'll tell me its path, that's it, constantly having to do that, I mean if you had something like this and it's just telling you whatever train is going to do…that'd be a help." (Interview 4)

However, the best solution would be a system that can provide the signallers with constant accurate information about its intentions far enough in advance to allow the signallers time to intervene if necessary.

5.6.3.6 Understanding

Signallers' understanding of ARS was probed during the semi-structured interviews. The levels of understanding varied but all signallers were aware that ARS matches trains in the area to the timetable and routes them accordingly.

"ARS stands for automatic route setting, so it will automatically set a route for a train that it recognises, basically, to a timetable. To its planned path. Don't really know much else really." (Interview 6)

"Whatever train, you've got the TTP, which is the timetable and all the service is actually loaded into that and once it comes in it'll start pulling off for that train" (Interview 4)

When asked about how ARS approached conflict resolution, there was considerably less knowledge among signallers.

"It normally brings everything to a stand and looks at every single train that's involved and eventually it'll work out which one's to go first, pull off for it and whilst doing that, it just brings everything to a stand." (Interview 3)

- Do you understand how ARS resolves conflicts?
"Not entirely, no…I understand that there are certain parameters programmed into ARS to decide how it'll make decision, depending on the class of trains and tidal flows, and that kind of thing." (Interview 5)

"I'm told that this bag of jellybeans does X and Y, but I'm not convinced it does because I've used it so often, and been involved with it so often that the bag of jellybeans doesn't actually bear fruit from what they say" (Interview 1)

> *"Yes, I do understand, but ARS doesn't understand….it's like two magnets pushing against each other, gets stuck, can't make a decision." (Interview 10)*

One SME commented on the importance of understanding ARS.

> *"I actually think fundamentally they need to know what they can do to maximise its performance potential. Or its potential to enhance performance" (Interview 2)*

These comments reflect a vague knowledge among signallers that ARS takes some factors into account but most do not have detailed knowledge of what these are or how ARS uses them. The comments also reflect dissatisfaction both with the time ARS takes to make a decision and the results of the process. The lack of understanding means that signallers cannot rationalise some of the decisions made by ARS and it limits the signallers' ability to maximise performance and work cooperatively with ARS. It also has an impact on their ability to predict the automation.

5.6.3.7 Predictability

Signallers' inability to predict the automation was a strong theme from the video archive analysis. There were frequent examples on the tapes of signallers stepping in because they were not sure if ARS would make the decision they wanted it to.

> *"I put the reminder on 281 signal for the fact that ARS once it's down to the platform could possibly route it off in front of that Mickle Trafford because ARS is unpredictable. You can never trust it. You don't know what it's going to do to be honest." (Video 5)*

> *"I don't know which is going to go first" (Video 7)*

The semi-structured interviews probed this issue further and found that signallers did not think ARS consistent in its decisions.

> *"In all your research, that's the one thing you'll actually find with signallers, that ARS at times does things that you don't expect it to do." (Interview 2)*

> *"You can't really…you wouldn't say 100% say what it was going to do if you sat there watching it knowing the trains were running late." (Interview 8)*

> *"Even with things running to time it does unexpected things." (Interview 9)*

> *"You see dents on the workstation [from signallers banging their heads in frustration]. You sit here head butting the workstation thinking 'why the hell did it do that?'" (Interview 3)*

Some signallers admitted that they believed ARS must be consistent but they were unable to predict it because they didn't have sufficient understanding to be able to do so.

> *"I'd say it has to be consistent; however, I don't always understand the processes, factors that ARS is taking into consideration when it makes that decision...it doesn't always appear consistent, but one would imagine as a computer it's making that decision it must be consistent." (Interview 5)*

> *"You could have the same scenario two days on the trot, i.e. a late running train up the suburban and it'll pull off out of the platform but the next day it'll run the late running train…there probably is a reason for it, there might be something I hadn't realised the system's done." (Interview 4)*

Signallers also find that ARS surprises them at times.

> *"It constantly surprises you"*
> *- Even after 17 years?*
> *"Yeah, it does….sometimes you'll look and think 'I've no idea why it's done that." (Interview 5)*

> *"Oh yeah, it's constantly surprising" (Interview 4)*

Unpredictability makes the automation harder to control, and thus makes the signallers' job more difficult.

> *"Without being able to predict what it might do you get stitched up." (Interview 2)*

> *"If I'm to be proactive I need to know what it's going to do before it does it." (Interview 1)*

The lack of predictability means that signallers must constantly be aware of the areas in which they are unsure of ARS and be prepared to take over, either by manually routing one train or by turning ARS off. In this way the lack of predictability affects ARS usage.

> *"It's predictably unpredictable. You know what it might not do." (Interview 10)*

> *"It's just too unpredictable. If it was consistent you'd sort of know more what it was going to do, but because it's unpredictable the only safe thing to do there is to have all of Gospel Oak off until this little patch of trains is gone." (Interview 8)*

> *"With conflicts…you will find that signallers will stick reminders on to make sure it does it." (Interview 1)*

There was an exception to the agreement on the lack of predictability, and that came from a signaller interviewed in the least advanced IECC included in the interviews.

The ARS in this particular IECC does not attempt conflict resolution but just routes according to the timetable. This wasn't a decision made during commissioning but rather reflects a lack of funding to programme the ARS at junctions.

> *"You know what's coming"*
> *- You can predict it?*
> *"Yeah"*
> *- So you find ARS very predictable?*
> *"Oh yeah, definitely." (Interview 7)*

ARS could still surprise the signaller in this case; he related a recent event where ARS routed a train in an unexpected direction. Initially he thought it was a mistake by the automation but on checking the timetable he discovered the route for that train had been changed and ARS had routed it correctly.

> *"It had done the right thing but I was expecting it to do something completely different" (Interview 7)*

The same signaller felt strongly that his ability to predict ARS was very important.

> *"Oh yeah, got to know what it's going to do…I can't imagine being a signaller if you didn't know what the trains were actually going to do and where they were going to go, just integral part of the job isn't it, knowing what's going on. It'd be awful sitting here not knowing what it's going to do next." (Interview 7)*

The interview results indicate that there are two main reasons why ARS is unpredictable. The first is the lack of feedback, but this would not be a great problem if ARS was perceived as consistent. However, the second reason ARS is not predictable is that the attempts it makes to regulate are based on complex algorithms which are not transparent to the signallers and so make it very difficult for them to predict the outcomes. This means that signallers cannot themselves work out what ARS will do, and ARS is not capable of feeding this information back to them. These two factors combined make ARS a difficult system with which to work cooperatively.

5.6.3.8 Trust

The observation study indicated that trust in automation is related to the level of intervention. This study probed signallers' general trust in ARS. Signallers were wary about saying they trusted ARS 100%.

> *"Not implicitly, but I do trust it generally." (Interview 5)*

> *"When you've worked with this system for a while you begin to distrust it. It'll do something for no apparent reason. You've got a train coming in, you know*

> *it's booked in first and it'll pull off for an outgoing train for no reason at all. Don't know why, just does it." (Interview 4)*

> *"If I was really honest, no. Your ability to say can you trust it is that you can not always hold your hand on your heart and say it was going to do what it said on the tin." (Interview 2)*

However, signallers do completely trust the interlocking which prevents both them and ARS from implementing any unsafe routings. This removes the danger from using ARS, as the most it can do wrong is make a bad regulating decision or wrong route a train.

> *"I trust the interlocking 100%, the ARS I wouldn't say 100%" (Interview 8)*

> *"It's the interlocking that I trust" (Interview 5)*
> *- Do you trust ARS?*
> *"Yes…there's a net underneath you see, called SSI[18]" (Interview 1)*

The lack of predictability impacts on the signallers' trust.

> *"I think you should be able to predict it, because then you could trust it and perhaps leave it on more. But when its unpredictable like this and you sit here all day it is just on off on off all the time." (Interview 8)*

> *"When you've worked with this system for a while you begin to distrust it. It'll do something for no apparent reason. You've got a train coming in, you know it's booked in first and it'll pull off for an outgoing train for no reason at all. Don't know why, just does it." (Interview 4)*

It seems that although signallers do not find ARS completely trustworthy they have strategies to cope with this and work with it. They identify what areas it can be trusted in and pay careful attention to all other areas.

5.6.3.9 <u>Summary</u>

It is clear from the results that ARS has not lived up to signallers' expectations of it and that their training on it is limited. Research has found that differences between operator expectations and actual system performance can strongly influence automation use (Merritt & Ilgen, 2008). It is therefore important to ensure that signallers begin working with ARS in possession of a correct mental model. However, in reality, it is the experience they gain from using the system that contributes most to their understanding of how ARS works. The development of this knowledge through experience depends on the visibility and observability of the automation (Dekker,

---

[18] Solid State Interlocking (Appendix C).

2004; Endsley, 1996; Parasuraman & Riley, 1997), but ARS is widely referred to as 'dumb' as it does not feed back valuable information to the signaller. It is not surprising then that there are gaps in the signallers' knowledge, particularly with regard to the more complex functions of ARS (i.e. conflict resolution). Without this knowledge it becomes very difficult to predict ARS and signallers stated that they cannot always predict what ARS will do in a given situation. Factors including understanding, predictability and feedback were found to differentiate high and low interveners in the observation study (Chapter 4). This lack of ability to understand and predict the system is likely to have the effect of making working with ARS more difficult, and research suggests that it may reduce the success of the overall socio-technical system (Hopkin & Wise, 1996).

These issues have made it difficult for signallers to develop an accurate mental model of the system and may contribute towards a lack of trust in the automation (Dzindolet et al., 2003). However, this lack of trust was tempered by their experience with it and knowledge of its limitations which influence trust calibration (Madhavan & Wiegmann, 2007). This means that signallers know when they can trust ARS to make the correct decision and when they cannot. From a safety point of view signallers can trust ARS because all its decisions must go through the interlocking which is widely regarded as foolproof.

The gaps in signallers' knowledge of the system can be explained by the lack of training and the reliance on experience, but the improvement of their knowledge may yield results in their ability to control and work with the automation. Ideally, future systems would rely less on training and experience by improving the visibility and observability of the system and thus increase the operator understanding and ability to predict the automation.

5.6.4 Interaction with Automatic Route Setting

The strategies signallers employ to monitor and control the automation are reported here with the perceived effect on workload and responsibility.

5.6.4.1 Monitoring

Automation has reduced the amount of manual route setting by signallers and correspondingly increased the time spent just monitoring the system. The interviews probed the types of things signallers looked for when monitoring the system. Signallers spoke of four main monitoring strategies they employ; monitoring individual trains, monitoring route setting, monitoring hot spots or pinch points, and maintaining an overview of the whole workstation. The interviews also gathered data on the monitoring of plain line (i.e. pieces of straight line which have no junctions or crossovers), use of CCF for monitoring, and the need to pre-empt ARS.

Monitoring individual trains

Signallers learn through experience which trains are likely to cause problems and they will monitor these carefully. They also monitor late running trains which may conflict with others and watch for any trains which do not appear to be running as expected.

> *"I suppose really there's certain train numbers coming down, anything that was a problem train, watch that." (Interview 8)*
>
> *"Unusual train running pattern, i.e. that it's not running as fast as I would expect it to. Sometimes just the system reaction." (Interview 1)*
>
> *"I will watch individual trains if it's going to conflict with another train obviously like over there crossing from the fast to the slow at Rainhill and I know there is going to be a problem there I'm going to watch it" (Interview 6)*

Monitoring for these trains and situations allows time for the signallers to step in to prevent situations developing further, with the objective of reducing or preventing delay.

Monitoring Route Setting

Signallers monitor that routes are being set for trains, and that each train is running on two green signals wherever possible.

> *"You just quickly make sure that all the signals are off that should be off" (Interview 8)*
>
> *"How ARS is actually clearing signals. Whether it is actually clearing signals" (Interview 5)*
>
> *"I'm looking for the finer points that the ARS doesn't know about maybe, or doesn't understand. The basic system for running trains in ARS is to run trains with a minimum of two green signals at all times. Now if you see a train, 4L87, he's got two greens at the moment, that's fine. Now he's down to one, wait for the route to call so he goes back on to two. Now if there's something wrong, that route won't call...it is doing the job correctly, but for a little bit too long as I see it, it was running on one green instead of two. It should have called the route earlier." (Interview 3)*

This monitoring strategy is about maximising the efficiency of the signalling rather than picking up or preventing more major problems.

Monitoring 'hot-spots'

All workstations have junctions or regulating points[19] where decisions can be made on the order of trains. Signallers indicated that these areas are monitored carefully.

> *"Anywhere we've effectively got a pinch point….just monitor what ARS is up to in that area." (Interview 5)*

> *"I think you establish in the early time working any workstation, establish the criteria where you think the crunch points are, pinch points are." (Interview 1)*

In terms of junctions, the interviewed signallers indicated that they tend to monitor the approach to the junction rather than the junctions itself.

> *"The approach…it's about presentation, it's about what's about to present itself." (Interview 2)*

Hot spots may also be depots.

> *"Coming out of Whittlesea car sheds, if you didn't monitor that, control that, then it would mess the evening peak up all the time." (Interview 4)*

Monitoring these areas is probably the most critical as these locations are where decisions are made, either by the signaller or ARS, which affect the signalling performance.

Maintaining an overview

Signallers indicated that they maintain a general awareness of their whole area of control. They appear to achieve this by setting an expectation of what they should see next time they look at the signalling screens. Any deviation from this expectation is quickly noticed and can be investigated further.

> *"You look at the general picture, an overview of all trains on the area" (Interview 5)*

> *"You see the whole screen, and you know if something's not moved. It's a sort of, I've seen that, capture, gone, I've seen that, capture, that now should not look like that, that should look like this. And there's this sort of mental, it should look like this and it doesn't. Why doesn't it?" (Interview 1)*

> *"When ARS pulls for something, particularly if it's something you're not expecting, then yeah, it's probably going to catch your eye straight away." (Interview 7)*

---

[19] See Appendix C for a description of regulating points.

> *"You're looking for where change has not occurred. But your attention is drawn to it." (Interview 1)*

It is not known whether this overview is maintained through peripheral vision or scans of the whole area of control, or a combination of both.

Monitoring Plain Line

Signallers stated that they tend not to monitor areas of plain line[20] to the same extent as other areas on their workstations.

> *"You wouldn't really worry, you don't worry about what ARS is doing in those sections so much, you worry about what the train is doing, but you're not as concerned with ARS because you really make an assumption that it is clearing the routes it needs to clear. It's plain line, there's nothing in front of it". (Interview 5)*

> *"These bits here where you've got several green lights[21]…you don't pay no attention, you pay less attention to them, because something could happen, one of those signals could black out[22]. So again, you do attend to it, but not in such detail. And you automatically look, you see a row of green lights…you can glance at it and carry on. If there was one black in there, you would pick it up very quickly." (Interview 3)*

Monitoring of plain line is clearly a much lower priority for signallers and there may be very little direct monitoring of these areas. Signallers are likely to pick up any problems in these areas through their maintenance of an overview of the workstation.

Monitoring CCF

Some of the information signallers are looking for can be more effectively gathered from other sources. Late running trains is a good example of this; the information can be collected much more easily from CCF than from the signalling system which would require querying the routing of the train on the general purpose (GP) screen[23] and then comparing this with the simplifier for the workstation. On CCF each train is colour coded according to its delay.

---

[20] Plain line is a section of track with no points. Therefore any train travelling over a piece of plain line can only go forward; there is no potential to change its route.

[21] "Green lights" refers to automatic signals. These differ from controlled signals as they routinely display a green aspect and only show red if a train has just passed through into the section. They can be regarded as having a constant route set and therefore the signaller has very little interaction with them. They would usually only be used on plain lines.

[22] This refers to failures of signals, either failure of the lamp out on track or of the indication on the workstation. In this event, the indication on the workstation shows a black aspect rather than red, yellow, or green.

[23] The GP screen gives details on alarms, trains approaching the control area, and responses to signaller queries. See section 5.4.1 for a description of the GP screen.

> *"We usually, mainly we use this [CCF], you're constantly looking at that. If everything is green it's on time, you're not worried about it." (Interview 4)*

Use of CCF varies greatly between signal boxes and individuals. Some signallers have been observed using CCF for a large proportion of their monitoring while most only refer to it occasionally. However, it is clear that CCF contains information pertinent to the signalling task and this might be better integrated with the main signalling screens.

5.6.4.2 <u>Querying</u>

Signallers have the ability to query ARS about individual trains. This gives them information on the timetable ARS has for that train and the current status, for example, if it is giving priority to another train. This is quite a useful tool particularly when ARS has not set an expected route for a train.

> *"The best policy is always to interrogate the system and say why aren't you pulling off" (Video 4)*

> *"Signallers question paths quite a bit, and to find out platform information it's quite good actually, tell you where the train's going to be routed into the platform. Also, sometimes you use it to find out what ARS is going to do in a particular area, it's not always apparent." (Interview 5)*

> *"We can question it as to what it's timetabled next move is going to be, which is quite handy to have." (Interview 7)*

But signallers do not always understand the information they get from ARS.

> *- What does that mean?*
> *"I have got no idea…It does annoy, this system. That to me doesn't mean anything." (Video 4)*

> *"Sometimes you've got information that a train was standing at x or y because of another conflicting move and the train that was conflicting was miles away, nowhere near it. And yet it saw it as something that was close. Don't know why." (Interview 1)*

In addition, ARS can say it is giving priority to a train when queried but later change its mind, so signallers have learned not to always trust the replies from ARS.

> *"If something's got priority it's not to be trusted; if it's holding it for a certain time you know you can trust it." (Interview 8)*

> *"Sometimes you'll ask it why it isn't routing the freight and it says it's waiting for another train and 30 seconds later it set the route" (Interview 9)*

"It won't necessarily tell you the truth. That's probably the best way of looking at it." (Interview 3)

Sometimes ARS does not provide the information signallers are looking for.

"I had one earlier when there was absolutely no other trains about but the train that was leaving only had a yellow to the junction, and it wouldn't pull no further so I was like asking it why and it was saying that it had tried, it had already requested the signal, but that was the one it had already pulled…it was going to stop it on the red on the junction, and when I asked what it was doing it said 'well I pulled this bit'. Basically as if to say 'what more do you want?'…It just came up with useless information really, on that occasion" (Interview 8)

In this case the signaller was trying to establish why ARS would not set further route for a train. The ARS can be queried for this information but the feedback on this occasion only gave information on the route already set. This was not what the signaller was looking for, but it was not possible to obtain any further information from ARS.

Querying ARS is one way which signallers can learn more about it and understand it better.

"I think if signallers make more use of that then they get more confidence in using ARS" (Interview 7)

As discussed in Section 5.6.3.4, even with the limited information provided it may be more beneficial if this information was displayed constantly, rather then signallers having to query ARS for it. This would make the information more accessible and help signallers develop a better understanding of how ARS works.

5.6.4.3 Directability

There are a number of methods the signallers can use to interact with and control the automation.

Manual Control

The first method for signallers to control ARS is to take trains out of ARS control and route them manually, or to manually set routes for trains before ARS gets a chance to.

"I just like to take them out if I can to give me more control over it" (Video 1)

> *"I'll make him non ARS. That's my personal choice because that way I've got complete control over the train." (Video 8)*

A more drastic method of taking control is to turn off one or more ARS sub-areas[24]. This means that any trains within that area will have to be routed manually while the sub-area is off.

> *"Switch it off if there's some late running" (Interview 8)*

Signallers may choose to do this if there is congestion in an area to ensure that they have total control over the movement through that area. One unwanted movement by ARS can ruin a plan.

Reminders

Reminders appear to be the most common method employed to control ARS. Reminders are a legacy from the older signalling systems where they took the form of a physical cap over a lever or button. Their purpose was to 'remind' signallers not to set a certain route, usually in the case of some form of line blockage, for example, a failed train or engineering works. Their purpose is primarily as a safety device. A reminder function was provided on IECC systems to serve the same purpose and as well as reminding the signaller not to set a route it also prevents the ARS from setting the affected route. Therefore, setting a reminder in a particular location can be a very effective way of preventing ARS from calling a route.

> *"Just put a reminder on the up main there, just in case something comes along. Don't want ARS routing across." (Video 1)*

> *"I put the reminder on 281 signal for the fact that ARS once it's down to the platform could possibly route off in front of that Croft because ARS is unpredictable." (Video 5)*

> *"We do use reminders on here quite extensively, yes" (Interview 5)*

> *"I look at a situation and if I want a particular train to go in before one comes out then I'll actually put a reminder on there or turn ARS off so it won't pull off for outgoing trains. A lot of us do that." (Interview 4)*

These comments reflect widespread use of reminders to control ARS. The preference for reminders seems to be because they are considered easier to use.

> *"It's a lot easier, it's so much easier just to stick a reminder on and hold it rather than lock the points or take it out of ARS, because you could forget*

[24] Each workstation is divided into a number of areas for the purposes of ARS. Each of these is called a sub-area and may be turned on or off as the signaller wishes.

about it. So the easy way out is just to stick a reminder on, take it off, leave it and come away from it. Laziness if you like." (Interview 3)

Signallers need to be quick at applying reminders or ARS will set the route before they apply the reminder to stop it.

"He should follow 1F14. So what I'll do, reminder there, reminder there, damn, too late. The route has just called and caught me." (Video 8)

There was only one interviewed signaller who didn't consider reminders the easiest way to control ARS.

"It's easier to turn the sub-area off. It's just one click of a button….I don't believe in putting reminders everywhere, you just do the routes as you want to do them." (Interview 8)

This comment was from one of the less busy workstations. In the larger IECCs turning off a sub-area can potentially affect more trains and this may explain why turning sub-areas off is not the preferred interaction method on busy workstations.

However, this mixed use of reminders may decrease their value as a safety device, and both SMEs pointed out that they should not be used for controlling ARS.

"Reminders are only used, primarily as a safety value, or as a specific reason that you need to do something, not for routing or pathing reasons." (Interview 1)

"Reminders are not designed for that. Once I put a reminder on I knew what it was for." (Interview 2)

Only one interviewed signaller was concerned by this use of reminders, but he was quite disturbed about reminders being used for regulating purposes instead of purely as a safety device.

"We are conditioning people to remove collars…collars are being used for the purpose of not delaying trains and they should be just for safety purposes.." (Interview 10)

This is a very valid concern. The routine use of reminders as regulating tools may degrade their status as a safety device. The danger is that signallers may forget which reminders are in place to protect track workers, particularly if applied in an area where reminders are frequently used to control ARS. Signallers may also become conditioned to removing or override reminders to let trains pass and under pressure the possibility exists to do this in error. Clearly reminders are a valued tool for controlling ARS and banning their use for this purpose would greatly reduce the directability of the system for a large number of signallers. One possible solution is to provide another tool which achieves the same results but has a different name and appearance, for example an inhibitor.

Special Timing Patterns

Special Timing Patterns (STPs) are pre-programmed timetables for common routes across the workstations. If a train arrives on a workstation and ARS cannot find a timetable for it in the database, the signaller has the option of putting the train into an STP rather than routing it manually. These are commonly used, particularly for freight trains. The signaller simply types the headcode of the train into ARS followed by the 'name' of the STP; ARS will then route that train according to that STP.

> *"It's looked after itself all through the patch and hasn't bothered me again, which is an excellent way of reducing your workload." (Video 3)*

> *"You have the special timing patterns for traffic that doesn't belong to us, so out of region traffic we can put in a programme that will take it from Bingham say, to Hednesford." (Interview 3)*

However, signallers didn't always know all the STPs available to them, and in some cases they would have liked to have more STPs programmed.

> *"The ones that perhaps you don't often use you wouldn't know…and when you get to certain workstations like Grindleford where you could have 20 or 50 codes if you don't use them for weeks and weeks you're not going to remember them so you just refer to the sheet." (Video 3)*

> *"We've been given some codes that have turned out to be used very little and other actions which are quite repeated which you could successfully use a special timing pattern codes for haven't been put in" (Video 3)*

STPs have been demonstrated to be a useful tool but could be greatly improved if signallers had more control over the design and modification of STPs.

Contingency Plans

Contingency plans are similar to STPs but apply to all trains travelling over a certain area rather than to just one train. They were envisioned as helping the signallers in the event of a restriction of infrastructure in a particular area. In this situation they could be invoked to route all trains around the restriction. In practice, no signallers seem to have ever used one and the majority are not aware of what contingency plans are available for use. The reason for this appears to be that they are not fine tuned enough to be of real use. In practice they do not match events on the workstation because these events can be so complex and variable. Even if they do match, the signallers are largely unaware of what contingency plans are available, and whilst they are dealing with a restriction of the infrastructure they do not have the time to look up manuals.

> *"I put a contingency plan in for the down sub, so I don't want any Oakham trains, any Wymondham trains, or anything of that nature to come down that road because that junction is out. So if I put a contingency plan in, I would actually block the whole line and not just that bit [the crossover needed by the Oakham and Wymondham trains]…what I'd need is a contingency plan that says I can run you, but I can't run Ts, Hs, Os and Ss [i.e. headcodes of trains travelling the same route]." (Interview 1)*

In this quotation one SME has described a system whereby contingency plans could be built up to match a situation rather than a rigid plan. This may be considerably more beneficial than the current system.

Keyboard vs. Trackerball

There is a choice of interaction method with ARS; either via the keyboard or using the trackerball. The majority of signallers prefer the trackerball.

> *"The way you interact with the keyboard…I'm not computer literate…the way it's been designed, the big failing with it is that it was designed by the technicians at Derby…they did a really good job with it, but it's designed from the engineer's point of view, from the designer. "It's not designed with the user in mind. So, to do things on the keyboard are a lot longer than they need to be." (Video 8)*
>
> *"When I first heard of trackerballs I thought 'oh, that'll be difficult' but it's not. It's so easy, so so easy" (Video 8)*
>
> *- Generally you prefer the trackerball?*
> *"Yes…it's quicker" (Interview 4)*
>
> *"Unfortunately you're encouraged to be lazy with IECC interface because of the cursor and the keyboard. Which one do you use? You use the trackerball because it's easier and lazier. A monkey can use the cursor, the trackerball, but it takes a bit of intelligence to use the keyboard." (Interview 1)*

With practice the keyboard routes may become a quicker and more convenient method to interact with the system. However, it is rare for signallers to use this method, perhaps because it requires detailed knowledge of the names of each route on each workstation. The use of the trackerball places less of a burden on the signallers' memory, but may also represent a less advanced form of interaction with the automation.

Proactive Control

The ability to proactively control the ARS was frequently spoken about. The poor feedback from the system constrains signallers' ability to control proactively, but signallers regarded good signalling as seeing problems developing and dealing with them in advance.

> *"I think you've got to be sort of six trains ahead of yourself…that's part of being a signaller…you can't just work by what the next train is that's running up, you can see a problem coming miles away and you know that you've got to do something about it" (Interview 7)*

> *"A lot of people sit here and 'oh look at what ARS has done' but the majority of us sit here and if you held your hand up you know a lot of the things it's going to do that you don't want it to do, so there's a lot of times where you should be intervening during the day" (Interview 8)*

> *"I can decide what to do with the system. You've got control really." (Interview 10)*

One of the SMEs was concerned that signallers do not take full advantage of ARS allowing them to be proactive.

> *"I think that we've got the problem in our culture in that signallers are basically reactive and not proactive. That's my biggest problem…there is an aspect of being proactive on a panel….but I would say that ARS give you more of an opportunity to do that. Because you can actually set the system up to operate…well ahead of actually doing the operation." (Interview 1)*

In one sense ARS forces signallers to be proactive as they must intervene early to stop ARS setting unwanted or wrong routes. In another sense it can force them to be reactive as the unpredictability of the system can result in unexpected routes being set which causes a situation that the signaller must retrospectively attempt to fix.

5.6.4.4 Workload

The perceived effect of ARS on workload was an interesting topic. Generally signallers felt that ARS reduced their workload and made their job easier.

> *"With this equipment, if you switched it all off, you really would have your work cut out, but I have done it on weekdays before, and you soon get caught out delaying trains…you do need it if you're going to work this area" (Interview 8)*

> *"It's easier for the signaller workload wise, without any doubt…in these sort of areas, you couldn't work without it. Probably one man wouldn't cope with it, not in the height of the peak….off-peak you wouldn't be able to take your eyes off the screens" (Interview 7)*

It may not reduce workload as much during disruption.

> *"It has made it easier. The difference with ARS, when it's working manually you've got to do everything. When it's working automatically you've got to step in to stop it doing everything. When you're busy you then find you've got to reminder, reminder, reminder everywhere to stop things happening and then if you want it to happen you've got to override the reminder. That creates as much work in itself as everything being manual." (Video 10)*

> *"Overall, out of 100% of one day, I would say it's far easier. You get a disturbance and it's far harder." (Interview 1)*

> *"The majority it's eased it, but only when things are going well. It's higher workload when there's problems." (Interview 10)*

The interviews suggest that there is a lower level of workload below which the benefit from ARS is not realised.

> *"I've got…four trains on the panel, over there, there's more traffic over there so the ARS is a good tool to be used….over here I can't see the point in it." (Interview 6)*

ARS has allowed one signaller to work a much larger area than in the past; typically one IECC workstation will have replaced several lever frames or NX panels. In the semi-structured interviews the workstations had replaced between two and five signal boxes.

> *"I wouldn't say it's reduced our workload…ARS allows management to give a signaller a bigger area." (Interview 4)*

Most of the reduction in workload seems to be in the physical domain. The effect on the signallers' cognitive workload is less clear.

> *"I think all it's actually really done is made the signaller not to have to physically do a lot more routing. Clearing of routes...it's lowered the physical; it's not lowered the mental" (Interview 2)*

> *"The workload is totally different. From being a physical element, it becomes a mental element." (Interview 1)*

> *"I wouldn't say harder, there's just another facet on it that you've always got to check." (Interview 10)*

The overall effect on workload therefore is not clear cut. The general consensus seems to be that there is a definite reduction in physical workload during normal operations. The effect on mental workload, and during disruption, is not as clear and it seems that workload in these areas has probably increased, but it is not known how much of this increase is due to the larger area of control and how much to the presence of ARS.

5.6.4.5 Responsibility

The introduction of automation such as ARS has the potential to blur the lines of responsibility. Signallers were asked who was ultimately responsible for running trains. The majority felt they were and considered it good practice to ensure that trains were signalled as well as possible between themselves and ARS.

> *"The signaller is responsible for running the trains, ultimately" (Interview 5)*

> *"I find it a bit embarrassing saying 'ARS did it'. Because there's things ARS can do that you can't prevent. There's other characteristics that ARS has got that you learn, and learn to deal with. And in my opinion I think I'm learning with it. It shouldn't do what it does, but I'm dealing with it." (Video 8)*

Although the majority of signallers felt that ARS was simply an aid and did not bear any responsibility, some were more circumspect preferring to allow ARS to take the blame and considering their role as only stepping in when a situation developed which ARS could not cope with alone. This attitude was expressed by one signaller on the video tapes in particular and another during the interviews, and other interviewees complained of the attitude among their colleagues.

> *"As far as I'm concerned, this is the system we operate to because the thing is, if you start pre-selecting routes on the system there are people who program the system, the timetable into the system, to my mind then, if you start trying to be three or four steps ahead of the system and you cause delay, then that's got to go down to you, as a signalman. Otherwise it's ARS." (Video 4)*

> *"It's better not to step in…if ARS is going to make a move and it's a stupid move and I stop it the delay goes down to me…I'd have to say ARS is responsible or there's no point in having it." (Interview 9)*

> *"When I first came here…one of the guys who was here told me…because management brought this system in, they've got to take the delay. ARS is my slave. It saves me having to set every single route monotonously. It saved me that, but it's my slave. I'm not its slave. But there's some people…" (Video 8)*

> *"Quite often to be honest I have sort of sat and watched some people and I think 'Yeah, ARS is in charge over there'…signallers tend to use it to their advantage, don't they? Blame ARS. You know, a lot of that goes on" (Interview 7)*

Leaving ARS to operate autonomously is not always a good plan however.

> *"We have a signaller up here…he's quite famous for it, he was on this workstation and he wandered off with his back to the workstation, just made some comment about how fantastic ARS was, and it like signalled him into oblivion over here. And it took him about the rest of the shift to get back to normal." (Interview 7)*

This refusal to take responsibility for system performance by some signallers seems to be due more to organisational factors, such as management attitudes and the attribution of delay to the originator, rather than difficulties with working with ARS.

5.6.4.6 Summary

The introduction of automation has changed the level and nature of the signallers' interaction with the system. In particular, the level of monitoring has increased. Where previously signallers only monitored for trains coming on to their patch of railway and failures, IECC signallers now also monitor the automation and how well it is operating with the aim of intervening where necessary. They use a variety of techniques for this including monitoring the progress of individual trains but also closely monitoring 'hot-spots'. Their monitoring strategies allow them to pick up on areas where they need to intervene. The lack of predictability of the automation makes this more difficult but signallers can query the automation to get some insight into its intentions. However, this does not always provide useful or understandable information.

There are also a variety of techniques signallers use to interact with the automation. They use different methods to control the automation depending on suitability to both the railway in their area and the specific situation and personal preferences also seem to play a part in their choice of interaction method. Similar to Swedish signalling systems (Kauppi et al., 2006), signallers in this study indicated that they cannot easily control the automation and have resorted to using reminder devices to constrain it. However, reminder devices were not intended for regulating purposes and their use in this context may degrade their effectiveness as a safety device.

In terms of workload, signallers were very clear that ARS has reduced their physical workload by removing the need to set routes for each train. However, as suggested in the literature (Macdonald, 1999), this has allowed management to assign larger control areas to signallers in IECCs. Mental workload therefore has probably not had a corresponding reduction, both due to the increased control area which the signaller must monitor and the need to 'think ahead' of ARS. Megaw (2005) has discussed the potential for technology to increase cognitive workload in general, but no reference to this concept was found in relation to automation of control systems. Overall workload, although considerably lower during normal operations, suffers an additional burden during disruption due to the need to constrain ARS as well as dealing with the disruption itself. This study therefore suggests that the label 'clumsy automation' may be applied to ARS in this respect (Woods, 1996).

In general, signallers considered themselves, not ARS, responsible for running the railway. There were comments of the opposing nature however, where signallers believed that it was not their job to deviate from what ARS was programmed to do. They considered their job to be applying rules and procedures in the event of disruption. The lines of responsibility therefore appear to be ambiguous and this may make operators less likely to intervene (Mosier et al., 1994).

5.6.5 Discussion of Method

5.6.5.1 Video archive analysis

The video archive analysis was a first step in gathering data on the specific issues with ARS. It was based on eight previously recorded videos of unstructured interviews with signallers at their workstations. The strength of this approach was that the interviews had not primarily been aimed at ARS and so any comments that arose were natural and extremely unlikely to be biased. There was also an element of verbal protocol as the signallers explained what they were doing to the researcher, and why. This gave some insight into signallers' strategies when dealing with ARS. Overall, the videos were a rich and useful source of data and provided a good starting point for researching the issues signallers have with ARS.

5.6.5.2 Semi-structured Interviews

The semi-structured interviews were extremely effective and gathered a large amount of information on signallers' views on ARS. Carrying out the interviews at the workstation gave the signaller the opportunity to illustrate any examples they gave, and this was utilised by all participants. Although a subjective methodology, several interviews were carried out to obtain a range of views. The results were analysed in the light of the knowledge gained from participant observation and direct field observations, thus increasing the validity.

5.7 Conclusions

A great deal of information with regard to system monitoring was gathered, in particular from the semi-structured interviews. It appears there are a variety of strategies signallers use when monitoring the system, including monitoring individual trains, routes set, 'hot-spots', and maintaining an overview of their control area. The importance signallers place upon maintaining awareness of the system is evidenced by their unwillingness to leave it unattended for any length of time. However, the increased level of monitoring associated with ARS does not appear to have been an entirely welcome change as some signallers expressed regret at the reduction in direct involvement in signalling trains.

Signallers' interaction with the system was also examined and the variety of methods they use to interact were identified. The choice of method of interaction varies depending on the circumstances and the personal preferences of the signaller, but reminders were widely used to control the automation. This is a misuse of reminders as they are intended as a safety device and their use for regulating purposes may reduce their effectiveness as a safety device. Their use in this context indicates the lack of other powerful tools to direct automation (Lenior et al., 2006).

Signallers' understanding of the automation was found to be generally quite low. Although they had a reasonable understanding of the basic operation of ARS under normal circumstances there was considerably less understanding of ARS's operations during disruption. This lack of understanding stems from both the unstructured training signallers receive on ARS operations and the lack of feedback or observability of the automation. The result is poor predictability of the automation and signallers are obliged to intervene more frequently as they are unsure of what the automation will do if they do not intervene. This situation must be improved by supporting the development of accurate mental models (Sheridan, 2002). There may be a consequential effect on trust which makes it difficult for signallers to work cooperatively with the automation, a situation compounded in some cases by the blurring of responsibility between the automation and the signaller.

While there are a number of issues with ARS, including incorrect programming, inability to regulate, poor feedback, and poor data entry, they are not sufficient to prevent the majority of signallers using the system. This is in contrast to previous research which suggested that such competency issues are the strongest predictor of automation usage (Muir & Moray, 1989). It may be that in real world systems, operators have the time and opportunity to calibrate their trust to an extent which allows them to identify when the automation can be trusted. Most successful operation is achieved by ARS when there are no problems on the infrastructure and all trains have a timetable provided for them. During disruption, when well designed, ARS can be of benefit if it is capable of keeping other parts of the workstation moving reliably. It is not useful in the area where disruption is occurring and in most cases must be switched off. The introduction of ARS has clearly reduced the physical workload of signallers as it has removed the need for signallers to manually set routes for each train. However, this has meant that a larger control area can be assigned to each signaller and as they are expected to maintain an awareness of this larger area it is likely that their mental workload has increased. There are a number of situations which ARS is unable to manage alone and the signallers must be aware of these as they develop and be prepared to step in and take over. Furthermore, during disruption the signaller is required to constrain ARS while dealing with the disruption and this represents an additional burden during situations where workload may already be high.

Overall, ARS seems to have been a positive addition to the signalling environment enabling one signaller to control a large area and by and large they themselves are positive about it. However, it is far from perfect and there are a number of areas upon which future automation could improve, for example, specific competency/ performance issues, improved feedback, and improved control interfaces. However, the imperfections in the current system do have the advantage of keeping the signaller involved and this is a positive trait which should be desirable in any future system.

5.8 Chapter Summary

This chapter has presented the results of qualitative investigations into ARS. Descriptive data were gathered on system performance issues, knowledge of ARS, and monitoring and interaction with ARS. Key findings include identification of specific competency issues, a lack of support in mental model generation, types of monitoring strategies, and a lack of directability. An important finding was the use of safety related devices (reminders) to control the automation.

The qualitative data provided a rich picture of ARS but could not quantify the effect of automation on performance and workload. An experimental study was designed to investigate these relationships, in addition to data on signaller gaze and behaviour, and SA. This experiment is discussed in Chapter 6.

CHAPTER 6: LEVEL OF AUTOMATION EXPERIMENT

6.1 Chapter Overview

This chapter presents the design and results of an exploratory simulator experiment examining the effect of level of automation (LOA) and disruption on a number of dependent variables in a signalling context. Three LOA were examined and two levels of disruption. The dependent variables were workload, behaviour observations, gaze fixations, performance, and SA. Key results indicate that workload, during both normal and disrupted phases of the experiment, decreased as the level of automation increased. However, the high automation condition saw the greatest average increase in workload between the normal and disrupted phases of the experiment. Higher and more consistent performance scores were achieved in the high automation condition as compared to the other conditions.

6.2 Introduction

Building on the previous studies, an experiment was designed to gather empirical data on the impact of automation. Signalling simulators are available for use in most IECCs and though they are generally used for training it was thought to be also possible to use them for research. The level of automation can be varied on a simulator by turning off the automation (i.e. manual control). When ARS is not available for use, signallers can use a lower form of automation, Auto-routes, in which trains can be automatically signalled along a particular route. The simulators therefore meant it was possible to design an experiment to compare three LOA. Figure 6-1 illustrates the position of this experiment as the final piece of quantitative data collection for this research.

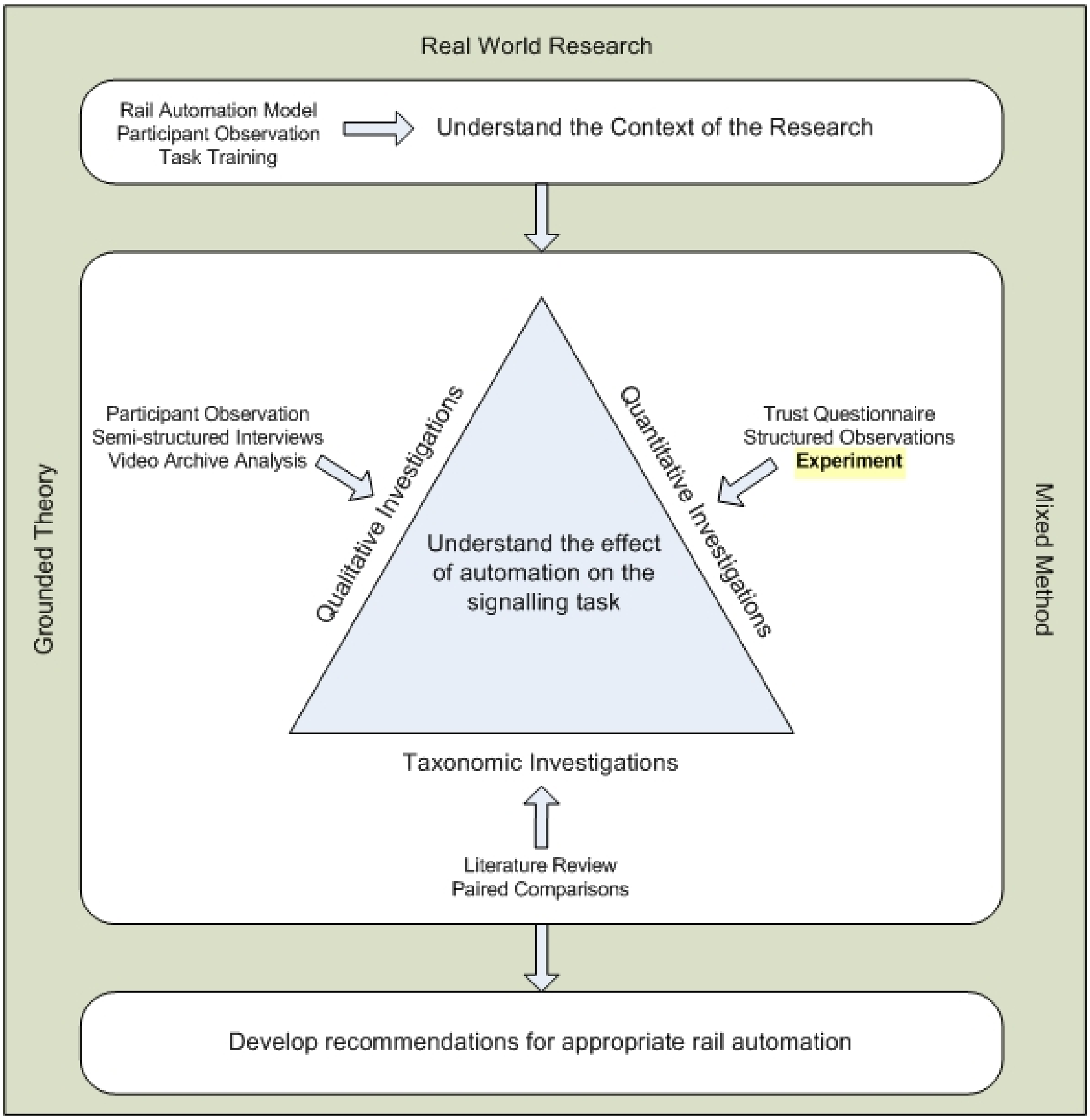


Figure 6-1: Position of the Level of Automation Experiment in the Research Framework

The previous interview study indicated that ARS is most effective during normal operations but can become a hindrance to the signaller during disruption; that is, it conforms to what has been labelled "clumsy automation" (Sarter et al., 1997). This piece of research aimed to gather targeted information on the effect of automation during both normal and disrupted operations. Workload and performance were identified as the main variables of interest. The previous studies indicated that workload was generally reduced by ARS, although perhaps less so during disrupted running. This corresponds with findings from other research (Kaber et al., 2000; Kantowitz, 1994). Performance was not investigated in any of the previous studies in this research as the methods employed did not facilitate effective measurement of performance. However, the simulator used for this experiment provided an objective performance score for each scenario. Performance in terms of mission effectiveness is a popular metric for evaluating and comparing levels of automation (Donmez, Pina, & Cummings, 2009). Without improving performance, or at least maintaining performance, it is difficult to see the benefit in automation. Data on performance are therefore an important part of evaluating an automated system. The coding scheme developed in Chapter 4 to support structured observations of signallers was also employed in this study. The observation study looked at the differences in individual signaller behaviour. This experiment provided the opportunity to examine the differences in signaller activity and behaviour under different levels of automation.

Throughout this research data were collected on monitoring. The observation studies indicated that signallers engage in different types of monitoring behaviour under different conditions, and the interview data generated some hypotheses on types of monitoring in which signallers may engage (e.g. monitoring hot-spots or monitoring route setting). However, the information gathered from these studies was quite high level and more precise data on monitoring strategies are required to help inform the design of future automation interfaces. In pursuit of this, eye tracking equipment was used to determine eye fixations under the different conditions. Eye tracking equipment allows the experimenter to track the movement of a participant's eyes within a domain. Duchowski (2003, p. 3) stated that we may presume that this follows the participant's path of attention and hence gives "some insight into what the observer found interesting". Eye tracking assumes that visual attention is predominantly represented by the participant's foveal focus (i.e. the direct focus of the eyes) and neglects information which may be processed using peripheral vision. Thus, eye tracking equipment can help identify what pieces of information are important to the operator and how their monitoring strategies differ under the different experimental conditions. This information can be used in the design of interfaces or training programmes (Ottati et al., 1999).

A piece of research examining signaller SA and its measurement was ongoing at the time of this experiment and as part of that research a questionnaire was developed to be administered at the end of an experimental scenario (Golightly, Balfe, & Sharples, 2009). Research in domains such as telerobotics and automobile control has shown that SA varies with level of automation (Endsley & Kiris, 1995; Kaber et al., 2006). The results predominantly indicate that operator SA is better during intermediate levels of automation, although there is some indication that high levels of automation may improve SA if the automation is used to relieve a high operator workload (Endsley & Kaber, 1999). The interviews conducted with signallers found that they had difficulty understanding and predicting the automation suggesting the possibility that automation may have a negative affect on signaller SA. This experiment offered an opportunity to test this hypothesis using the questionnaire developed for signallers. However, the first step in measuring SA in a new domain is the identification and thorough understanding of the elements necessary to build and maintain SA in that domain (Endsley & Rodgers, 1994) and only one previous study was found which had made any attempt to measure SA in a signalling environment (Wilson et al., 2001). The conceptual understanding of SA in a signalling context was not therefore very advanced, and the questionnaire administered during this study was a pilot of the method.

Trust also emerged as a theme in both the previous studies. However, trust was not investigated as part of this study; as expert signallers were recruited their levels of trust in the system would already have been well established and it was unlikely that the independent variables would be powerful enough to affect these. Nor was it desirable to manipulate participants' trust in a system they work with daily. As the experiment already had five dependent variables a decision was made to exclude trust.

It is important to note that this was an exploratory experiment and it involved use of new techniques and procedures. For example, the simulators had never been used in this context previously and the eye tracking equipment employed to determine gaze fixations had not previously been used in this environment. Therefore, it was not expected that all the measures would be fully successful, but the knowledge gained in applying them would help inform future research in this area.

6.3 Method

6.3.1 Participants

Six participants took part in this study. All were male signallers from Liverpool Street IECC. Signaller participation was arranged in advance, although it proved extremely difficult to procure signallers for the experiment due to staff shortages in the signal box in which the experiment was based. The embarrassment of wearing the head mounted eye tracking equipment also made recruitment of participants difficult. For this reason, the number of participants was limited to six. The participants had a minimum of 5 years experience in the signal box and thus were expert signallers with familiarity of the signalling area used in the experiment.

6.3.2 Apparatus

6.3.2.1 Liverpool Street Simulator

The existing simulator in Liverpool Street IECC was used for this experiment (Figure 6-2). This simulator is typically used for training new recruits and to assess existing signallers. All the workstations in the IECC are accurately represented on it and Stratford, as the most complex workstation, was chosen for the experiment. Although not an exact physical replica of the real workstations, this simulator functions in an identical manner to a real workstation and has the same number of screens and identical input devices (i.e. trackerball and keyboard). The simulator gives a percentage score based on performance compared to the timetable.

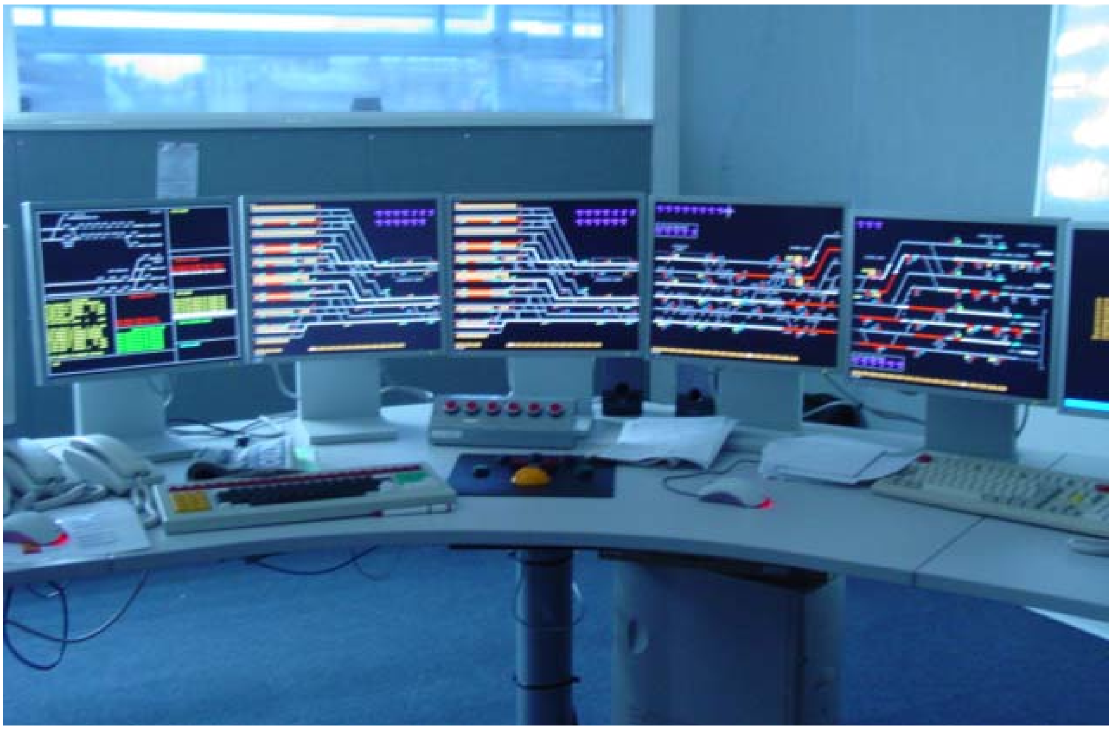


Figure 6-2: Liverpool Street Simulator

6.3.2.2 Integrated Workload Scale

The Integrated Workload Scale (IWS; Appendix N) was used to measure participants' perceived workload. This is a nine point scale developed specifically to measure perceived mental workload in the signalling environment. A high score on the IWS indicates a high workload. Pickup, Wilson, Norris, Mitchell, and Morrisroe (2005) report that this tool has proven to be a valuable measure of peaks and troughs in workload over time or within a set of scenarios. They also report that the tool is acceptable to signallers, having been developed specifically for use in the signalling environment, and maps well onto expected workload measured using other techniques. It was constructed using the Thurstone technique and so the ratings can be used as interval data. Participants were provided with a laminated copy of IWS and asked to verbally rate their workload on this scale at 2min intervals throughout the experiment.

6.3.2.3 Head Mounted Eye Tracker

Participants' gaze fixations were determined using eye tracking equipment. An SMI iViewX HED head mounted eye tracker was used in this study (Figure 6-3; Appendix O).

Figure 6-3: Head Mounted Eye Tracker

The eye tracker is attached to a bicycle helmet which holds the equipment stable on the participant's head. It is connected to a laptop via two USB cables and collects data in the form of a video of the participant's field of view with a red cross-hair indicating the participant's gaze location (Figure 6-4). Following data collection, this video can be loaded into The Observer XT for analysis.

Figure 6-4: Screenshot of Eye Tracking Data

6.3.2.4 <u>The Observer XT</u>

The Observer XT (Noldus, 2007) software was used to analyse the data gathered from both the behaviour observation and the gaze fixations described. This software is used for the logging and analysis of observation data. Figure 6-5 shows a screenshot from The Observer XT software. The coding scheme was pre-programmed with the codes for the signaller observation, gaze position, and IWS scores. The signaller observation and IWS Scores were coded live during the experiment and the eye tracking video was later loaded and coded in the same event log.

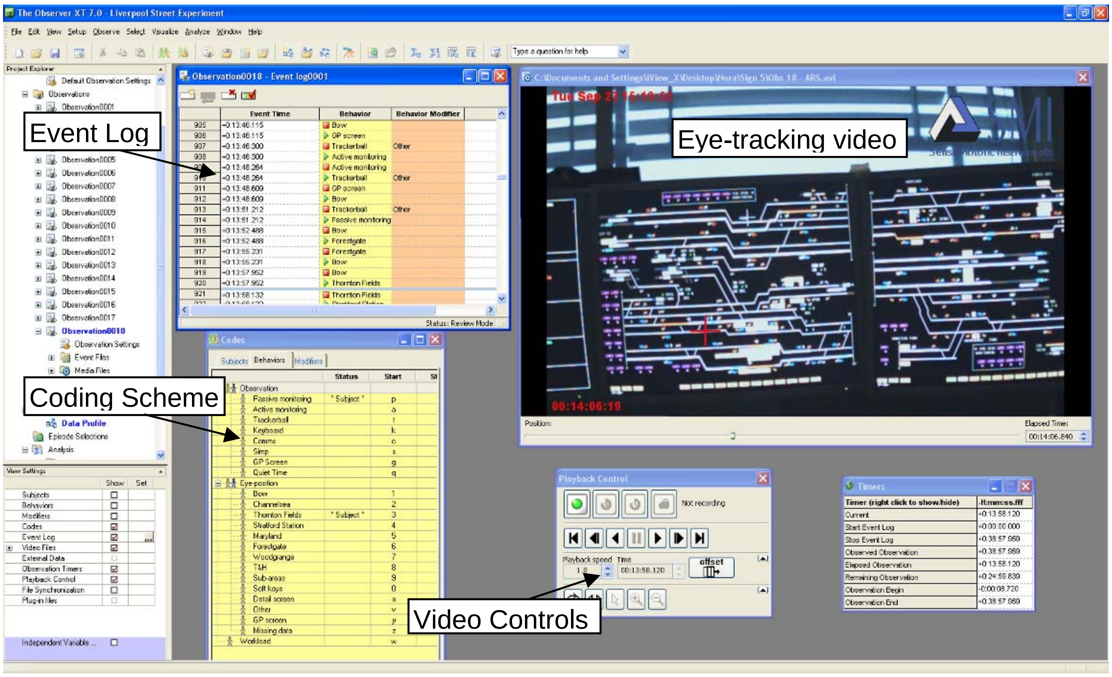


Figure 6-5: The Observer XT Screenshot

6.3.2.5 <u>SME Performance Scale</u>

A three point performance scale was created for an SME to rate signaller performance throughout the experiment at the same intervals as the IWS scores were collected (Appendix P). This scale was provided in a laminated format for the SME to refer to during the experiment. The SME also made notes of the activity on the workstation which were intended to provide rich data to contextualise some of the other more objective measures.

6.3.2.6 <u>Situation Awareness Questionnaire</u>

A questionnaire containing two measures was administered to assess SA (Appendix Q). The first measure was a simplified version of SAGAT (Endsley, 1995) which was administered only at the end of the simulation, so was not disruptive. The second part of the questionnaire was a rail contextualised version of 3D-SART (Situation Awareness Rating Technique) (Jones, 2000). The simplified SAGAT consisted of blank representations of the two overview screens (i.e. main signalling screens) of the workstation used in the experiment and the rail contextualised 3D-SART consisted of three questions asking the participant to rate the complexity, attentional demand and understanding of the simulation.

6.3.3 Design

Table 6-1 shows the experimental design. A part-counterbalanced repeated measures design was used in which three LOA were examined; ARS, Auto-routes, and Manual. Each condition lasted for 30min and used the same scenario based on the same section of the timetable. After 15min disruption was introduced.

Order	*Group A*		*Group B*		*Group C*	
1st	ARS	Normal	Manual	Normal	Auto-routes	Normal
		Disrupt		Disrupt		Disrupt
2nd	Auto-routes	Normal	ARS	Normal	Manual	Normal
		Disrupt		Disrupt		Disrupt
3rd	Manual	Normal	Auto-routes	Normal	ARS	Normal
		Disrupt		Disrupt		Disrupt

Table 6-1: Experimental Design

In order to balance the potential learning effect the participants completed the three conditions in different orders. However, a learning effect was not anticipated as the participants were expert signallers who operate this timetable and area on a daily basis and are competent to deal with any disruption which may occur. The experiment was not fully counterbalanced as this would have left only one participant in each group.

The Independent Variables for the experiment were:

- Level of Automation
- Level of Disruption

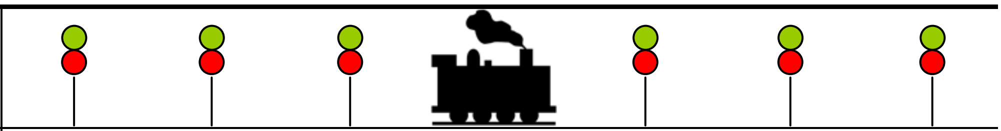
The Dependent Variables for the experiment were:

- IWS Scores (perceived workload)
- Signaller Behaviour
- Gaze position
- Performance Scores (generated by the simulator)
- SME Performance Scores
- SA

6.3.3.1 Level of Automation

Three levels of automation (LOA) were examined in this experiment; ARS, Auto-routes, and Manual. In the 'ARS' condition ARS was operating and available for use and all trains were running in ARS. In the 'Auto-routes' condition all of the ARS sub-areas[25] were switched off, and therefore ARS was not available for use, but signallers could set up auto-routes. These are a lower form of automation whereby a route set by the signaller remains permanently set and all trains arriving on that section of track are automatically signalled along that route. Any trains taking a different route require the signaller to cancel the auto-route and set the alternate route. Finally, in the 'Manual' condition the signallers were required to route all the trains manually without any automated assistance.

The route setting for each level of automation is described in Table 6-2[26]. Provided ARS has access to a timetable for a train it will set the route in front of the train. It will always set two green signals in advance of the train if possible. The signals behind the train remain red, unless another train requires that route. In the Auto-routes LOA all the signals remain green, unless a train has passed into the section the signal protects. In this case, the signal changes to red to protect the train by preventing other trains from passing into the same section. In the Manual LOA, the signaller is required to set all routes and this controls the colour of the signals. The signal immediately behind the train will still be red to ensure that the train is protected.



ARS	Red	Red	Red	→ Direction of travel	Green	Green	Red
Auto-routes	Green	Green	Red		Green	Green	Green
Manual	As set	As set	Red		As set	As set	As set

Table 6-2: Route Setting for Each Level of Automation

[25] ARS sub-areas divide the area of control into small pieces; ARS can be turned on or off in each area.

[26] This illustration uses a two aspect signalling system for simplicity. The signalling on the simulator was four aspect. See Appendix C for an explanation of two, three and four aspect signalling.

6.3.3.2 Level of Disruption

A form of disruption was introduced half way through the experiment, meaning there were two levels of disruption, normal and disrupted. Choice of disruption was a key part of the experimental design as a noticeable effect on workload was required. Many disrupted conditions on the railway involve a high degree of communication and/or knowledge and application of the rules. It was necessary to control communications as far as possible to ensure that they did not affect the results. It was also desirable to avoid application of the rules as this held ethical considerations in the event of mis-application of the rules by any participant. For these reasons, disruption in the form of closing a section of track, a platform at Stratford Station, was chosen as it minimises communication and application of the rules. The participants were required to route trains around the closed platform and regulate this change to the service.

The disrupted condition was always second in the experiment; the participants encountered 15min of normal running and then 15min of disruption. It was not possible to vary the order as disruption has consequential effects and even if the platform had reopened, the signaller would still be required to regulate around the resulting delays.

6.3.3.3 Behaviour Coding Scheme

The participants were observed during each scenario to note their activity. The same method was used as in the earlier observations presented in Chapter 4 but with one important difference; rather than coding manually at 5s intervals; the signaller activity was coded live using a software package. This allowed for much more accurate time intervals. Fewer codes were required than during the live observations as planning tools and communications were limited in the simulated environment.

The codes used were:

- Active monitoring – as with the previous observations, active monitoring was coded when the signaller was sitting up while viewing the signalling screens.
- Passive monitoring – this was coded when the signaller was sitting back while viewing the signalling screens.
- Trackerball – use of the trackerball.
- Keyboard – use of the ARS keyboard.

- Communications – any communications pertaining to the signalling task. An SME was present to deal with these.
- Simplifier – use of the paper based simplified timetable.
- GP Screen – looking towards the GP screen.
- Quiet Time – not involved in the signalling task.

The use of software also allowed the purpose of trackerball use to be logged. The following interventions were coded:

- Route Set – setting a manual route.
- Cancel route – manually cancelling a route.
- Sub-area on – turning on an ARS sub-area.
- Sub-area off – turning off an ARS sub-area.
- Set auto-route – setting an auto-route.
- Cancel auto-route – cancelling an auto-route.
- Change view – the workstation has several views, two overviews and eight detail views, but only four screens. These four screens may display any of the ten views and the signaller uses the soft keys to move between them.
- Use of a reminder appliance – applying or removing a reminder appliance.
- Unknown – any use of the trackerball for which the reason was not clear.

6.3.4 Pilot

The experiment was piloted twice before data collection began. The outcomes of each pilot and subsequent changes are detailed in the following sections.

6.3.4.1 Pilot One

The study was initially piloted with a signalling SME from the NR Ergonomics team. This team member had previously worked and managed the signalling centre where the trial took place so was familiar with the infrastructure and traffic patterns, although not up to date. The SME was used for piloting because problems were anticipated with sampling for the real experiment due to staff shortages and reluctance to participate by some staff in the signal box. The use of the SME preserved the largest possible sample size for the experiment.

The pilot revealed a number of issues, specifically:

- The disruption half way through did not create a large enough change in workload;
- Each scenario was too long, resulting in fatigue by the end of the three scenarios;
- The eye tracking equipment was not sufficiently accurate to pick up exact locations of fixations;
- It was difficult to control the experiment with only one researcher;
- Difficulties using the eye tracking equipment;
- Three different scenarios were too complicated;
- The SA test had been based on the incorrect maps.

These are addressed in the following sections.

Workload Adjustment

The IWS scores were captured for each of the three scenarios (Figure 6-6). It became clear early on in the pilot that the perceived workload was not greatly increased following the disruption. The disruption for the pilot was the closing of the electric line to the next workstation (Ilford). This meant that all traffic on the electric line travelling in that direction had to be routed across to the main line. The 'ARS' condition showed an increase in workload immediately after the disruption was introduced. This was as the participant identified all the areas requiring reminder devices to stop ARS routing over the affected line, and as he applied these reminders. However, once these had been applied the workload dropped off. Both the 'Auto-Routes' and 'Manual' conditions showed only small increases in workload, although these were sustained.

Following the pilot it was decided that a larger and more sustained effect on workload following the introduction of a disruption was desirable. An additional day was spent with the pilot participant testing effects of different disruptions. The following scenarios were tested, all with ARS on (Figure 6-7):

1. Closing platform 8 at Stratford;
2. Closing platform 8 at Stratford during the start of the evening peak;
3. Closing platform 8 at Stratford slightly later in evening peak;
4. Closing platform 8 at Stratford slightly later in the evening peak with two additional trains.

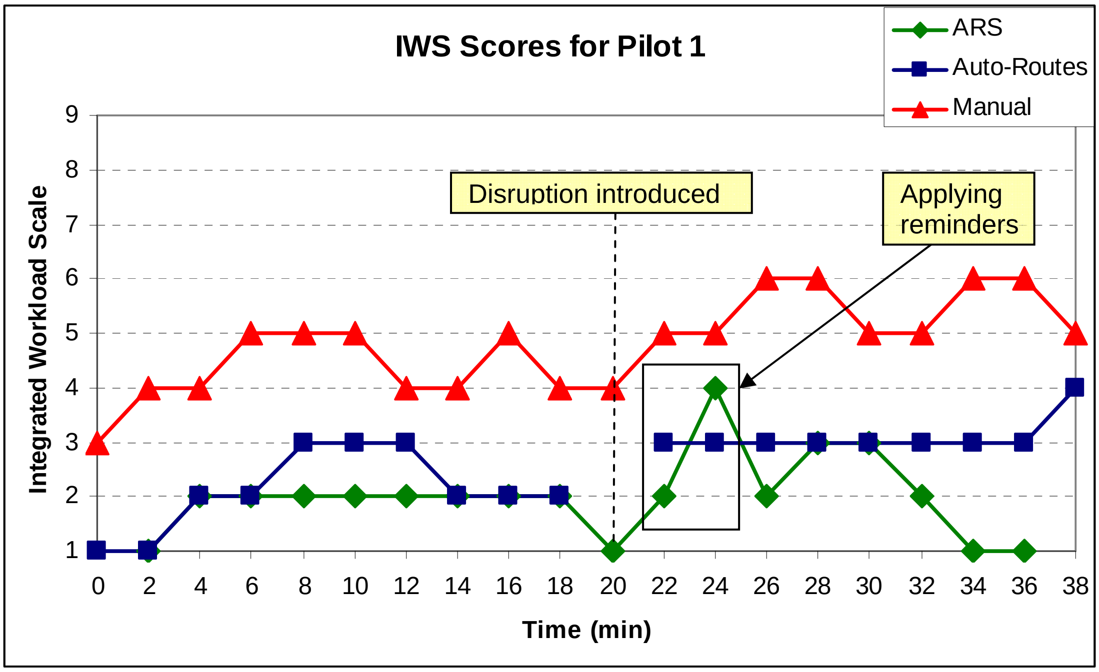


Figure 6-6: IWS Scores for First Pilot

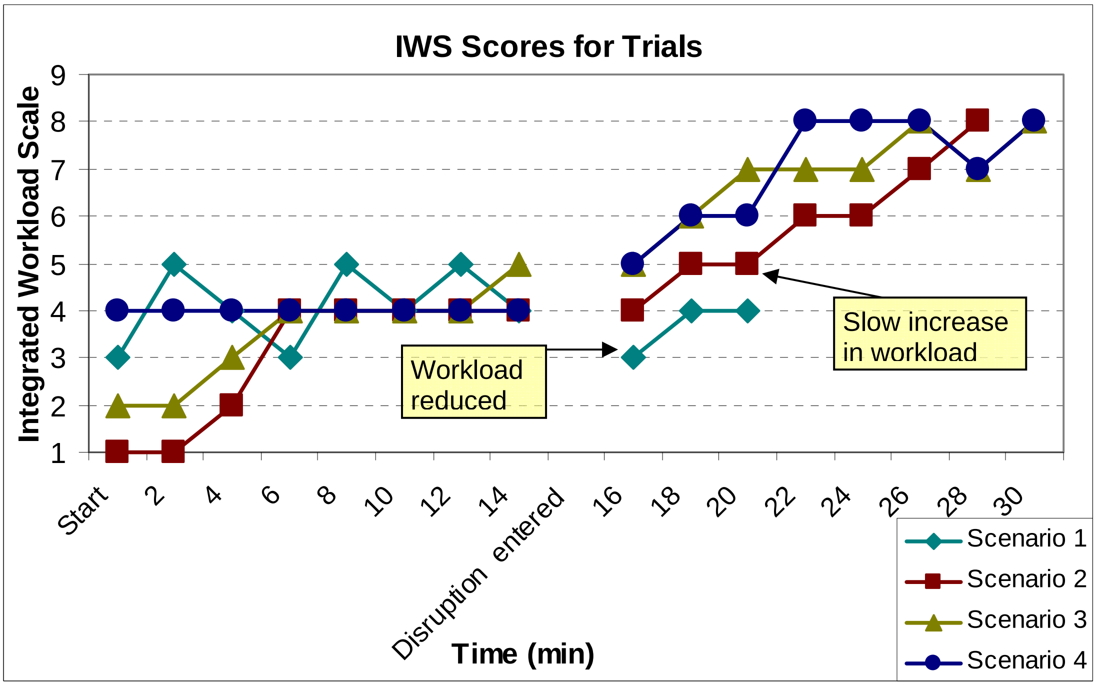


Figure 6-7: IWS Scores for Workload Adjustment Tests

During these tests it became apparent that the major problem with manipulating the workload was the decision to exclude any disruption containing elements of communication. Regulation is the signaller's main task during disruption, but although disruption alone can raise the workload, the associated communications draw their attention and concentration away from the regulation task and without this effect it was difficult to greatly increase workload.

Despite this finding, a decision was made to still withhold communications from the experiment. There were a number of reasons for this. Most importantly, the original reasons still stood. This experiment was intended as a pilot in the area and it aimed to look solely at the interaction between the operator and the automation. Communications would add another dimension into the experiment and the results would become more difficult to interpret. Secondly, this experimental design avoided any issues with signaller competence by keeping the actions within the capabilities of the interlocking; that is, the scenario chosen meant that the interlocking supported all decisions made by the signaller and no unsafe decision could be enforced. If communications were to be introduced into the experiment it would open up the possibility of a signaller breaking, or not following, the rules. There may then be ethical issues surrounding reporting of the signaller’s competence. In addition to the original reasons, time pressures meant that fundamentally redesigning the scenario of the experiment was not an attractive option. Finally, Figure 6-7 indicated that it was possible to manipulate the workload upward and keep it there for the duration of the second half of the experiment.

Time Adjustment

During the pilot each scenario was 40min long. It was apparent that this was slightly too long resulting in participant fatigue and a decision was made to reduce each scenario by 10min. This meant that each scenario would yield 15min of normal running data and 15min of disruption. No difference in the quality of the data collected was anticipated as a result of this decision.

Eye tracking accuracy

If the participant moved significantly during the experiment, the accuracy of the eye tracking was lost. This was due to the parallax error induced when the participant moved from the position they were in when the equipment was calibrated. Instead of being accurate to within 2-3mm as is possible if the participant remains in the same position for the duration of the experiment, the accuracy appeared to be more of the order of 2-3cm. It was possible to estimate this because data gathered during the manual condition in the pilot showed the eye tracking cursor preceding the trackerball cursor but with an offset of up to 2-3cm. It is reasonable to assume that when setting routes the signaller does look at the signal head as he/she moves the cursor to it.

It may have been possible to request the participants to sit as still as possible for the duration of each scenario, or to use a harness to restrain them, but as previously noted in the observation studies, signaller position may give an indication of the level of attention the signaller is paying to the monitoring activity. The experiment was aiming to look at the differences between three levels of automation and eye tracking was only one measure being used. Restraining the signaller would have affected the other data collection. It would also have caused discomfort to the participants and perhaps influenced their actions by increasing the artificial air of the experiment. For this reason a decision was made to allow the signallers to move as they wished and accept the parallax error induced.

There was an initial intention to record very accurate eye tracking data which would allow an analysis of the differences in monitoring between LOA in terms of detailed fixations on signals or tracks. The results from the pilot indicate that such a level of detail was impossible due to the parallax errors. As a result, the proposed coding for the eye tracking data was greatly simplified, dividing the screens into the eight areas shown earlier in Figure 6-27 and Figure 6-28. This had the added advantage of making the data considerably easier to code in The Observer XT but still held the potential to yield information on the differences between the conditions, for example in the difference in attention paid to the identified areas. It was still likely that some of the data collected would not be usable, but this is not unusual with eye tracking (Morimoto & Mimica, 2005). As this was an exploratory experiment and the first time eye tracking had been used in a signalling environment in the UK, the limitations of the equipment within this experiment were accepted.

Additional researcher

It was found to be difficult to run the experiment efficiently with only one researcher as they were required to set-up, calibrate and record the eye tracking, code signaller activity data live, administer IWS and the SA pilot test whilst also ensuring that the experiment kept on track. A decision was made to utilise a second researcher to administer IWS and the SA pilot and to be responsible for timekeeping, leaving the first researcher free to concentrate on the eye tracking and data coding. A third researcher, a signalling SME, gathered additional performance data and handled any communications with signallers required as part of the experiment.

Use of the Eye Tracking

The pilot of the experiment was the first occasion the eye tracking equipment had been used in a signalling environment by this research group. Unsurprisingly, a number of problems were encountered including:

- Problems calibrating the equipment;
- Remembering to press record after calibrating and setting up the experiment;
- Remembering to save the data after each scenario;
- Remembering to focus the camera on the screen;
- Getting the participant in a good position for calibration;
- Lead loosening on the helmet and corrupting data.

Due to these problems much of the eye tracking data were lost, and that which was gathered was not very high quality. However, it was a very useful learning experience and prompted the production of a checklist to ensure that these issues would not arise in the subsequent experiment (Appendix R).

Reduction in scenarios

As it was found to be difficult to increase the workload following the disruption a decision was taken to use the same scenario for each of the three conditions. This did risk a learning effect through the three conditions, but the order of conditions was balanced to take account of this. This decision also meant that the data would be more comparable as even minor changes to the timetable could have a major effect, but under the new design exactly the same scenario would be encountered in all conditions.

SA questionnaire

The maps used in the SA questionnaire were found not to accurately match the area being simulated as they had been based on old diagrams. The SAGAT element of the questionnaire therefore did not work in this pilot.

6.3.4.2 <u>Pilot Two</u>

The experiment was re-piloted to ensure that it ran as expected with the changes made following the first pilot. The participant on this occasion was the trainer in Liverpool Street IECC. While he continues to work as a signaller his main duties involve training new recruits on the simulator. He was not eligible as a participant in the experiment as he had been involved in the setting up of the experiment. Thus the full sample pool was still preserved for the experiment.

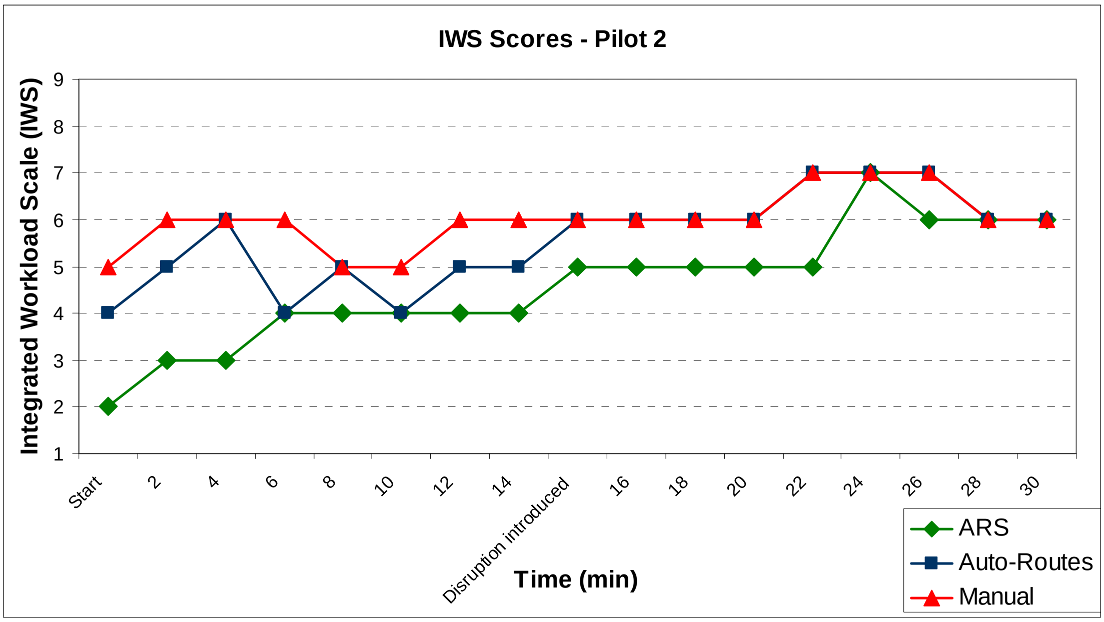


Figure 6-8: IWS Scores for Second Pilot

Figure 6-8 describes the IWS results for the second pilot. An increase in workload can be clearly seen following the introduction of disruption. It was still not as large as desired but it is unlikely that a large increase could be manipulated without the introduction of communications. The increase was at least sustained. Other issues encountered during the first pilot, including eye tracking difficulties, participant fatigue, and researcher pressure were not repeated in the second pilot. However, the new SA questionnaire was not prepared in time for this pilot and so was not included.

6.3.5 Procedure

Six signallers participated in the experiment and three researchers were used to gather the data. The first researcher used a laptop to code signaller behaviour in The Observer XT. The second researcher administered the verbal IWS and the paper based SA test, and was responsible for timekeeping. The third researcher, a signalling SME, sat in the adjoining room and gathered performance data. The SME also handled any communications with the participants required as part of the experiment. The experimental set-up is shown in Figure 6-9.

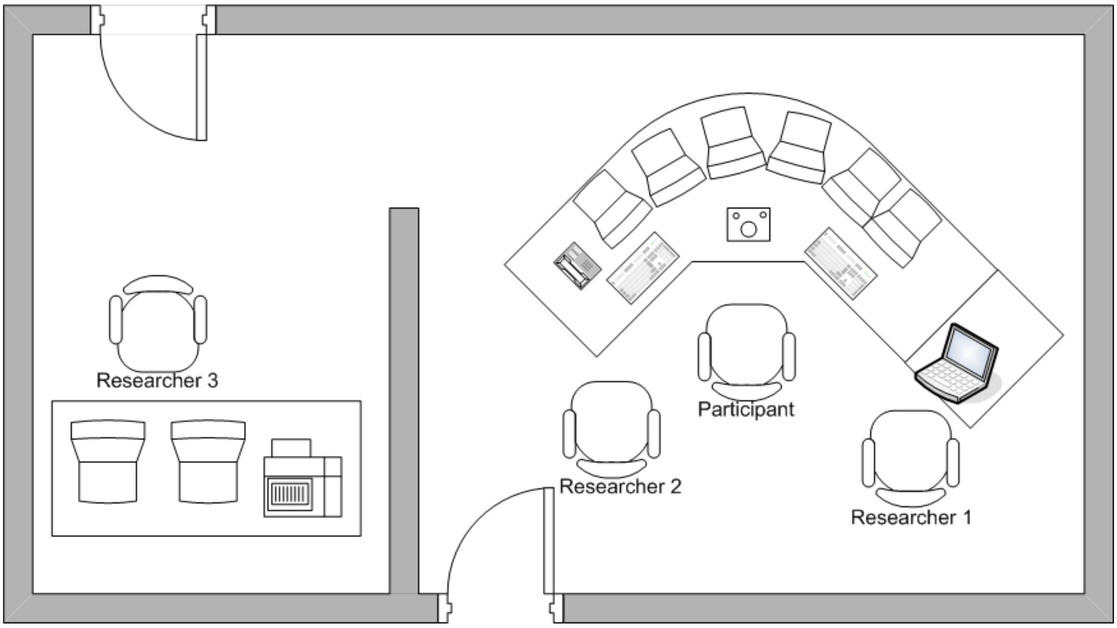


Figure 6-9: Experimental Set-up

The researchers typically arrived at the signal box in the morning, loaded and paused the experimental scenario on the simulator and set up the eye tracking equipment. The participant was then invited into the simulator room and the experiment was explained to him. He was asked to read the briefing sheet and sign the consent form (Appendix S). The participant then took his place at the simulator and was asked to sit as he would when signalling. The eye tracking helmet was placed and secured on his head (Figure 6-10) and calibrated to the middle screen of the simulator.

Figure 6-10: Signaller at Simulator with Eye Tracking Helmet

Once the eye tracking was calibrated and tested, recording began on both the eye tracking data and the observation data. The participant was asked to clap his hands in front of his face so that the two data files could be synced within The Observer XT during later data coding. Following the clap, the simulator was un-paused and the experiment began. The signaller began signalling as he normally would; the first researcher used The Observer XT package to code his behaviours live, the second researcher requested IWS scores at 2min intervals, and the third researcher, assessed the signallers' performance on the assessor's workstation (Figure 6-11).

Figure 6-11: Assessor's Workstation

At the mid-point of the experiment, the third researcher announced the closure of Platform 8 at Stratford Station to the signaller. The remaining half of the experiment was therefore under disrupted conditions. Data collection continued as before.

At the end of the experiment the simulator was paused and the eye tracking and behaviour coding were stopped. The screens were switched off and the second researcher administered the SA questionnaire while the first researcher downloaded performance data from the assessor's workstation. Following completion of the questionnaire, the participant was given a break, although in most cases the eye tracking remained in place for ease of re-calibration, and offered a cup of tea or coffee. Once rested, the same procedure was followed for the second and third scenarios for which the level of automation was changed according to the group to which each participant was assigned.

Following successful completion of all three scenarios, the signaller was thanked for his time and received a gift for his participation.

6.4 Results

6.4.1 Workload

The participants were asked to verbally rate their workload on the IWS Scale every 2min. The results are presented here as a graph showing the average workload scores for each level of automation (LOA) at each 2min interval. It is clear from Figure 6-12 that the ARS LOA was consistently rated lowest and the Manual LOA was consistently rated highest. The Auto-routes LOA initially showed increased workload scores which quickly tapered off. This was due to the necessity to set up the auto-routes. Once these were established the workload fell and remained reasonably consistent until the disruption was introduced. All three LOA showed an increase in perceived workload following the introduction of disruption. The ARS LOA showed a steep increase in workload immediately after the introduction of disruption. This was as the signaller applied reminder devices both to remind himself not to route any trains through the blocked station platform, and to prevent ARS doing so.

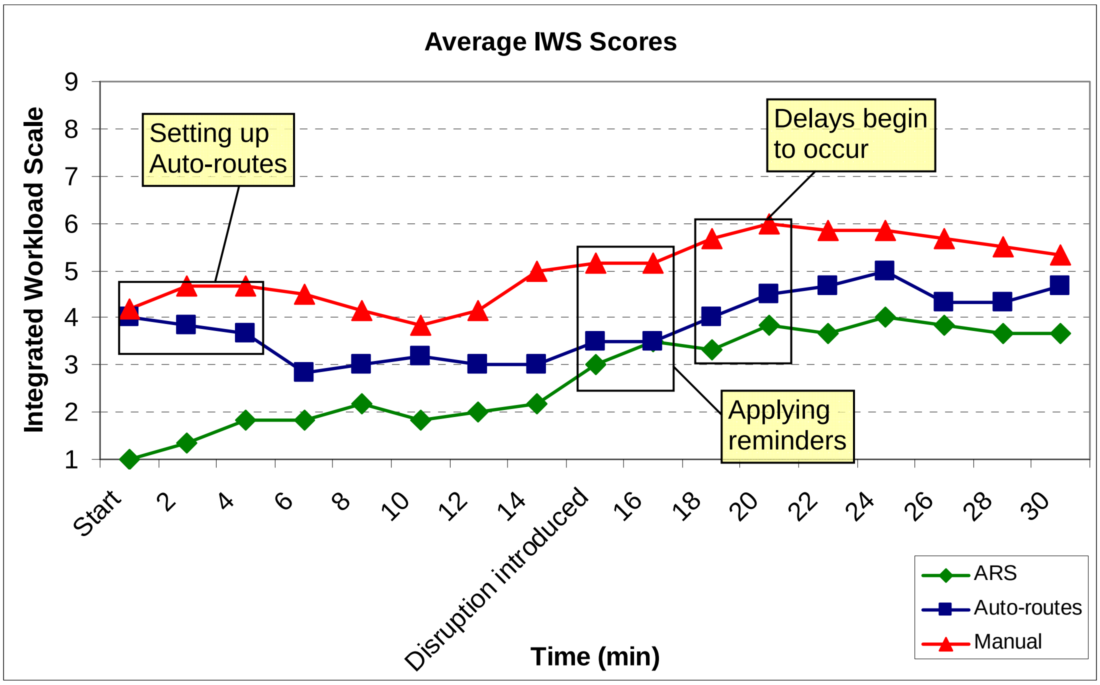


Figure 6-12: Mean IWS Score for Each Condition

Table 6-3 shows the mean score during the normal running phase of the experiment (i.e. the first 15min of each scenario) and the mean score during the disrupted phase (i.e. the last 15min of each scenario) for each of the three LOA. The difference between these is also given.

	Mean Normal IWS	*Mean Disrupted IWS*	*Difference*
ARS	1.77	3.69	+1.92
Auto-routes	3.31	4.38	+1.06
Manual	4.40	5.63	+1.23

Table 6-3: Mean IWS Scores during Normal and Disrupted Conditions

A 2 x 3 ANOVA was run on the IWS data to determine whether there were any significant differences due to LOA or disruption. A significant main effect of LOA was found (F (2, 30) = 12.431, p<.001); Tukey's post-hoc comparison revealed this difference was between the ARS (M = 2.74, SD = 1.01) and Manual (M = 5.02, SD = 0.69) LOA (p<.001). The Tukey post-hoc did not show a significant difference between ARS and the lower form of automation, Auto-routes (M = 3.82, SD = 0.69). A significant difference was also found between the normal and disrupted conditions (F (1, 30) = 14.216, p<.001). No interaction effect was found.

6.4.2 Performance

Two performance measures were used; a simulator score based on the delay minutes caused by the participant and those recovered by the participant, and a SME rated performance score.

Figure 6-13 describes the simulator generated performance score of each signaller for each LOA. It shows that performance was the most consistent across different signallers for the ARS LOA. This was also consistently the highest performance, followed by Auto-routes and finally Manual control, both of which showed a large variation between signallers. A one-way ANOVA was run on these data and a significant main effect of LOA was found (F(2, 15)=9.903, p<.005). A Tukey post-hoc test showed differences between the Manual group (M(Manual) = 75.17, SD(Manual) = 3.97) and both automation groups (M(Auto-Routes) = 81.83, SD(Auto-Routes) = 5.00, p<.05; M(ARS) = 84.83, SD(ARS) = 1.94, p<.005).

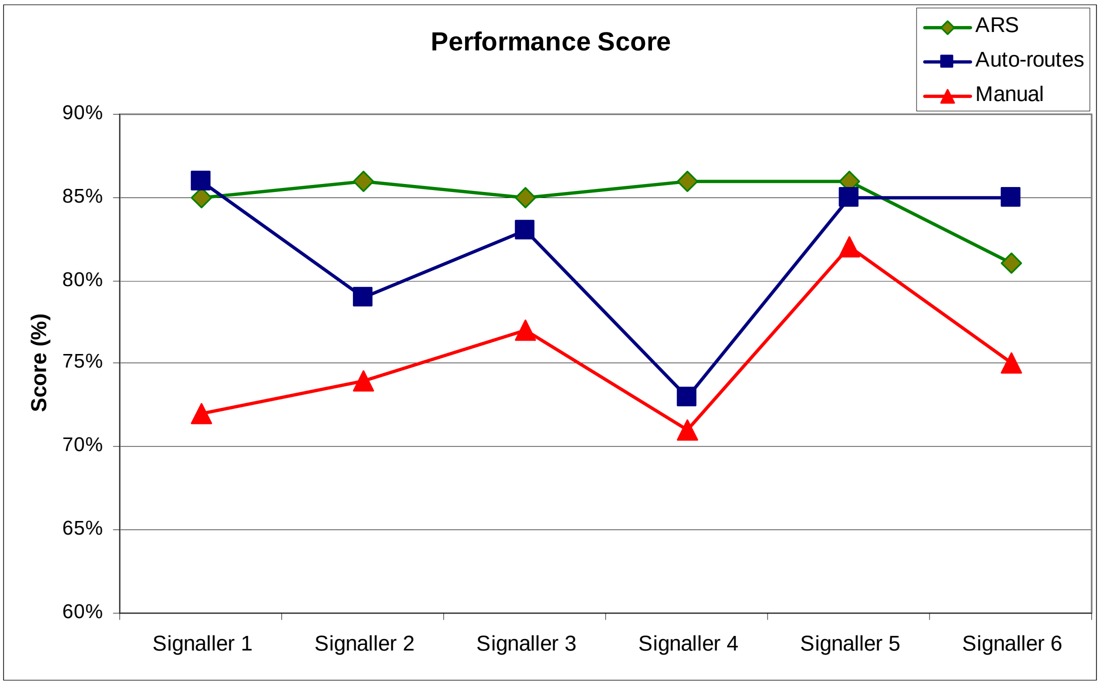


Figure 6-13: Performance Scores

The SME performance scores became corrupted due to a misunderstanding regarding the scale. Participants 1 and 2 were marked on a three point scale, but the remainder were marked using a nine point scale. Although an attempt was made to compensate for this by dividing the remaining performance scores by three, it seemed likely that the consistency of the scale had become corrupted as the SME had allocated a '3' on the original scale considerably more often than he allocated a '7', '8', or '9' on the new scale. It was not clear how this corruption occurred as the SME had a printed scale provided during the experiment. One possible explanation is that the SME mistakenly used the nine point IWS scale, and therefore may have been rating his perception of the participants' workload rather than performance. For this reason the SME performance scores were not analysed.

Similarly the qualitative comments of the SME were not as useful as anticipated. It was hoped that some form of analysis would be possible on these data but the internal consistency of the data collected was not very high. This was compounded by the SME's position in an adjoining room. Although the actions of the signaller were visible to the SME through the Assessor's workstation it was not possible for him to determine the reasons behind the signallers' actions and thus much of the benefit of an SME commentary was lost.

6.4.3 Behaviour Observation

6.4.3.1 Active Monitoring

The following graphs show the amount of time spent actively monitoring; first during normal running (Figure 6-14) and second during disrupted running (Figure 6-15). The results for each signaller are shown on each graph for each of the three LOA.

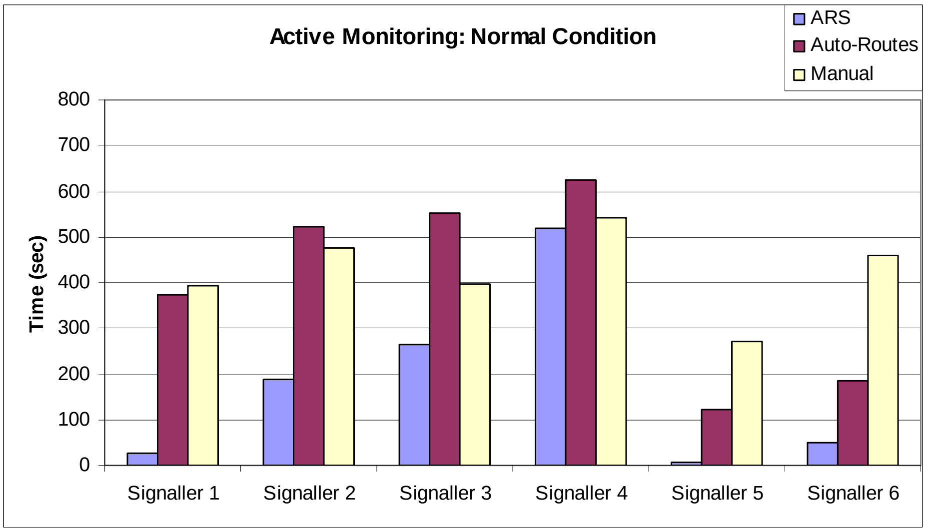


Figure 6-14: Active Monitoring Results: Normal Condition

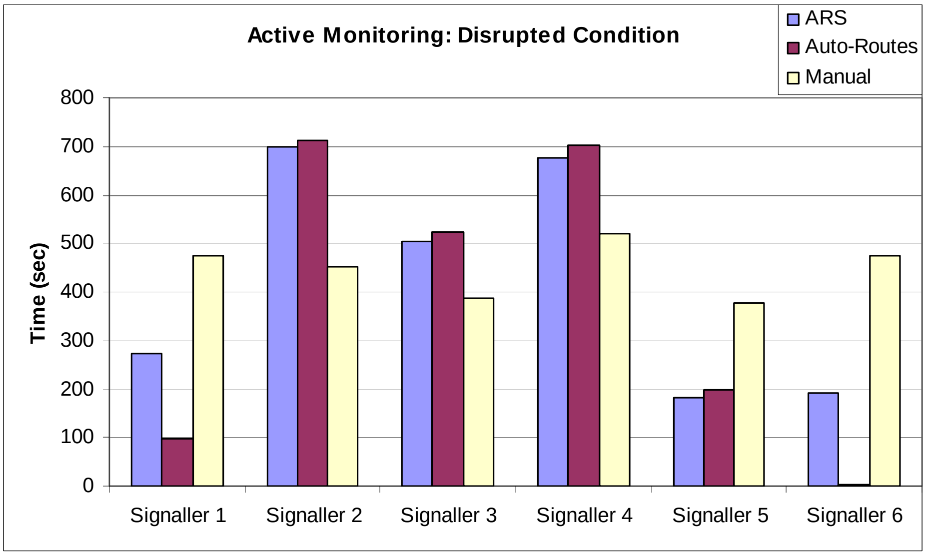


Figure 6-15: Active Monitoring Results: Disrupted Condition

A 2x3 ANOVA found no significant differences for LOA or level of disruption, but active monitoring was lowest during the ARS LOA, and was very variable between different signallers during this condition. Figure 6-16 shows the mean time dedicated to active monitoring for all conditions with the mean IWS scores overlaid.

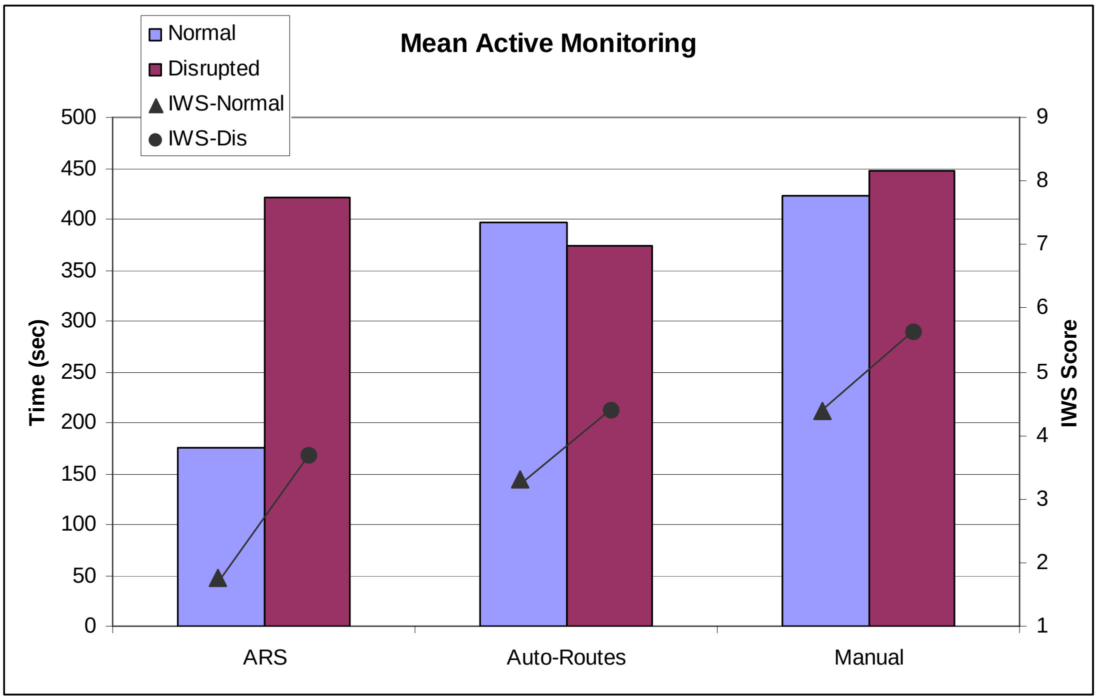


Figure 6-16: Mean Active Monitoring and IWS Scores

6.4.3.2 <u>Passive Monitoring</u>

Passive monitoring was coded in the same manner used during the observation studies (i.e. when the signaller sat back from the workstation). Figure 6-17 and Figure 6-18 show the passive monitoring results for each signaller during the normal condition and disrupted condition respectively.

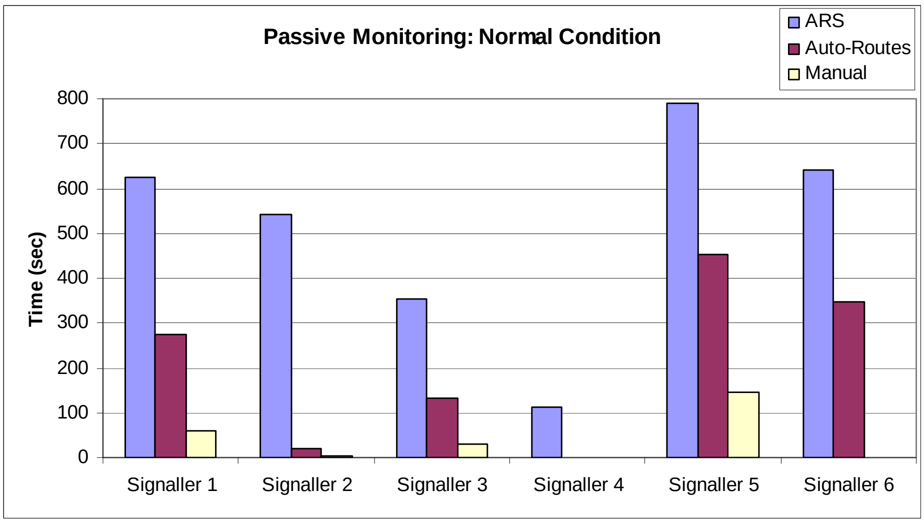


Figure 6-17: Passive Monitoring Results: Normal Condition

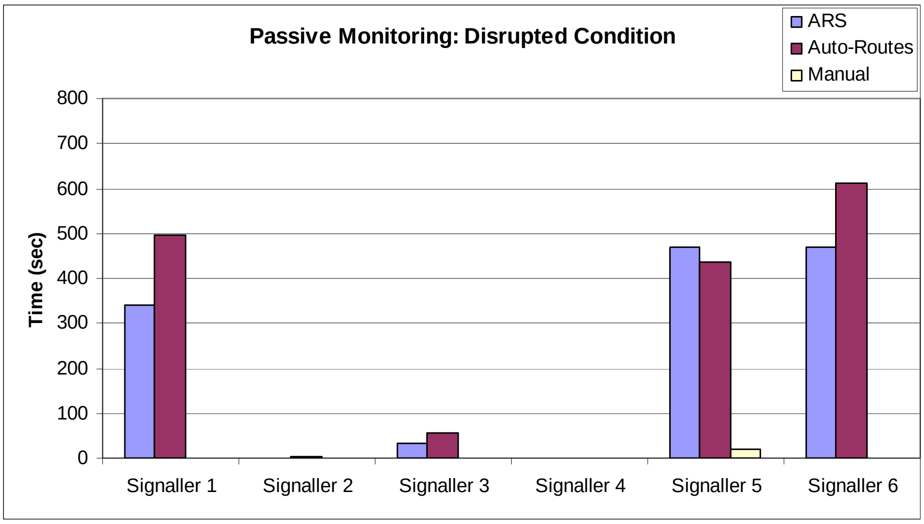


Figure 6-18: Passive Monitoring Results: Disrupted Condition

A 2x3 ANOVA revealed a significant main effect of LOA for passive monitoring, (F = 9.562 (2, 30), p<.05). Tukey's post-hoc comparison revealed this difference was between the Manual LOA and both automated conditions (M(Manual) = 21.5, SD(Manual) = 43.29; M(ARS) = 365.17, SD(ARS) = 272.95, p<.001; M(Auto-routes) = 235.58, SD(Auto-routes = 222.74, p<.05). Passive monitoring therefore increased with the LOA. Figure 6-19 describes the mean time dedicated to passive monitoring for all conditions and the corresponding IWS scores.

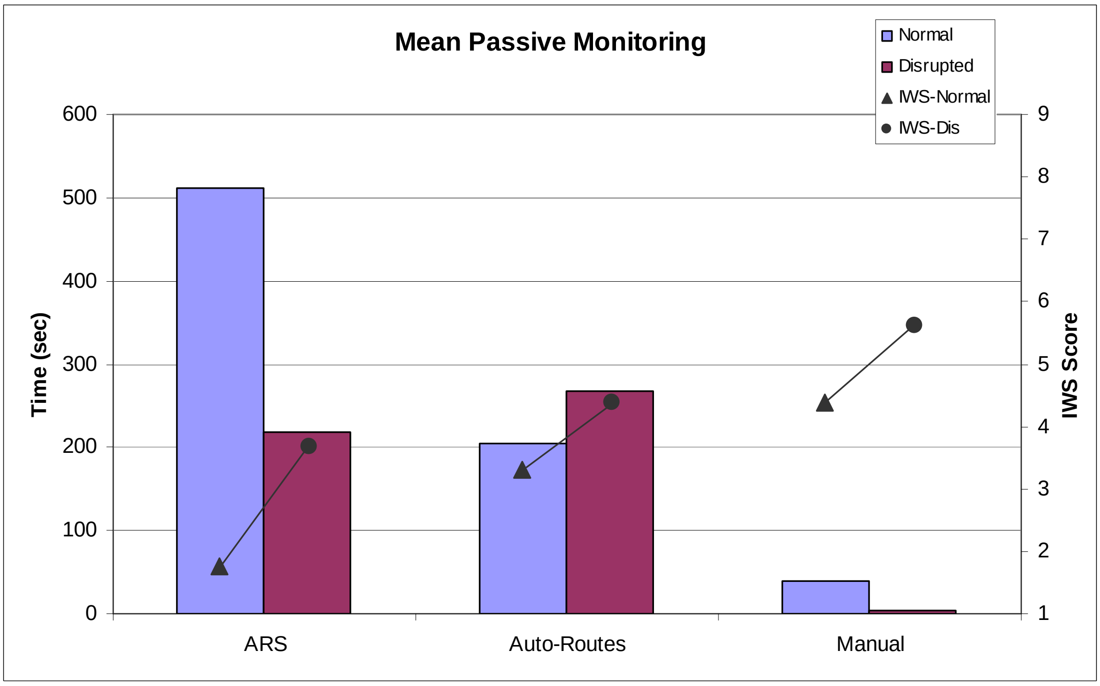


Figure 6-19: Mean Passive Monitoring and IWS Scores

6.4.3.3 Use of the Trackerball

Figure 6-20 and Figure 6-21 show the results for the use of the trackerball for the normal and disrupted conditions respectively.

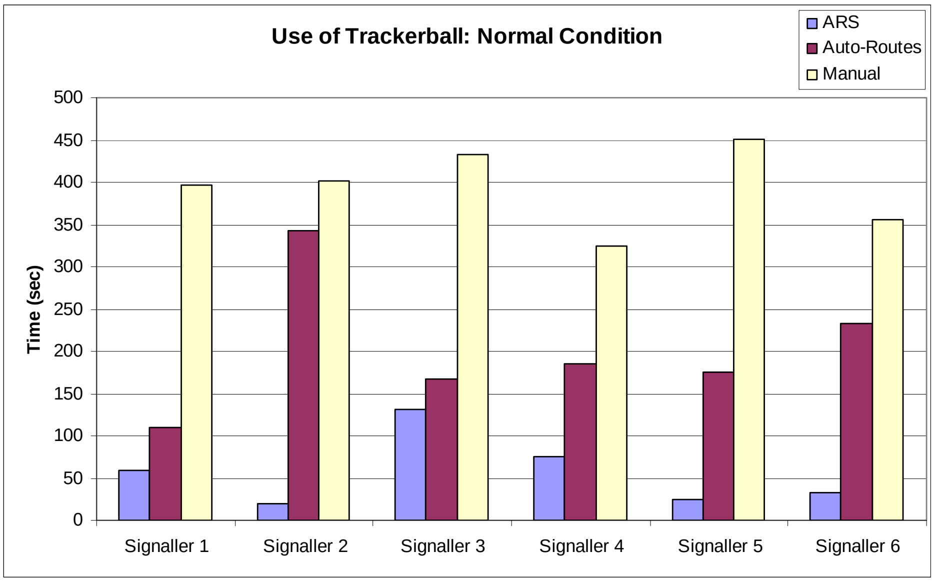


Figure 6-20: Use of Trackerball Results: Normal Condition

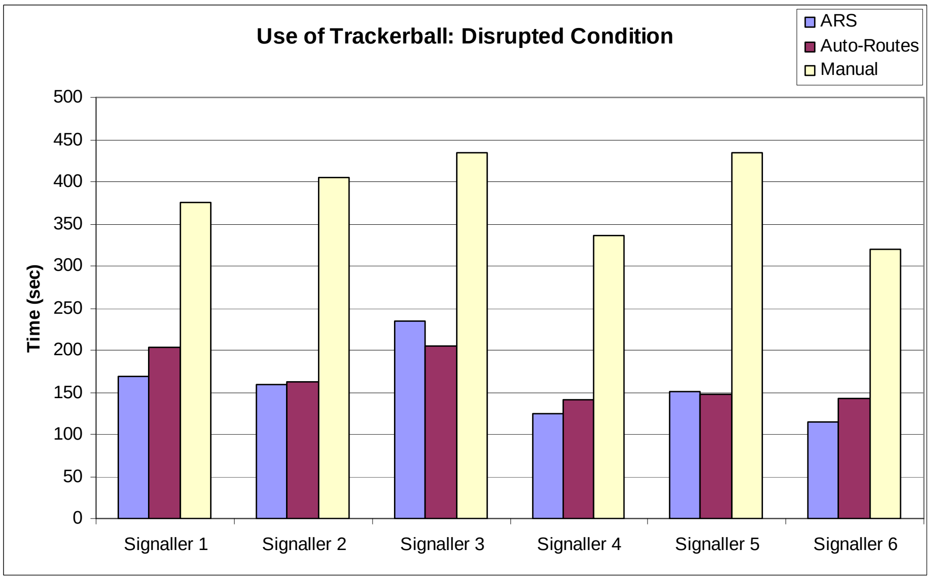


Figure 6-21: Use of Trackerball Results: Disrupted Condition

A 2x3 ANOVA revealed a significant effect of LOA in trackerball use ($F(2, 30) = 99.410$, $p<.001$). A Tukey post-hoc comparison showed this effect was between all LOA (M(ARS) = 107.75, SD(ARS) = 66.58; M(Auto-Routes) = 184.75, SD(Auto-Routes) = 59.76; M(Manual) = 388.92, SD(Manual) = 45.95; p(ARS/Manual)<.005, p(ARS/Auto-Routes)<.005, p(Auto-Routes/Manual)<.001). There was a significant interaction due to the increase in trackerball use when disruption was introduced (F(2, 30) = 6.190, p<.01).

The purpose of the trackerball interventions was also coded. Figure 6-22 describes the mean and standard deviation of the different types of intervention coded for each LOA.

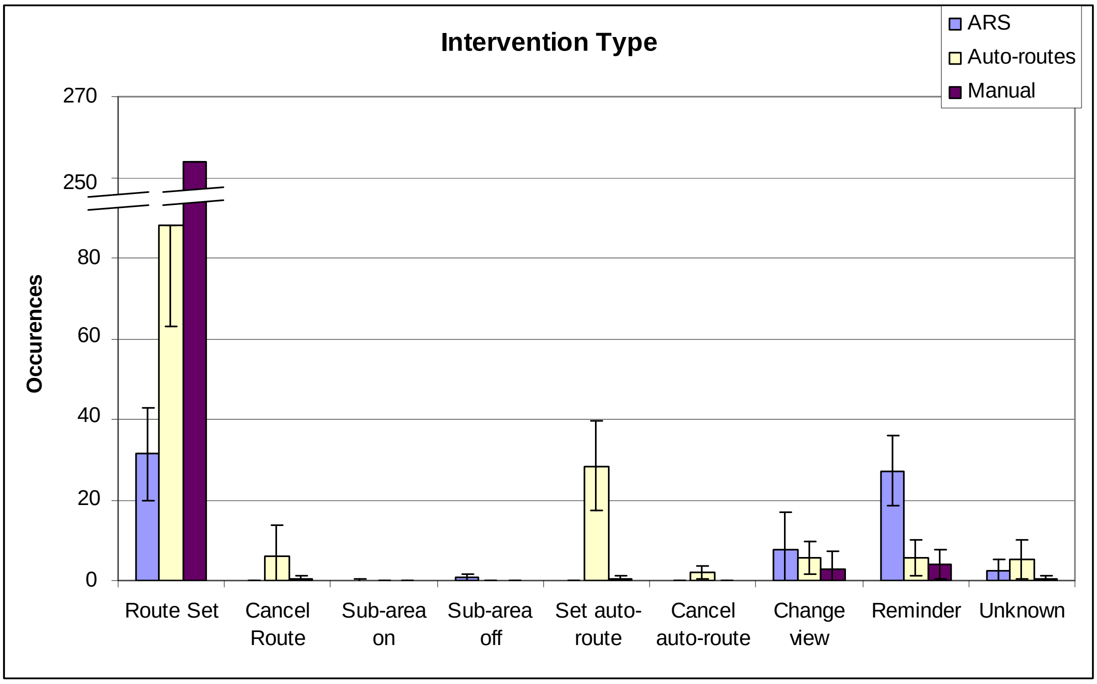


Figure 6-22: Mean and Standard Deviation of Intervention Types

A series of one-way ANOVAs were run on the data and four significant differences were found. A significant difference was found for route setting (F(2,15) = 58.627, p<.001), and a Tukey post-hoc revealed that this difference was between the two automated conditions (M(ARS) = 31.33, SD(ARS) = 11.34, p<.001; M(Auto-routes) = 81.17, SD(Auto-routes) = 25.20, p<.001) and the Manual condition (M(Manual) = 254.00, SD(Manual) = 57.84). A significant difference was also found for setting auto-routes (F(2, 14) = 39.193, p<.001) and this was between the Auto-routes condition (M(Auto-routes) = 28.40, SD(Auto-routes) = 11.17) and both other conditions (M(ARS) = 0, SD(ARS) = 0, p<.001; M(Manual) = 0.33, M(Manual) = 0.82, p<.001). The third significant difference was cancelling auto-routes (F(2, 15) = 7.500, p<.01), and Tukey's post-hoc again showed this was between the Auto-routes condition (M(Auto-routes) = 2, SD(Auto-routes) = 1.79) and both other conditions (M(ARS) = 0, SD(ARS) = 0, p<.005; M(Manual) = 0, SD = 0, p<.005). The final significant difference was applying reminder devices ($F(2, 15) = 27.257$, $p<.001$), and this was between the ARS condition (M(ARS) = 27.17, SD(ARS) = 8.73) and the Manual and Auto-routes conditions (M(Manual) = 4.00, SD(Manual) = 3.69, $p<.001$; M(Auto-routes) = 5.83, SD(Auto-routes) = 4.45, $p<.001$).

6.4.3.4 Use of the Keyboard

Figure 6-23 and Figure 6-24 describe the observed use of the ARS keyboard during the experiment.

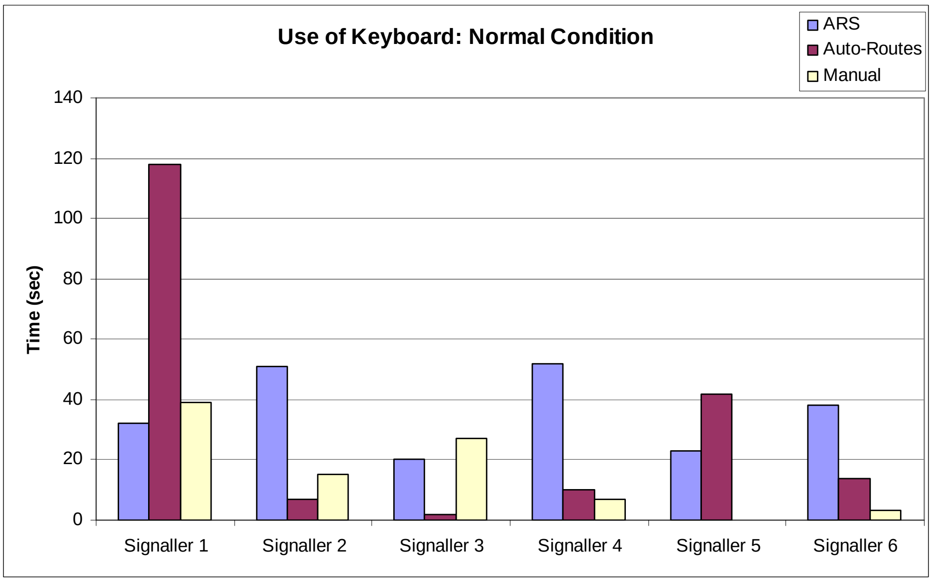


Figure 6-23: Use of Keyboard Results: Normal Condition

Signaller 1 showed increased use of the keyboard during the Auto-routes LOA; this was most likely due to an unintended fault on the simulator at the start of the simulation which he was attempting to rectify using the keyboard. There was no further effect of this fault.

Although codes were programmed into The Observer XT to collect data on the type of keyboard inventions, as was achieved for the trackerball data, it proved too difficult to identify what each intervention was during the experiment and the vast majority of keyboard use was coded as ‘unknown’. It was therefore impossible to do any analysis on the types of keyboard intervention.

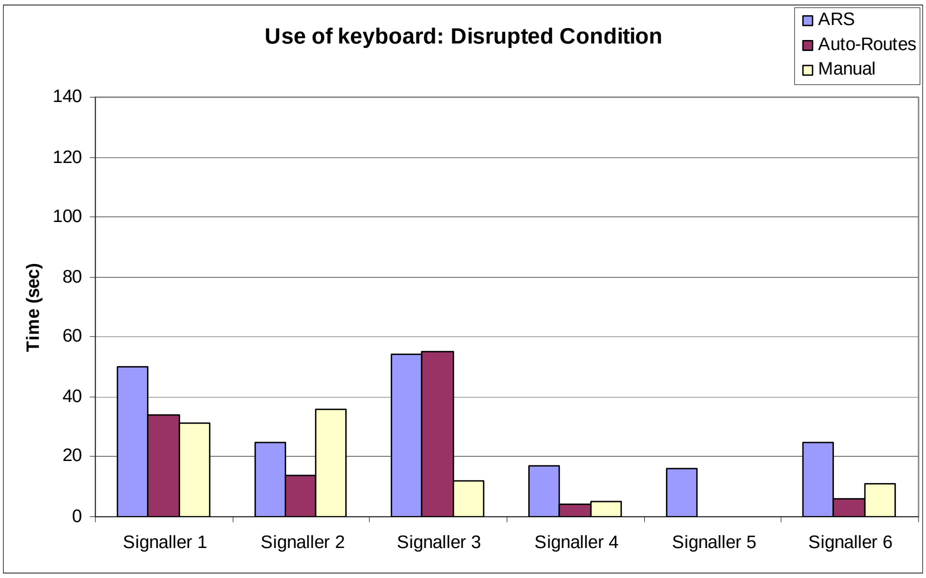


Figure 6-24: Use of Keyboard Results: Disrupted Condition

As can be seen from the graphs, use of the keyboard occupied little of the signallers' time, and no significant differences were found either between LOA or disruption.

6.4.3.5 <u>Quiet Time</u>

Figure 6-25 and Figure 6-26 describe the signallers' observed level of quiet time during the experiment.

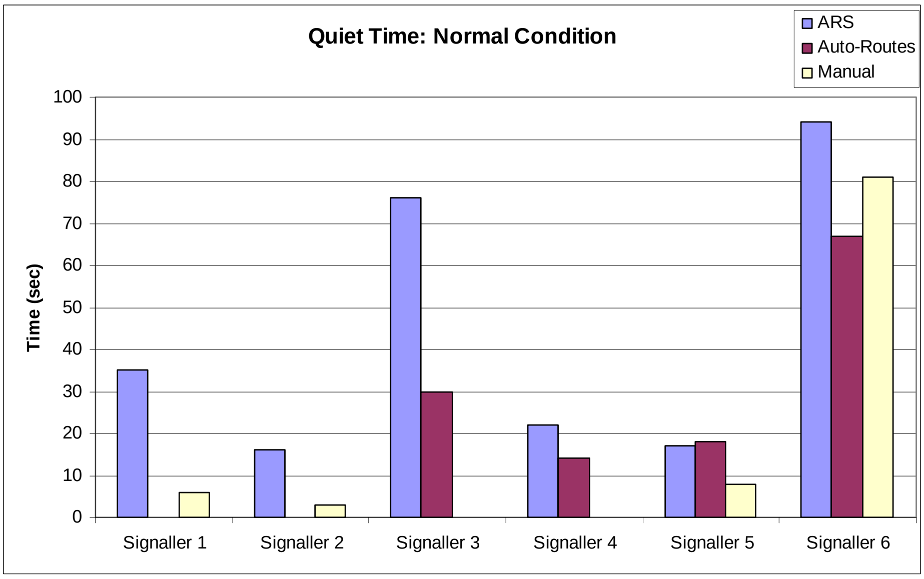


Figure 6-25: Quiet Time Results: Normal Condition

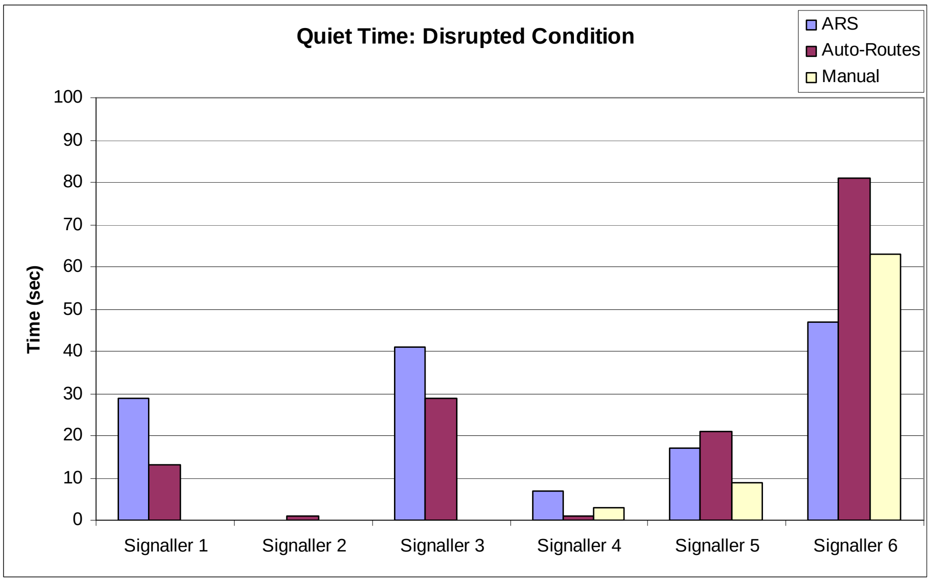


Figure 6-26: Quiet Time Results: Disrupted Condition

Considerable variation can be seen between participants and between LOA for quiet time, but a 2X3 ANOVA found no significant differences.

6.4.4 Eye Tracking

Gaze fixations of eight areas on the two signalling overview screens were coded; these are the screens used most frequently for signalling purposes. These areas were determined in consultation with a SME with expert knowledge of this area. Figure 6-27 and Figure 6-28 show the coding scheme for both overview screens. The eight main areas are outlined, as well as the sub-area icons and soft keys which were also coded. In addition, fixations on the general purpose (GP) screen, the detail screens, missing data and ‘other’ were coded. These codes proved useful for determining when the calibration was lost during the experiment as high values for any of these codes, particularly ‘other’, indicated that it was likely calibration had been lost.

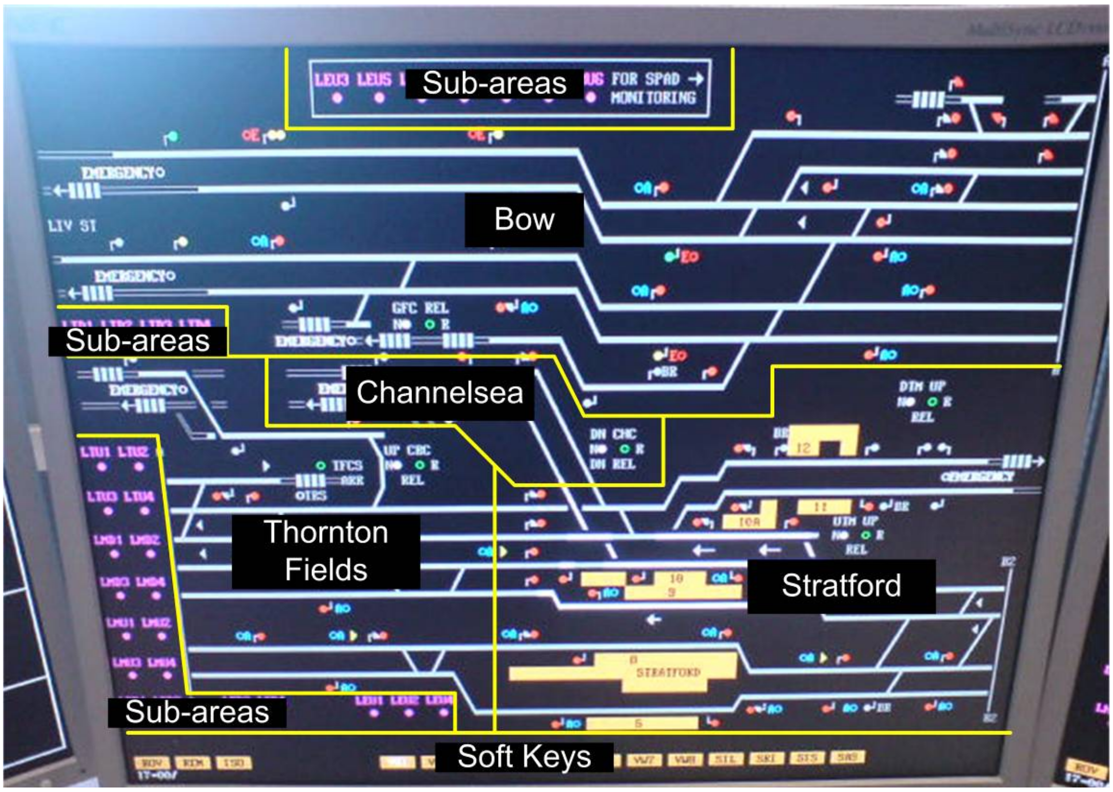


Figure 6-27: Eye Tracking Coding Diagram - Screen 1

The eight areas coded were Bow, Channelsea, Thornton Fields, Stratford, Maryland, Forestgate, Woodgrange Park, and Tottenham and Hampstead. Signalling diagrams are not always easy to follow, so for clarity, a train travelling from London would enter this workstation on one of the four lines at Bow (top left of Screen 1). It would travel from left to right across Bow which links to Thornton Fields. From Thornton Fields the train would travel through the Stratford area before moving to Screen 2 and Maryland to Forestgate and then on to Woodgrange Park. Most commonly trains would exit the workstation at the far right of Woodgrange Park but alternatively they may be routed through Tottenham and Hampstead exiting the workstation to the far right of that area. This describes the most common route for trains from London across the workstation, and the reverse describes trains travelling towards London. In addition, trains may enter the workstation from the Channelsea area, the bottom left of the Tottenham and Hampstead area, and a depot in the Thornton Fields area.

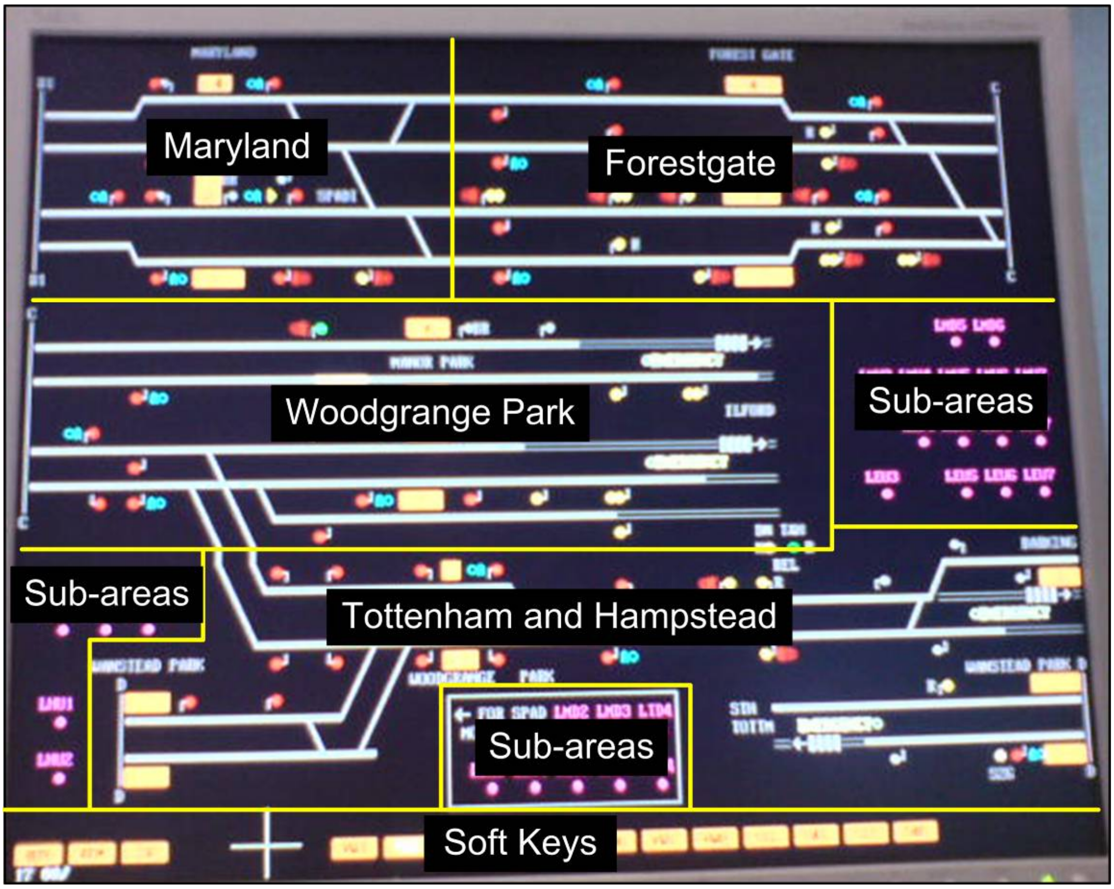


Figure 6-28: Eye Tracking Coding Diagram - Screen 2

The eye tracking videos were imported into The Observer XT for analysis. As expected, a considerable amount of eye tracking data were lost during the experiments, primarily due to the adverse lighting conditions and the difficulty of maintaining calibration throughout the experiment. In order to determine which data were useful, each minute of data was plotted for each experimental condition. Those sessions with high missing data or ‘other’ were eliminated from further analysis. In the case of high ‘other’ data, this probably indicated that the eye tracking was not calibrated, but those with high missing data may have simply had trouble picking up the participants’ eyes but still remained calibrated. However, it is impossible to be certain so the data were eliminated from the analysis as a precaution. The data from 11 of the 18 scenarios were included in the analysis (61%).

Each video was played in The Observer XT software package and the crosshair position logged throughout using the coding diagrams. Coding these videos was an extremely laborious and mundane task. The coding scheme was developed and one set of data coded from the pilot. Following this it became clear that the time and effort required to code all the data gathered during the experiment required additional resources. Hence, a research assistant was engaged to code the data gathered. This assistant also helped with the data collection and so was familiar with the experiment. Once all the data had been coded total durations in each area could be calculated for each video as well as the transfer of attention between areas (i.e. where gaze moved to from each area).

Link diagrams showing the transfer of attention from areas of the screens and heat diagrams showing the proportion of time signallers' gaze rested in each area were constructed from the data. Data showing the number of times the signaller transferred gaze from one area of the display to another, a lag sequential analysis, were obtained from The Observer XT as well as total durations in each area. Link diagrams were constructed from the lag sequential analysis to illustrate the most common visual gaze path of the signallers in the experiment (Figure 6-29 and Figure 6-30). The arrows in these diagrams indicate where participants' gaze most commonly moved from each observed area. The proportion of time the signallers' gaze dwelled in each area is also illustrated in these diagrams by the size of the yellow circle in each screen section. There were only minor differences in the data for different levels of disruption so the following diagrams are based on the total scenarios.

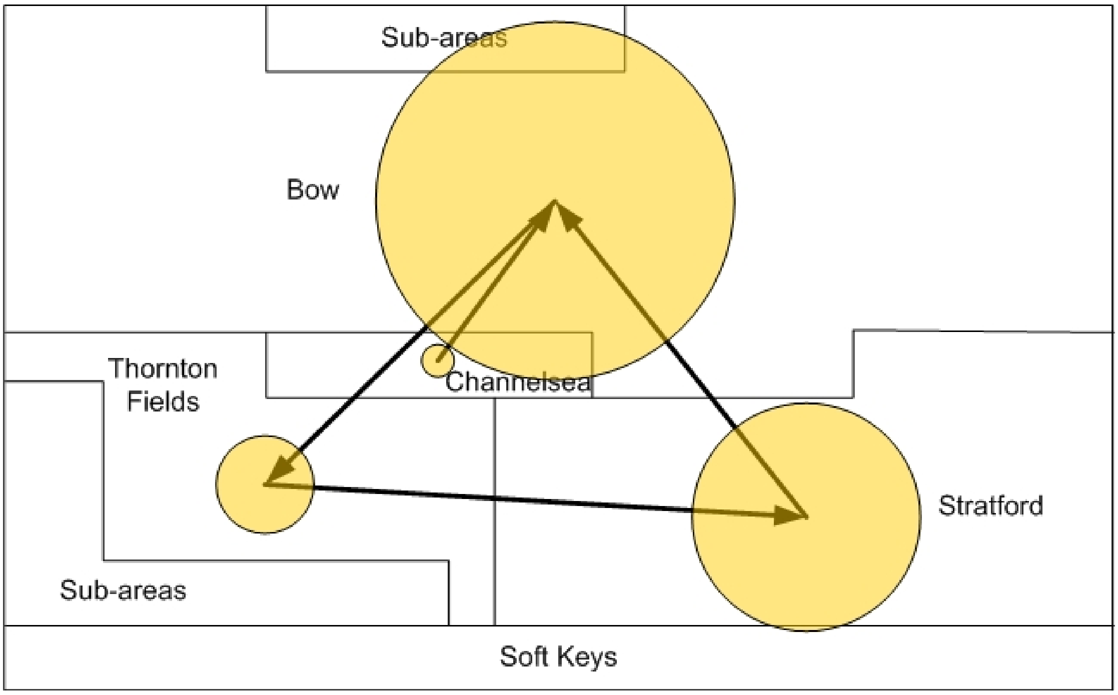


Figure 6-29: Screen 1 Link and Heat Diagram

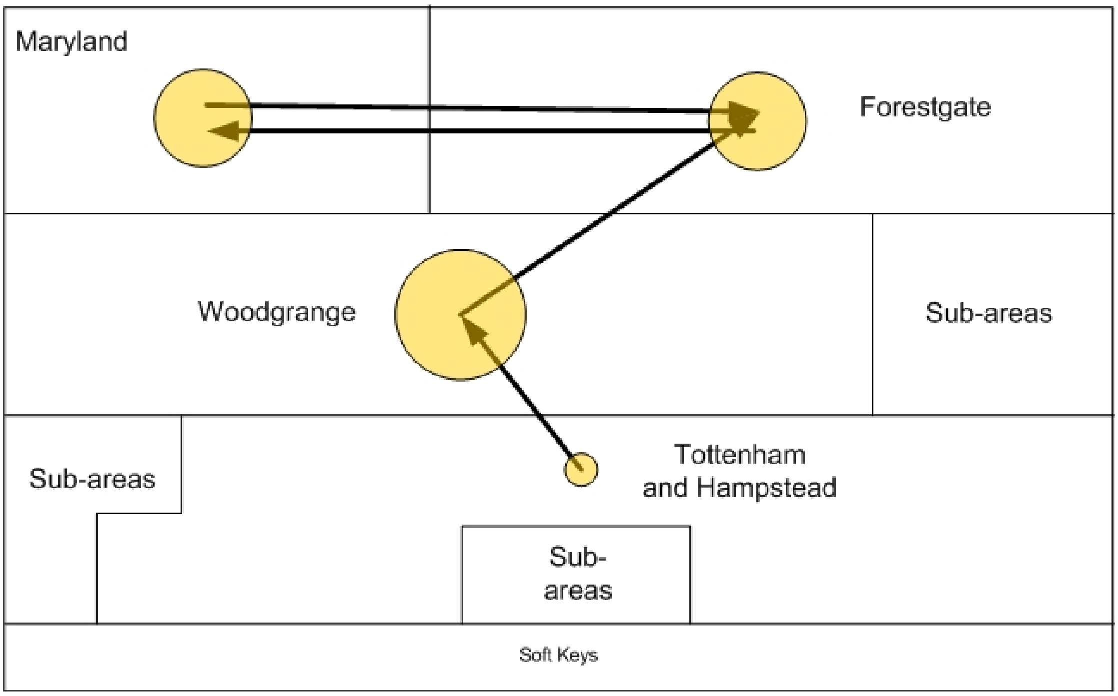


Figure 6-30: Screen 2 Link and Heat Diagram

A series of 2 x 3 ANOVAs were performed on the duration spent on each screen area to determine whether the LOA or the level of disruption condition had a significant effect on the length of time signallers dwelt on each area. Only one area (Bow) showed significant results, and this was both in the case of LOA ($F (2, 16) = 3.636$, $p<.05$) and level of disruption ($F (1, 16) = 6.546$, $p<.05$). A Tukey post-hoc comparison showed that the difference in LOA was between the Manual condition (M = 30.44, SD = 5.10) and the two automated conditions (M(ARS) = 36.78, SD(ARS) = 11.55, $p<.05$; M(Auto-routes) = 35.96, SD(Auto-routes) = 5.43, $p<.05$). A greater average time was spent monitoring Bow during the disrupted conditions (M = 37.24, SD = 8.77) than during normal conditions (M = 31.69, SD = 6.62).

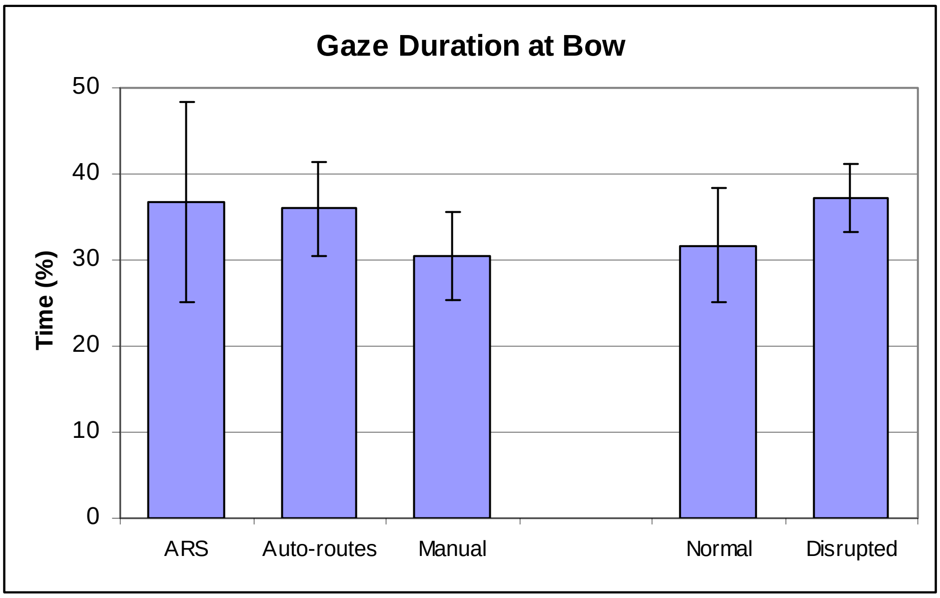


Figure 6-31: Gaze Duration at Bow

6.4.5 Situation Awareness

Neither SA measure showed any significant difference between the levels of automation; simplified SAGAT (X^2 = 0.9, df = 2, p>.05), and 3D SART (X^2 = 3.9, df = 2, p>.05). The SAGAT analysis is based on the amount of data recalled only. The actual positions of trains on the workstation at the end of each experimental scenario was not recorded, and so it is impossible to determine the accuracy of the information that was recalled. The level of detail in the results was very variable with some signallers indicating only the position of trains with an X, while others gave the first two digits of the headcode (i.e. indicating the priority and destination of the trains), and others still were about to give the full headcode. No correlation could be found between either the LOA or the order of scenarios to account for these differences.

6.5 <u>Discussion</u>

6.5.1 Workload

A steep increase was seen in the ARS condition when disruption was introduced and the signaller began to apply reminder devices. It was also necessary to apply these reminders in the other LOA (Auto-routes and Manual), but the same steep increase in workload is not seen on those graphs. The increase seen on the ARS LOA graph is likely to represent the signaller becoming more involved in the signalling and processing more information to support awareness. The ARS condition still showed the lowest workload scores overall, so on the basis of this experiment it cannot be labelled ‘clumsy automation’.

The difference between the normal and disrupted conditions indicates that the disruption introduced was sufficient to generate a significant difference in workload and validates the form of disruption used. The difference found in the LOA condition shows that ARS does significantly reduce workload. Automation may reduce workload within the four functional dimensions (Parasuraman et al., 2000) and previous research has indicated that automation is most successful at reducing workload when applied to the information acquisition and action implementation functional dimensions (Kaber et al., 2006). It is not possible to state with certainty how ARS reduces workload; however, complex decisions were not required in the normal running scenario, and ARS was inhibited or turned off in the area where disruption was introduced. This implies that it was not useful in decision making. The interview data suggested that ARS reduces workload by relieving the manual task of setting routes. Nevertheless, the overall reduction in workload is an important finding as the strengths of ARS can often be forgotten in the light of its weaknesses. These data also provide the basis for a preliminary estimation of the effect on workload of introducing ARS into a system and will be used in future to help determine how much additional infrastructure may be acceptable if ARS is provided.

6.5.2 Performance

The results indicate that performance was significantly improved with the assistance of automation and ARS showed the most consistent performance across all signallers. Unfortunately, it was only possible to measure performance for the whole experiment so the difference in performance between normal and disrupted conditions could not be investigated. It is possible that the performance decrement in the Manual condition came from the physical workload of setting routes manually rather than reduced quality of decision making. The volume of route setting required may have resulted in differences in efficiency between the manual and automated conditions. There was not a significant difference between the ARS and Auto-routes LOA, which supports this theory as the signaller was still required to make decisions on train routing in the Auto-routes LOA.

6.5.3 Behaviour Observation

6.5.3.1 Monitoring

When the two levels of monitoring (active and passive) were differentiated in the observation study it was noted that active monitoring was associated with interventions and this study provides further evidence that active monitoring has a strong association with route setting, as high levels of active monitoring were sustained by all participants in the Manual condition. Active monitoring in the ARS condition also rose to comparable levels with both other conditions after the introduction of disruption. This was as the participants became more involved with route training around the blockage and further indicates the link between active monitoring and interaction. It can be seen that there was little difference between the normal and disrupted conditions in the Auto-Routes and Manual LOA but the ARS LOA shows a large increase and a corresponding increase in workload scores.

Passive monitoring was almost exclusively confined to the automated conditions, and a sharp reduction could be seen in the ARS condition when disruption was introduced. Interestingly, the average passive monitoring observed in the Auto-routes condition actually rose following disruption. This is in contrast to workload scores, which rose in the Auto-routes condition following disruption. However, this was not a sharp rise, and taken with the result on passive monitoring, it can be said that Auto-routes is a more stable condition than ARS. The significant difference found between the Manual LOA compared to the ARS and Auto-routes LOA for passive monitoring was as would be expected as the physical necessity to set routes left little time to sit back.

The results from this study appear to support the hypothesis that signallers regulate their workload by engaging in passive monitoring. When the circumstances on the workstation became more demanding, signallers reduced their passive monitoring and engaged in more active monitoring. This provides validation for the method by which active and passive monitoring were identified and coded.

6.5.3.2 Intervention

The number of observed interventions were found to be significantly different between conditions. This is unsurprising as the requirement to set routes manually in the Manual condition would have greatly increased the number of interventions in that condition. Even during disruption, the ARS condition was significantly lower than Manual as ARS continued to set routes for trains in the unaffected parts of the workstation. This demonstrates the value of ARS in disturbed conditions.

Four significant differences were found between LOA for types of intervention. Most of these differences are as expected. There was a difference between all groups for route setting, since the requirement to set routes increased as the level of automation decreased. Auto-routes were not used in the ARS and Manual LOA so there are obviously significant differences between the Auto-route LOA and the others. There was a significantly greater use of reminders in the ARS LOA, which is interesting as they were also required in both other conditions to protect the platform area following the introduction of disruption. The significant increase in the use of reminders in the ARS condition reflects their use as a control mechanism for ARS, a fact which has previously been picked up in both the observations and interviews. However, the disruption in this experiment was not anticipated to have greatly increased the use of reminders in this context and it is interesting that the effect has appeared. This demonstrates how extremely common it is for signallers to use reminders in this way.

It is also interesting to note those interventions for which no significant differences were found. The sub-areas were only relevant to the ARS condition as they may be used to switch ARS on or off in areas of the workstation. No significant differences were found due to the extremely low use of this mechanism by signallers. The experiment therefore provides evidence for the non use of this control mechanism, perhaps because signallers are unsure of which area is controlled by which sub-area and prefer to apply reminders as protection with which they are more comfortable. Neither were there any significant differences with regard to changing of screen views. The lack of difference between the changing of these views suggests that the information requirements of the signallers remained broadly constant in this sense regardless of the LOA.

6.5.3.3 Quiet Time

Quiet time reduced in both the Manual and ARS conditions following the introduction of disruption. However, a reduction was not seen for the Auto-routes condition. This seems to indicate that signallers felt they had the same amount of free time in normal and disrupted running and provides further evidence of the robustness of this form of automation. It is also noteworthy that quiet time was present for the Manual condition. Despite the demands of route setting, participants did spend a small amount of time not involved with the system. This is in contrast to passive monitoring during the Manual condition which was rarely engaged in and may indicate that participants devote monitoring resource during automated conditions, but do not feel it is necessary when controlling manually, presumably because they are in control of any changes. On this basis, it can be assumed that information is processed during passive monitoring.

6.5.4 Eye Tracking

There was an initial aim to use the eye tracking equipment to gather data on specific elements of the infrastructure participants fixated on during the different conditions. This information was expected to be useful in determining signaller monitoring strategies and hence help guide future interface design (Ottati et al., 1999), and was anticipated to provide some validation for the monitoring strategies suggested in the interviews. Unfortunately, the difficulties encountered with the equipment, first noted in the pilot, meant that the level of accuracy was not sufficient to support this aim. Such difficulties are not uncommon and are well documented in eye tracking research (Morimoto & Mimica, 2005). The coding of the data was therefore reduced to the broad areas of the screen fixated. Although less specific, the results still contain some useful data.

The results illustrate that the majority of the signallers' time was spent on the first of the signalling screens (66% of time compared to 32% on Screen 2). This might be explained by the time of the simulation (evening peak) when the majority of trains were coming from London and had to be regulated through Stratford. This hypothesis could be confirmed by another experiment in the morning peak, in which the division of time between the two screens would be reversed. However, an alternate explanation may be that Screen 1 is the more complex screen and thus demands more of the signallers' attention. Bow emerged as an important area for monitoring purposes. The heat diagrams indicate that this was the main area dwelt upon by signallers. The statistics showed that signallers monitored this area more during the automated conditions, during which time they were comparatively more free to distribute their time as they wished, and also during the disrupted running. Bow appears to be an important area for regulation purposes. It is also interesting to note that the results indicate that signallers do change their monitoring strategy according to the circumstances on the workstation.

The link diagrams, illustrated by the arrows in the diagrams, illustrate the most frequent visual path of the signallers across the workstation. It can be seen that a logical path through adjacent sections of track was followed, except between screens. There was some movement between screens but it was more common for participants to scan within a screen. There was a suggestion that signallers would jump between regulating areas on the screens rather than following a logical path, but this does not appear to be the case. Although this finding does not contribute towards understanding of the effect of automation, it potentially has implications for future workstation design. By attempting to ensure that signalling screen diagrams have a logical progression designers can aid monitoring.

6.5.5 Situation Awareness

The results of the freeze probe measure indicated that much of the data could not be recalled by the signallers, and may indicate that these data are not routinely held in memory by the signallers. It is interesting that all signallers do not appear to retain data on train positions and names, even during manual routing. It was expected that signallers would hold this information in memory as the train headcodes contain information on priority and pathing which is vital to correct regulation. It may be that the constant presence of this information on the signalling screens means that signallers do not encode the information in their own memory, but rather they remember where to look for it. Research in air traffic control has found a similar result in that air traffic controllers have a better knowledge of aircraft location than call sign (Durso & Dattell, 2004). The main conclusion from the piloting of this SA measure is that further research is necessary to determine what is appropriate to measure with regard to SA in a signalling context.

6.5.6 Discussion of Method

Introducing disruption which generated a large and sustained increase in workload proved difficult. This was a limitation of the experiment and reflects the elimination of other variables, primarily communications, which ordinarily would contribute towards workload. Communications were omitted from the experiment for a number of valid reasons, but it is clear that they have a major impact on workload and any future studies of this nature should give careful consideration as to whether they are necessary or not.

The experiment was very ambitious in terms of the number of dependent variables, and this meant that data collection was difficult and time consuming. The eye tracking was particularly labour intensive and this will be discussed separately below. Some of the measures were more productive than others, with Integrated Workload Scale (IWS) and the simulator performance scores being the most revealing. While giving their IWS scores signallers frequently made comments on why their score had changed. Unfortunately this had not been anticipated and no facility had been made to record these data. Future experiments using IWS to measure subjective workload should incorporate a method to gather this type of data as it would add to the analysis. Although the simulator generated performance score was useful, it was only possible to obtain a score for the whole simulation and so could not be used to examine the effect of disruption. The behaviour observation data were also useful, but the SME performance scores were not. This was due to the unstructured nature of the data collected which meant that its use during the analysis was extremely limited. It is apparent that the SME needs to be in the same environment as the signaller to collect this information. It is recommended that future experiments of this type have much more structured data collection from SMEs. The data gathered were useful during the IWS and behaviour analyses in providing some context as to the state of the workstation at different points in each scenario, but the data were not sufficient to support an analysis itself. The SA measure did not show any difference between the conditions, but this experiment was intended as a pilot for this measure and the usefulness of the data was more in refining thoughts on how to measure SA in a signalling context.

6.5.6.1 Eye Tracking

This research is believed to be the first to use head mounted eye tracking in the signalling environment in the UK although Network Rail intends to make further use of the equipment in this context to support a number of other projects and goals. There are a number of research questions which the equipment could potentially support. One of these is the monitoring strategies of signallers and this research attempted to begin to examine this question. For example, when re-signalling an area and designing workstations to control it there are frequent discussions about how best to represent the area. In the past there have been instances of designs put forward which were rejected by SMEs because they cut across a regulating location which must be monitored closely to make correct decisions. The eye tracking equipment may facilitate easier recognition of these areas and a visualisation for the engineers of which areas should remain grouped together.

A limitation of the equipment is that accuracy is lost when participants move about during the experiment due to parallax error. The eye tracking cursor may be offset from the actual gaze location if the participant moves from the position in which the equipment was calibrated. It is still possible to gain some data but not at a high degree of accuracy. It was possible to design the data collection and analysis around this limitation for this experiment, but future research may require more accurate data and these are difficult to obtain. Use of the equipment is also limited by the time taken to analyse any data collected. Data analysis took approximately 3 hours for each hour of data, and this was with only 12 codes. Finer grained analysis would take far longer, and this limits the use of the eye tracking equipment to research projects which can dedicate time to the analysis. Such time is not likely to be available for use of the equipment in supporting commercial projects such as re-signalling schemes. Lighting conditions can also make the equipment difficult to use, and these may be difficult to adjust in operational environments such as signalling centres.

Overall the eye tracking equipment is very resource intensive, both in terms of data collection and data analysis and interpretation. In order to maximise the potential of the equipment it is recommended that it is the sole focus of any such research in future.

6.5.6.2 Use of The Observer XT

The Observer XT package was used to code the behavioural observation live during the experiment and the eye tracking was subsequently coded using the same software. Although a powerful tool, it has a number of shortcomings. First, the coding scheme cannot be changed once data entry has begun. This means that any unexpected behaviours cannot be accounted for in the analysis. Secondly, the software can freeze during observations. This is usually only for a short time; however, on a couple of occasions during this study the software froze and refused to accept any codes. Hence, a couple of minutes of data were lost.

Although a useful tool for data collection, provided a complete coding scheme has been developed in advance and it does not freeze during data collection, the analysis power of The Observer XT is not greater than that of Excel, despite this being a selling point. Indeed for some of the data analysis (the lag sequential analysis) the raw data were exported from The Observer XT and analysed in Excel as The Observer XT could not generate a sufficiently large matrix to display the results.

6.6 Conclusions

The key findings of the experiment are summarised in Table 6-4.

	ARS	*Auto-routes*	*Manual*
Workload	Lowest workload; Highest proportional increase following disruption	Most stable (i.e. smallest change during disruption)	Highest workload
Observed Monitoring	Largest change following disruption; Highest passive monitoring; Variable	Variable	Least variable; Least passive monitoring
Observed Trackerball Use	Lowest use; Increased use following disruption; Use of reminders	Stable following disruption	Highest use; Stable following disruption
Observed Quiet Time	No differences found		
Gaze Fixations	Increased monitoring of Bow area	Increased monitoring of Bow area	Reduced monitoring of Bow area
Performance	Highest and most consistent performance		Lowest performance
Situation Awareness	No differences found		

Table 6-4 Key Findings of the Level of Automation Experiment

This experiment has shown that ARS does lead to a reduction in workload compared to lower levels of automation, but the reduction is not as large during disrupted running. In addition, performance was highest and most consistent when working with ARS. Auto-routes showed the most stable workload scores throughout, but performance was variable between different signallers. There were few differences in signallers' behaviours between the conditions apart from the obvious such as increased use of the trackerball during the Manual LOA and increased passive monitoring when using ARS. However, the amount of time dedicated to monitoring varied between signallers more for the automated conditions compared to the manual. This appears to indicate that individual signallers engage in different strategies during the automated conditions. The eye tracking equipment showed that signallers also changed their monitoring strategy under disrupted running for both the ARS and Auto-routes LOA compared to the Manual LOA choosing to spend more time looking at Bow, where regulating decisions would have to be made.

Overall the findings of the experiment support the use of signalling automation. However, the advantages of ARS over Auto-route functionality are not as great as might be assumed given the differences in complexity and cost.

6.7 Chapter Summary

This chapter presented the final study undertaken for this research, a level of automation experiment using a high fidelity signalling simulator. Three levels of automation and two levels of disruption were examined and themes including workload, monitoring, and SA were further examined. Key findings include the reduction of workload through the use of automation and consistency of performance with automation. The next chapter will summarise the findings from this and the other studies undertaken to evaluate ARS and provide guidance for future automated signalling systems under the principles of automation developed in Chapter 2.

CHAPTER 7: GENERAL DISCUSSION

7.1 Chapter Overview

This chapter discusses how the research undertaken supported the objectives of this thesis. The findings from the studies are drawn together and related to previous research on automation. Specific recommendations for future automation systems are given under the principles of automation drawn from the literature.

7.2 Introduction

The aim of this research was to study a real world automated system in order to understand the impact of automation on the human operators who work alongside it and identify how automation can be implemented to best support overall system performance. Figure 7-1 illustrates how the research conducted led to an understanding of the effect of automation on the signalling task, providing the basis for the recommendations given in this chapter.

The domain was rail signalling and the specific system under investigation was Automatic Route Setting (ARS) which automatically sets routes for trains using timetable information and uses algorithms to resolve conflicts arising between trains. A number of research methods were used to this end, including observation of signallers working with ARS, video archive analysis, semi-structured interviews, and an exploratory level of automation (LOA) experiment. This chapter will discuss the findings of the research, and show how it supported the objectives outlined in Chapter 1.

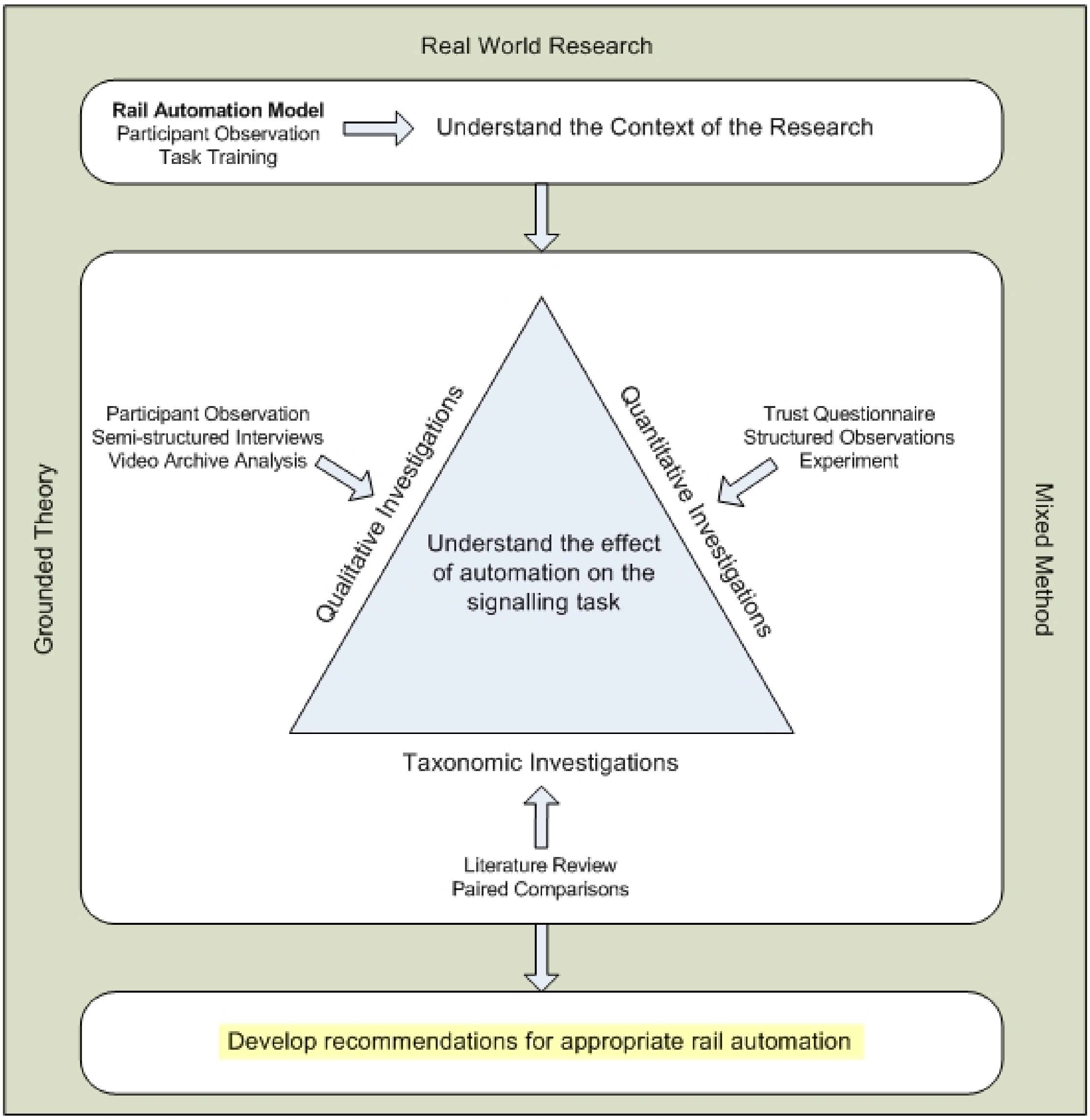


Figure 7-1: Research Framework

7.3 Discussion of Research Approach

Objective 1

To develop a theoretical framework within which to research and implement human oriented automation within rail signalling.

The conceptual framework (Figure 7-2) illustrates the approach taken to the research.

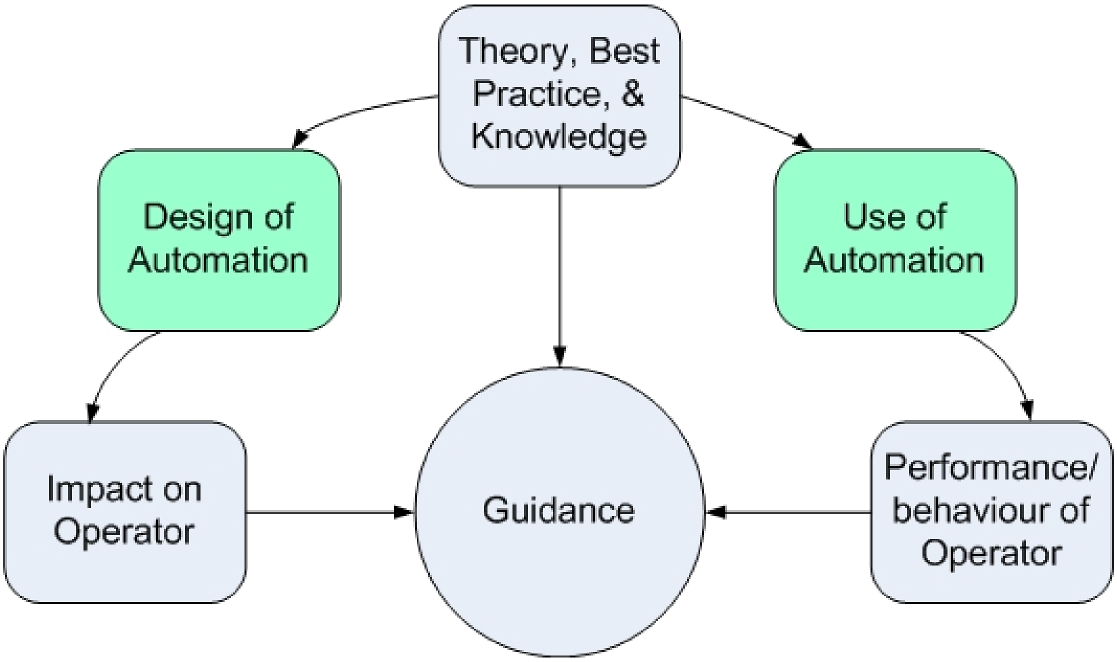


Figure 7-2: Conceptual Framework

A review of the literature was undertaken to identify relevant theories, best practice, and knowledge. The literature review focussed on the impact of automation design on the operator, in terms of workload and trust and on the use of automation, with the key themes of monitoring and SA. It introduced the automation of control systems and highlighted major research findings on the key themes. These are summarised in Table 7-1.

Theme	*Key Finding*
Trust	• There is a correlation between trust in and usage of automation. • High reliability and competence are fundamental requirements for trust in automation. • Operator self confidence and the usefulness of the automation also influence usage. • For complex systems, explicit feedback is required to develop trust. • Trust must be well calibrated to ensure optimal use of automation. • Accurate mental models are important to ensure correct calibration of trust. • Individual differences influence trust.
Situation Awareness	• Level 1 SA is higher during automated operation of information acquisition, suggesting that the use of information is more important for SA than gathering the information. • Level 2 SA is higher during intermediate levels of automation. • Level 2 SA may be improved by automation during high workload conditions. • Performance during automation failures is better with higher SA. • Well designed automation has the potential to improve operator SA. • SA is also affected by high workload conditions.
Workload	• Automation can reduce workload during normal operations. • Automation of information acquisition and action implementation has a greater effect on workload. • Monitoring of automation may increase workload. • Automation may increase workload during incidents.
Monitoring	• The subjective workload associated with monitoring may be high. • 'Passive monitoring' may reduce awareness as compared with 'active control'. • Although a reliable laboratory result, no evidence of a vigilance decrement has been found in real world systems. • Alleged complacency may be due to the calibration levels of trust. • Operators develop strategies to monitor automation effectively.

Table 7-1: Summary of Key Findings from the Literature Review

The conceptual framework focussed this research on the use of automation and design of automation and both empirical study and archive study were used to gather data on the performance and behaviour of operators using automation and the impact of the design on the operator. Production of guidance on the appropriate implementation, level, and design of new automation systems was the ultimate goal of the research and this guidance was generated on the basis of the 12 principles of automation developed from the literature and the research carried out.

7.4 Discussion of Research Findings

Objective 2

To study current use of automation within rail signalling and understand the effect automation has on the signalling task, including:

- *How signallers monitor the system;*
- *How signallers interact with the system;*
- *Signallers' understanding of how the system works;*
- *Overall system performance.*

The research undertaken for this thesis built on the themes identified in the literature review (i.e. trust, workload, SA, and monitoring) but other themes also emerged strongly, such as directability, performance and wider organisational issues. These themes will be discussed first in the following sections. The findings within these themes will be summarised with regard to the above objective.

7.4.1 Trust

Trust was investigated both during the observation study and the interviews with signallers. Key dimensions identified from the literature including reliability, competence, understanding, predictability, and robustness (Madsen & Gregor, 2000; Muir & Moray, 1996; Rempel et al., 1985) were investigated over the course of both studies. It is well established that trust in automation is influenced by low or variable reliability (Dzindolet et al., 2003; Lee & Moray, 1994). However, ARS is a highly reliable safety critical system and this fact allowed other factors influencing trust to emerge.

The competence of the system was found to be low, particularly during disruption (i.e. it is not robust). Signallers could and did give many examples of incorrect regulating decisions made by ARS, particularly with late running trains or complex junctions. The system cannot cope with infrastructure restrictions at all and had to be constrained when disruption was introduced during the second half of the experiment and all routing in the area of the closed platform achieved manually. However, the ratings of perceived competence on the trust questionnaire did not differentiate between high and low interveners, despite evidence that competence is the greatest predictor of the operators' overall trust (Muir & Moray, 1989). Other factors, including feedback, reliability, understanding, and predictability, yielded significant differences between high and low interveners, providing further evidence for the correlation between trust in and usage of the automation (de-Vries et al., 2003). In the literature, as discussed in Chapter 2, no research was found which had established empirical links between reported understanding, prediction, and faith and automation use. Therefore, this may be the first time such an empirical link has been found. These links were also found in a real world setting, rather than laboratory, increasing the validity of the research.

Feedback from ARS was also found to be very poor, resulting in low understanding and low predictability of the automation. As signallers cannot predict what the automation will do in all situations they do not feel they can trust it to set routes and frequently step in to ensure trains are routed in the correct order. In the observation study, the differences found between high and low interveners in terms of feedback, understanding and predictability confirm the importance of good mental models in the development and calibration of trust (Sheridan & Parasuraman, 2006). The lack of such models was emphasised during the interviews when all signallers could give recent examples of having been surprised by actions taken by ARS, even those who had been working with the system for some time.

The development of accurate mental models is supported by observable and understandable automation (Hopkin & Wise, 1996; Lenior et al., 2006; Parasuraman & Riley, 1997) and this must be achieved through good feedback in complex automation systems. Even when queried, ARS does not provide reliable or easily interpretable information. Over the course of the research it became clear that the designers of ARS envisioned a near autonomous system with full decision capability (i.e. Level 5 on the Rail Automation Model) but the inability of the system to perform competently outside of routine running means that the decision making functional dimension is considerably lower than was aimed for. The outcome is a system which the operator struggles to understand and predict and results in higher levels of intervention than may be necessary in order to control ARS adequately.

7.4.2 Situation Awareness

Situation awareness has been found to vary with the level of automation in the aviation domain, although the direction and degree of this variation depends on the type of automation and the level of workload operators are working under (Endsley &

Kaber, 1999; Endsley & Kiris, 1995). Situation awareness has not previously been examined in detail in a UK signalling context and so the factors which contribute to good SA have not yet been identified. This meant that it was not possible to measure SA as a part of this research as significant work is first required to identify a valid method of measurement. However, a pilot measure for SA was incorporated into the LOA experiment. The freeze probe measure required participants to recall train position and headcode but signaller recall was poor and no difference was found between conditions. It seems likely that because this information is constantly displayed to signallers they do not maintain it in their memory, but rather develop future oriented SA involving the routes that they need to set and the potential conflicts that may arise. Until a validated measure of SA is developed it will not be possible to measure how it is affected by automation.

7.4.3 Workload

The literature review suggested that workload tends to decrease when automation is employed (Endsley & Kaber, 1999) but that a significant mental workload may be involved in monitoring (Warm et al., 1996). Observations of signallers and discussions with subject matter experts suggest that signalling workload can be divided into at least four areas: physical workload associated with setting routes, mental workload associated with memory burden of routing and platforming requirements for individual trains, physical communications workload, and mental workload associated with regulation.

The introduction of ARS was intended to reduce signaller workload and allow them to control a larger area (Burrage et al., 1991) and it is likely that ARS does reduce the first two of the workload types outlined above. This is because it sets the majority of the train routes in areas where it is operational and it holds information on the routing and platforming requirements of individual trains. However, ARS has little or no effect on communications, and it seems likely that it increases the mental workload associated with regulation. The interviews indicated that ARS is not totally competent at regulation, particularly under disrupted conditions, and so signallers working with ARS consider regulation to be a primary duty. However, ARS introduces an additional cognitive burden as signallers consider how it will react in a situation and act appropriately to counter this if necessary. Hence there is a potential for a mental workload increase. Such a potential has been previously noted by Megaw (2005) but no specific reference was found in the literature on control system automation.

The findings of the LOA experiment clarified the overall magnitude of workload differences between manual and ARS conditions. A significant reduction was seen for the ARS condition as compared to the manual, reflecting the removal of the requirement for physical route setting. There was not a significant difference between ARS and Auto-routes, suggesting that similar workload reductions could be achieved with a much simpler form of automation. Unfortunately, it was not possible to directly measure the reduction in each of the four types or workload outlined above, or in the functional dimensions in which reductions were achieved (Kaber et al., 2006). However, the interview data collected suggest that the reduction in workload while working ARS is primarily due to the removal of the requirement to physically set routes (i.e. action implementation).

Automation may increase workload during incidents (Kantowitz, 1994). The LOA experiment found that workload did increase in the ARS condition following the introduction of disruption but was still lower than the Manual and Auto-routes conditions. This contrasts with the findings from the interviews which suggested that ARS does hinder the operator during disruption. During the LOA experiment, the disruption introduced half way through each scenario was chosen following numerous attempts to try to ensure an increase in workload would be experienced. Despite the effort, the workload increases were not as large as desired. Indeed, some participants did not show any increase in workload scores at all following the disruption. Any incident causing delay or re-routing of trains would normally prompt numerous communications with station staff, train staff, control staff, etc. all seeking information on the cause and effects of the delay. This multi-tasking appears to be the cause of much workload associated with disruption and the interaction between it and the control of the automation may be the cause of the workload increase reported by interviewed signallers. There may be a difference in this respect between different implementations of ARS; some signallers reported that ARS helped during incidents by keeping other parts of the workstation moving, and this was the case in the experiment, but other locations report that ARS must be switched off during disruption. It is therefore not possible to say with certainty that ARS is a ‘clumsy automation’ system (Wiener, 1989; Woods, 1996); however, when not well implemented it certainly has the potential for clumsiness.

7.4.4 Monitoring

The observation study initially noted that observed monitoring behaviour could be divided into two distinct categories, active and passive. Signallers appeared more relaxed when engaged in passive monitoring and this was typically for longer periods of time and interspersed with periods of quiet time (i.e. distractions). In contrast, signallers appeared more engaged with the signalling task while actively monitoring; the average time spent actively monitoring was considerably shorter, and tended to be interspersed with interventions. It seems likely that this variation of monitoring behaviour is a strategy to cope with variation in the demands on a workstation. When required, the signaller becomes more involved and actively seeks out information whereas when the demands lessen they relax and simply maintain an overview, waiting for a situation to arise which requires their attention. The LOA experiment provided some validation of this theory; as workload increased in the ARS LOA following the introduction of disruption, active monitoring increased and passive monitoring decreased.

No such descriptions of monitoring behaviour were found in the literature, but a difference has been investigated between ‘active control’ and ‘passive monitoring’ (Endsley & Kiris, 1995; Metzger & Parasuraman, 2001). These studies suggest that individuals engaged in monitoring passively process information and this may be poorer than active processing during control activities. In contrast, this research suggests that monitoring of automation is not always a passive activity, and interventions were frequently observed immediately following periods of passive monitoring. It seems likely that passive monitoring is less demanding than active monitoring, or there would be no benefit for signallers to engage in it, but the interventions suggest that they are still focussed on the task.

A variety of monitoring strategies were reported during the interviews with signallers. These are developed through experience and include maintaining an overview of the workstation, monitoring the progress of individual trains which ARS is known to have difficulties with, and monitoring route setting to ensure that trains have two green signals in front of them wherever possible. These strategies allow signallers to recognise situations requiring their attention quickly and to direct their attention where it is required.

The eye tracking equipment gave some insight into the manner in which gaze is transferred across the workstation. The link diagrams presented in Chapter 8 illustrate that the signallers in the experiment predominantly followed a logical path across the workstation, rather than jumping between hot-spots. The eye tracking also showed an increase in monitoring of one area of the workstation (Bow) following disruption for both the ARS and Auto-routes LOA. This was a key area for regulation on the approach to the closed section of track and the presence of automation allowed the signallers to dedicate more attention to this area. These findings are relevant to interface design, illustrating the importance of presenting track diagrams in a logical format and in ensuring that key regulating locations are presented in a manner which facilitates monitoring.

The literature review revealed some work which had identified monitoring strategies similar to those outlined above (Vicente et al., 2004), but most of the research in the area of monitoring was regarding vigilance and complacency (Parasuraman, 1987; Parasuraman et al., 1993). Although not investigated as main themes, neither of these arose as an issue in signalling. During the interviews, signallers highlighted how important they felt it was to monitor the automation, and the observation study objectively recorded the high frequency of signaller monitoring behaviour. It is likely that this is because of the dynamic nature of the signalling task and the involvement of other parties. The demands on the signaller are constantly changing, keeping them involved and minimising the chance of missing anything. However, if he/she does miss something, it is likely that a train driver will ring to prompt him/her. The lack of a vigilance decrement or complacency issue, even in those IECCs where ARS runs well, supports the theory that these issues do not arise in real world dynamic systems (Moray, 2003; Moray & Haudegond, 1998).

7.4.5 Performance

There are a number of issues with ARS which can affect performance including poorly input timetable information, incorrect programming of the weighting factors and algorithms determining priority in the event of a conflict between trains, and disruption in the area of control or surroundings. These all affect the system competence, and addressing such issues should increase the overall trust and use of the automation (Muir & Moray, 1989).

However, the LOA experiment found that system performance was higher and more consistent with ARS as compared to manual operation. Because ARS operates in a number of different sub-areas, it can be switched off in one area of a workstation in the event of disruption in that area. Signallers can then focus their attention on that one area while leaving ARS to continue to run the rest of the workstation. This was the case in the experiment. However, this is only possible if ARS is well programmed and can be trusted to run the rest of the area. During the interviews, some signallers regarded this ability as a key strength of ARS while others stated they were not happy to work in this manner as ARS unsupervised could go on to cause enormous problems in other areas of the workstation. Good planning and programming allow signallers to trust ARS to manage parts of the workstation and therefore play a key role in the performance of the system overall.

7.4.6 Directability

The ability to interact with and direct the automation arose as a key issue during the research. There are a number of options for a signaller to intervene in the routing of a train. They can turn the ARS sub-area off, they can take a single train out of ARS control, or they can apply a reminder device. The latter two options are preferred with reminders by far the most common method employed. Turning off ARS sub-areas is generally only used when there is significant disruption as this means all trains in that part of the workstation must be routed manually. Reminder appliances are quick and easy to apply and they prevent ARS setting a route to or from the signal they are placed over. This is an easy way to constrain ARS. The reminder can remain in place until the signaller is happy for the route to be set and then removed; ARS will then set the route if it is available. This method is frequently used to resolve conflicts between trains at junctions. The signaller places a reminder in front of one train to ensure that the other is routed first. It is a quick and effective method of controlling ARS. However, this is not the purpose of reminders. They are primarily intended as a safety device to protect staff working on or near the track. Their purpose is both to prevent a route being set in that area and to remind the signaller of the presence of staff in the area. Use of reminders to control ARS degrades their effectiveness as a safety device. The use of reminders in this context indicates the lack of powerful methods to direct the automation.

The LOA experiment found that signallers had to manually take over routing in the area of disruption. Their attention was then focussed on this part of the workstation, and ARS was allowed to run the remainder. More powerful tools could be employed which would allow the signaller to instruct the automation on trains that should be routed around the disruption and free up signaller resource to concentrate on the overall performance of the train service in their area. This is the ‘control by replanning’ suggested by Kauppi et al. (2006).

7.4.7 Organisational Issues

Alongside the more technical problems with the ARS system there are some softer issues, most particularly the unacknowledged change in the role of the signaller. Although ARS adds a new element to the signalling task, and changes it in some fundamental ways, the organisational approach is much the same. This is particularly reflected in the training of IECC (ARS) signallers which is no different to the training of NX panel signallers. Any training on ARS is informal and ad hoc. Information on ARS and how it works is passed on from generation to generation in an informal manner. The result is signallers who do not know how to work with the system, although this should also be considerably more intuitive, and cannot achieve optimal levels of operation.

The train planning team are responsible for devising and inputting timetables. As with any automated system, to work effectively ARS must have access to a complete and accurate timetable (Sheridan, 1996). This is in contrast with non automated signalling where the signaller can compensate for any missing or inaccurate timings or route codes. Good timetable information is therefore a prerequisite for ARS performance, but the planning staff are remotely located and have little knowledge of ARS. The result is that they are relatively unaware of the importance of their work to ARS and it falls to signallers to compensate.

The introduction of ARS also clouds the responsibility of the signaller for delays. The organisation tends to take the view that the signallers are in charge and should take responsibility for any delays occurring in their area. However, if ARS sets an unexpected route which the signaller could not have anticipated this attitude does not seem fair. On the other hand, without the burden of responsibility there is little incentive for the signaller to remain involved with a system which can run itself, albeit not optimally. There is no standardised view on this issue, but the most common and sensible is that the signaller takes responsibility for any delays, unless ARS implements a decision which has never been encountered before. Once such a problem has occurred, the details are passed to the signallers and it is their responsibility to ensure that ARS does not make the same mistake again.

It is clear that ARS has suffered from a lack of integration into the railway system as a whole. For future automation systems to succeed there must be appropriate support and buy-in from all parties who may affect it, or be affected by it.

7.4.8 Summary

This research set out to study the current implementation of rail signalling automation and understand its effect on the signalling task. Signaller monitoring was investigated throughout the research and different levels of monitoring were identified. The research also identified different types of monitoring behaviour, but no evidence of a vigilance decrement was found. Signaller interaction with the system was found to vary between individuals, and this appears to be related to their trust in the automation, specifically with the reported level of feedback, understanding, predictability and faith. Signaller workload was found to be lower when working with ARS, and performance was increased, but a number of issues have also been identified. In particular the complexity of, and lack of feedback from, ARS makes the system difficult to work with and has knock-on consequences on overall system performance. The key findings of the research are summarised in Table 7-2. Recommendations were developed from these findings and will be stated in the following section.

Theme	***Key Findings***
Trust	• ARS was rated highly for reliability. • Predictability and competence were rated much lower. • Feedback was poor, and signallers did not have complete mental models of ARS. • Trust is related to the level of intervention a signaller displays. • Understanding and prediction were found to influence the level of intervention.
Situation Awareness	• Factors suitable for measuring SA in signalling have yet to be identified.
Workload	• Signallers reported that ARS reduces workload, but less so during disrupted conditions. • The reduction in workload appears to be in the physical rather than mental domain. • ARS may increase mental workload. • The experiment found that ARS decreased workload, but the reduction was less during disrupted conditions.
Monitoring	• At least two levels of monitoring are regularly employed by signallers, active and passive, and it is likely that these are related to perceived workload. • Signallers reported a variety of monitoring strategies; monitoring the progress of individual trains, monitoring route setting, monitoring 'hot-spots', maintaining an overview, and monitoring CCF. • Signallers follow a logical path across the workstation when monitoring. • No evidence was found of a vigilance decrement or complacency.
Competence	• ARS cannot be considered to regulate trains.
Performance	• Poorly input information (i.e. planning) affects ARS performance. • ARS programming also has a strong effect on performance and how often the signaller must intervene. • ARS does not perform well under disrupted conditions (i.e. it is not a robust system). • ARS facilitates highest performance and most consistent performance across different signallers.
Directability	• Use of reminder appliances was the preferred method of intervention for most signallers. • There is a lack of powerful tools to direct the automation.
Organisational Issues	• No standardised training for ARS. • Information passed on to new signallers is haphazard. • Responsibility for wrong routings or incorrect regulation is a grey area when ARS is employed. • Planning do not fully understand their importance.

Table 7-2: Summary of Key Findings

These findings inform the recommendations and these will be presented under the principles of automation in the next section.

7.5 Recommendations for Future Automated Signalling Systems

Objective 3

To develop recommendations for development and implementation of automation in future rail signalling systems.

As Network Rail continues to invest in the British rail network further automation is likely. This section uses the principles of automation extracted from the literature and presented in Table 2-6 to structure recommendations for the design of future automated rail signalling systems.

7.5.1 Reliable

The mechanical reliability of ARS is very high, as would be expected of an automated system employed in a safety critical environment. During the design stages, ARS was required to meet high standards regarding its safety integrity level and this ensured a highly reliable product.

The reliability of signalling automation systems should continue to be extremely high. This means that signallers can rely on the system to continue to operate and in turn engenders trust.

7.5.2 Competent

The competence of the ARS system is considerably lower than its reliability. Signallers who work with ARS have many examples of incorrect decisions made by ARS. The low competence of the system is likely to be due its attempts to make regulation decisions using complex algorithms but without the accurate real time information on the railway which is more readily available to signallers. The hard red lines within ARS also leave no room for the context sensitive conclusions at which signallers may arrive.

Given the low competence of ARS in some complex areas, careful consideration should be given as to whether full ARS is necessary or whether a simpler and more predictable form of automation, such as running trains strictly to timetable order or on a first come first serve basis, is more appropriate. The rail automation model highlighted the high levels of decision making automation, which do not appear to be supported by high levels of automated analysis. Lowering the level of decision making automation may improve operator trust in the automation and help them to calibrate their trust correctly.

The rail automation model also highlighted the lack of support for communications in automated signalling systems. This is one area where increased automated support could provide major benefit. Future signalling systems should aim to reduce the communications burden by automatically transmitting pertinent information to relevant parties.

7.5.3 Visible

The visibility of ARS is relatively good quality, with the same information displayed as on non automated signalling system (i.e. NX panels) available to the signallers. The introduction of ARS did not therefore result in information hidden from the signaller. However, the weighting factors within ARS on which regulating decisions are based are not visible and this has an impact on the observability of the system. In a broader signalling context, the information relevant to signalling decision making is spread across different systems and this lack of integration reduces the overall visibility.

Future automation should integrate information which is currently spread across different systems. In particular, delay information should be incorporated on the main signalling screens. This could be achieved either through annotation of existing train headcodes, or via graphical displays of train movements. However, care must be taken not to clutter the display with too much information.

7.5.4 Observable

Observability of ARS was found to be very low. The signaller receives no information on its intentions and is only aware of decisions after they are made. Although there is an ability to query ARS as to its intentions, this requires the signaller to request the information for each train in question and there is no guarantee that the response will correspond to the actions finally taken. This is because ARS is constantly re-evaluating and may come to a different conclusion in a subsequent calculation. This ability of ARS to 'change its mind' reduces the observability still further. The low observability has come across strongly as a weakness with which signallers struggle and which must be addressed in future systems.

Observability can be improved through better feedback from the automation. This should be both in terms of giving the signaller insight into how the automation works and in terms of providing the signaller with explicit, relevant, concrete, and easily understood information on its analysis and future intentions. The former will improve the signallers' understanding of how the system works and enable them to work more effectively with it. The latter allows the signaller to be confident in the automations actions and will remove the current tendency to intervene when it may not be required.

7.5.5 Understandable

Understanding of ARS was found to be quite poor, probably as a result of the low observability. Signallers reported surprise at the decisions of the automation and found it difficult to understand the decisions even in retrospect. The low observability and understandability of the system make it more difficult for the signaller to work cooperatively with the system and result in interventions to control the automation which may not be necessary.

Improved feedback will facilitate the development of more accurate mental models allowing signallers to understand and predict the automation. More basically, simpler automation, such as has been recommended to improve competence, would also be more understandable.

7.5.6 Directable

There are currently a number of methods signallers may employ to interact with ARS, including taking trains out of ARS control or turning ARS off in parts of the workstation. Although these provide the signaller with the ability to intervene and take control of the railway, they do not allow him/her to direct ARS. Reminders are commonly used for this purpose as applying a reminder to a signal prevents ARS setting a route from that signal. As discussed previously, reminders are not intended for this purpose being primarily a safety device to ‘remind’ signallers that there is a safety reason why they may not operate a signal or set of points, for example because there are staff working on the track beyond it.

In the short term, an alternative to reminder devices, perhaps called inhibitors, should be provided. These should offer the same functionality as reminders but a different form and would be used for regulation purposes, leaving reminder devices solely for safety purposes. Longer term, other methods of directing ARS should be introduced, such as allowing the signaller to give individual priorities to trains or controlling how ARS makes regulating decisions (i.e. changing the mode in which ARS operates between conflict resolution), running trains to timetable order, and running trains on a first come first serve basis. It should also be possible to ‘lock’ ARS into a particular decision to ensure that a train is routed as the signaller wishes. Improving the directability of the automation has the potential to improve overall system performance.

7.5.7 Robust

The operating envelope within which ARS is capable is relatively small; it is not a robust system. It is only fully competent during normal running or minor delays. In addition, although it does reduce signaller workload during normal circumstances, workload is not reduced by the same magnitude during disrupted circumstances. There is also the potential for ARS to hinder operations during disruption as it continues to route trains as normal. The signaller is then required to inhibit ARS as well as dealing with the cause of the disruption, and this may be difficult when large amounts of communications are necessary.

To address the issue of robustness, better methods of working during disruption should be developed. This is quite a challenge, but one idea is to allow the signaller to dictate a train path across the workstations which can be temporarily applied to relevant trains, for example to route them around a blockage. Such a system would be similar to the existing auto-routes functionality and would mean that once the relevant routes are programmed the signaller would be free to concentrate on the fault and its ramifications.

7.5.8 Accountable

The system was not designed to be accountable, as evidenced by the low observability of the system. However, the work systems in place in Network Rail make the signaller responsible for errors made in routing trains, even if the error is made by ARS. Although different signal boxes have different practices, the most common places the responsibility on the signaller to ensure ARS routes trains correctly unless a situation occurs which has not previously been encountered. This distinction is driven by the poor observability and understandability of ARS which makes it impossible for the signaller to anticipate its every action. It is likely that this problem stems from the designers' original vision of an autonomous system which required little or no operator input.

It is recommended that designers and managers regard automated systems as a tool to assist the signaller. Accepting this fundamental assumption will facilitate development of a system which supports the operator rather than one which attempts to act autonomously.

7.5.9 Error Resistant

The ARS itself is not error resistant but the wider signalling system is, due to the interlocking. The interlocking provides error protection for both the signaller and the ARS system. To some extent, this removes the requirement for ARS to provide protection against signaller errors. However, the interlocking only protects against unsafe actions. Errors resulting in delay or wrong routings may still occur. Some such errors are due to the incorrect input of data from the train planners.

Interlocking systems should continue to support the safe operation of signalling systems. Where possible, methods should be employed to detect potential errors in data input, for example where data codes are missing which prevents ARS from routing a train, and highlighted to the planning staff or operator.

7.5.10 Error Tolerant

The interlocking provides such a high degree of protection against safety related errors that the need for error tolerance is reduced. Although trains may be sent along the wrong route, the interlocking still ensures that there are no safety consequences to incorrect routings. However, there is little tolerance to the resulting delay. The railway is a very constrained system and a wrong routing will inevitably result in delay. It could be argued that ARS does warn the signaller of a manual route set contrary to the timetable as the train headcode will turn from blue (i.e. in ARS) to pink (i.e. out of ARS). If the signaller has indeed set the route in error, this may allow the signaller time to take action to minimise the resulting delay.

The use of interlocking systems should continue to minimise the requirement for error tolerance in automated signalling systems. Future systems should also help the operator realise the consequences of actions, perhaps by provision of an analysis of delays resulting from different options.

7.5.11 Proactive Control

Currently, ARS does not make it easy for the signaller to proactively control the signalling system although it could be argued that effective management of a signalling system including ARS requires the signaller to proactively control. The information required for the signallers to control ahead easily is not provided by ARS and signallers may be forced into behaving reactively when ARS makes a surprising decision. However, signallers can work proactively in the short term by constraining ARS and forcing it to work as they wish, by using reminders, turning off sub areas, and putting trains in and out of ARS as necessary.

For new automated systems, Kauppi et al. (2006) suggested that the most effective method of controlling signalling automation is to change the plan from which it works, and this has been successfully implemented in the Netherlands (Lenior et al., 2006). One method of achieving this is a train graph (Figure 7-3). Train graphs plot train progress on a graph of location against time. This display method presents information to the operator clearly and facilitates the development of new plans (Kauppi et al., 2005).

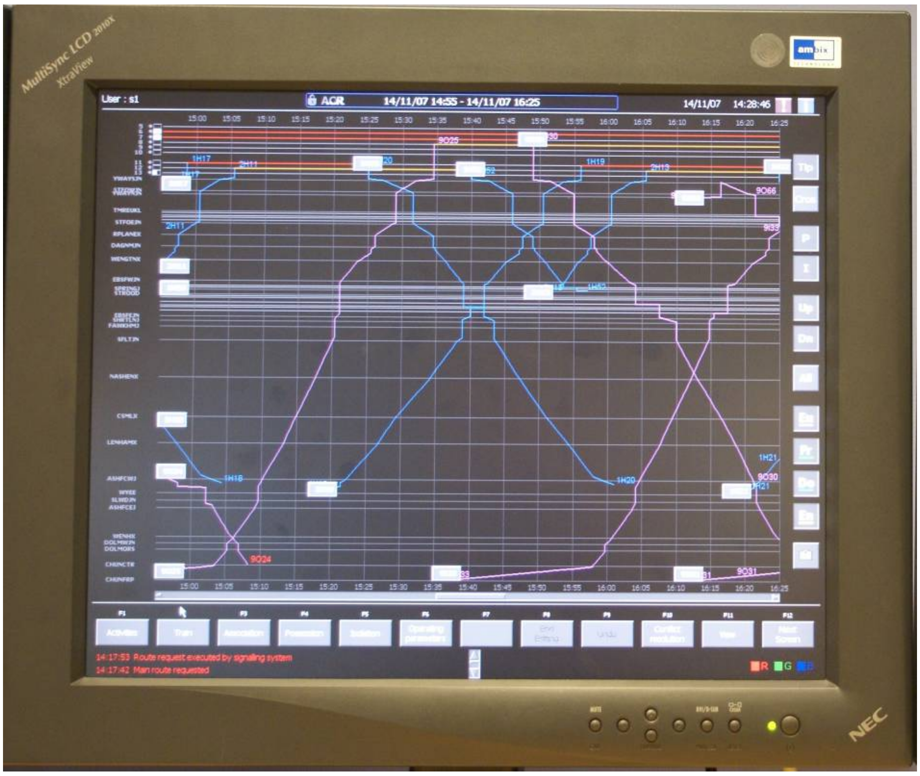


Figure 7-3: Example of a Train Graph

It is recommended that future automated systems in the UK should utilise train graphs as they have the facility to integrate large amounts of information in an easily interpretable format. They also allow the signaller to visualise the future state of the railway, including conflicts, and the consequences of changes to train running. If used in conjunction with improved tools to direct the automation, the signaller can work in a proactive manner, generating solutions to conflicts well before they occur.

7.5.12 Skill Degradation

There is no evidence that ARS allows signallers' regulation skills to degrade over time. Signallers still monitor the system closely and frequently step in to route trains manually. It is likely that the weaknesses of ARS in terms of competence, observability and understandability ensure that the signallers maintain skills which otherwise might degrade over time. However, there have been reports of signallers having a lesser knowledge of the train timetable when working with automation.

Should a more competent automated system be introduced in the future skill degradation may become more of an issue. However, a move towards a proactively controlled system may help avoid the issue of skill degradation. Any future system should ensure that the skills required to operate it are maintained, either through frequent application or structured training.

7.6 Limitations of the Research

The potential scope of work to study automation in rail signalling was very wide, and it was not possible to examine all aspects of human automation interaction in great depth. Therefore the thesis has concentrated upon trust, workload, and monitoring. Other potential human factors issues in automation presented study problems. Situation awareness posed a particular problem because there is not yet a sufficient understanding of what constitutes SA in a signalling context. It proved impossible to measure clearly the effect of automation on signaller SA.

A second limitation was the number of participants in the level of automation experiment. As the experiment necessarily used expert signallers who had experience working on the workstation in the study, the sample size was small. The experiment itself was also quite complex and required participants to wear head mounted eye tracking equipment, which further reduced the number of willing volunteers.

The usefulness of the eye tracking data collected was also limited; difficulties were encountered with the use of the equipment in a signal box environment and although these were in the main overcome, accurate calibration of the equipment remained an issue. This limited the analysis of the data, which was constrained to an analysis of the segment of the screen rather than a more detailed analysis of particular elements of infrastructure monitored.

7.7 Chapter Summary

This chapter has demonstrated how the research presented and discussed in the previous chapters has met the objectives set in Chapter 1. The conceptual framework and review of the literature formed a theoretical framework to support the research. The use of automation was studied and findings on monitoring, interaction, trust, and workload were presented. Finally, the results of these investigations were presented in terms of the principles of automation and recommendations for future automation were given. The final chapter summarises the recommendations, describes the impact of the research, and suggests areas for future research.

CHAPTER 8: CONCLUSIONS

8.1 Chapter Overview

This final chapter summarises the recommendations generated by the research and describes the impact of the research thus far. Future research is also suggested to build on this work. This includes further investigation of the importance of mental models in understanding and predicting automation, and how this relates to trust and usage of the system.

8.2 Recommendations

The research undertaken and described in this thesis examined the existing implementation of automation in UK rail signalling in order to understand the use and design of that automation and generate recommendations for future systems. The recommendations generated are based on the 12 principles of automation and are summarised in Table 8-1.

Principe	*Recommendation*
Reliable	Reliability of automated signalling systems should continue to be extremely high.
Competent	Consideration should be given to lower levels of decision making automation. Communication of information should be better supported through automation.
Visible	Decision relevant information should be integrated on one display.
Observable	Clear and explicit feedback should be provided on the automation's analysis and intentions.
Understandable	Improved feedback should facilitate the development of operator mental models. Simpler automation would have the advantage of being more easily understood.
Directable	Inhibitors should be introduced to replace the use of reminders to regulate trains. New methods of controlling the automation and how it resolves conflicts should be introduced.
Robust	Methods of directing the automation during disruption should be introduced.
Accountable	Designers, managers, and operators should regard ARS as a tool to assist the signaller and ARS should support the signaller in controlling the railway.
Error Resistant	Where possible, errors in the timetable data should be highlighted by the automation.
Error Tolerant	Interlocking should continue to provide support. Analysis of the consequences of actions should be provided.
Proactive Control	Train graphs should be used to support proactive control of the railway.
Skill Degradation	Skills should be maintained, either through frequent use or simulator training.

Table 8-1: Summary of Recommendations

Use of the principles of automation, and the specific recommendations provided under each one, in the design of new automated signalling systems should result in a more powerful human machine system which works cooperatively to increase performance.

8.3 Impact of the Research

The work carried out for this thesis has already contributed towards the improvement of rail signalling automation in future by influencing a new specification from Network Rail which requires automation systems to give greater insight into their actions. In addition, the company responsible for developing ARS (DeltaRail) has adopted the two key principles of observability (leading to predictability) and directability and will use these to guide future developments. The principles will also be applied as human factors requirements for new automated signalling systems introduced to the UK network within a major new programme defining the future operating strategy.

The observation framework developed during this research is now used by EPSRC Rail Research UK (RRUK) researchers in their research on signaller activities. It has also been further developed for use in workload assessments. The same five basic behaviours (monitoring, intervention, planning, communications, and quiet time) are used but the rigorous collection of data every 5s is not. Data are now collected in 1min intervals, with subdivisions within that to the nearest approximate 5s. This relieves the demand on the observer. The requirement for the observer to remain withdrawn from the signallers being observed has also been removed and this allows for probing of the reasons behind observed behaviours. Some granularity has therefore been lost but has been replaced by the facility to gain some insight into signaller strategies. This has proven a useful tool in conducting workload assessments, and is set to become a standard part of the Network Rail workload toolkit.

The findings from this research also have implications beyond rail signalling; the observation study and trust questionnaire found correlations between several dimensions of trust, including feedback, reliability, understanding, and predication, and the operators' level of intervention. This is in agreement with existing research in terms of feedback and reliability (Dzindolet et al., 2003; Wiegmann et al., 2001), but the empirical link between understanding the automation and predicting its future actions and trust in the automation may be new. The research indicated that automation has been most successful at reducing workload in the area of action implementation, as suggested by existing literature (Kaber et al., 2006). A new theory of active and passive monitoring has been generated by this research, which proposes that operators vary their level of monitoring according to the conditions of the area under their control, but crucially they do actively process information, even during passive monitoring. This is in contrast to research which suggests that processing of information is negatively affected by monitoring behaviour (Metzger & Parasuraman, 2001). Finally, no evidence was found to support the existence of a vigilance decrement or complacency in a real world automated system (Moray, 2003).

8.4 Future Work

The interviews with signallers indicated that understanding of the automation is likely to be key to its optimal use. The literature in the area has shown a link between feedback from the automation and its use (Dzindolet et al., 2003), but this was in quite simple systems. Future research evaluating the link between understanding of the factors the automation takes into account in decision making and use of automation would be very interesting. The research should aim to address the key question, "what level of understanding of the underlying algorithms, which may be quite complex, is necessary for improved usage?"

With respect to rail signalling systems, a key emerging question in Network Rail is to determine what span of control is appropriate for a signaller working with automation. This is a complex question and is dependent on a number of factors, including the type of automation available, the competence of that automation and its ability to operate during disruption, and the acceptable drop in performance during disruption. As recommended in this thesis, automation which allows the signaller to plan ahead and removes the need for manual route setting is likely to reduce workload, even during disruption and so could allow for an increased span of control as compared to the current ARS system. Automation which can be trusted to continue operating (either due to its predictability or competence) during disruption would also contribute towards allowing a larger span of control as the signaller can concentrate on the disrupted area allowing the automation to control the remaining area. Finally, if major disruption were to become relatively rare (the railway is also working towards increased reliability of its assets) then it may be acceptable for a larger drop in performance during disruption, perhaps even cancelling large parts of the train service. This would reduce the requirement to staff the signalling centres for disrupted conditions, and an increase in control area per signaller could be achieved. Obviously, to answer these questions there is a great deal of research still required to determine more definitely the effect of automation on workload, the effect of different interfaces on workload, and to develop methods to measure both of these.

Due to the technical difficulty of manipulating the 'cleverness' of ARS it is not possible to examine the differences in system performance between working with a simple and complex ARS. Much of the research carried out point to the complexity of ARS as a factor in signallers' failure to understand and predict the automation. The results of the experiment indicate that there is not a significant difference in workload or performance between ARS and the considerably less complex Auto-routes. Future research could investigate this relationship further and examine whether losing some of the complexity in the conflict resolution algorithms has a positive impact by facilitating the signallers' control of the system better, or a negative impact because of the loss of decision making power.

8.5 Summary

This thesis has presented the research performed to investigate the impact of automation on rail signalling. Both the use and design of automation systems were considered using a variety of methods, and guidance for new automated signalling systems was generated from the data collected. This guidance will help direct the design and implementation of the next generation of signalling systems in the UK.

REFERENCES

Anders, G. (2001). Pilot's attention allocation during approach and landing-eye and head-tracking research in an A330 full flight simulator. In R. Jensen (Ed.), *Proceedings of the International Symposium on Aviation Psychology (ISAP)*. Columbus, OH.

Atoyan, H., Duquet, J.-R., & Robert, J.-M. (2006). Trust in new decision aid systems. In *Proceedings of the 18th International Conference on Association Francophone d'Interaction Homme-Machine* (pp. 115-122). New York: ACM Press.

Bagheri, N., & Jamieson, G. A. (2004). Considering subjective trust and monitoring behaviour in assessing automation-induced "complacency". In D. A. Vicenzi, M. Mouloua & P. A. Hancock (Eds.), *Human Performance, Situation Awareness and Automation HPSAA* (pp. 54-59). Mahwah, NJ: Erlbaum.

Bainbridge, L. (1983). Ironies of automation. *Automatica,* **19**(6), 775-779.

Barber, B. (1983). *The logic and limits of trust*. New Brunswick, NJ: Rutgers University Press.

Beck, H. P., Dzindolet, M. T., & Pierce, L. G. (2007). Automation usage decisions: Controlling intent and appraisal errors in a target detection task. *Human Factors,* **49**(3), 429-437.

Bentley, R., Hughes, J. A., Randall, D., Rodden, T., Sawyer, P., Shapiro, D., et al. (1992). Ethnographically-informed systems design for air traffic control. In *ACM 1992 Conference on Computer-Supported Cooperative Work: Sharing perspectives* (pp. 123-129). New York: ACM Press.

Billings, C. E. (1991). *Human-centered aircraft automation: A concept and guidelines* (No. NASA Technical Memorandum 103885). Springfield, VA: National Technical Information Service.

Billings, C. E. (1997). Issues concerning human-centered intelligent systems: What's "human-centered" and what's the problem? [Electronic Version]. Retrieved April 24, 2005 from www.ifp.uiuc.edu.

Billings, C. E., & Woods, D. D. (1994). Concerns about adaptive automation in aviation systems. In M. Mouloua & R. Parasuraman (Eds.), *Human performance in automated systems: Current research and trends* (pp. 263-269). Mahwah, NJ: Lawrence Erlbaum.

Bisantz, A. M., & Seong, Y. (2001). Assessment of operator trust in and utilisation of automated decision-aids under different framing conditions. *International Journal of Industrial Ergonomics,* **28**(2), 85-97.

Bristol, N. (2005). *Shared mental models: Conceptualisation and measurement.* University of Nottingham, Nottingham.

Burrage, F. W., How, F., Waller, J., Weightman, C. I., Weedon, R. L., & Leach, M. E. (1991). *Railway Control Systems*. London: A&C Black.

Chipman, S. F., Schraagen, J. M., & Shalin, V. L. (2000). Introduction to Cognitive Task Analysis. In J. M. Schraagen, S. F. Chipman & V. L. Shalin (Eds.), *Cognitive Task Analysis* (pp. 3-23). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Christoffersen, K., & Woods, D. D. (2001). How to make automated systems team players. In E. Salas (Ed.), *Advances in Human Performance and Cognitive Engineering Research* (Vol. 2, pp. 1-12). New York: JAI Press/Elsevier.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin,* **104**(2), 163-191.

de-Vries, P., Midden, C., & Bouwhuis, D. (2003). The effects of errors on system trust, self-confidence, and the allocation of control in route planning. *International Journal of Human-Computer Studies,* **58**(6), 719-735.

Dekker, S. W. A. (2004). On the other side of promise: What should we automate today? In D. Harris (Ed.), *Human Factors for Civil Flight Deck Design* (pp. 183-198). Aldershot, Hampshire: Ashgate Publishing.

Dekker, S. W. A., & Woods, D. D. (2004). MABA-MABA or abracadabra? Progress on human-automation co-ordination. *Cognition, Technology & Work,* **4**(4), 240-244.

DeltaRail. (2008). Automatic Route Setting System. Retrieved 15 April 2009, from http://www.deltarail.com/what_we_do/signalling_control_systems/ars.html

Dishart, D. C., & Land, M. F. (1998). The development of the eye movement strategies of learner drivers. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 419-430). Oxford, England: Elsevier.

Dixon, S. R., & Wickens, C. D. (2006). Automation reliability in unmanned aerial vehicle control: A reliance-compliance model of automation dependence in high workload. *Human Factors,* **48**(3), 474-486.

Donmez, B., Pina, P. E., & Cummings, M. L. (2009). Evaluation criteria for human-automation performance metrics. In *Performance Evaluation and Benchmarking of Intelligent Systems* (pp. 21-40). New York: Springer US.

Duchowshi, A. T. (2003). *Eye tracking methodology: Theory and practice*. London: Springer-Verlag.

Durso, F. T., & Dattell, A. R. (2004). SPAM: The real-time assessment of SA. In S. Banbury & S. Tremblay (Eds.), *A cognitive approach to situation awareness: Theory and application* (pp. 137-154). Aldershot, UK: Ashgate.

Durso, F. T., & Sethumadhavan, A. (2008). Situation awareness: Understanding dynamic environments. *Human Factors,* **50**(3), 442-448.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies,* **58**(6), 697-718.

Endsley, M. R. (1995). Measurement of situation awareness in dynamic systems. *Human Factors,* **37**(1), 65-84.

Endsley, M. R. (1996). Automation and situation awareness. In R. Parasuraman & M. Mouloua (Eds.), *Automation and Human Performance: Theory and Applications* (pp. 163-181). Mahwah, NJ: Lawrence Erlbaum.

Endsley, M. R. (2000). Direct measurement of situation awareness: Validity and use of SAGAT. In M. R. Endsley & D. J. Garland (Eds.), *Situation awareness analysis and measurement* (pp. 147-173). Mahwah, NJ: Lawrence Erlbaum Associates.

Endsley, M. R., & Kaber, D. B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics,* **42**(3), 462-492.

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors,* **37**(2), 381-394.

Endsley, M. R., & Rodgers, M. D. (1994). Situation awareness information requirements for en route air traffic control. In *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting* (pp. 71-75). Santa Monica, CA: Human Factors and Ergonomics Society.

Fitts, P. M. (1951). *Human engineering for an effective air-navigation and traffic-control system*. Washington, DC: Ohio State University.

Golightly, D., Balfe, N., & Sharples, S. (2009). *Measuring situation awareness in rail signalling.* Paper presented at the Third International Conference on Rail Human Factors, Lille, France (3-5 March).

Hammersley, M., & Atkinson, P. (1995). *Ethnography: Principles in Practice*. London: Routledge.

Harris, W. C., Hancock, P. A., Arthur, E. J., & Caird, J. K. (1995). Performance, workload, and fatigue changes associated with automation. *The International Journal of Aviation Psychology,* **5**(2), 169-185.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 139-183). Amsterdam: North-Holland.

Hayes, N. (2000). *Doing Psychological Research*. Maidenhead: Open University Press.

Hellstrom, P., Frej, I., Gideon, A., & Sandblad, B. (1997). *Algorithms and control systems for computer-aided train dispatching.* Paper presented at the World Congress on Railway Research, Florence, Italy (16-19 November).

Hignett, S., & Wilson, J. R. (2004). The role for qualitative methodology in ergonomics: A case study to explore theoretical issues. *Theoretical Issues in Ergonomics Science,* **5**(6), 473-493.

Hoc, J. M. (2000). From human-machine interaction to human-machine cooperation. *Ergonomics,* **43**(7), 833-843.

Hollnagel, E. (2001). Human-oriented automation strategies. In *Proceedings of the World Congress on Safety of Modern Technical Systems* (pp. 443-452). Cologne: TUV-Verlag, Cologne.

Hopkin, V. D., & Wise, J. A. (1996). Human factors in air traffic system automation. In R. Parasuraman & M. Mouloua (Eds.), *Automation and Human Performance: Theory and Applicatiosn* (pp. 319-336). Hillsdale, NJ: Erlbaum.

Jacob, R. J. K., & Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In R. Radach, J. Hyona & H. Deubel (Eds.), *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research* (pp. 573-605). Oxford, England: North Holland.

Jian, J. Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics,* **4**(1), 53-71.

Jiang, X., Khasawneh, M. T., Master, R., Bowling, S. R., Gramopadhye, A. K., Melloy, B. J., et al. (2004). Measurement of human trust in a hybrid inspection system based on signal detection theory measures. *International Journal of Industrial Ergonomics,* **34**(5), 407-419.

Jones, D. G. (2000). Subjective measurement of situation awareness. In M. R. Endsley & D. J. Garland (Eds.), *Situation awareness: Analysis and measurement* (pp. 113-128). Mahwah, NJ: Lawrence Erlbaum Associates.

Kaber, D. B., & Endsley, M. R. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science,* **5**(2), 113-153.

Kaber, D. B., Onal, E., & Endsley, M. R. (2000). Design of automation for telerobots and the effect on performance, operator situation awareness, and subjective workload. *Human Factors and Ergonomics in Manufacturing,* **10**(4), 409-430.

Kaber, D. B., Perry, C. M., Segall, N., McClernon, C. K., & Prinzel, L. J. (2006). Situation awareness implications of adaptive automation for information processing in an air traffic control control-related task. *International Journal of Industrial Ergonomics,* **36**(5), 447-462.

Kantowitz, B. H. (1994). Pilot workload and flight deck automation. In M. Mouloua & R. Parasuraman (Eds.), *Human performance in automated systems: Current research and trends* (pp. 212-223). Mahwah, NJ: Lawrence Erlbaum.

Kauppi, A. (2006). *A Human-Computer Interaction Approach to Train Traffic Control.* Uppsala University, Uppsala.

Kauppi, A., Wikstrom, J., Hellstrom, P., Sandblad, B., & Andersson, A. W. (2005). Future train traffic control: control by re-planning. In *Rail Human Factors: Supporting the Integrated Railway* (pp. 296-308). Aldershot: Ashgate.

Kauppi, A., Wikstrom, J., Sandblad, B., & Andersson, A. W. (2006). Future train traffic control: Control by re-planning. *Cognition, Technology & Work,* **8**(1), 50-56.

Krisler, B., & Alterman, R. (2008). Training towards mastery: Overcoming the active user paradox. In *Proceedings of NordiCHI 2008: Using Bridges* (pp. 239-248). New York: ACM.

Kvist, T., Hellstrom, P., Sandblad, B., & Bystrom, J. (2002). Decision support in the train dispatching process. In J. Allan, R. J. Hill, C. A. Brebbia & G. Sciutto (Eds.), *Computers in Railways VIII* (pp. 626-633). Boston: WIT Press.

Lee, J. D. (1991). The dynamics of trust in a supervisory control simulation. In *In Proceedings of the Human Factors Society 35th Annual Meeting* (pp. 1228-1232). Santa Monica, CA: The Human Factors Society.

Lee, J. D. (2008). Review of a pivotal Human Factors article: "Humans and automation: Use, misuse, disuse, abuse". *Human Factors,* **50**(3), 404-410.

Lee, J. D., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics,* **35**(10), 1243-1270.

Lee, J. D., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies,* **40**(1), 153-184.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors,* **46**(1), 50-80.

Lenior, D., Janssen, W., Neerincx, M., & Schreibers, K. (2006). Human-factors engineering for smart transport: Decision support for car drivers and train traffic controllers. *Applied Ergonomics,* **37**(4), 479-490.

Lenior, T. M. J. (1993). Analyses of cognitive processes in train traffic control. *Ergonomics,* **36**(11), 1361-1368.

Lewandowsky, S., Mundy, M., & Tan, G. P. A. (2000). The dynamics of trust: Comparing humans to automation. *Journal of Experimental Psychology: Applied,* **6**(2), 104-123.

Liu, Y., Fuld, R., & Wickens, C. D. (1993). Monitoring behaviour in manual and automated scheduling systems. *International Journal of Man-Machine Studies,* **39**(6), 1015-1029.

Luke, T., Brook-Carter, N., Parkes, A. M., Grimes, E., & Mills, A. (2006). An investigation of train driver visual strategies. *Cognition, Technology & Work,* **8**(1), 15-29.

Macdonald, W. A. (1999). Train controller interface design: Factors influencing mental workload. In *Proceedings of the International Conference on Human Interfaces in Control Rooms, Cockpits and Command Centres* (pp. 31-36). London: Institute of Electrical Engineers.

Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human-human and human-automation trust: An integrative review. *Theoretical Issues in Ergonomics Science,* **8**(4), 277-301.

Madsen, M., & Gregor, S. (2000). Measuring human-computer trust. In G. Gable & M. Viatle (Eds.), *Proceedings of the 11th Australasian Conference on Information Systems* (pp. 53-64). Victoria: AIS.

Megaw, T. (2005). The definition and measurement of mental workload. In J. R. Wilson & N. Corlett (Eds.), *Evaluation of Human Work* (3rd ed., pp. 525-551). London: Taylor & Francis.

Merritt, S. M., & Ilgen, D. R. (2008). Not all trust is created equal: Dispositional and history based trust in human-automation interactions. *Human Factors,* **50**(2), 194-210.

Metzger, U., & Parasuraman, R. (2001). The role of the air traffic controller in future air traffic management: An empirical study of active control versus passive monitoring. *Human Factors,* **43**(4), 519-528.

Miller, C., Funk, H., Wu, P., Goldman, R., Meisner, J., & Chapman, M. (2005). The playbook approach to adaptive automation. In *Proceedings of the Human*

Factors and Ergonomics Society 49th Annual Meeting (pp. 15-19). Santa Monica, CA: Human Factors and Ergonomics Society.

Moray, N. (2003). Monitoring, complacency, scepticism and eutactic behaviour. *International Journal of Industrial Ergonomics,* **31**(3), 175-178.

Moray, N., & Haudegond, S. (1998). An absence of vigilance decrement in a complex dynamic task. In *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting* (pp. 234-238). Santa Monica, CA: Human Factors and Ergonomics Society.

Moray, N., & Inagaki, T. (2000). Attention and complacency. *Theoretical Issues in Ergonomics Science,* **1**(4), 354-365.

Moray, N., Inagaki, T., & Itoh, M. (2000). Adaptive automation, trust, and self-confidence in fault management of time-critical tasks. *Journal of Experimental Psychology: Applied,* **6**(1), 44-58.

Morimoto, C. H., & Mimica, M. R. M. (2005). Eye gaze tracking techniques for interactive applications. *Computer Vision and Image Understanding,* **98**(1), 4-24.

Mosier, K. L., Skitka, L. J., & Korte, K. J. (1994). Cognitive and social psychological issues in flight crew/automation interaction. In M. Mouloua & R. Parasuraman (Eds.), *Human performance in automated systems: Current research and trends* (pp. 191-197). Mahwah, NJ: Lawrence Erlbaum.

Muffett, R. J. (2007). Ergonomics assessment of lever operations in mechanical signalling. In J. R. Wilson, B. Norris, T. Clarke & A. Mills (Eds.), *People and Rail Systems: Human Factors at the Heart of the Railway* (pp. 257-266). Aldershot: Ashgate Publishing Limited.

Muir, B. M. (1987). Trust between humans and machines, and the design of decision aids. *International Journal of Man-Machine Studies,* **27**(5-6), 527-539.

Muir, B. M. (1994). Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics,* **37**(11), 1905-1922.

Muir, B. M., & Moray, N. (1989). Operators' trust in and use of automatic controllers. In *In Proceedings of the 22nd Annual Conference of the Human Factors Association of Canada* (pp. 163-166). Ontario: Human Factors Association of Canada.

Muir, B. M., & Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics,* **39**(3), 429-460.

Noldus. (2007). The Observer XT. Computer Software. Amsterdam, The Netherlands.

Norman, D. A. (1983). Design rules based on analyses of human error. *Communications of the ACM,* **26**(4), 254-258.

Norman, D. A. (1990). The 'problem' with automation: Inappropriate feedback and interaction, not 'over-automation'. *Philosophical Transactions of the Royal Society of London,* **B327**(1241), 585-593.

Olsen, W. A., & Sarter, N. B. (1999). Supporting informed consent in human-machine collaboration: The role of conflict type, time pressure, and display design. In *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting* (pp. 189-193). Santa Monica, CA: Human Factors and Ergonomics Society.

Ottati, W. L., Hickox, J. C., & Richter, J. (1999). Eye scan patterns of experienced and novice pilots during visual flight rules (VFR) navigation. In *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting* (pp. 66-70). Santa Monica, Ca: Human Factors and Ergonomics Society

Pan, X., Gillies, M., & Slater, M. (2008). Male bodily responses during an interaction with a virtual woman. In H. Prendinger, J. Lester & M. Ishizuka (Eds.), *Intelligent Virtual Agents* (pp. 89-96). Heidelbery: Springer Berlin.

Parasuraman, R. (1987). Human-computer monitoring. *Human Factors,* **29**(6), 695-706.
Parasuraman, R. (2000). Designing automation for human use: Empirical studies and quantitative models. *Ergonomics,* **43**(7), 931-951.
Parasuraman, R., Molloy, R., Mouloua, M., & Hilburn, B. (1996). Monitoring of automated systems. In R. Parasuraman & M. Mouloua (Eds.), *Automation and Human Performance: Theory and Applications* (pp. 91-115). Mahwah, NJ: Lawrence Erlbaum Associates.
Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance Consequences of Automation-Induced 'Complacency'. *International Journal of Aviation Psychology,* **3**(1), 1-23.
Parasuraman, R., Mouloua, M., & Molloy, R. (1996). Effects of adaptive task allocation on monitoring of automated systems. *Human Factors,* **38**(4), 665-679.
Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors,* **39**(2), 230-253.
Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans,* **30**(3), 286-297.
Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically support cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making,* **2**(2), 140-160.
Parasuraman, R., & Wickens, C. D. (2008). Humans: Still vital after all these years of automation. *Human Factors,* **50**(3), 522-520.
Pickup, L. (2006). *Understanding and Assessment of Mental Workload in Railway Signalling.* University of Nottingham.
Pickup, L., & Wilson, J. R. (2007). Mental workload assessment and the development of the operational demand evaluation checklist (ODEC) for signallers. In J. R. Wilson, B. Norris, T. Clarke & A. Mills (Eds.), *People and Rail Systems: Human Factors at the Heart of the Railway* (pp. 215-224). Aldershot: Ashgate Publishing Limited.
Pickup, L., Wilson, J. R., Norris, B. J., Mitchell, L., & Morrisroe, G. (2005). The integrated workload scale (IWS): A new self-report tool to assess railway signaller workload. *Applied Ergonomics,* **36**(6), 681-693.
Pidgeon, N. (1996). Grounded theory: Theoretical background. In J. T. E. Richardson (Ed.), *Handbook of Qualitative Research Methods* (pp. 75-85). Leicester: BPS Books.
Rasmussen, J. (1983). Skills, rules, and knowledge: Signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics,* **SMC-13**(3), 257-266.
Rasmussen, J., & Vicente, K. J. (1989). Coping with human errors through system design: Implications for ecological interface design. *International Journal of Man-Machine Studies,* **31**(5), 517-534.
Reinach, S. J. (2006). Toward the development of a performance model of railroad dispatching. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (pp. 2042-2046). Santa Monica: Human Factors and Ergonomics Society.
Rempel, J. K., Holmes, J. G., & Zanna, M. P. (1985). Trust in close relationships. *Journal of Personality and Social Psychology,* **49**(1), 95-112.
Renyard, W. D., Billings, C. E., Cheaney, E. S., & Hardy, R. (1986). *The development of the NASA Aviation Safety Reporting System* (No. NASA Publication 1114). Washington, DC: NASA.

Riley, V. (1994). A theory of operator reliance on automation. In M. Mouloua & R. Parasuraman (Eds.), *Human performance in automated systems: Current research and trends* (pp. 8-14). Mahwah, NJ: Lawrence Erlbaum.

Riley, V. (1996). Operator reliance on automation: Theory and data. In R. Parasuraman & M. Mouloua (Eds.), *Automation and Human Performance: Theory and Applications* (pp. 19-35). Hillsdale, NJ: Erlbaum.

Robson, C. (2002). *Real World Research* (2nd ed.). Oxford: Blackwell.

Roth, E. M., Malsch, N., & Multer, J. (2001). *Understanding how train dispatchers manage and control trains* (No. DOT/FRA/ORD-01/02). Washington, DC: US Department of Transportation, Federal Railrod Administration.

Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin,* **100**(3), 349-363.

Ruff, H. A., Narayanan, S., & Draper, M. H. (2002). Human interaction with levels of automation and decision-aid fidelity in the supervisory control of multiple simulated unmanned air vehicles. *Presence: Teleoperators and Virtual Environments,* **11**(4), 335 - 351.

Salmon, P., Stanton, N. A., Walker, G. H., & Green, D. (2006). Situation awareness measurement: A review of applicability for C4i environments. *Applied Ergonomics,* **37**(2), 225-238.

Sandblad, B., Andersson, A. W., Bystrom, J., & Kauppi, A. (2002). New control strategies and user interfaces for train traffic control. In J. Allan, R. J. Hill, C. A. Brebbia & G. Sciutto (Eds.), *Computers in Railways VIII*. Boston: WIT Press.

Sarter, N. B., Mumaw, R. J., & Wickens, C. D. (2007). Pilots' monitoring strategies and performance on automated flight decks: An empirical study combining behavioural and eye-tracking data. *Human Factors,* **49**(3), 347-357.

Sarter, N. B., & Woods, D. D. (1997). Team Play with a powerful and independent agent: Operational experiences and automation surprises on the airbus A-320. *Human Factors,* **39**(4), 553-569.

Sarter, N. B., Woods, D. D., & Billings, C. E. (1997). Automation surprises. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (2nd ed., pp. 1926-1943). New York: Wiley.

Scerbo, M. W. (1996). Theoretical perspectives on adaptive automation. In R. Parasuraman & M. Mouloua (Eds.), *Automation and Human Performance: Theory and Applications* (pp. 37-63). Mahwah, NJ: Lawrence Erlbaum.

Sheridan, T. B. (1996). Speculations on future relations between humans and automation. In R. Parasuraman & M. Mouloua (Eds.), *Automation and Human Performance: Theory and Applications* (pp. 449-460). Hillsdale, NJ: Erlbaum.

Sheridan, T. B. (1998). Rumination on automation. *Annual Reviews in Control,* **25**, 89-97.

Sheridan, T. B. (1999). Human supervisory control. In A. P. Sage & W. B. Rouse (Eds.), *Handbook of Systems Engineering and Management* (pp. 645-690). New York: John Wiley & Sons.

Sheridan, T. B. (2002). *Humans and Automation*. New York: John Wiley & Sons.

Sheridan, T. B., & Parasuraman, R. (2006). Human-automation interaction. In R. S. Nickerson (Ed.), *Reviews of human factors and ergonomics* (Vol. 1, pp. 89-129). Santa Monica, CA: Human Factors and Ergonomics Society.

Sheridan, T. B., & Verplank, W. L. (1978). *Human and Computer Control of Undersea Teleoperators*. Cambridge, Massachusetts: Massachusetts Institute of Technology.

Shorrock, S. T., & Straeter, O. (2006). A framework for managing system disturbances and insights from air traffic management. *Ergonomics,* **49**(12-13), 1326-1344.

Sinclair, M. A. (2005). Participative Assessment. In J. R. Wilson & N. Corlett (Eds.), *Evaluation of Human Work* (3rd ed., pp. 83-112). London: Taylor & Francis.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics* (5th ed.). Boston: Pearson Education Inc.

Timmer, P., & Stork, A. (2000). *Domain analysis and planning for network management centre design* (No. Report v 1b): Ergonomics and HCI Unit, University College London.

Vicente, K. J., Mumaw, R. J., & Roth, E. M. (1998). More about operator monitoring under normal operations: The role of workload regulation and the impact of control room technology. In *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting* (pp. 229-233). Santa Monica, CA: Human Factors and Ergonomics Society.

Vicente, K. J., Roth, E. M., & Mumaw, R. J. (2001). How do operators monitor a complex, dynamic work domain? The impact of control room technology. *International Journal of Human-Computer Studies,* **54**(6), 831-856.

Vicente, K. J., Roth, E. M., & Mumaw, R. J. (2004). Operator monitoring in a complex dynamic work environment: A qualitative cognitive model based on field observations. *Theoretical Issues in Ergonomics Science,* **5**(5), 359-384.

Warm, J. S., Dember, W. N., & Hancock, P. A. (1996). Vigilance and workload in automated systems. In R. Parasuraman & M. Mouloua (Eds.), *Automation and Human Performance: Theory and Applications* (pp. 183-200). Mahwah, NJ: Lawrence Erlbaum.

Waterson, P. E., Older Gray, M. T., & Clegg, C. W. (2002). A sociotechnical method for designing work systems. *Human Factors,* **44**(3), 376-391.

Wickens, C. D. (1992). *Engineering Psychology and Human Performance* (2nd ed.). New York: HarperCollins.

Wickens, C. D. (2008). Situation awareness: Review of Mica Endsley's 1995 articles on situation awareness theory and measurement. *Human Factors,* **50**(3), 397-403.

Wickens, C. D., & Carswell, C. M. (1997). Information Processing. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (2nd ed., pp. 89-129). New York: John Wiley & Sons.

Wickens, C. D., & Dixon, S. R. (2007). The benefits of imperfect diagnostic automation: A synthesis of the literature. *Theoretical Issues in Ergonomic Science,* **8**(3), 201-212.

Wickens, C. D., Gordon, S. E., & Liu, Y. (1998). *An Introduction to Human Factors Engineering*. New York: Longman.

Wiegmann, D. A., Rich, A. M., & Zhang, H. (2001). Automated diagnostic aids: The effects of aid reliability on users' trust and reliance. *Theoretical Issues in Ergonomics Science,* **2**(4), 352-367.

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

Wiener, E. L. (1987). Application of vigilance research: Rare, medium, or well done? *Human Factors,* **29**(6), 725-736.

Wiener, E. L. (1989). *Human factors of advanced technology ('glass cockpit') transport aircraft* (No. Technical Report 117528). Moffett Field, CA: NASA Ames Research Center.

Wiener, E. L., & Curry, R. E. (1980). Flight-deck automation: Promises and problems. *Ergonomics,* **23**(10), 995-1011.

Wilson, J. R., Cordiner, L., Nichols, S., Norton, L., Bristol, N., Clarke, T., et al. (2001). On the right track: Systematic implementation of ergonomics in railway network control. *Cognition, Technology & Work,* **3**(4), 238-252.

Woods, D. D. (1996). Decomposing automation: Apparent simplicity, real complexity. In R. Parasuraman & M. Mouloua (Eds.), *Automation Technology and Human Performance* (pp. 3-17). Hillsdale, NJ: Erlbaum.

Woods, D. D. (1997). Human-centered software agents: Lessons from clumsy automation. In J. Flanagan, T. Huang, P. Jones & S. Kasif (Eds.), *Human Centered Systems: Information, Interactivity and Intelligence* (pp. 288-293). Washington, DC: National Science Foundation.

Young, M. S., & Stanton, N. A. (2002). Attention and automation: New perspectives on mental underload and performance. *Theoretical Issues in Ergonomics Science,* **3**(2), 178-194.

APPENDICES

APPENDIX A: SITE VISITS

Site	*Type*	*ARS present*	*Number of Visits*
Leicester	NX Panel	No	1
Croft	Lever Frame	No	1
Liverpool Street	IECC	Yes	20
Ashford	IECC	Yes	10
Eastbourne	Lever Frame	No	1
Knottingley	NX Panel	No	2
Marylebone	IECC	Yes	5
Beddingham	Level Crossing	No	1
Hednesford	Lever Frame	No	1
York	IECC	Yes	12
Slough	IECC	Yes	1
Swindon B	IECC	Yes	1
Upminster	IECC	Yes	1
Tyneside	IECC	Yes	5
Bournemouth	VDU	No	1
Wembley	NX & VDU	No	1
Yoker	IECC	Yes	1
Edinburgh	IECC	Yes	4
Manchester South	VDU	No	1
Stoke	VDU	No	1
Trowse	VDU	No	1
Wimbledon	NX	Yes	1
Carlisle	NX	No	2
Woking	NX	No	1
Farnham	Lever Frame	No	1
Three Bridges	NX	No	1
Moorthorpe	Lever Frame	No	1
Glasgow	NX	No	1
Victoria	NX	No	1
Rugby	VDU	No	2
Gloucester	NX	No	1
Paisley	NX	No	1
Glasgow WSSC	VDU	No	3
Sandhills	IECC	Yes	1

APPENDIX B: SUMMARY OF RESEARCH METHODS

Method	*Sample size/Time*	*Description*	*Purpose*	*Chapter*
Participant Observation 1: Embedding in Organisation	3 years	Knowledge and understanding gained through placement in Network Rail.	Develop an understanding of research context and the factors influencing automation usage.	All
Direct Field Observations	81 box visits	Observe signallers at work and converse with them regarding their work.	Developing and maintaining an understanding of the research context and complexities of the rail signalling environment.	All
SME Discussions	3 years	Continuing discussions with a core of five subject matter experts, both operational and engineering based, facilitated by placement within Network Rail.	Develop an understanding of the research context and interpret and expand on findings from other methods.	All
Participant Observation 2: Role Training	2 weeks	Three weeks spent at signaller training school.	Develop a thorough understanding of the research context.	All
Structured Observations	24 obs. (90 min. each)	Observations of signallers working with ARS using a structured coding scheme	Gather quantitative data on signaller behaviours whilst working with ARS.	4
Questionnaire	21 signallers	Questionnaire examining signallers' trust in ARS across a number of different dimensions of trust.	Correlate reported trust levels with observed differences in the use of ARS.	4
Video archive analysis	8 videos (average 3 hours/video)	Analysis of pre-recorded videos of interviews with signallers.	Gather information on signallers' use and opinions of and interaction with ARS.	5
Semi-structured Interviews	2 SMEs 8 signallers	Semi-structured interviews carried out at workstations.	Gather information on the knowledge, attitudes and opinions of signallers towards ARS.	5
Level of Automation Experiment	6 signallers	Simulator experiment examining the differences in level of automation.	Determine the effect of automation on the workload, performance, gaze position, and observed behaviour of signallers.	6
Paired Comparisons	22	Paired comparison technique for rating concepts.	Determine the relative importance of key requirements identified from the literature.	2

APPENDIX C: RAIL ENVIRONMENT

Overview

This appendix describes the environment in which the research was conducted. The UK rail industry is briefly described and an introduction to railway operations is given. The main focus of this appendix is the description of signalling systems and procedures. The three main generations of signalling system are described from the early lever frames developed in the 1800s through to the IECC systems developed in the 1980s. Finally, details are given on ARS and its method of operation.

UK Rail Industry

Development of the Industry

The rail industry in the UK dates back to the 18th century when horse drawn carts were run on wooden rails around coalfields. With the arrival of iron rails and steam power in the 1800s railways began to appear around the country. Initially small private companies built and operated separate rail lines but slowly these became amalgamated, until by the 1920s, there were four main railway companies – Great Western Railway; London, Midland and Scottish Railway; London and North Eastern Railway; and Southern Railway. To aid the war effort the railways were taken under government control during the Second World War and in 1948 they were nationalised. Thereafter, the whole of the UK railway system was owned and operated by British Rail, a government owned company; which meant that the control and development of the whole rail network was centralised. A programme of modernisation was embarked upon regenerating tracks and stations and introducing electrification. However, as the UK road network developed rail travel became less popular and no longer made a profit. In the 1960s the government asked Dr Richard Beeching to re-organise the railway resulting in many unpopular line closures made on the basis of current profitability of routes.

This decline in the railways continued until the industry was privatised in 1997 and divided British Rail into a number of different companies. These included Railtrack, which was given ownership of the rail infrastructure and was responsible for maintaining and operating the infrastructure. Although Railtrack was responsible for maintenance and renewal of the railway, in practice these tasks were carried out by private companies under contract. Railtrack embarked on a programme of investment in the railway to upgrade and enhance the network reversing the decline of the railway. A number of franchises for train operating companies (TOCs) and freight operating companies (FOCs) were also created to run passenger and freight services respectively.

Following a number of high profile incidents, both operational and financial, Railtrack was placed into administration in 2001 and was subsequently dissolved. Network Rail was set up as a replacement in 2002 but rather than being a private company paying dividends to shareholders it was set up to operate under the same model as a private company but without shareholders; that is, a not-for-dividend company rather than a not-for-profit company. This means that any profits made by Network Rail can be reinvested in the railway.

Network Rail is the company that currently owns, maintains, and operates the rail infrastructure in the UK. Network Rail’s core business is moving trains through the infrastructure according to the train paths sold to the TOCs and FOCs. Maintenance and renewals activities are required to keep the railway operational, and currently there are a number of large scale projects underway to expand the capacity of the network. On one site visit, in an area where major work is taking place the local manager stated that in 30 years working for the railway he had seen many tracks torn up but this was the first time he had seen a brand new railway being laid. Increasingly maintenance activities, which were carried out under contract by private companies following privatisation of the industry, have been brought back in house by Network Rail.

Railway Operations

The railway is regulated by the Office of Rail Regulation (ORR) who aim to ensure that the rail network is managed efficiently and ensure health and safety objectives are met. The Railway Safety and Standards Board (RSSB) maintains standards for safety across the industry. They are responsible for producing and updating the Rule Book which contains the regulations for all activities on the railway.

The structure of the industry requires the infrastructure owner (i.e. Network Rail) to sell train paths to the TOCs and FOCs, who then sell their services to passengers and customers requiring freight services. TOCs and FOCs bid for franchises to run trains in certain areas and these are awarded by the government. Compensation for any delays to these train paths must be paid by the originator of the delay. Thus if a train breaks down the TOC is responsible for any subsequent delay and must pay compensation, but if delays are caused by management of the infrastructure Network Rail is required to pay compensation. A system of delay attribution has been established to determine the cause of delays, and this has influenced the strategies used to move trains around the network. Currently the delay is attributed under a system known as Public Performance Measure (PPM). This means that any train arriving at its destination less than 5 minutes late does not incur a penalty, but if a regional train arrives more than 5 minutes late or a long distance train arrives more than 10 minutes late compensation will have to be paid by the originator of the delay.

The timetable for railway operation is generated by Network Rail's planning department after train paths have been bought by the TOCs and FOCs. The timetable is only changed twice a year so the opportunity to change train timetables is limited. In the shorter term, changes to the timetable are facilitated through a short term planning system. This can mean that a path for a train can be created in the very short term and the signallers will have to implement the routing for this train.

As mentioned, the railway must be maintained and renewed and this requires track access for engineering workers. There are a number of methods of gaining access to the track including:

- Red Zone working: track work is carried out on a line open to traffic and track workers are responsible for detecting approaching trains and ensuring they are clear of the track when they pass;
- Green Zone working: the area in which work is being carried out is closed to traffic and thus track workers are protected;
- Protections for short term work known as T12s and T2s: the signaller ensures that no trains are routed over a particular section of track. Additional methods of protection may also be used on track but the signaller remains in control of the railway;
- Possessions (T3): used for longer term work; the signaller ensures that no trains are routed into a particular portion of the railway. Another person (PICOP) is put in charge of that portion of the railway during the possession

and he controls any movement of engineering trains or on track machinery during the possession.

Railway Staff

Signallers are required to set routes for trains and train drivers drive the trains over those routes. There are a number of other personnel who are involved with running the railway. Control staff manage larger areas of the railway than signallers and take a more strategic view of the entire railway. The TOCs, FOCs and Network Rail all have control staff who must coordinate in order to make strategic decisions such as train cancellations. They are also responsible for managing faults on the network. Electrical Control Room Operators (ECRO) are responsible for controlling electrical power supply in those areas of the rail network which are electrified. Maintenance staff require access to the railway to maintain and renew the track and signalling equipment. Mobile Operations Managers (MOMs) are usually the first line of response to any incidents on the railway including trespass, accidents, minor faults, etc. All of these people must coordinate in the everyday running of the railway.

For example, if a piece of equipment fails the signaller or train driver are likely to be the first to notice. They must then contact control to report the failure. Control organise for an inspection, either by a MOM or maintenance staff depending on the type of failure, and start to plan a strategy for the train service until the equipment is repaired. Signallers are responsible for implementing this plan. The staff inspecting and repairing the equipment must co-ordinate with the signaller, and the ECRO, if the failure is in an area of electric traction, to provide protection from trains for them while they are on track. Thus, to ensure the railway continues to run, all of these roles must work together.

Information and Communications Systems

There are a number of publications and systems used on the railway to disseminate information. This section gives a brief introduction to the most commonly used in signal boxes.

Publications

There are a number of publications routinely used in the rail industry. The *Rule Book* contains the regulations for all activities on the railway and has individual modules instructing different groups of rail workers on principles and procedures in their area of work. There are two *timetables* issued, one for passenger services and one for freight services. Individual signal boxes have *Box Special Instructions* detailing any areas in which their local operation operations differ from the regulations in the Rule Book. The *Sectional Appendix* gives details on the infrastructure in a specific area. The *Weekly Operating Notice* (WON) gives information on engineering arrangements and operating restrictions on a weekly basis.

Simplifier

The simplifier is a paper based version of the timetable used by signallers showing the trains booked on each workstation. It shows the time and route of each train booked to travel over the infrastructure controlled by that workstation. Figure C- 1 shows a simplifier.

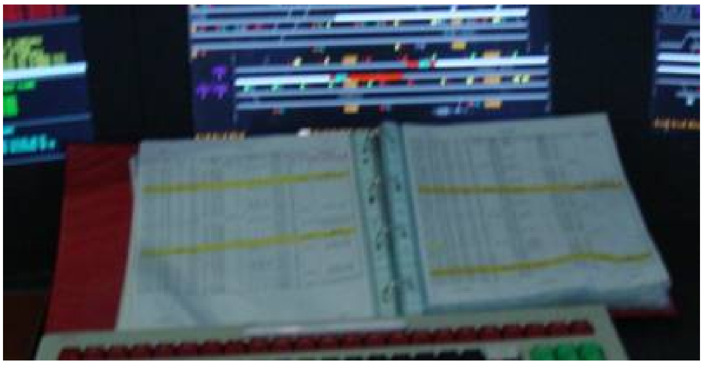


Figure C- 1: Simplifier

TOPS

Total Operations Processing System (TOPS) is a system which uses track circuit occupations to determine the location and progress of each train on the rail network.

TRUST

Train Running System on TOPS (TRUST) is a frequently used computer based tool which gives information on actual train running as compared to the timetable. It is a text based tool and allows the user to look up individual trains to see their current running. This gives signallers information on when trains can be expected to arrive on their workstation. Figure C- 2 shows a TRUST screen showing the running times for a train.

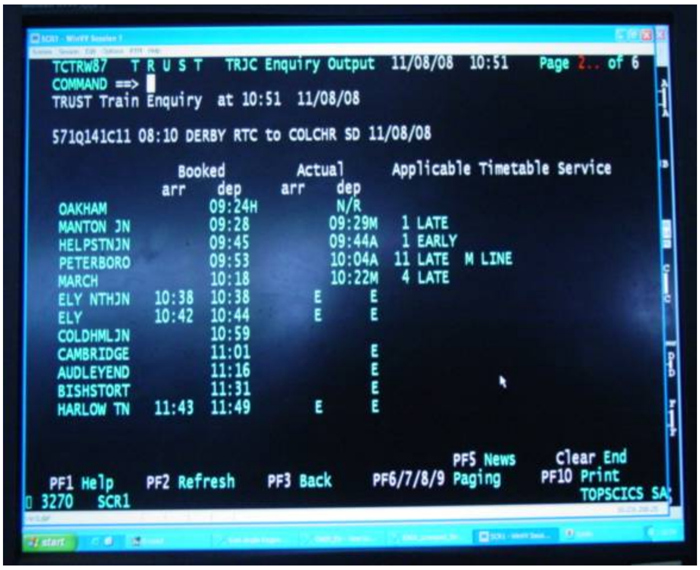


Figure C- 2: Trust Screen

CCF

Control Centre of the Future (CCF) is a map based computerised tool showing train running information (Figure C- 3). Large areas can be viewed on CCF and each train in that area is colour coded according to its status (i.e. right time or current delay). Although developed for use by controllers, this system allows signallers to get an overview of how trains surrounding their area of control are running. As with TRUST individual trains can be queried for more detailed text based information on running and delay times.

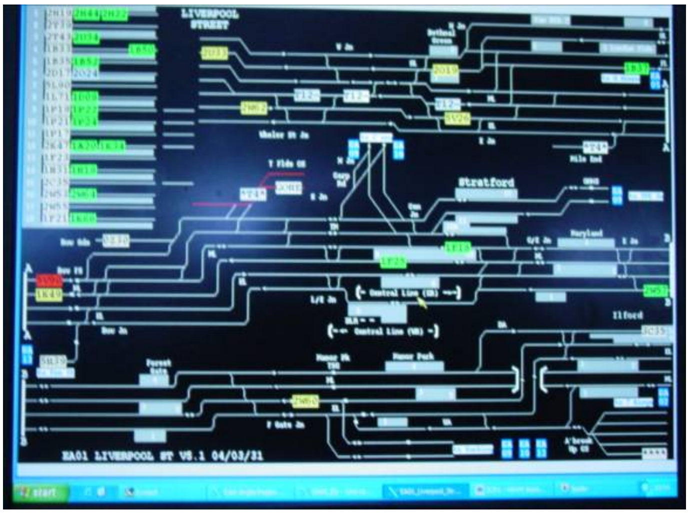


Figure C- 3: CCF Screen

Train Register Book

A train register book (TRB) is present in most signal boxes. In older manual boxes this is used to register each train passing through the area as well as any incidents or occurrences in that area. In more technologically advanced signal boxes the passing of trains is registered automatically, and the TRB is used only to record incidents and occurrences.

Communications Systems

Although communication with drivers is achieved through the signalling system by changing signal aspects to tell drivers they can proceed, some verbal communications are also required at times, particularly in the event of failure situations. These are primarily achieved through Signal Post Telephones (SPT) which are telephones located at each signal. Telephones may also be located at points. Calls from these telephones and from other operational staff working alongside a signaller are made and received from the signallers' telephone concentrator (Figure C- 4). Train drivers in some areas also have a radio system, Cab Secure Radio (CSR), in their cabs from which they can call the signaller in emergencies (Figure C- 5). CSR also contains the facility to send preset text messages between the driver and signaller for enquiry regarding routine events such as a train waiting at a red signal.

Figure C- 4: Telephone Concentrator

Figure C- 5: CSR

Signalling

Moving trains through the infrastructure is achieved through signalling systems. There are a number of types of signalling systems, but the primary goal of each is to maintain separation between trains. In the early days of the railway signalling systems were not necessary as trains had limited routes and ran much more slowly. As the weight and speed of trains increased the time taken to slow to a stop also increased. Eventually this reached the point where a train could not necessarily stop in time upon sighting an obstruction in front. This meant that if a train was to break down a following train was likely to collide with it. A system to instruct train drivers on the safety of proceeding was necessary and signalling was developed for this purpose.

At its essence, signalling divides the railway up into sections and only one train is allowed into a section at a time. This is known as 'Block System'. Each of these sections has some form of signal to pass information on to the driver about the availability of the section in front and the system is controlled by a signaller.

As changes in signalling were generally driven by developing technologies the enabling technologies for each generation of signalling system are first outlined before definitions of the signalling principles behind each system and a description of the operation of a typical signal box are given.

Early Signalling

Initially each section of the railway had a responsible citizen, usually a policeman, at its entrance. The policemen gave hand or flag signals to trains to proceed into the section and simply timed how long it had been since the last train before letting the next train in. This was known as 'Time Interval Block'. The flaw in this system was that the policemen had no way of knowing whether a train had broken down in the section and so collisions could still occur. Safer methods of signalling were developed as technology facilitated it and Time Interval Block is no longer used as a signalling principle. However, this initial use of policemen as signallers means that signallers today are still frequently referred to as 'Bobby'.

Lever Frames

Lever Frames are the oldest form of signalling system still in use on the railway today. Figure C- 6 shows a typical lever frame box. Each lever is connected to entities controlled by the signaller such as a signal or set of points and the signaller has a map of his control area displayed over the levers.

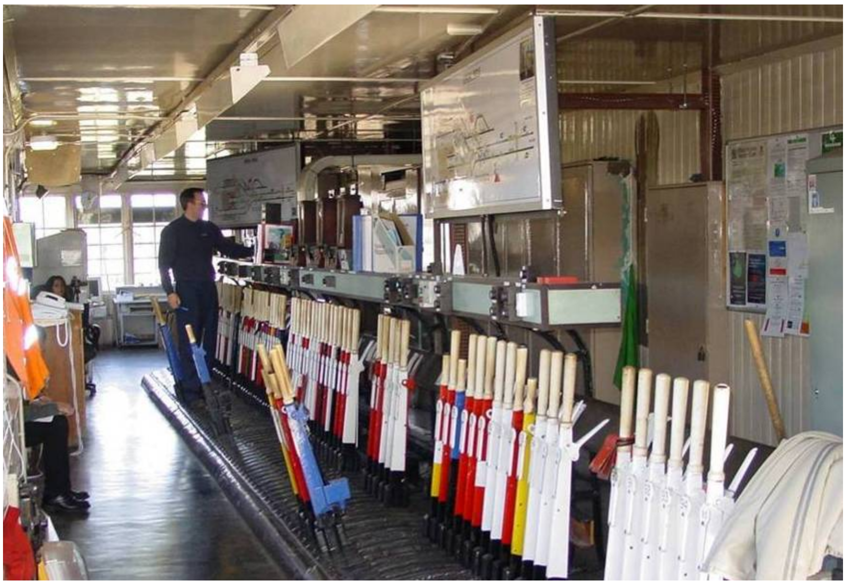


Figure C- 6: Lever Frame

Technologies

Semaphore Signals

Out on the track semaphore signals were originally used to communicate with the train drivers. A horizontal signal indicates that the driver must stop but a signal sitting at 45 degrees from horizontal indicates permission to proceed into the next section (Figure C- 7). The position of the signal is known as the aspect. Setting a signal to show a proceed aspect is called ‘clearing the signal’ or ‘pulling the signal off’.

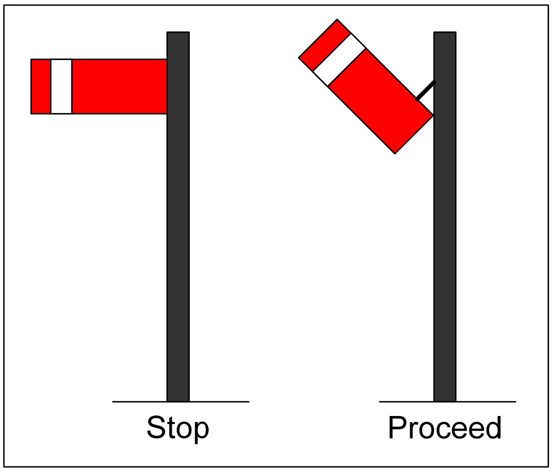


Figure C- 7: Semaphore Signals

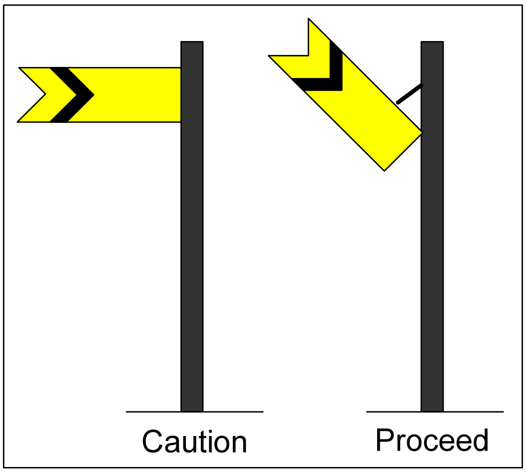


Figure C- 8: Distant Semaphore Signals

As trains can take some distance to stop ‘distant signals’ were used to indicate the position of the signal being approached to allow the train driver time to stop (Figure C- 8).

Points

As the railway became more complex, trains also required points to be set to allow them to take different routes. The position of the points is said to be normal for the route straight ahead (Figure C- 9) and reverse for the secondary route (Figure C- 10).

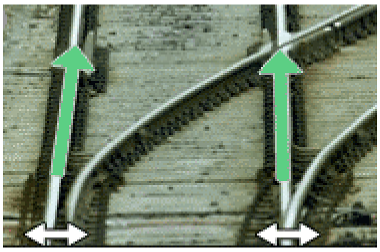


Figure C- 9: Points in Normal Position

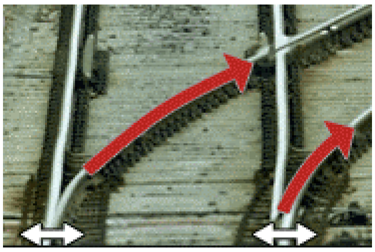


Figure C- 10: Points Set to Reverse

The signaller became responsible for setting the points in the correct positions for each train. In lever frame boxes this was achieved by the signaller pulling a lever physically connected to the points to ‘swing’ them. In more modern boxes the moving of points has been mechanised.

Reminders

Reminder appliances are used as a reminder to a signaller that a particular route or set of points should not be operated. They take different forms in different types of signal boxes but the purpose is always the same. On lever frames, reminders take the form of a metal collar which fits over the lever and prevents the signaller from pulling it. While a reminder device is applied the signaller is not able to pull the lever to clear that signal or change the position of the points. An example of when they would be used is if the signaller is protecting some track workers. A reminder would be placed on any signals which if showing proceed could allow a train to travel over the protected portion of track.

Block Bell

The major facilitating technology for lever frame boxes was the electric telegraph. This allowed the signaller at the start of a section to communicate with a signaller at the end to check whether trains had passed through the section successfully. A typical block bell is shown in Figure C- 11. Signallers use coded messages similar to Morse code to communicate train movements and infrastructure state with each other.

Figure C- 11: Block Bell

Figure C- 12: Block Instrument

Block Instrument

A typical block instrument is shown in Figure C- 12. Block instruments show status of a section of railway. It typically has 3 settings:

- Normal – no trains in the control area;
- Train on Line – a train or other obstruction is present in the section;
- Train Accepted – the line is clear for a train to proceed.

The signalling system will not allow the signaller to set a route until the relevant block instrument is set to 'Train Accepted'. When a train is present on a piece of line the block instrument must be set to 'Train on Line'

Interlocking

The term interlocking refers to the systems developed to ensure that conflicting routes are not set. This is achieved by connecting the signalling equipment together in such a way that the levers can only be operated in a certain order. The interlocking used in lever frame boxes is mechanical in nature. It works by metal bars attached to the levers in the signal box. When a lever is pulled the metal bar is positioned so that it blocks other levers which, if pulled or released, would endanger that route. The interlocking is designed using logic tables to describe the releases and locks associated with each lever.

Absolute Block

Lever frame boxes work under a principle known as Absolute Block (AB). This principle states that only one train may be in one section of railway on one line at one time. This principle forms the basis of all signalling systems.

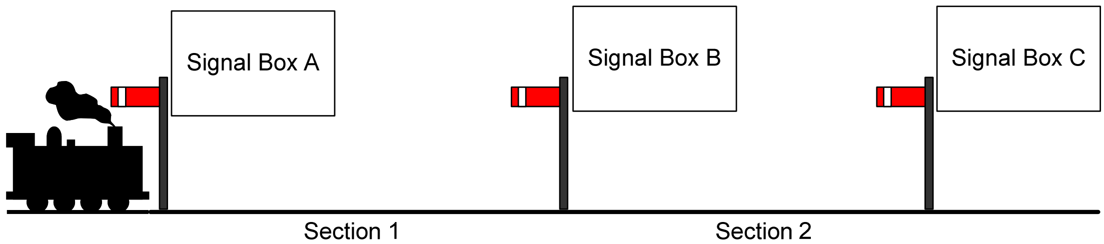


Figure C- 13: Absolute Block Signalling

Signallers receive information regarding approaching trains in the form of a request via the block bell from the preceding signaller. The diagram in Figure C- 13 can be used to describe a very simplified version of absolute block during normal operations. In this example, Signaller A would contact Signaller B to request a route for the train. If the line is clear Signaller B turns his block instrument to “Train Accepted” and this releases the interlocking for Signaller A to set and clear the route through section 1. Once the train has passed the signal at the entrance to section 1 Signaller A must set his block instrument to “Train on Line”. Signaller B then contacts Signaller C via the block bell to request the further route. Once Signaller C has set his block instrument to “Train Accepted” Signaller B can set and clear the route through section 2. Once the train has passed the entrance signal to Section 2 signaller B contacts signaller A to let him know that the train has left Section 1. Signaller A then turns his block instrument to “Normal”.

Entry-Exit Panels

The lever frame form of signalling dominated until the 1950s when Entry-Exit (NX) panels were developed (Figure C- 14). NX panels were a major leap from lever frames. They fundamentally changed the interface signallers use to control trains and enabled much larger areas of the railway to be controlled by each signaller. The development of NX panels was made possible by the use of track circuits for train detection.

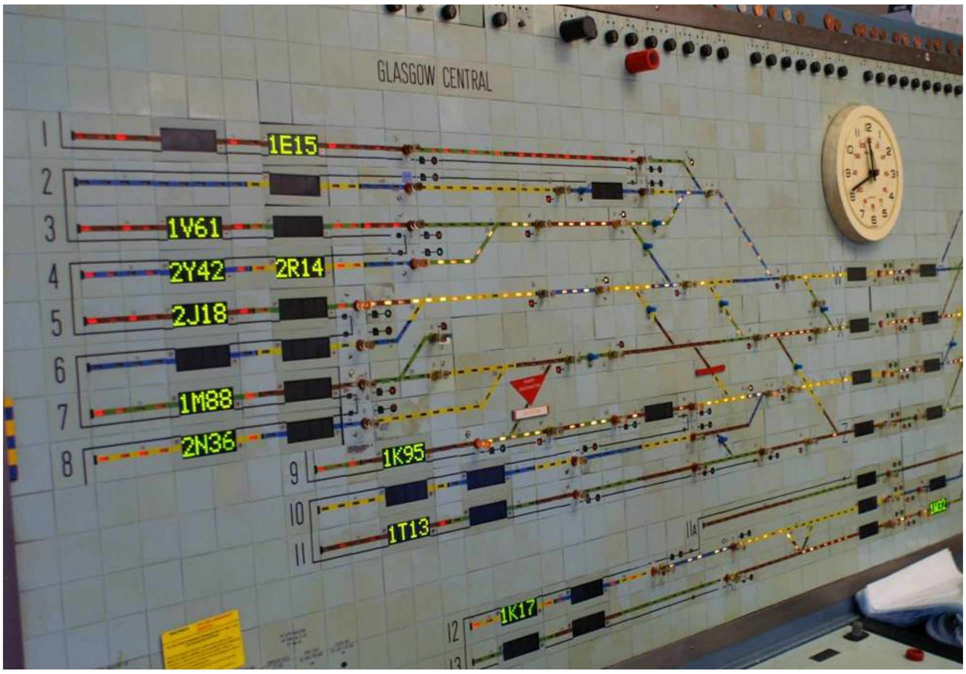


Figure C- 14: NX Panel

<u>Technologies</u>

Colour Light Signals

Colour light signals operate on the similar principles to traffic lights. A red means stop, yellow indicates that the next signal will be red and the driver should be prepared to stop at it (serving the same function of a distant signal under AB principles), and a green signal means proceed. These are used in two aspect (red and green) signalling and three aspect signalling. High density railways use four aspect signalling which includes a double yellow aspect, indicating that the next signal will be at yellow (Figure C- 15). This facilitates shorter block sections allowing trains to travel closer together.

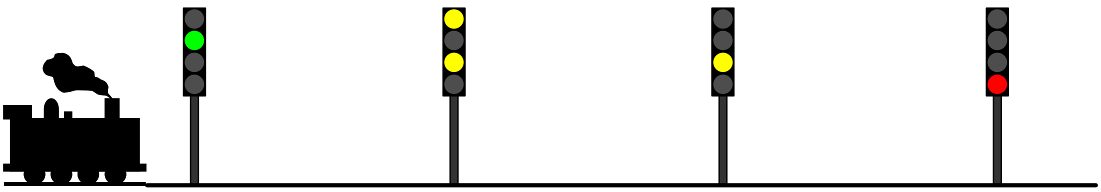


Figure C- 15: Four Aspect Signalling

Route Indicators

Route indicators are normally in the form of five white lights over signals and are used to inform drivers which route has been set for them. For simple junctions with only two routes forward the signal would be as shown in Figure C- 16. If the route set is straight on the five white lights above the signal would not light up, but if they are lit up it indicates to the driver that the route to the right is set.

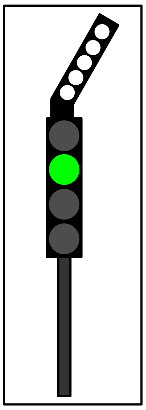


Figure C- 16: Simple Junction Indicator

Signals can have up to seven route indicators above a signal to indicate route set to drivers at complex junctions. If there are more than seven possible routes from a signal then each route is given a number and theatre lights are used to indicate which route is set (Figure C- 17).

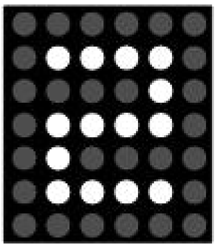


Figure C- 17: Theatre Lights

Track Circuits

Track circuits are used to detect the presence of a train on a particular section. These operate by running a small current through the rails operating a relay. When a train is present the electricity runs through the wheels of the trains and the track circuit relay is interrupted. Once the train leaves the section the track circuit relay operates again. When connected to a display system this development meant that signallers no longer had to physically see a train to know its position and thus the number of trains a signaller could keep track of was increased. In this sense, track circuits were fundamental in facilitating the development of new signalling control systems. Track circuits fail safe in the majority of circumstances as any interruption to the power supply would result in the track circuit showing occupied.

Train Describer

Train describers (TD) are identifiers for individual trains on the rail network. They take the form of a 4 digit alpha-numeric code. The first digit indicates the class of train, the second character is a letter indicating the route of the train, and the final two digits are simply used to distinguish that train from others following the same route on the same day. Thus, 2J18 would be a Class 2 train perhaps travelling between Glasgow and Edinburgh and would be the 18th train to follow such a route that day. These TDs are displayed to the signaller on the panel (Figure C- 18) and pass automatically between signal boxes.

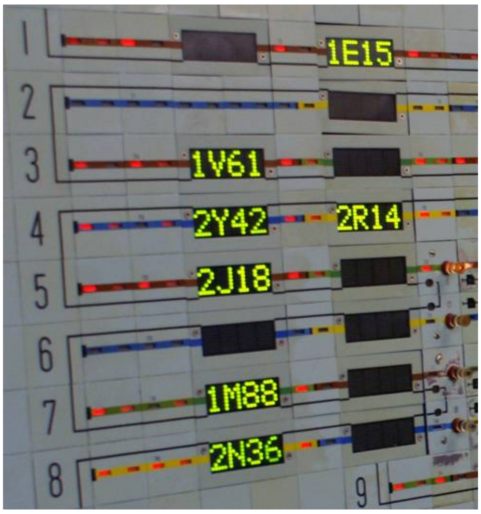


Figure C- 18: Train Identifiers Displayed on a Panel

Reminders

Reminders on NX Panels are similar to those of lever frames; they take the form of a small plastic collar which fits over the buttons on the panel preventing signallers from pressing that button.

Interlocking

NX Panels use Route Relay Interlocking (RRI) instead of mechanical interlocking. The principles of logic and the purpose are the same but the data are held in electrical circuitry rather than mechanical bars.

TORR

Train Operated Route Release (TORR) uses track circuit occupation to determine when a train has left a particular section. It then automatically releases the route that was set over that section, freeing the route to be set for another train. TORR can be regarded as an early form of automation freeing the signaller from the menial task of clearing routes after trains. Along with track circuits TORR was a major facilitating technology for future highly automated systems, such as ARS.

Track Circuit Block

Under AB the number of trains running over a line was limited by the number of signal boxes on that line as only one train could be in a section at a time and sections typically stretch between signal boxes. As demand increased it was desirable to have trains running closer together.

To achieve this 'Track Circuit Block' (TCB) was developed. TCB uses track circuits to determine the location of a train. The track circuit block system permits a signal to show a proceed aspect when:

- All track circuits, up to and including the overlap of the next stop signal, are clear, and
- All necessary points within the route are detected in the correct position for a train to pass safely

This means that a route for a train can only be set if the track circuits show that route to be clear and the points are proved to be in the correct position. Overlaps are a safety margin so that if a train goes slightly beyond a stop signal the route is guaranteed to be clear. As the information on train position is gathered by track circuits rather than signallers TCB allows one signaller to keep track of more trains. Another important facilitating technology was the mechanisation of points and the introduction of colour light signals. This meant that signallers did not have to be located close to these entities to operate them. TCB tends to be associated with colour light signalling although there is no reason why mechanised semaphore signals cannot be used. Track circuit block is also the method of signalling used in IECC signal boxes.

On NX panels the area of control is represented on the panel and instead of manually controlling each signal section and set of points using levers the signaller sets routes using buttons on the panel (Figure C- 19). The signal aspects and points appropriate to that route are changed automatically. Train detection is provided by track circuits and train positions are indicated by red lights on the panel. The development of this technology allowed one signaller to control a much greater geographical area.

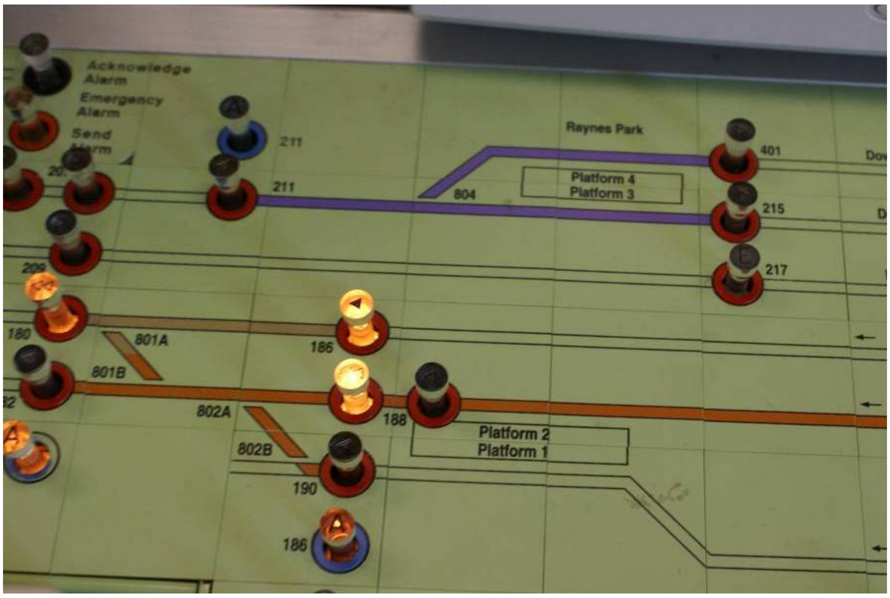


Figure C- 19: NX Panel Buttons

The signaller can see the TD of any approaching trains displayed in an approach berth on the panel. This prompts them to set a route for the train using the buttons on the panel.

Track circuits and train describers eliminated the need for signallers to physically communicate the movement of trains and so the block bell and instrument could be done away with (except if the TD does not automatically transfer to the next signal box). Signallers have an emergency alarm button on the panel which sends an alarm to adjacent signallers warning them of an emergency situation. Other communications when necessary are achieved by telephone.

VDU Signalling

In the 1980s the concept of NX panels was moved onto VDU screens and ARS was introduced in IECC boxes. IECC refers to the whole system used in those signal boxes, including the display systems and interlockings. ARS is only one part of the IECC system. There are other types of signal box which use VDU technology but only IECCs currently have ARS.

Figure C- 20: IECC Signal Box

Figure C- 20 shows an IECC signaller working. Instead of a panel showing the control area, signallers working with VDU screens have their entire control area displayed on up to two overview screens, typically leaving 2 screens to call up detailed pieces of those overviews. Signallers use a trackerball and keyboard to interface with the signalling screens. Most basic commands can be done with either the trackerball or the keyboard, although some less common commands require the keyboard alone or a combination of the trackerball and keyboard.

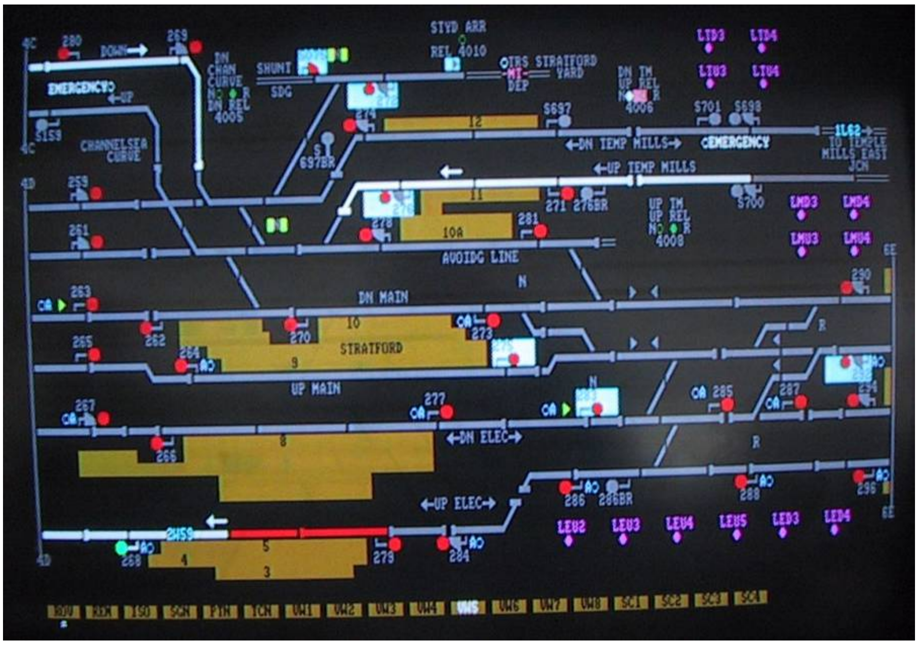


Figure C- 21: IECC Screen View

Figure C- 21 shows a typical overview screen. The indications are as follows:

- Orange blocks are station platforms.
- Grey lines are tracks.
- White lines indicate a route has been set over that portion of line.
- Signals are indicated by a round dot showing red, yellow or green depending on the aspect of the signal at the time. If the aspect is double yellow two yellow dots are shown.
- Blue dots represent ARS sub-areas.
- Blue or pink squares over signal heads (round dots) indicate a reminder has been applied to that signal.

The soft buttons along the bottom of the screen include some controls such as reminder appliances and also allow the signaller to choose a different screen view.

The method of signalling is identical to NX panels; again using TCB and the interlocking is very similar although it tends to be either Solid State Interlocking (SSI) or Computer Based Interlocking (CBI) both of which are software based interlockings. RRIs can still be used, but are less common.

Other Signalling Systems

The signalling systems described above are the most commonly used. There are other principles of signalling (e.g. Electric Token Block and Tokenless Block) and types of signalling system (e.g. One Common Switch (OCS) and Miniature Lever Frame) as well as hybrid systems combining two or more types of signalling principles or systems on the railway but they have not been covered here as their implementation is relatively limited.

Signalling Task

Aside from physical operation of the signalling equipment, the signalling task requires the signaller to route trains according to the timetable. This is a straightforward goal to achieve while the railway is running smoothly but if disruption occurs it can become considerably more complicated. There are many causes of disruption to the railway, including train failures, trespassers and vandals, track circuit failures, or simply late running trains. Disruption is an impediment to routing trains on time. In this case, signallers are called upon to ‘regulate’ the service.

Regulation

Although regulation of trains is frequently referred to within the rail environment there is no standard definition of what regulation entails. Discussions with Subject Matter Experts resulted in the following definition for regulation:

> "The planning and implementation of train paths over the available infrastructure in order to optimise the train service, mitigate the effects of disruption, and support recovery from disruption."

Signallers take a variety of factors into account when making regulating decisions. These include:

- The class of train - Each train in the timetable has a train class associated with it. For example, passenger trains are classes 1 and 2, with 1 being express passenger trains and 2 being ordinary class, freight trains are classes 3, 4, and 6, and empty coaching stock is class 5. In the past the lower the number of the class of train the higher priority that train had, so express passenger trains had priority over freight trains for example. Since privatisation and the introduction of PPM the class priority system has become obsolete but many signallers still use it as a rule of thumb to determine which train to route first;
- Next possible passing location – how soon can one train pass another;
- Train stopping patterns – if one train stops at all stations and another is an express it may be best to put the express train first, even if this means stopping the first train until the express can pass through the junction;
- Train speed – different trains have different running speeds and this will affect the signallers' decision;
- Delay already accumulated by trains – it may be preferable to stop an already delayed train and cause it further delay than to delay a train which would otherwise be on time;
- Route (or platform) availability – if a portion of the forward route or platform is not available for a train it may not be advisable to route it forward as this may block other trains from proceeding;
- Experience – signallers learn from previous occasions and may base routing decisions on avoiding situations which have caused problems in the past;

- Delay attribution and PPM – signallers are required to signal trains to reduce PPM, but with the way delay is attributed signallers may be inclined to signal trains to reduce the likelihood that a delay will be attributed to them regardless of whether this is in the best interest of the whole network.

As regulation involves the ordering of trains regulating decisions can only be implemented at regulating points (i.e. junctions or crossovers). Junctions are where two or more routes converge or diverge (depending on the direction of travel). Crossovers allow trains to cross between parallel lines. Plain line is a section of track with no crossovers. There are no regulating decisions to be made on these sections of track. These three types of track infrastructure are illustrated in Figure C- 22.

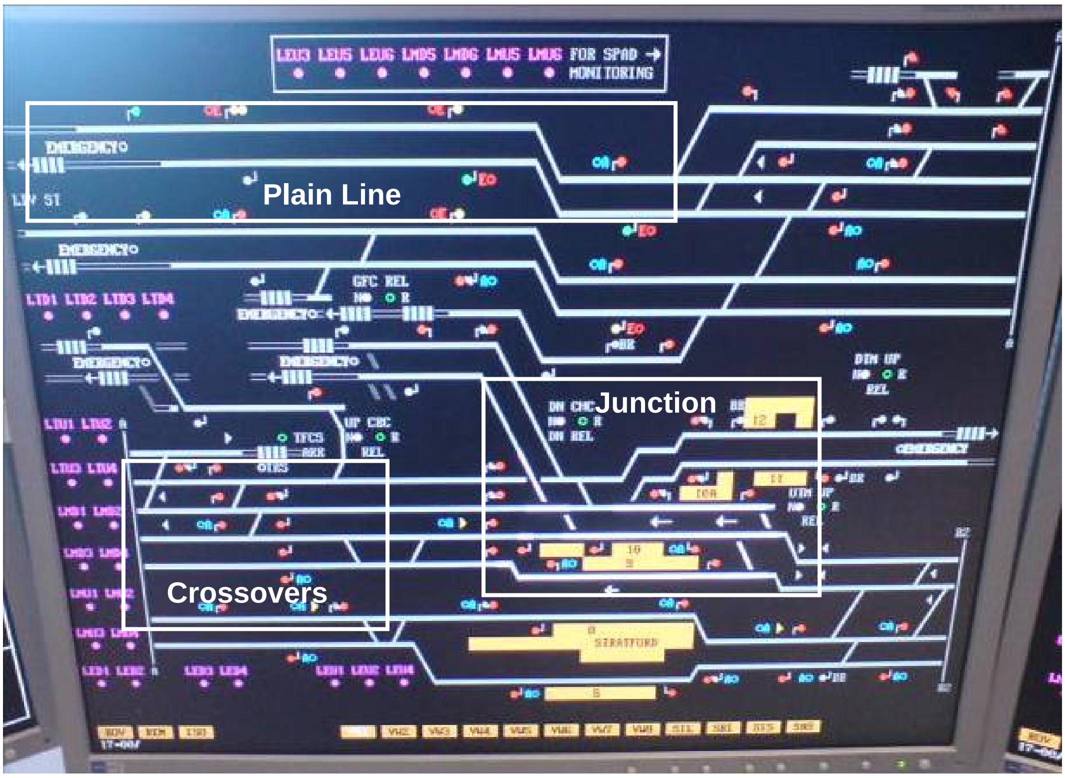


Figure C- 22: Junctions, Crossovers, and Plain Line

The signaller's objective is to minimise delay to trains over his/her patch of railway, although signallers should also take into account the potential effects of their decisions on other signallers in each train's route. Of course there is a limit to how far down the line these effects can be taken into account. There have been attempts in the past to develop an advanced automation system which would maximise the routing over trains over the entire rail network, but this proved extremely complex and too much data even for a computer to handle. The practicality of making routing decisions for trains in London on the basis of optimising them in Scotland must also be considered. As the railway is a dynamic system with numerous interfaces to the rest of the world there is no guarantee that the system will not look totally different by the time the train reaches Scotland. Nevertheless, experience may tell signallers that certain situations create problems further down the line and a good signaller may attempt to avoid these.

Railway Failures

Signallers must routinely deal with failures of parts of the rail system or disruption caused to the railway. The fact that the railway has so many interfaces to the rest of the world makes it difficult to insulate against failures.
Typical failures or disruption are outlined below:

TC Failure

Track circuits (TC) are not highly reliable and fail relatively often. As they are fail safe a track circuit failure means that a track section shows up occupied and the interlocking will therefore prevent any attempts to set a route through the affected area. Until the fault is rectified signallers must use verbal procedures to give drivers permission to pass through the affected section. Drivers may only travel at very slow speeds in this case and together with the additional time taken for the verbal communications this means a TC failure greatly reduces the number of trains that can travel over the affected area.

Line blockage

Lines may become blocked for a number of reasons, for example, train failures, points failures, accidents or incidents, trespassers, emergency engineering work. Any line blockage means that approaching trains may become stuck or may need to be re-routed around the blockage.

SPAD

Signals passed at danger (SPAD) are one of the most serious incidents on the railway. These occur when a train fails to stop at a red signal. This may be merely because the train has slightly overrun and occupied the track circuit in advance of a signal, or because the train driver has not realised that the signal is at red. If a train passes a red signal all trains in the area must be stopped.

Weather

Weather can cause problems for the railway with fog reducing visibility of signals so trains may have to travel at slower speeds, leaves on the track in autumn can prevent track circuits from proving, and ice reduces the traction on the rails.

Permissive Working

Permissive working allows more than one train to be in a signal section at one time. It may only be used in specific circumstances, such as for joining movements at platforms. Ground position lights (GPL) give the train driver permission to proceed into the section (Figure C- 23 & Figure C- 24), but they must do so at a very slow speed that will allow them to stop upon seeing an obstruction.

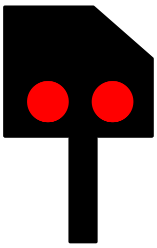


Figure C- 23: Ground Position Light Showing Stop Aspect

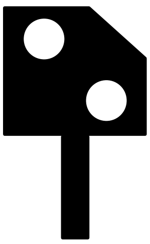


Figure C- 24: Ground Position Light Showing Proceed Aspect

ARS

Automatic Route Setting (ARS) has been in place since the late 1980s, and was first introduced in Liverpool Street Integrated Electronic Control Centre (IECC). DeltaRail who now develop ARS state that:

> “ARS optimally routes trains using timetable data, current train positions and an internal representation of the rail network. It can handle severely disrupted

service patterns and assist the signaller in the event of train or infrastructure failures.”

ARS has access to the central timetable services database (TSDB) and each day downloads the timetable for all the trains in the area it controls. It then uses codes from the timetable to determine the route and timings for each train. As each train enters the control area ARS automatically sets the route ahead of the train. ARS also incorporates algorithms to compare trains on the workstation to decide which to route first. Less advanced forms of route setting automation could either route strictly according to the timetable or operate on a first come first serve basis, but ARS attempts to regulate.

How ARS works

The information contained in this section has been gathered from visits to signal boxes, relevant Network Rail standards, guidance materials issued by the manufacturers of ARS, and discussions with signalling SMEs and senior signalling engineers. Even then it was not possible to obtain a clear picture of how ARS works and so a preliminary model was developed to guide discussions. The knowledge available within Network Rail was not sufficient to validate the model and so it proved necessary to meet with the engineers who originally designed the system. The validated model is shown and explained in this section.

Figure C- 25 shows the internal processes of ARS at a high level. When a train enters the control area ARS recognises the track circuit occupation, reads the TD of the approaching train and uses data from the TSDB to generate a list of the routes, or path segments, required by that train. It then compares this list to those generated for other trains in the control area and identifies which trains potentially conflict. There are three ways a train may conflict; they may travel over the same section of track in the same direction, travel over the same section of track in opposite directions, or travel over lines which cross. Trains with no potential conflicts are discarded, that is they are not again considered with respect to the routing of the new train. The new train can then enter into the cyclical processing of ARS.

Every 10 seconds ARS considers whether each of the trains in the control area requires a route to be set (ARS attempts to ensure that there are 2 green signals in front of every train where possible). If no routes are required the ARS will consider again in 10 seconds. If a route is required for a train ARS compares the train requiring a route with each of the trains identified earlier as potentially conflicting with it. At this stage it uses a set of parameters such as train class and priorities, current train delay, and predictive forward movements from the timetable to determine the weighted delay for putting each train first in this situation. If the calculations show that the train requiring the route has the least weighted delay in all the pairwise comparisons then ARS will check if the route required is available. If the route is available it will then request the route for that train. If the weighted delay is lower for another train then ARS will not request the route. The whole process is repeated every 10 seconds. ARS uses a complex set of parameters in these algorithms and these vary for each location on each workstation. This means that a bespoke ARS system must be designed for each workstation.

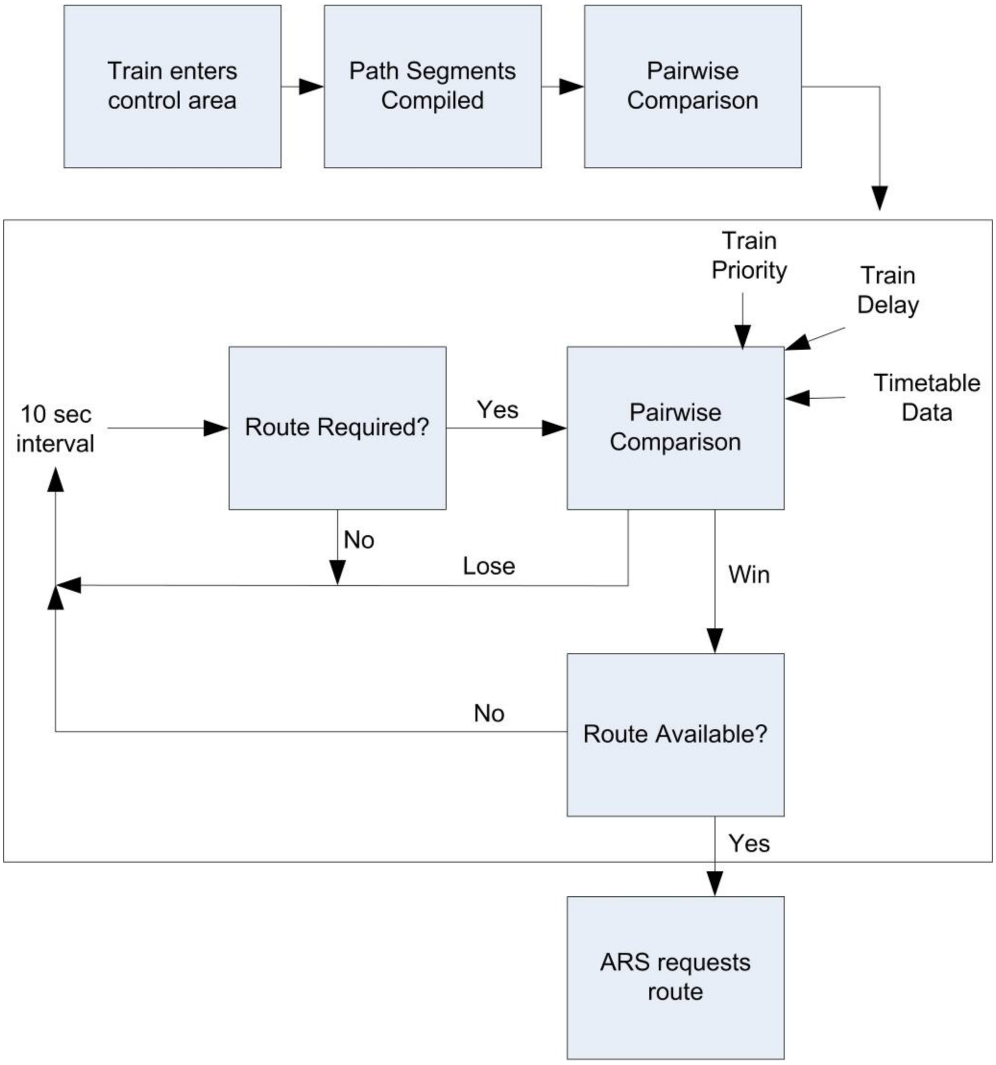


Figure C- 25: ARS Processes

The signaller has no insight into this process. The signalling screens only display when routes are set by ARS and although there is an ability to query ARS through the general purpose (GP) screen this information is not always informative to the signaller, particularly if they do not fully understand the processes ARS uses to make its decisions. It is impossible for ARS to give information on what it is planning to do as it does not make any decisions until it has to.

Not all trains are in ARS; this is most likely to be because there is no timetable or an incomplete timetable for them in the database. Trains which are not in ARS are shown in pink and must either be routed manually or put into a Special Timing Pattern (STP). STPs are pre-programmed routes over the workstation which can be applied to individual trains. If an STP is applied to a train the train is coloured brown and ARS will route the train according to the route instructions in the STP. Trains in an STP are given priority over all other traffic, which may not be practical as freight trains are the most common to require STPs and these are less likely to be the priority on the workstation. Signallers may also choose to take trains out of ARS. This allows signallers to maintain control over that train as it must be routed manually, although it can be put back in to ARS if the signaller wishes.

Reminder appliances exist for ARS as they do for signallers. If a reminder appliance is placed over a signal by the signaller ARS will not be able to call a route to or from that signal. The reminder also serves its traditional function of reminding the signaller not to call that route. Although intended as a safety device, reminders are frequently used by IECC signallers to control ARS as it is a direct and easy way to inhibit route calling.

Use of ARS

ARS was intended to be “as efficient as a good signalman” (Burrage et al., 1991) and with respect to routing trains according to the timetable this can be said to be the case. However, ARS is often referred to as “deaf and dumb” as it does not receive all the information the signaller does, particularly from voice communications, and the feedback from the system is relatively poor. In practice this means that ARS cannot cope effectively with failure situations and restriction of infrastructure and this falls upon the signaller.

Each workstation is divided into a number of "sub-areas" for the purposes of ARS control. Each of these sub-areas may be disabled by the signaller allowing him to control all trains in that area manually. Where localised problems exist, the signaller may disable the relevant sub-area and ARS may be able to provide benefit by keeping other areas of the workstation running while the signaller focuses on the failure.

The principles behind ARS allow the signaller to remain in charge. He can manually set routes for trains, restrict ARS working in an area (disable the sub-area), and take individual trains out of ARS control. As mentioned above, ARS also incorporates a function which allows the signaller to query ARS routing decisions and timetables for trains. The reply to these queries and information on alarms appears on the general purpose (GP) screen usually located to the far left of the signalling screens (Figure C-26). Typical responses to queries in ARS may be:

> "Sufficient route in advance; R1053C is last route found"
>
> "Train is currently routed off planned path: R7777 is last route found"
>
> "Train is not ARS controlled"

Figure C- 26: General Purpose Screen

Information on trains approaching the area of control is also displayed on the GP screen. The engineers designing ARS anticipated that signaller vigilance would fall when ARS was introduced and they introduced alarm systems to counter this effect (Burrage et al., 1991). Thus ARS produces alarms for a large number of events including unexpected occupation of a track circuit, train entering workstation, non-ARS trains, and lengthy occupation of a track circuit. In addition a large number of alarms are presented which are not directly relevant to the signaller but are more directed towards technicians, such as system response timing alarms. Apart from the more recently introduced SPAD alarms the audible tone for all alarms on the workstation is identical. This means that in practice signallers are bombarded with alarms and apart from periodically cancelling them they pay relatively little attention to them.

Any change to the algorithms or base data that ARS uses in its decisions requires a major data change. In 1997 when British Rail was broken up the research wing which had been responsible for the development, implementation and maintenance of ARS was also privatised and was bought by AEA Technologies. It has since been sold again and is now called DeltaRail. Therefore after privatisation Railtrack, and subsequently Network Rail, were required to pay for support for ARS which had formerly been free.

Summary

This appendix has given details on the signalling systems and technologies predominant in the UK railway. This information is essential in understanding the role automation currently occupies in the signalling environment.

APPENDIX D: VALIDATION OF PRINCIPLES OF AUTOMATION

A paired comparisons exercise was undertaken with 22 human factors professionals to validate the principles. Ten of the participants work in Network Rail Ergonomics team and the remaining 12 work in the Human Factors Research Group at Nottingham University. A PowerPoint presentation was prepared in which each principle was presented with each of the other principles. Participants were given an instruction sheet and a briefing sheet giving an explanation of each principle. They were then asked to work their way through the PowerPoint presentation. Each slide stated "An automated system should be…" and two principles were shown. Participants were required to choose one principle over the other to complete the sentence. A researcher noted their answers.

The responses were entered into a spreadsheet and analysed using the paired comparisons technique. A comparison matrix was generated detailing the number of times each principle was chosen over each of the other principles. From this, the probability for each principle was calculated and hence z-scores. The average z-score was calculated for each principle and they were plotted along a line, as shown in Figure D- 1.

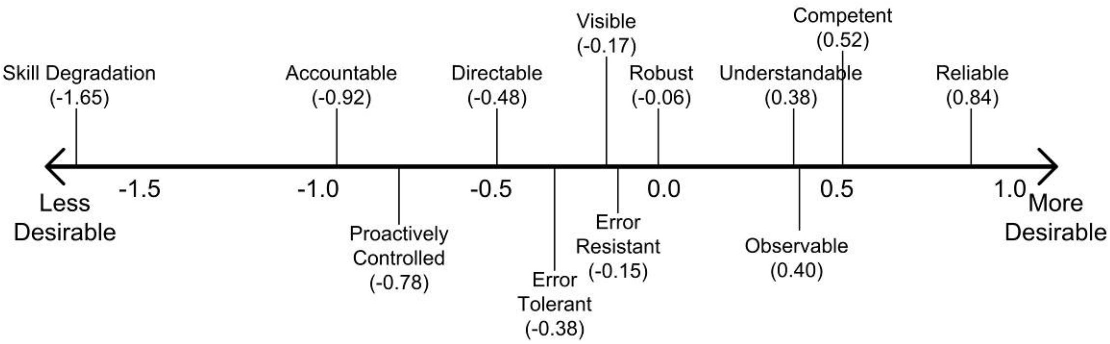


Figure D- 1: Scaling of Principles of Automation

The results suggest reliability is the most desirable principle and competence, observability and understandability also all score towards the top of the scale. The only principle which stands out at the negative end of the scale is skill degradation, suggesting perhaps it warrants removal. However, as the reliability and competence of an automated system increase the operator's requirement to intervene reduces and skill degradation becomes more likely and potentially more important as a principle. For this reason this principle was kept.

APPENDIX E: TRUST QUESTIONNAIRE

Below is a list of statements for evaluating trust between people and automation. Please circle the number which best describes your feeling or impression for each statement.

Note: 1=Strongly Disagree, 5=Strongly Agree

1. ARS is always available for use (Reliability)
   1 2 3 4 5

2. ARS is capable of performing under a variety of different circumstances (Robustness)
   1 2 3 4 5

3. It is easy to understand what ARS does (Understandability)
   1 2 3 4 5

4. ARS is capable of signalling trains as competently as a signaller (Competence)
   1 2 3 4 5

5. ARS gives explicit information on its intended actions (Explication of intention)
   1 2 3 4 5

6. I can count on ARS to do its job (Dependability)
   1 2 3 4 5

7. I have a personal preference for using ARS (Personal Attachment)
   1 2 3 4 5

8. I can predict what ARS will do from moment to moment (Predictability)
   1 2 3 4 5

9. If ARS makes a routing decision which I am uncertain about I have confidence that ARS is correct (Faith)
   1 2 3 4 5

10. I understand how ARS works (Understandability)
    1 2 3 4 5

11. ARS performs well under normal running conditions (Competence)
    1 2 3 4 5

12. ARS is very unpredictable, I never know what it is going to do (Predictability)
    1 2 3 4 5

13. I can rely on ARS to function as it is supposed to (Reliability)
    1 2 3 4 5

14. Even if I have no reason to expect that ARS will be able to deal with a situation, I still feel certain that it will (Faith)

    1 2 3 4 5

15. I understand why ARS makes the decisions it does (Understandability)

    1 2 3 4 5

16. ARS performs well under disturbed conditions (Competence)

    1 2 3 4 5

17. ARS is very consistent (Predictability)

    1 2 3 4 5

18. ARS will always make the same routing decision under the same circumstances (Reliability)

    1 2 3 4 5

19. I trust ARS

    1 2 3 4 5

APPENDIX F: ODEC SCORES

Entity	Guidance High	Guidance Med	Guidance Low	Ashford N Kent 1	Ashford w/s 4	Liverpool Street Shenfield	Liverpool Street Ilford	York York South	York Leeds East	Tyneside Newcastle	Tyneside Darlington
Stations: Terminus Platforms	9+	5-8	1-4	14	4	12	10	11	17	12	4
Junctions/regulating points	4+	3-2	1	12	5	9	13	4	5	3	18
Connections	35+	12-34	0-11	51	73	76	44	10	39	45	38
Level crossings non CCTV	9+	2-8	1	0	2	15	1	9	9	2	2
LX CCTV & MCB	4+	2-3	1	0	0	1	0	0	0	2	1
Single lines	4+	2-3	1	3	1	10	5	0	1	2	1
Controlled signals	75+	26-74	0-25	75	100	79	96	90	121	85	77
Automatic working signals	20+	10-19	1-9	100	61	96	46	10	4	11	81
Permissive Working	6+	3-5	1-2	1	4	9	0	11	18	12	4
Block system	3+	2	1	1	1	1	2	1	1	2	1
Special Box Instructions	15+	7-14	1-6	27	23	65	65	45		24	31
Non-timetabled trains: VSTP	40+	11-39	0-10	31	10	30	30		15	12	12
No. of Points traversed (normal)	10+	4-9	0-3	15	26	9	9	19	26	25	22
No. of Points traversed (complex)	11+	5-10	0-4	20	26	25	24	28	26	32	27
Ground frames	5+	4-2	1	0	0	5	0	12	0	24	21
Incidents & occurrences	10+	5-9	1-4	56	15	35	32	10	21	12	24
Infrastructure Failures	6+	3-5	0-2	28	15	28	26	10	25	24	24
PLODs	3+	2	1	0	0	0	0	1	0	0	0
HABDs	5+	4-2	1	0	0	2	0	5	0	1	3
No. LX Phone call requests (	20+	10-19	1-9	0	3	2	0	2	30	1	0
No. LX Phone call requests (	40+	20-39	1-19	0	4	10	0	6	60	6	1
Emergency/unplanned T3s	2+	1	0	8	1	1	1	1	1	1	1
Isolations (emergency or planned)	3+	2	1	3	3	4	4	8	1	1	1
Depots/Yards/Sidings	20+	10-19	1-9	150	94	48	68	20	100	60	1
Station Operations	10+	5-9	1-4	2	14	20	8	45	40	1	2
No. of Trains (max per hour)	45+	16-44	0-15	78	20	87	85	>45	60	24	40
No. of Trains (max per day)	395+	151-394	0-150	900	410	1140	1120	>400	1150	493	385
No. of delays (	68 - 100%	34 -67%	0-33%	5%	1%	3%	3%	<33%	5%	5%	5%
Min and Max class speed	60+	21-59	0-20	35	45	65	60		75	105	115

APPENDIX G: OBSERVATION STUDY CONSENT FORM

This study is designed to observe signallers interacting with the signalling system to determine individual differences in strategies.

Participation in this study is voluntary and there will be no consequences if you choose not to participate. Any information obtained from the research will be anonymous and will not be used for any purposes other than this study. The data obtained will be used to help design automation in the future.

The method used in this study will be real time data collection by a researcher sitting near the signaller. Written records will be made of certain behaviours, such as using the trackerball or looking at the simplifier. No recording devices will be used. The length of time for the observations will be 90 minutes.

Consent

1. I confirm that I have read and understand the above information and have had the opportunity to ask questions.

2. I understand that my participation is voluntary and that I am free to withdraw at any time.

3. I agree to take part in the above study.

______________________________	____________	______________________
Name of Participant	Date	Signature
______________________________	____________	______________________
Researcher	Date	Signature

APPENDIX H: OBSERVATION STUDY SAMPLE DATA

Time	Code	Notes
33:45	|	
33:50	|	
33:55	|	
34:00	CCTV	
34:05	PS	
34:10	MA	
34:15		
34:20	CS	Move a train?
34:25	|	coming in late
34:30	|	
34:35	MP	
34:40	CS	
34:45		
34:50	CT	on
34:55	|	Driver
35:00	|	wait for signal
35:05	|	
35:10	CS	
35:15	MA	
35:20	PS	
35:25		
35:30	MA	
35:35	TB	SR
35:40		
35:45	MA	
35:50	PS	
35:55		
36:00		"Proactive signaller"
36:05	MA	
36:10		platform changes
36:15		
36:20	TB	
36:25	PS	
36:30		
36:35	TB	SR
36:40	|	
36:45	|	
36:50	MA	
36:55	|	
37:00	|	
37:05	|	
37:10	|	
37:15	MP	
37:20		
37:25		
37:30	Q	Chat
37:35	MP	
37:40		
37:45	Q	
37:50	MP	
37:55		
38:00	Q	
38:05	MP	
38:10	Q	
38:15		

4

APPENDIX I: CONDITIONS DURING OBSERVATIONS

	Running problems	*% Trains in ARS (approx)*	*Trains taken out by signaller*	*Experience of signaller*	*Main reason for intervention*	*Main method of intervention*
Obs 1	Late running from earlier incident	>80%			Regulation	
Obs 2	None	>90%		21 years	Permissive Working	
Obs 3	None	90-95%	10%		Regulation	Turned off sub-areas Manual route setting Reminders
Obs 4	One minor mistake	95%		15 years	Regulation	Manual route setting
Obs 5	None	100%	1		Regulation	Reminders Manual route setting
Obs 6	None	>90%	0			Manual route setting Reminder STP
Obs 7	None				Regulation	Reminders
Obs 8	Line blockage				Route trains around blockage	Manual route setting
Obs 9	Dodgy track circuit	100%			Routing trains over track circuit Regulation	Manual route setting
Obs 10	None	100%	0		Boredom Regulation Set routes into sidings	Reminders Set manual routes
Obs 11	Congestion		1		Regulation	Reminders
Obs 12	1 failed train – no major effect	100%	0		Letting cars out of depot	
Obs 13	None	>90%	0	23 years	Regulating	Reminders Manual route setting
Obs 14	None	95%	1 – late running	17 years	Late running trains	1 sub-area turned off 1 reminder applied > manual route setting
Obs 15	Minor late running	100%	0	16 years	Regulate late running	Turn sub-areas off Manual route setting
Obs 16	None	98%	3	5.5 years	Regulation Platform Sharing	Take trains out of ARS Reminders

	Running problems	*% Trains in ARS (approx)*	*Trains taken out by signaller*	*Experience of signaller*	*Main reason for intervention*	*Main method of intervention*
Obs 17	1 broken train – no major effect	0%	n/a	7 years	Running workstation manually	Manual route setting
Obs 18	None	90%	0	17 years	Regulation Put trains into sidings	Manual route setting Reminders
Obs 19	Minor late running	90%	4	6 months	Regulation	STP Manual route setting Reminders
Obs 20	Some platform changes	100%	8/9 (freight)	>16 years	Regulation	Route set Reminders
Obs 21	None	90%	1	31 years	Stop ARS pulling routes	Turn off sub-areas Reminders
Obs 22	None	90%	4	13 years	Regulation	
Obs 23	None	99%	3	23 years	Regulation	
Obs 24	Minor late running	99.9%	2	23 years	Stop ARS making a mistake	Train out of ARS Turn sub area off

APPENDIX J: SMOOTHNESS OF OBSERVATION DATA

York South Workstation

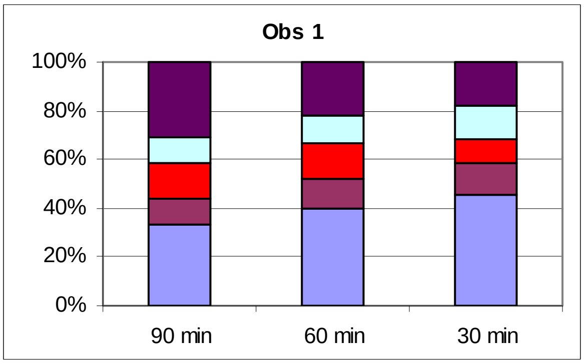


Figure J- 1: Smoothness of Data for Obs 1

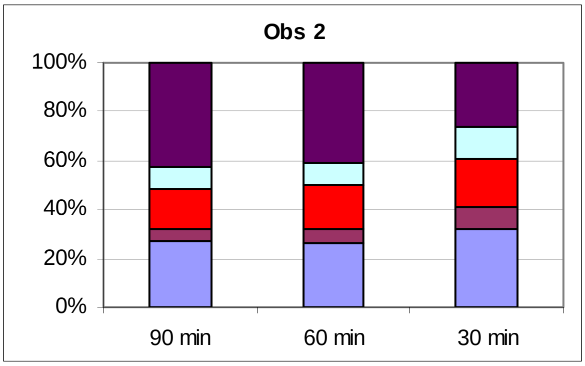


Figure J- 2: Smoothness of Data for Obs 2

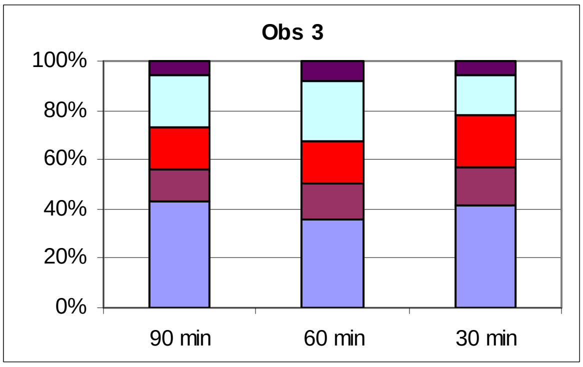


Figure J- 3: Smoothness of Data for Obs 3

Leeds East Workstation

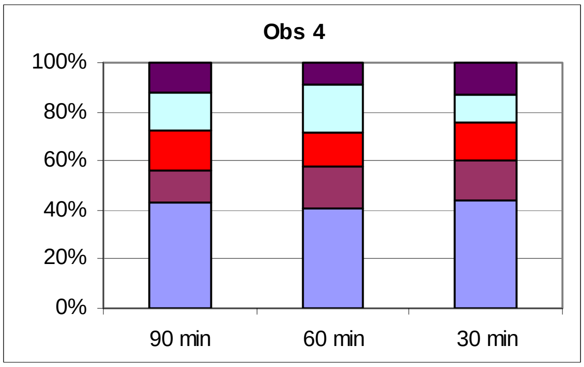


Figure J- 4: Smoothness of Data for Obs 4

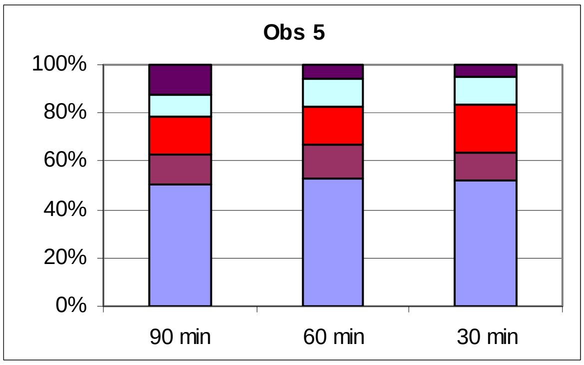


Figure J- 5: Smoothness of Data for Obs 5

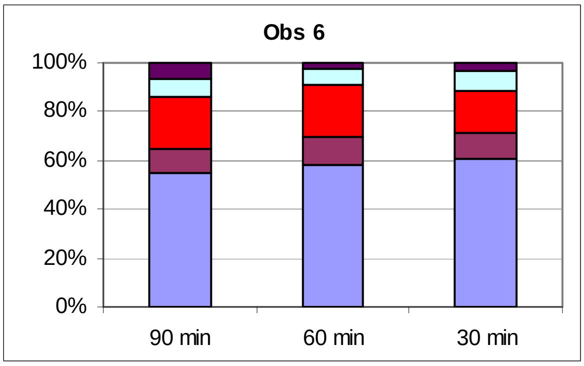


Figure J- 6: Smoothness of Data for Obs 6

Shenfield Workstation

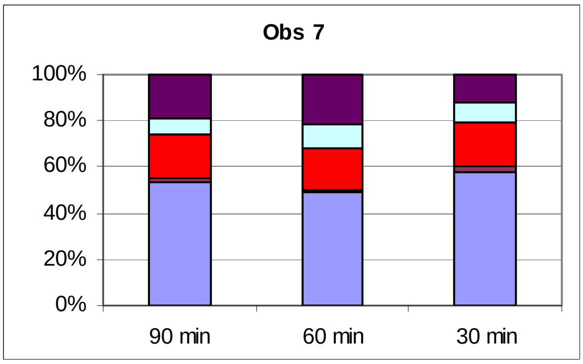


Figure J- 7: Smoothness of Data for Obs 7

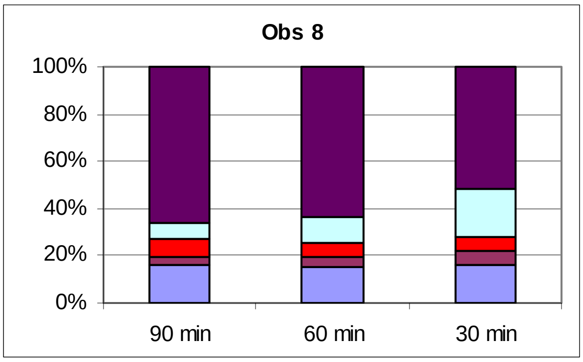


Figure J- 8: Smoothness of Data for Obs 8

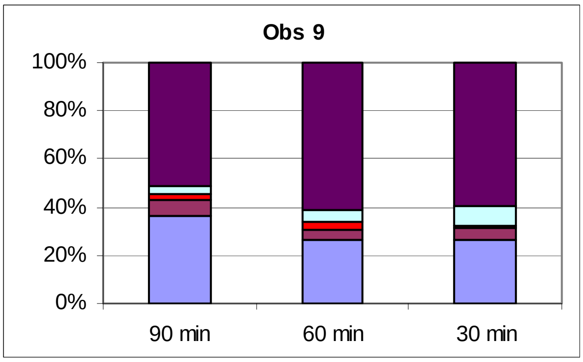


Figure J- 9: Smoothness of Data for Obs 9

Ilford Workstation

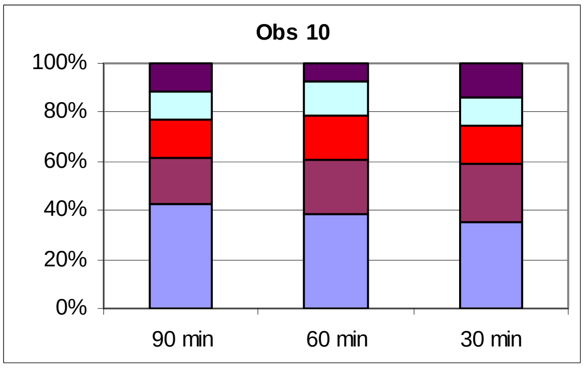


Figure J- 10: Smoothness of Data for Obs 10

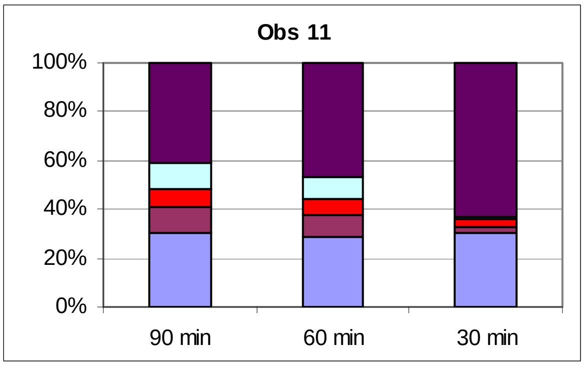


Figure J- 11: Smoothness of Data for Obs 11

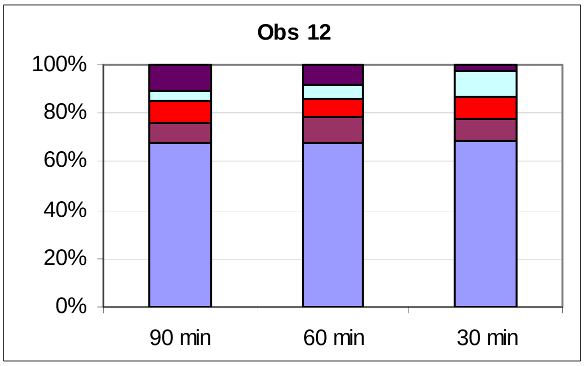


Figure J- 12: Smoothness of Data for Obs 12

North Kent Workstation

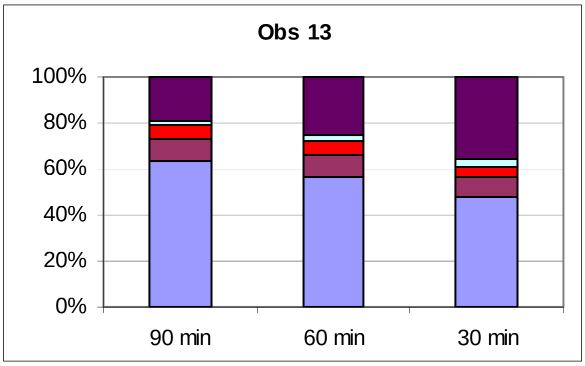


Figure J- 13: Smoothness of Data for Obs 13

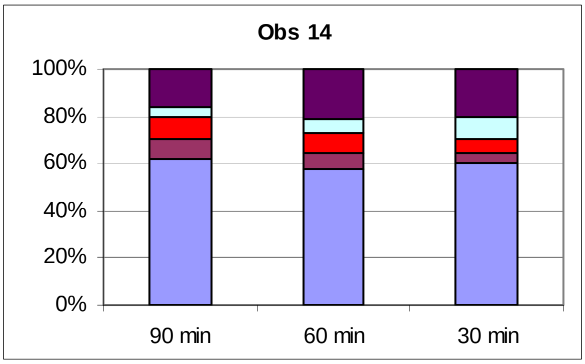


Figure J- 14: Smoothness of Data for Obs 14

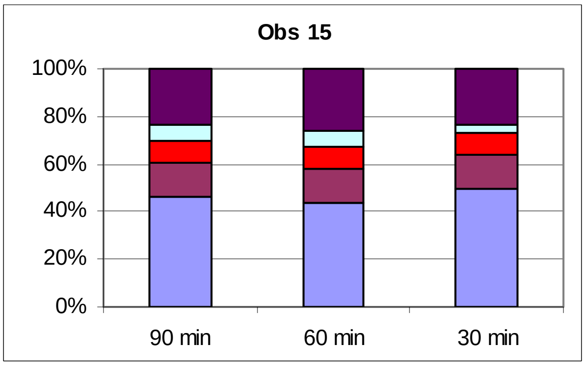


Figure J- 15: Smoothness of Data for Obs 15

Ashford Workstation

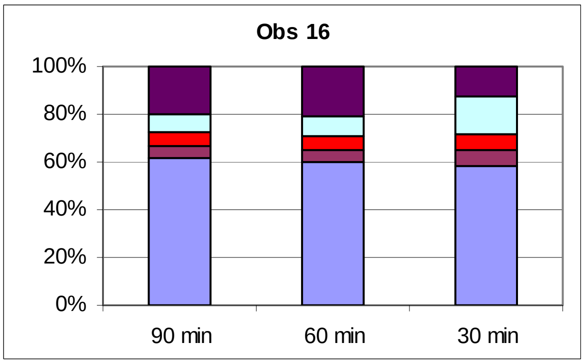


Figure J- 16: Smoothness of Data for Obs 16

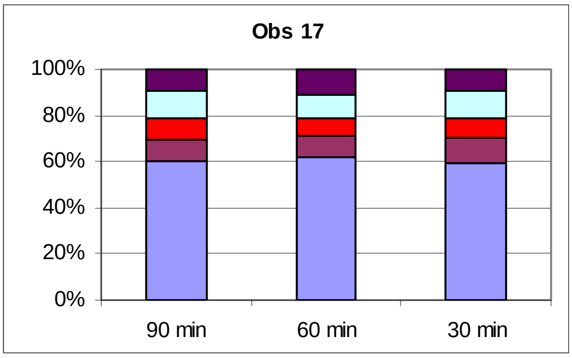


Figure J- 17: Smoothness of Data for Obs 17

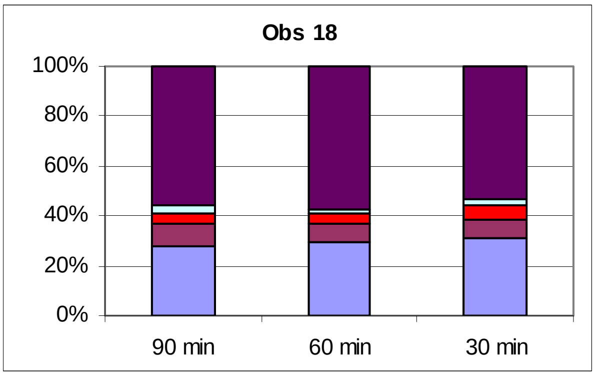


Figure J- 18: Smoothness of Data for Obs 18

Newcastle Workstation

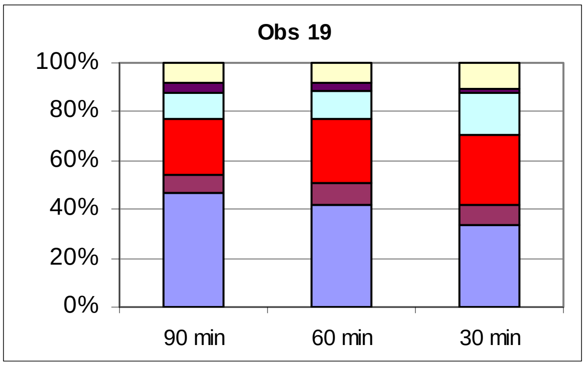


Figure J- 19: Smoothness of Data for Obs 19

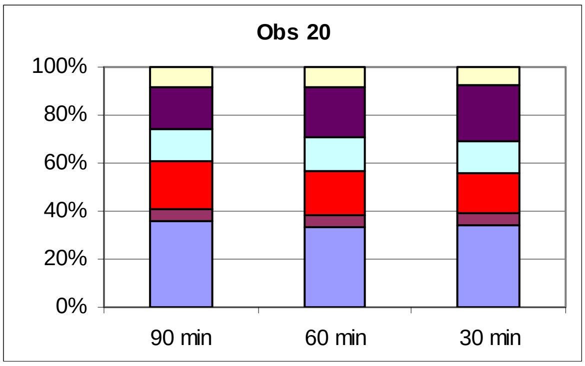


Figure J- 20: Smoothness of Data for Obs 20

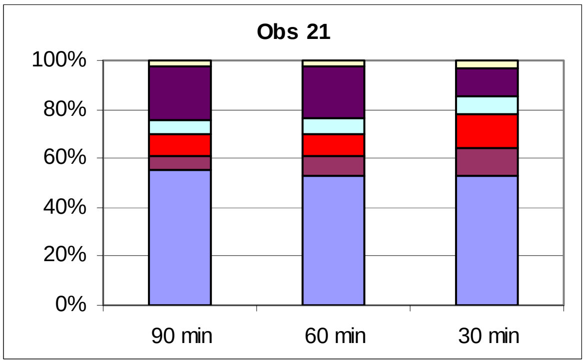


Figure J- 21: Smoothness of Data for Obs 21

Darlington Workstation

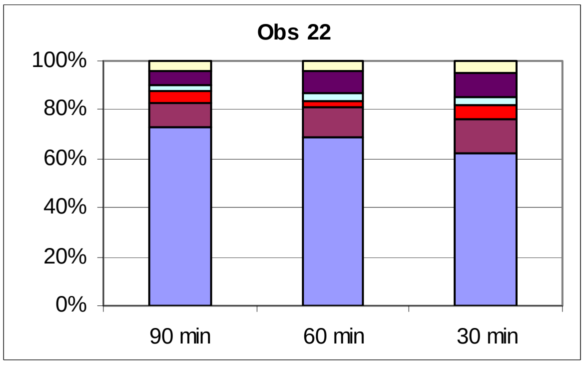


Figure J- 22: Smoothness of Data for Obs 22

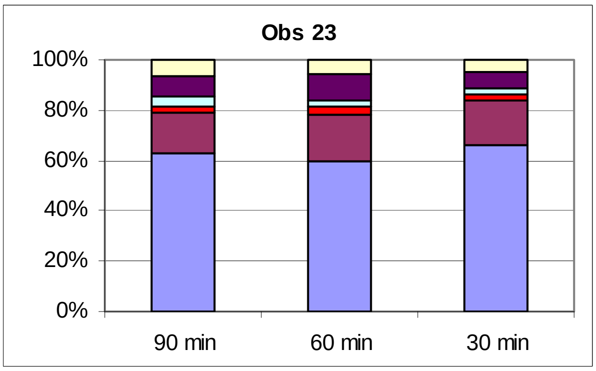


Figure J- 23: Smoothness of Data for Obs 23

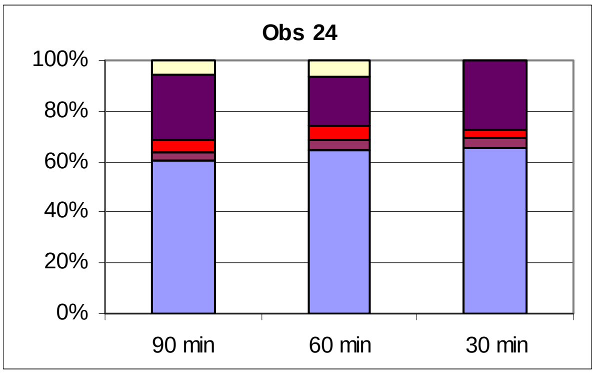


Figure J- 24: Smoothness of Data for Obs 24

Appendix K: Signaller Interview Questions and Probes

Opening Questions/General

- How is ARS running today?

- How long have you been working here/with ARS?
- Where did you work before?
    - NX
    - Lever Frame

- What do you think about ARS?
    - Do you like it?
    - Would you prefer to work without it?

- Do you trust ARS?
    - Without the interlocking?
    - Compared to another signaller?

- What do you think are the main problems with ARS?
    - When does it get into difficulties? (Engineering works, late running trains)

- What can ARS not do on this workstation?

- Can ARS cope with all situations?
    - Which ones does it have trouble with?
    - Why?
    - What are the problems?

- How could ARS work better in disturbed conditions?
    - What would you like it to do?

- Do you ever wish you didn't have ARS?
    - Does it ever get in the way?

- How is the programming on here?
    - Is there anything ARS always does wrong?

Regulation

- Can ARS regulate?
    - Is it as good as a signaller?

Responsibility

- Who is responsible for running the trains?

Understanding and Prediction

Understanding

- Can you explain how ARS works?

- Do you understand how ARS resolves conflicts?
    - The factors it takes into account
    - How do you think it resolves conflicts?

- Do you think ARS is consistent in its decisions?
    - Examples of when it isn't

Prediction

- Can you predict ARS?
    - Does it ever surprise you?

- Do you think it is important/good to be able to predict ARS?

Proactive Control

- Do you try to think ahead to control, or do you tend to react to things as you see them?
    - Why?

Use of Automation

Monitoring

- How do you monitor ARS?
    - What screens do you use?

    - What are you looking for?
        - Hot-spots
        - Individual trains
        - Length of route

    - Do you know how many trains are in your control area right now?

    - How long do you feel comfortable looking away for?

    - What do you check when you leave the workstation?
        - What information do they need?
        - Where do they get the information from?

    - Do you find monitoring difficult?
        - Hard to concentrate?
        - Hard to pay attention?

        - Boring?
        - Lose focus?
        - How do they get round this?

    - What sort of things catch your eye when you're monitoring?

    - If you were to glance at the screen, could you see pretty quickly if something was wrong?
        - Needed intervention
        - Potential conflict

Interrogate

- How often do you interrogate ARS?
    - Why
    - For what information
        - Do you always get the information you want?

Workload

- How is your job different with ARS?

- How has ARS changed the workload?
    - Higher/lower
    - Changed the tasks
    - ARS failure/worked workstation without ARS

- Is the job easier with or without ARS?

- Does ARS work the way you thought it would? Do the things you thought it would?

Organisation

- Apart from control, does anyone else affect the way you use ARS?

    - Does delay attribution affect the decisions you make?
    - If ARS does something stupid, does that delay go down to you?

    - What impact does planning have on ARS?

    - Do management affect your use of ARS?

Closing Question

- Anything else you want to say about ARS?

APPENDIX L: SIGNALLER INTERVIEWS CONSENT FORM

The purpose of these interviews is to understand how signallers are using ARS and what problems there are with it. The interviews are informal but will be recorded by the interviewer. Each interview should take about 1 hour.

Participation in this study is voluntary and there will be no consequences if you choose not to participate. Any information obtained from the research will be anonymous and will not be used for any purposes other than this study. The recorded data will not be made available to anyone else.

The data obtained will be used to help design automation in the future.

Consent

1. I confirm that I have read and understand the above information and have had the opportunity to ask questions.

2. I understand that my participation is voluntary and that I am free to withdraw at any time.

3. I agree to take part in the above study, and to be recorded.

____________________________	____________	______________________
Name of Participant	Date	Signature
Researcher	Date	Signature

<u>APPENDIX M: SIGNALLER INTERVIEWS SAMPLE DATA</u>

4.9 It's on, but I'm doing what I want before it does it, and it's only there as my backup if I get distracted.
I like to be in control of it, rather than a computer being in control of it where you don't know what it's thinking.

So it's a bit like your safety net

1.2 Yeah, basically. I got into a phone call and get talking for a few minutes, it takes over while I'm distracted, but other than that I'll sit here and do it all myself.

Do you just do these [redacted] *workstations, or the other IECC workstations?*

[redacted]

Would you prefer to work without ARS?

4.9 Not really, it's nice to know that you can go into a phone conversation if it's going to be long-winded knowing that it's doing it for you. It's nice knowing that, so, no, I like the fact that I'm doing it all myself, I don't switch it off, if I didn't like it I'd switch it off. But, it's just a safety net for me like you say, it's there, it's on, it'll do it's job if it has to, but other than that I'll pre-empt it and do it myself.

Do you trust it?

5.1 Pretty much so, over here [redacted]. Tend to know…I mean, I baby-sit it as well, because I program it on the TTP over there. I know how to run that, so if there's any long term flaws with it, I'm sort of, I know the fixes.

So you get in and change the timetable

So I can change it to what suits us

Do you think, is it ARS that you trust or is it the interlocking?

5.1 Well both really, I trust the interlocking 100%, the ARS I wouldn't say 100%, which is why, like I've just switched the sub-area off now because this one comes out and goes round the corner and em…

So why have you taken the sub-area out?

Appendix N: IWS Scale

IWS Scale

1	Not Demanding	Work is not demanding at all
2	Minimal Effort	Minimal effort required to keep on top of situation
3	Some Spare Time	Active with some spare time to complete less essential jobs
4	Moderate Effort	Work demanding but manageable with moderate effort
5	Moderate Pressure	Moderate pressure, work is manageable
6	Very Busy	Very busy but still able to do job
7	Extreme Effort	Extreme effort and concentration necessary to ensure everything gets done
8	Struggling to Keep Up	Very high level of effort and demand, struggling to keep up with everything
9	Work too Demanding	Work too demanding - complex or multiple problems to deal with and even very high levels of effort is unmanageable

APPENDIX O: IVIEW HED SPECIFICATION

Technology	Non-invasive, video-based eye tracking Monocular, pupil-CR, dark pupil tracking
Performance	Sampling rate eye movements: 50Hz (default), 200 Hz (optional) Tracking Resolution: <0.1° (typ.) Gaze position accuracy: <0.5° - 1° (typ.)
System	Operating System: Windows XP Workstation: Subnotebook or laptop
Headset	Lightweight bicycle helmet Interface weight: 79g Cable length: 5m and 2m (set of cables)
Auxiliary devices/ communications	Digital scene video recording in broadcast quality (720 x 576, MPEG-4) Socket based API interface via Ethernet (UDP) Compatible with SMI BeGaze™ Analysis Software Compatible with 3rd party video analysis packages (e.g. The Observer™ from Noldus)
Norm Compliance	CE, EMC, Eye Safety

APPENDIX P: SME PERFORMANCE SCALE

Performance Scale

1	Comfortable	Easily achieving all demands
2	Coping	Achieving all demands but some opportunities missed
3	Pressurised	All demands not met or frequent opportunities missed

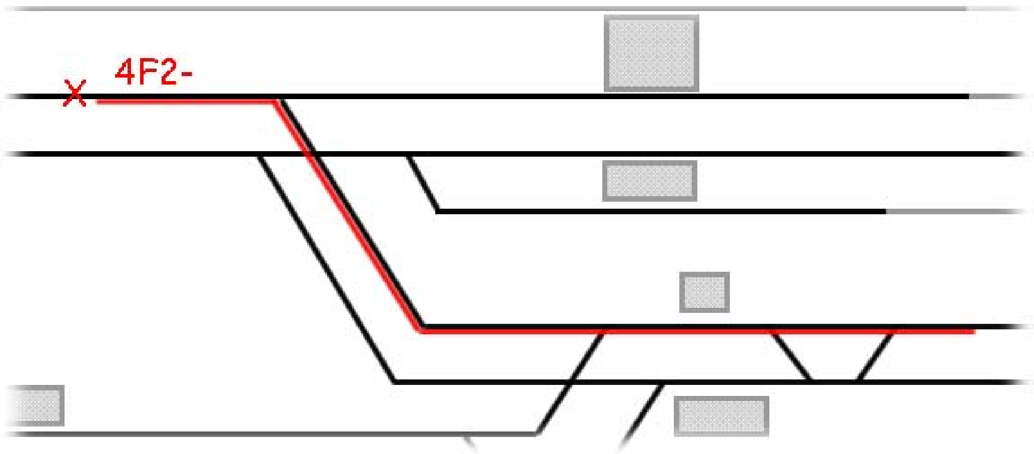
APPENDIX Q: SITUATION AWARENESS QUESTIONNAIRE

Situation Awareness Study – Part 1

Mark on the track layout diagrams (overleaf)

1. The position at the end of the simulation of as many trains as you can recall. Where possible, write the headcodes – if you can only remember part of the headcode, mark any gaps with a dash e.g. “4F2-“
2. Where possible, mark the routes these trains will follow. Mark different routes for different trains using the coloured pens provided.

You can see an example, below



Put your answers down in any order you want.

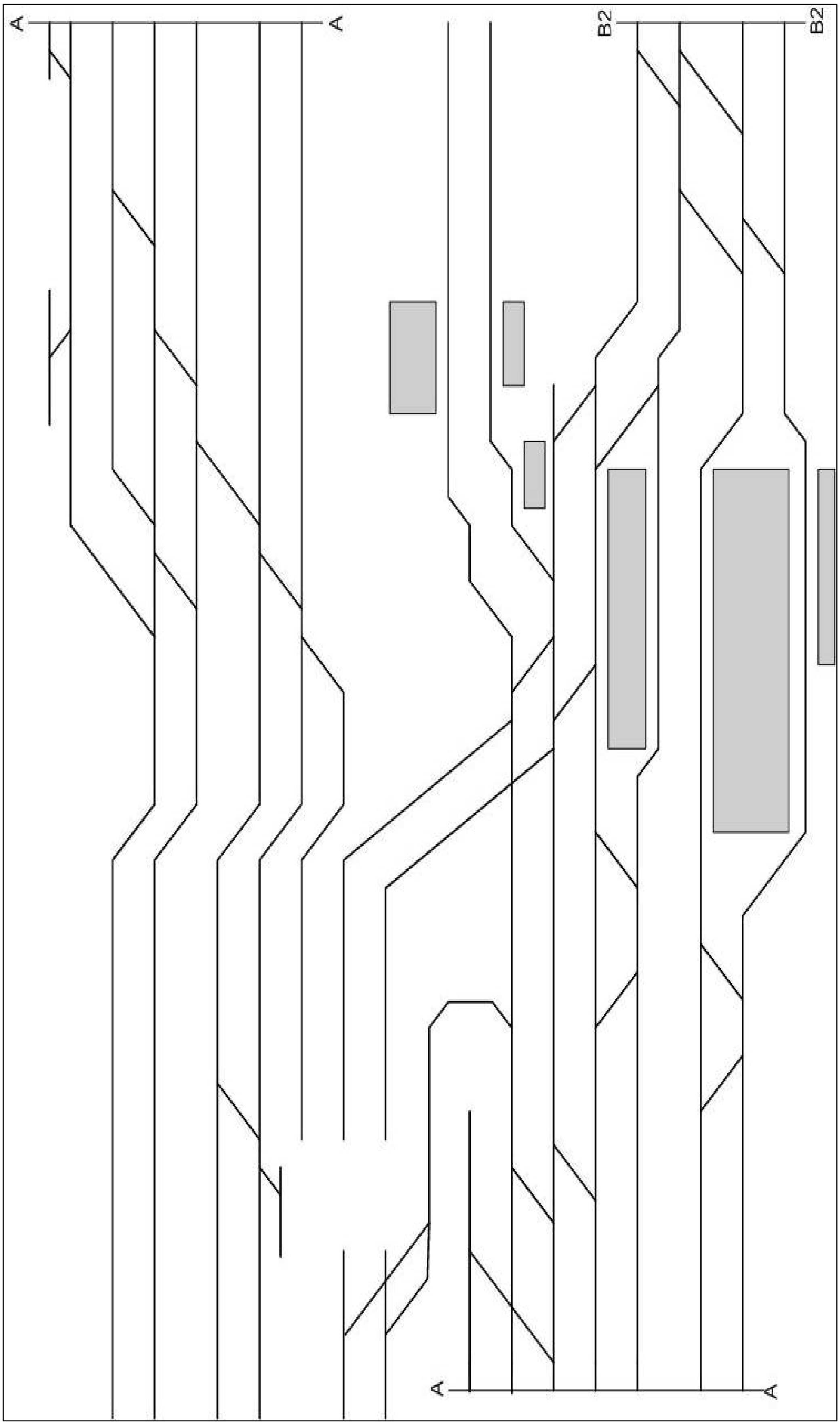
Try to answer the questions as quickly as possible – don’t worry about pinpoint accuracy.

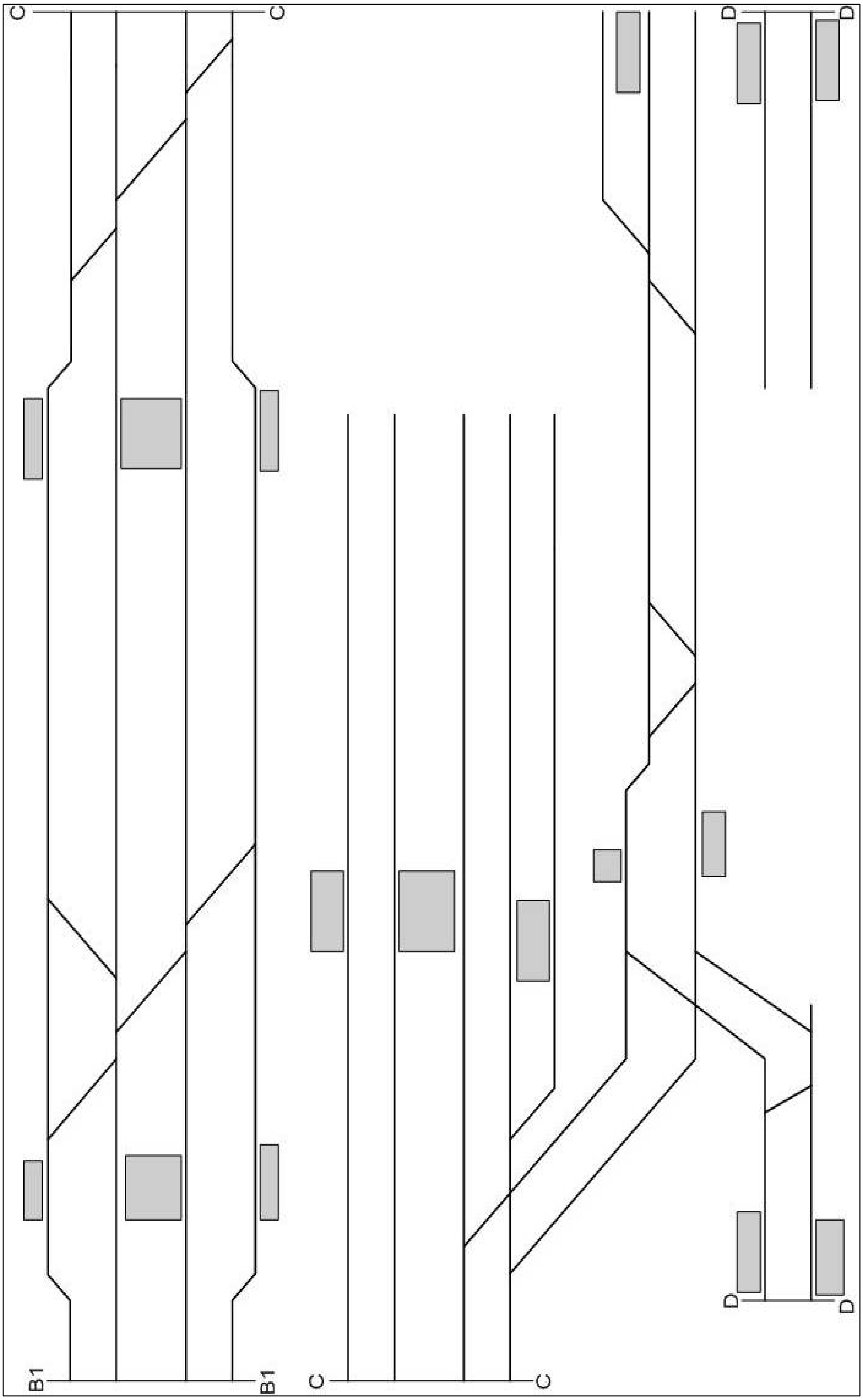
For admin

Participant number:

Date: Time:

Condition:


A
A
B2
B2
A
A


C
C
B1
B1
C
C
D
D
D
D

Situation Awareness Study – Part 2

Read the questions and below and mark your answer on the scale beneath, like this

Low_______________________X________High

Qu. 1. Can you rate the simulation for complexity? Was the simulation simple and straightforward (low) or highly changeable with many variables to consider (high)?

Low_________________________________High

Qu. 2. Can you rate the simulation for attentional demand? Did you have capacity to think about other things, or did your mind wander, (low) or did the simulation require a high degree of concentration and use all your mental capacity (high)?

Low_________________________________High

Qu. 3. Can you rate the simulation for understanding? Did you feel that what was happening was unfamiliar or that your understanding was incomplete (low) or did you feel you had complete and accurate knowledge of everything that was happening on the panel (high)?

Low_________________________________High

APPENDIX R: EXPERIMENT CHECKLIST

Before Experiment	
Start simulator	
Plug in, and turn on switch for laptop	
Explain study and sign consent form Note that the right hand buttons tend to stick Familiarise with IWS	
Note participant number, scenario and order	
Put on eye-tracking helmet	
Make sure to use correct mirror	
Check connections	
Start up laptop and open HED and Observer	
Calibrate eye-tracking	
Focus camera	
Press record in Experiment Centre	
Start Experiment	
Press record in Observer	

During Experiment:

- Record eye-tracking
- Record activity in Observer
- IWS Scores
- Performance data

After Experiment:	
Stop recording in Observer and save (also a backup)	
Stop recording eye-tracking and save	
Administer SA test	
Save scenario on simulator	
Get data from simulator	

APPENDIX S: EXPERIMENT CONSENT FORM

The purpose of these experiments is to understand the impact of automation (ARS and auto-routes) on the signalling task and the signaller. The data obtained will be used to help guide the development of future signalling automation systems.

You will be asked to complete three scenarios on the Stratford workstation, using different levels of automation in each, using ARS, using auto-routes but not ARS, and fully manual signalling. Each scenario is 40 minutes long. You will be provided with a simplifier for each scenario and are asked to signal the trains as you normally would.

Several types of data will be collected during each scenario:

1. Eye-tracking data – you will be asked to wear a bicycle helmet which has eye-tracking equipment attached to it. This collects information on where you are looking throughout the experiment. It is not harmful in any way and should not be distracting to you either.
2. Workload scores – you will be provided with a scale at the start of the experiment and you will be asked to rate your workload on this scale at 2 minute intervals throughout.
3. Performance scores – minutes lost by signaller, minutes gained by signaller, overall percentage score, routes cancelled by signaller. All these data are provided by the simulator.
4. Activity – your activity will be recorded throughout the experiment. It will be recorded as one of five categories – monitoring the signalling screens, using the keyboard or trackerball, looking at the simplifier, communicating with drivers or other signallers, and not involved in the signalling task.
5. Situation awareness – following each 40 minutes scenario you will be asked to fill out a short questionnaire assessing your situation awareness.

In addition, with your consent, the experiment will be video-taped. These data will be used only to review the experiment. The recorded data will not be made available to anyone else.

Participation in this study is voluntary and there will be no consequences if you choose not to participate. Any information obtained from the research will be anonymous and will not be used for any purposes other than this study.

Thank you for your time and co-operation with this effort.

Consent

1. I confirm that I have read and understand the above information and have had the opportunity to ask questions.

2. I understand that my participation is voluntary and that I am free to withdraw at any time.

3. I agree to take part in the above study, and to be recorded.

Name of Participant	Date	Signature
Researcher	Date	Signature

APPENDIX T: DATA SETS

The following data sets are included on the accompanying CD:

Chapter 4: Structured Observations of Signallers Data Sets

- Observation Data
- Observation Data from Second Observer
- Trust Questionnaire Data

Chapter 5: Signaller Interviews Data Sets

- Interview 3 transcript

Only one transcript is included as a sample due to concerns with maintaining the anonymity of participants.

Chapter 6: Level of Automation Experiment Data Sets

- Observed Behaviours
- Observed Interventions
- IWS Scores
- Performance Scores
- Eye-tracking Data

Appendix C: Validation of Principles of Automation

- Paired Comparisons Data