Effective, Design, Configuration, and Use of Digital CCTV

Hina Uttam Keval

A thesis submitted in partial fulfilment
of the requirements for the degree of

Doctor of Philosophy

of

University College London

Department of Computer Science
University College London

April 2009

I, Hina Uttam Keval, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

HKeval

____

Hina Uttam Keval

To my parents and sisters.

Abstract

It is estimated that there are five million CCTV cameras in use today. CCTV is used by a wide range of organisations and for an increasing number of purposes. Despite this, there has been little research to establish whether these systems are fit for purpose. This thesis takes a socio-technical approach to determine whether CCTV is effective, and if not, how it could be made more effective. Human-computer interaction (HCI) knowledge and methods have been applied to improve this understanding and what is needed to make CCTV effective; this was achieved in an extensive field study and two experiments. In Study 1, contextual inquiry was used to identify the security goals, tasks, technology and factors which affected operator performance and the causes at 14 security control rooms. The findings revealed a number of factors which interfered with task performance, such as: poor camera positioning, ineffective workstation setups, difficulty in locating scenes, and the use of low-quality CCTV recordings.

The impact of different levels of video quality on identification and detection performance was assessed in two experiments using a task-focused methodology. In Study 2, 80 participants identified 64 face images taken from four spatially compressed video conditions (32, 52, 72, and 92 Kbps). At a bit rate quality of 52 Kbps (MPEG-4), the number of faces correctly identified reached significance. In Study 3, 80 participants each detected 32 events from four frame rate CCTV video conditions (1, 5, 8, and 12 fps). Below 8 frames per second, correct detections and task confidence ratings decreased significantly.

These field and empirical research findings are presented in a framework using a typical CCTV deployment scenario, which has been validated through an expert review. The contributions and limitations of this thesis are reviewed, and suggestions for how the framework should be further developed are provided.

Acknowledgements

The journey one must make to earn a PhD can be very lonely. I am fortunate to have had the support of many people from academia and industry along the way. I am sincerely indebted and thankful to all of them. This work could not have come together without their help.

I would like to acknowledge that this PhD research was funded by the European Physical Science Research Council (EPSRC) 2004–2007, grant no. EP1P50064811.

I am most grateful to my dedicated supervisors Prof. M. Angela Sasse and Dr Simon Prince. Angela has given me great confidence, encouragement, guidance, and a tremendous amount of support throughout my research studies. I would like to thank my second supervisor, Simon, who kindly gave me advice and constructive comments on the empirical parts of my research. A very big thank you goes to Hendrik Knoche, my official PhD mentor, colleague, and friend. Hendrik provided me with guidance on my video quality experiments and taught me a great deal about video, imaging and data analysis. Thank you to Dimitris Miras who provided feedback on my early experimental ideas and offered his expertise on video quality. I would also like to thank Dr Paul Cairns (formerly of UCLIC and now at York University) who was so willing to offer his time to review my research papers and offer guidance on data analysis for my empirical experiments – even on short notice.

I would like to thank Hendrik Knoche and Sven Laqua who helped me program my experiments. I am also very grateful to all of my colleagues in the Department of Computer Science who provided their opinions on my research ideas and gave their time to review papers and drafts: Dr John McCarthy, Dr Jens Reigelsberger, Dr Vinoba Vinayagamoorthy, and Rae Harbird.

Special thanks are also due to Jim Aldridge (former senior Home Office video expert, and now consultant at Effective Pictures, Ltd) for providing me with a huge amount of insight and offering his expertise on the technical problems surrounding CCTV. Several other individuals also helped me refine my ideas through discussions, long email queries, and telephone calls: John Wood (CCD Design and Ergonomics), Prof. Alf Linney (UCL), Dr Ian Nimmo (User Centered Design Services, LLC), Dr Martin Maguire (ESRI), Dr Iain Darker (ESRI), and my colleagues at the Home Office Scientific Development Branch (HOSDB): Simon Walker, Neil Cohen, Jay Gatusso, and Dr Ken Brown-Maclennan.

Finally, I would like to thank my family, friends and my dear Edward Elton for giving me the strength, encouragement, moral, and financial support throughout – it really went a long way.

Contents

List of Figures

List of Tables

Chapter 1

1.1 Introduction

Closed circuit television (CCTV) plays a significant role in protecting the public and assisting the police in the investigation of crime. The UK is one of the most watched countries in the world (McCahill and Norris, 2003; Phillips, 1999). It is estimated that there are five million CCTV cameras in use today, and this number is likely to rise in the future (Gill, 2006). Even though the exact number of CCTV systems deployed in the UK is unclear, *"the extent of CCTV coverage and the government's funding of new systems have increased dramatically over the last decade,"* (Armitage, 2002, p 6) yet there is little substantive research evidence to show that CCTV works (Armitage, 2002). The research presented in this thesis examines the social and technical problems affecting CCTV performance, provides a framework for understanding the causes of problems, and guidance on how to overcome them in practice.

Over the last two decades, CCTV systems have evolved as a result of several changes which have taken place over time:

1. Changing social perceptions and attitudes towards security.
2. Evolving technology platform.
3. Wider range of CCTV users.
4. Wider range of security tasks.

Change 1: Changing social perceptions and attitudes towards security.
Social perceptions and attitudes towards security have changed, and over time society has become increasingly security conscious. This change has also been a result of the mass media coverage on crime. People have changed their views as a result of terrorism, gun crime, child abductions, etc. and have adopted a more proactive role in ensuring their own safety. One way this has been achieved is through investment in CCTV systems. Security is now considered essential for the protection of both people (e.g., within businesses and for the general public) and their property.

Change 2: Evolving technology platform.
Over the last thirty years, CCTV technology has advanced and evolved through three key phases: 1) analogue; 2) digital; and 3) intelligent CCTV systems.

Phase 1: Analogue CCTV ~1965 to the mid-1990s.
CCTV was initially designed to record analogue video signals, received from a number of CCTV cameras, directly onto Video Home System (VHS) tapes using a Video Cassette Recorder (VCR) in a closed loop. At the time, tape-based video security systems were perceived to be an easy to use and affordable solution for security; however, in practice, they are not particularly easy to use when compared to their digital counterparts. The main limitations of analogue CCTV systems are the high

level of maintenance (e.g., constant tape changes), limited functionality (e.g., analogue video is very difficult to edit, meta-data cannot be attached to particular video scenes), and low quality video. Analogue CCTV systems are also difficult to integrate with other security applications, such as automatic face recognition systems.

Phase 2: Digital and networked CCTV ~late 1990s to present.
Around 1997, second-generation digital CCTV systems were designed to overcome the shortcomings of analogue ones. Digital CCTV systems convert analogue video signals into digital data, and then record them onto a hard disk. The capability of digital video has allowed data to be streamed cheaply in comparison to analogue video. Digital CCTV has also meant that users can do a lot more with data and a lot more efficiently. For example, digital CCTV video can be streamed across a network and viewed in real-time for monitoring purposes, linked to other technologies to provide the police with intelligence, and copied easily onto other media.

Phase 3: Intelligent CCTV ~late 1990s to present.
Third-generation CCTV systems were often described as intelligent CCTV because they support human operators by combining it with software to help human observers to detect and respond to crime more quickly. Depending on the particular software used, digital CCTV systems may detect unusual patterns of events (anomaly detection systems), automatically recognise faces (automatic detection systems), and track illegal drivers and their vehicles through the use of Automatic Number Plate Recognition (ANPR) systems.

In this thesis, the effectiveness of both analogue and digital CCTV systems (first and second generation systems) is examined to identify specific performance issues for security tasks carried out by human observers. Particular emphasis is placed on identifying problems with digital CCTV systems, which are widely deployed and used for a wide range of security tasks. Despite their widespread use, there is currently very little Human-Computer Interaction (HCI) knowledge on the performance and usability of such systems. A UK national evaluation of CCTV found that CCTV is not effective for key public security tasks (Gill and Spriggs, 2005; Gill et al., 2005). However, no studies have examined *why* CCTV is not effective. The key contribution of this thesis is to apply HCI knowledge: 1) to identify the specific socio-technical factors which reduce the effectiveness of CCTV; 2) to provide guidance for improving the effectiveness of current CCTV systems and; 3) provide knowledge for researchers and designers to use in the development of future CCTV systems.

Change 3: Wider range of CCTV users.
A socio-technical system is described by Shepherd (2001) as: *"… a complex grouping of interrelated parts, and can include human beings and machines. The parts interact to serve a purpose"* (p 8). Like any security system, CCTV needs to be considered as a socio-technical system. Throughout this thesis the term 'socio-technical' is used and is defined as the interaction between people and technology within the workplace. Specifically, a CCTV security system consists of technical parts and social sub-systems (e.g., CCTV stakeholders, such as the police, control room operators, forensic experts, courts, the public, etc.).

Not all stakeholders directly interact with a CCTV system, but they may, nonetheless, have some form of influence on the system and its performance. The shift from analogue to digital CCTV systems has increased utility, for an ever-expanding range of CCTV owners and more heterogeneous CCTV users. CCTV owners purchase and own the system, and are therefore responsible for its operation and performance; CCTV users perform tasks with the system and use CCTV video and images to respond to incidents and investigate them (e.g., police, operators, forensic experts, etc.).

Until recently, CCTV video and images have been utilised by trained and experienced CCTV users for security tasks such as detection and investigation:

1. The Criminal Prosecution Service (CPS) use CCTV to prosecute criminals in court.
2. Forensic experts analyse and process CCTV on behalf of the police to investigate crime.
3. CCTV operators view real-time CCTV video on monitors in a control room and perform four main security observation tasks: monitor, detect, identify and recognise (Aldridge, 1994).

Digital and networked CCTV systems are being increasingly deployed by government, commercial organisations, and private individuals. Furthermore, there is now a wide range of CCTV users with varying skills and experience in using CCTV. One of the reasons untrained CCTV users are performing security tasks is because of the wide range of capabilities digital CCTV systems offer. For instance, it is possible to distribute digital CCTV video to a wider audience via the Internet on a Personal Computer (PC) and televisions (within the home and at work). Following are three examples of situations in which CCTV is being utilised by untrained CCTV users:

1. In America, a pilot scheme allows web users to access and view real-time CCTV video to monitor the Texas-Mexico border for illegal crossings. If they witness an illegal crossing, they have the opportunity to alert the authorities (Web Users to Patrol, 2006).
2. In the UK, residents of an east London housing project can view digital CCTV video from their television sets by subscribing to a community safety channel (Rights Group, 2006). In this set-up, if a viewer sees a crime being committed, or identifies a person breaking an anti-social behaviour order (ASBO), they can alert the police by phone. Viewers can access images of known criminals and troublemakers by viewing a ‘rogue’s gallery’ prepared by the police on their TVs.
3. Members of the general public also own and use a CCTV system; therefore they are considered CCTV owners and users. In an unusual case, a man had erected his own CCTV system to support the British Transport Police in dealing with crime and acts of vandalism at his local railway station (Rail Safety Man, 2008).

In such cases, it is important that the security tasks can be performed reliably by untrained CCTV users at minimum error. The use of a best-practice framework for CCTV configuration and use can facilitate reliability and enable these tasks to be performed effectively.

Change 4: Wider range of security tasks.
Before the 1960s, CCTV security systems were first utilised within the retail environment to combat shoplifting, and to detect crime in real-time. In the 1960s, a large CCTV system was installed within the London Underground transport network, not to deter and detect crime, but to support train drivers and station staff in safety tasks. Before introducing CCTV to the London Underground, the entire underground network was manned by security guards to manage customer safety.

Throughout the UK, there are a number of CCTV control room systems which record video surveillance and allow operators to observe public safety and crime in real time on several video monitors. As well as monitoring and recording incidents within enclosed public spaces, CCTV has been applied to many related functions, such as monitoring traffic (Ney and Pichler, 2002), parking, Central London Congestion Charge; street cleaning (McCahill and Norris, 2003), and managing traffic incidents. There are many evident benefits to using CCTV, e.g. faster response to major incidents and congestion, more up-to-date and accurate road/traffic information for media and traffic management services.

The wide range of capabilities digital CCTV systems offer and its pervasiveness mean that digital CCTV systems can now be used for many tasks and in many contexts. The social and technical changes described have led to a number of different products that meet the needs of a diverse set of CCTV users, who use it for a wide range of goals. The research presented in this thesis examines the social and technical factors that reduce the effectiveness of CCTV, particularly in situations where trained and untrained CCTV users interact with CCTV and other technologies for security observation tasks.

1.2 Research Problems

The research conducted by Gill et al. (2005) determined that CCTV systems are not effective in terms of reducing crime activities and the public's fear of crime. The focus of this study was on the criminological aspects of CCTV, rather than issues associated with the technology. In this thesis, the effectiveness of CCTV was investigated from a socio-technical perspective to identify why CCTV is ineffective and what can be done to improve effectiveness. More specifically, the research in this thesis seeks to identify the factors that reduce performance when security observations are carried out using analogue and digital CCTV systems. This investigation was conducted to understand whether first and second generation CCTV systems are fit for purpose, whether deployments are worth the investment, and then provide guidance to improve CCTV system design configuration and use.

In the absence of this understanding, there are two key problems: 1) first and second generation CCTV systems will continue to operate ineffectively if CCTV owners and practitioners are unaware of how to assess the performance of their systems; 2) researchers and developers of CCTV systems will continue to research, design and develop CCTV systems without understanding the failures of previous systems. The following two research problems were addressed in this thesis from a socio-technical perspective using HCI research methods and techniques:

1.2.1 Research Problem 1

Gill et al. (2005) determined that CCTV is not effective by investigating the social implications of CCTV across several public CCTV control rooms. The main focus of this research looked at the social factors which reduced the effectiveness of CCTV. There were some findings which found that CCTV was not working well as a result of technological failures. Gill's study did not however focus on both the social and technical factors jointly. While a number of studies have examined operator factors in various control room environments, such as power plant control rooms (Norros and Nuttinen, 2005), air traffic control centres (Bentley et al., 1992; Twidale, Randall and Bentley, 1994), ambulance control rooms (McCarthy, Wright, Healey, Dearden and Harrison, 1997; Blandford and Furniss, 2005) and within London Underground control rooms (Luff, Heath and Jirotka, 2000), further research is needed to examine the effectiveness of CCTV from a socio-technical perspective.

Current guidance on the design of control centres, such as BS EN ISO 11064 (ISO, 2004), addresses physical ergonomics, however additional and specific guidance is needed to provide CCTV practitioners (i.e., CCTV consultants) and owners support on the design of operators' tasks and specific video technology. The standard (ISO, 2004) disregards the impact of the technology on people such as the level of audio and video information required to process, impact of environment noise on task performance and so on. Thus, awareness and guidance is needed on the cognitive elements of control room design where CCTV technology is used by human observers.

To understand CCTV technology from this perspective, research is needed to understand the *context* in which analogue and digital CCTV systems are used in order to improve their effectiveness and to better inform researchers and designers in the development of future CCTV systems. Security practitioners will also benefit from insights on how existing CCTV systems can be improved. The first half of this thesis, Part 1, investigates this problem by exploring how CCTV is used, how tasks are performed, and what factors affect task performance.

1.2.2 Research Problem 2

The quality of images recorded by both digital and analogue CCTV systems varies considerably. In fact, *"anecdotal evidence suggests that over 80% of the CCTV footage supplied to the police is far from ideal, especially if it is being used for primary identification or identities are unknown and identification is being sought, for instance, by media release"* (Gerrard et al., 2007, p 12).

A detailed review of previous CCTV research literature (see Chapter 5, Section 5.3) revealed that there is little CCTV guidance on the *digital* video quality requirements for achieving effective task performance. The only guidance currently available, the UK Home Office CCTV Operational Requirements (Cohen, Gatusso, and MacLennan-Brown, 2007), lack specific details on digital and network CCTV systems. In addition, it is not clear how these recommendations were derived and there is no indication for certain that they have been derived from user studies. In the absence of detailed and tested guidance, CCTV system owners have no way to understand the affect of using low-quality

CCTV video for security observation tasks. The use of low-quality CCTV video was identified as an issue in the control room study conducted by Gill et al. (2005), as well as in the extensive CCTV control room study conducted for this thesis (see Study 1 in Chapter 4).

A number of studies in the area of human centered multimedia have been conducted to assess the impact of using low-quality video (i.e. video played back at 1fps, resolution below 352x152, compressed at 32 Kbps) for different video-based tasks (e.g., video conferencing, e-learning, and mobile and TV applications) – this literature review can be found in Chapter 6. There has been just one study which has examined the human performance limits of CCTV when video quality is compromised for cost (van Voorthuijsen et al., 2005). As a result, there is a need for further research to determine what quality is needed to make the performance of security observation tasks more effective.

The second half of the research presented in this thesis examines the impact of low-quality CCTV video for security observation tasks performed by human observers. The purpose of this research is to produce a best-practice framework for CCTV security practitioners and owners to assist them in improving the design and configuration of CCTV systems.

1.3 Research Scope and Approach

CCTV has been studied from a number of different perspectives by researchers from a number of different fields. Thus, there is a wide variety of research relevant to the problem field and each discipline has approached the problems associated with CCTV from their own disciplinary perspective. As CCTV security research touches upon topics from a number of research disciplines, the research problem cannot be fully examined using a discipline-specific approach. Therefore, the research presented in this thesis approaches both research problems 1 and 2 by applying multidisciplinary knowledge and methods (i.e., HCI, computer science, psychology, and sociology).

Based on this multidisciplinary research approach, the objectives for this thesis are as follows:

1. Understand the context of CCTV and the different CCTV stakeholders.
2. Identify the socio-technical problems that reduce the effectiveness of CCTV.
3. Identify how and why these socio-technical problems affect different types of CCTV users.
4. Identify one specific socio-technical problem that can be further investigated through empirical research.
5. Develop a set of user requirements for the effective configuration of a CCTV design in which security tasks are performed by CCTV users.
6. Incorporate these user requirements into a best-practice framework for CCTV developed to provide CCTV practitioners and owners with methodological steps and specific guidance for the effective design configuration and use of digital CCTV.

Previous research on CCTV has mainly examined its effectiveness in terms of reducing crime and the public's fear of crime – thus, examines the effectiveness of CCTV from a sociological and criminological perspective. Although some insights into the use of technology for CCTV operator tasks are given (see Gill et al., 2005), further research is needed to identify the: 1) technical failures with CCTV and other associated technology and 2) impact of these technological failures on the CCTV user's performance in security tasks. Research therefore is required to examine other factors which negatively affect CCTV performance such as peer-to-peer communication, video quality, uses of multiple technologies for a single task etc.

Another area of CCTV research that has received much attention is the 'big brother' debate. Despite the strong support for CCTV, many believe that it has been taken too far in the UK (see Taylor, 2002). The research presented in this thesis does not examine the privacy concerns associated with CCTV or the impact it has on society. Furthermore, the research in thesis does not address the topic of intelligent CCTV.

Following a detailed critical review of the literature relevant to the problem space (see Section 1.1), two major themes have emerged for research:

1. Firstly, it is necessary to understand the key components of a CCTV security system and the current failures of the technology from the CCTV user's perspective (those individuals who directly interact with the system). Exploratory research is conducted within a large number of CCTV control rooms (*n* = 14). This security setting was chosen in order to understand the context of CCTV, the range of CCTV stakeholders, security tasks performed by users, the tools used to support tasks, and the socio-technical factors which reduce the effectiveness of CCTV. CCTV systems are specifically studied within CCTV control rooms, as a number of different technologies are used in these environments and there is considerable audio (radio and telephone) and video interaction between different CCTV stakeholders such as police officers, shop staff, and the public.
2. The second half of this thesis examines one particular performance issue which reduces the effectiveness of CCTV: the use of low-quality CCTV video for security observation tasks. This particular issue was identified in the control room field study (see Study 1 in Chapter 4). In addition, this issue was also identified in previous research (Gill et al., 2005; Bromby, 2002; and Mead, 1998). The impact of lowering CCTV video quality when performing identification and detection tasks with users was examined empirically (see Chapters 8 and 9). This research is entirely novel and is necessary in order to provide evidence-based and practical recommendations when storing and streaming CCTV video. The overall aim is to improve the effectiveness of CCTV for security observation tasks.

The approach taken for this research considers both the social and technical aspects of CCTV security, and follows a task-oriented methodology for the field and empirical research reported in this thesis. HCI knowledge and methods are applied to improve CCTV practice, in particular its design configuration and use, taking into account its context of use.

The research detailed in this thesis is deliberately split into two parts: 1) real-world field research and 2) experimental research. Field research (see Study 1) was conducted to understand the real-world problems with CCTV in practice and then, one specific problem (video quality) was selected to demonstrate through two empirical experiments how to identify video quality requirements for two observation tasks (see Study 2 and 3).

1.4 Research Goals

There are three research goals for this thesis. Goal 1 investigates in detail why CCTV is not effective through field context research (i.e., in several CCTV control rooms). The aim of this portion of the research is to understand the general context of CCTV and identify the wide range of performance issues within a real-world setting.

Research goals 2 and 3 have been specifically formulated to empirically investigate one of the most serious performance issues identified in Study 1: the use of low-quality digital CCTV video for security observation tasks. These goals are also motivated by anecdotal comments gathered from CCTV experts during the course of the field research detailed in Chapter 4, as well as from previous research (Gill et al., 2005; Bromby, 2002; and Mead, 1998).

1.4.1 Research Goal 1

Research goal 1 is a high-level goal formulated to investigate the social and technical problems associated with digital CCTV and other technologies used by human operators *within a real-world context*. This is a broad research goal to: gain an understanding of the general context of CCTV, identify the underlying factors which reduce the effectiveness of CCTV, and the technical failures of CCTV technology when used for security observation tasks performed by human operators.

1.4.2 Research Goal 2

Based on the findings of the field research, as well as findings of previous face identification studies (see Bruce et al., 1999; Burton, Wilson, Cowan and Bruce, 1999; and Henderson, Bruce and Burton, 2001), research goal 2 was specifically formulated to identify the minimum video compression level required for an observer to effectively identify targets from CCTV images when a good quality photograph of the target - who is unknown to the observer is available for the task. This research goal examines the effectiveness of low-quality CCTV video when used for a specific human observation task – face identification, when using a digital CCTV system operating under a limited storage or networking budget.

1.4.3 Research Goal 3

Research goal 3 was formulated to identify the minimum video frame rate required for an observer to detect events from CCTV video. Similar to research goal 2, this goal examines the effectiveness of low-quality CCTV video for another security task – event detection under the similar circumstances described above. These two tasks: face identification and event detection were chosen as they are two commonly performed CCTV tasks performed by human observers.

1.5 Thesis Structure

As has been detailed, this thesis addresses two major aspects of CCTV in HCI and security research: 1) the context of CCTV and the broad task performance issues which reduce the effectiveness of CCTV and 2) the effectiveness of digital CCTV video for security observation tasks. Thus, this research required a literature review be conducted for two distinct areas of CCTV security: CCTV in control rooms and CCTV video quality for user tasks. It is necessary to present both literature reviews and research contributions for these studies. Part 1 (Chapters 2, 3 and 4) presents the field research in CCTV security; this includes the background literature on control rooms, field methods used in HCI, and the CCTV control room study. Part 2 (Chapters 5, 6, 7, 8 and 9) presents the CCTV video quality research; this includes the background literature on previous video quality studies, HCI methods used for assessing video quality with users, and the two empirical experiments conducted with users on video quality. The remaining Chapters (10, 11 and 12) detail the contributions of this thesis and conclusions.

Part 1: Control Room Research

Chapter 2 presents a review of the relevant research literature addressing task performance issues for operators using CCTV and other technologies within control rooms. The literature reviewed was obtained through research conducted within the disciplines of HCI, computer supported cooperative work (CSCW), and sociology.

Chapter 3 describes the key field research methods used for understanding how users perform their work and how they use different tools and systems within control room environments. This chapter also outlines the methodological approach taken for the control room field study (Study 1).

Chapter 4 presents Study 1, the CCTV control room field study. This field study was conducted over eight-months and was carried out across a large number of CCTV control rooms in the UK (*n* = 14). Contextual inquiry research was used to carry out the study, which involved interviewing CCTV control room managers and operators and observing operators whilst they performed their core tasks within the control room. This field work was conducted in order to understand the context of CCTV security with the aim of identifying the social and technical factors which reduce the effectiveness of CCTV.

Part 2: CCTV Video Quality Research

Chapter 5 provides a background literature review of CCTV technology and its uses in security and surveillance applications. It also outlines the main differences between analogue, digital and networked CCTV systems, and provides a background to digital video compression, outlining the main challenges with using low-quality CCTV video and images for security observation tasks.

Chapter 6 provides a critical literature review of the previous studies on video quality which have investigated the impact of lowering video quality on users' task performance when interacting with video. The aim of this review was to compare and contrast previous studies and also to highlight their methodological limitations in order to provide a basis for the empirical studies carried out in this research.

Chapter 7 provides a critical review of the different methods used in HCI research for evaluating video quality with users. This review is presented to provide a basis for the methodology used for the two empirical video quality studies (Study 2 and 3) carried out in this thesis.

Chapter 8 presents the first empirical video quality study (Study 2), which investigates the impact of excessive video compression on a face identification task performed by human observers. This experiment was motivated by the fact that there is no: 1) digital CCTV video guidelines which provide recommendations for the minimum video compression levels required for security observation tasks; 2) research conducted to explore the relationship between video quality and user task performance for CCTV applications – also, there is 3) research evidence (Study 1 and previous research studies) which has found that very low- quality CCTV video is being recorded across several CCTV control rooms. The impact of this issue is assessed to determine the most effective video quality level for a face identification task.

Chapter 9 presents the second empirical video quality study (Study 3), which investigates the impact of lowering the frame rate of CCTV video on a detection task with untrained CCTV users. This study was also motivated by the factors described for Study 2 (see above).

Chapter 10 presents the contributions of this thesis. This contribution is a best-practice framework (TEC-VIS) which describes the effective design configuration and use of digital CCTV. The framework is designed to support CCTV security owners and practitioners with high-level guidance on carrying out user analysis and system requirements analysis for a CCTV deployment. Specific guidance is also provided in the form of recommendations, which guide CCTV practitioners and owners through the design and configuration of their CCTV systems: camera environment, CCTV observer's workstation environment, and the set-up of a digital video recorder. The framework is exemplified with a scenario that describes a CCTV deployment. The framework presented in Chapter 10 is the final version (version 2) and version 1 which was reviewed externally is given in Appendix H.

Chapter 11 critically discusses the contributions made by this thesis. This critique is based on an external review of TEC-VIS (see Appendix J) by three experts in the HCI, CCTV security, and HCI fields. The purpose of the review was to identify areas in which TEC-VIS could be improved, with the aim of developing it into a coherent and complete framework, so that it can be applied widely to all CCTV security systems and used by all CCTV stakeholders.

Chapter 12 details the conclusions of the research presented in this thesis. The research goals are revisited and the research contributions (substantive and methodological) are highlighted. Finally, the research conducted for this thesis is critically reviewed and an outline of the future work required to advance HCI research on CCTV effectiveness is presented.

PART 1: Control Room Research

Chapter 2

Previous Control Room Studies

To date, little research has been done to investigate how CCTV users utilise CCTV and related technologies – particularly modern technologies, to perform security observation tasks. One study (Gill et al., 2005; Gill and Spriggs, 2005) identified failings within control rooms where CCTV technology is used, and mention a number of difficulties CCTV operators experienced. These difficulties however need to be examined in more detail. This chapter critically reviews previous studies conducted within control rooms, purposely to identify the gaps in the research. This is necessary to: 1) identify what research has been done within control rooms in which CCTV technology is used for security observation tasks; 2) identify what further research needs to done in this area; and 3) formulate the research problems and goals for Study 1 (Chapter 4) defined in this thesis, which will then form the basis for the methodology and approach taken in examining the effectiveness of CCTV.

2.1 Introduction

A CCTV control room is a central hub for security control and coordination activities performed by several CCTV operators, who are responsible for monitoring and reacting to events they observe on real-time CCTV video displayed on video monitors.

The rapid development in the security and surveillance technology market has led to new systems being added and created more tasks and processes for CCTV operators. There is little knowledge about how CCTV users perform security observation tasks, how they use the technology, and what problems they encounter.

Gill et al. (2005) conducted a study on behalf of the Home Office to measure the impact of CCTV on crime; in both the reduction of crime and the public's fear of crime at a number of public-space CCTV control rooms (*n* = 14). The study identified a number of problems operators experienced when performing tasks in the control room. This study is reviewed in the next section. A number of other field studies have been carried out within control rooms where security is not the primary purpose, but uses similar technology, such as:

1. Plant control room (Norros and Nuttinen, 2005).
2. Air traffic control centres (Bentley et al., 1992; Twidale et al., 1994).
3. Ambulance control rooms (McCarthy et al., 1997; Blandford and Furniss, 2005).
4. London Underground control rooms (Luff and Heath, 2000).

Process control rooms, such as industrial manufacturing and nuclear plant control rooms operate quite differently from security control rooms. Process control room operators are responsible for monitoring safety critical processes and events, whereas security control room operators are responsible for monitoring crime and traffic related incidents. In this chapter, the most relevant control room study

(Gill et al., 2005 and Gill and Spriggs, 2005) is reviewed first, followed by other related control room studies (transport and emergency control rooms) where CCTV technology is used. These non-security control room studies were specifically chosen for this review, as operators at these control rooms carry out tasks similar to those carried out by CCTV operators, and most use CCTV and map-based tools to locate scenes.

2.2 Security Control Room Studies

Gill et al. (2005) evaluated 13 CCTV projects which were set-up under the Home Office Crime Reduction Programme (comprising of 14 separate CCTV systems). At six control rooms, surveillance video was recorded on the more traditional analogue S-VHS tapes, six recorded video digitally on a computer hard drive, and one recorded video on digital tape. Various aspects of the control room operations were examined such as: ownership, design, management, work practices, communication, operator pay and training, as well as the processing of CCTV evidence. The study did not examine the impact of specific technologies on operator performance when carrying out tasks. Despite this, a number of findings which related to the technical aspects of a control room system which could have affected operator performance were identified:

1. At a majority of the control rooms, digital tools (e.g., intelligent CCTV, digital radio, user interfaces etc.) were not used to support operators in their tasks.

2. In comparison to analogue CCTV systems, digital systems could be searched more quickly, which saved police time when looking for evidence. The average search time required to search digital CCTV was 18 minutes and 40 minutes to search analogue CCTV video.

3. At seven control rooms, CCTV cameras were unable to fully adapt to the levels of light. In addition to low light levels, some CCTV cameras were positioned too close to lights, which over-exposed camera lenses and created strobing and glare in the images. In fact, four out of the seven residential cameras suffered from inappropriate lighting levels at night, two of which were so dark the images were *"virtually useless"* (Gill et al., 2005, p 27). The problems with camera lighting at night affected operators when carrying a monitoring task in real-time and also affected the police when analysing post-event recordings.

4. A number of control rooms were found operating a high camera-to-operator and camera-to-monitor ratio, which meant that too few eyes were looking at too many cameras and monitors. This reduced the *"…probability of spotting an incident or providing usable recordings"* (Gill et al., 2005, p 14).

5. In the control rooms which recorded analogue CCTV (n = 6), video was recorded at low-quality as a result of tapes were being re-used far too often.

6. In the control rooms which recorded video digitally, the video quality was also poor. The problems with video quality were apparent at eight of the control rooms (both analogue and

digital systems). Six control rooms recorded video at 1-2 fps and two control rooms recorded at 1 frame every 3-5 seconds. At vast majority of the control rooms, Gill et al. found that the video quality was too low to aid the police in their investigations or even be used as evidence in court.

7. Operators and management had limited knowledge of digital technology (recording settings) in the case of two systems which consequently compromised the effectiveness of the system.

For both the analogue and digital systems, it was found that the biggest determinants for using these recording rates was down to equipment, cost, and the advice given by CCTV consultants. The problems with the use of low-quality video for an investigative task was a result of CCTV managers making uninformed purchasing decisions when deploying digital CCTV recording systems. For instance, managers would choose one system over another based on the available budget, rather than identifying the goals and requirements for the system.

There is a lack of knowledge regarding CCTV video quality requirements when used for security observation tasks, and in the absence of such guidance, security system owners make decisions which have a negative impact on the system performance. Thus, this issue is investigated in this thesis through field and empirical research (see Chapters 4, 8, and 9).

2.3 Non-Security Control Room Studies

Luff and Heath (2001) carried out naturalistic observations within in one particular type of control room: station control rooms in the London Underground. In-depth field work which involved making audio-visual recordings were carried out in the major stations of the London Underground. The study specifically examined how operators maintained situation awareness: how they monitored their surrounding domain and the activities of their colleagues. The results showed that operators made use of a wide range of communication and information technologies: radios, passenger announcement systems, train information, emergency control, and alarm systems. The technology most used was the banks of video monitors that displayed CCTV video in real-time (4-12 monitors per bank with up to 80 CCTV cameras in total). These monitors displayed video of various public areas within the London Underground such as train platforms, corridors, passageways etc. Control room operators (station supervisor as described by the authors) were responsible for utilising technology to monitor suspicious and troublesome behaviour within the station and track targets smoothly from one camera to another.

The observations revealed that operators were not able to monitor platform scenes effectively, and this was because the images were not always clear due to a number of technical observational difficulties. Operators struggled with their tasks due to the low-quality CCTV images observed on-screen, and this was due to the following problems:

1. Limited lighting provided near to the CCTV cameras.
2. Dirty CCTV camera lenses from train break dust.
3. Dirty video monitors within the control room due to a lack of cleaning.
4. Monitors kept on the same view for long periods of time which caused a burnt-out effect.

As a result of poor maintenance of the CCTV systems, operators struggled to monitor scenes. For example, they were unable to distinguish whether a crowd was stationary or moving. Furthermore, operators were unable to follow targets smoothly from one camera to another due to camera blind spots created as a result of hundreds of years of gradual station development. Luff and Heath believed that the problems operators faced when dealing with fragmented images was *"inevitable in the video coverage of a large and complex station"* (p 158).

Operators reported that they found it difficult to visualise scenes at the end of train platforms. In fact, certain individuals such as beggars and buskers discovered some of the camera blind spots, and frequently placed themselves outside of the CCTV camera views. The analysis of these tasks revealed that operators used various technologies to monitor fragmented and difficult to see station scenes. Given these problems with surveying scenes, it is not surprising that operators struggled.

Following 12-months of field work, Luff and Heath made a number of general design suggestions to improve operator task performance specifically when working on within a London Underground control room:

1. Extend the existing London Underground systems so that they are more integrated. This was proposed to reduce the need for: maintenance, need different controls/connections, and information handling.
2. Configure the technology appropriately to: allow for a selection of a sequence of images so that incidents can be easily tracked, monitored, and to support the invocation of appropriate 'next actions.'
3. Introduce automated surveillance technology to detect events such as crowding.
4. Make relevant information about the station accessible to operators in different locations of the control room to improve the efficiency in locating scenes on CCTV.
5. Implement a touch-screen interface allowing operators to quickly select camera views on their video monitors without delaying the situation assessment and response.

Although these design recommendations can be applied to the design of all types of CCTV control rooms, Luff and Heath fail to discuss an important aspect of carrying out tasks within such a complex work environment: *context* and its impact on the design of systems. CCTV owners and managers therefore could benefit with additional recommendations which take into account the various contextual factors surrounding the operator's work system (see Chapter 3, Section 3.3.2.2, Table 3.1).

In another transport control room study (Chen, Choi, Ruiz, Shi and Taib, 2005), in-situ observations, interviews and questionnaires were carried out within a traffic control room to gain insights into how operators' performed their work. The focus of the field work was to understand how operators utilise multiple hardware equipment and software applications when managing mobility and public safety related incidents within a control room. The results were used to formulate use scenarios[1] and then implement a mock-up application to improve operator performance when dealing with traffic incidents.

The main limitation with this study is the lack of background given on the research methodology and data analyses (this is discussed in more detail in Chapter 3, Section 3.5). Furthermore, in review of the study findings, it is unclear what tasks operators performed in the control room and what their work environment was like, as no discussion is given on the field study observations. Chen et al. instead placed more focus on the subjective findings gathered from the interviews and questionnaires and used this data to inform the design of a new mock-up system. Interviews with 14 control room operators revealed a number of the subjective preferences with the regards to the current information system they used to manage traffic incidents and specifically to locate scenes:

- 75% of operators were *"happy with their working environment"* (p 2), but several said that there should be a better way to integrate applications to allow critical data to be retrieved more easily.
- 70% of operators would have liked a map-based search tool for accessing geographical information.
- 33% of operators preferred a personalised way to filter information.
- 70% expected simpler procedures to login and log out of their systems.
- 60% thought a speech interface would be beneficial and 21% thought a multimodal (speech, gesture, touch, and others) would be beneficial.

These findings were then examined with 11 control room operators using a questionnaire. The two main issues reported in the questionnaire related to the problems with the contact pages (database), as the data was: 1) out-of-date and 2) inconsistent. The third biggest issue was the slow search speed at which the system operated; and this slow speed was *"potentially due to an indirect result of ineffective navigation tools"* (p 2).

[1] Use scenarios (also known as use case scenarios or usage scenarios) is a description of a system's behaviour as it responds to the user's actions.

Once the field research was complete, a mock-up user interface (web based) was created to validate the findings. The aim of the mock-up system was to examine whether operators could be better supported when handling incidents.

The design involved integrating a number of company policies into the flow of information within the user interface. This integration was included in the design to reduce operators' cognitive load when performing tasks with the system. A browsing navigation was proposed rather than a stand-alone search function so that 'entry point' information can be found using an advanced search function. Other features included the use of: 1) expansion of fields when completing information (this reduced the need to scroll excessively); 2) completion markers (in the form of checkboxes) to allow operators to visualise completed actions; and 3) error handling (the checkboxes trigger warnings when actions are incomplete).

The mock-up user interface was evaluated with six operators using a real-life traffic incident scenario. Operators were timed in their tasks which required the retrieval and handling of contact information with both the existing contact system and the mock-up system. In addition to task performance, operators were asked to rate the system under several criteria using a 5-point Likert scale. The results of the evaluation were described by the authors as 'preliminary.' The overall findings from the evaluation revealed that the mock-up system was more efficient (37% overall improvement in task time completion), despite operators receiving no training with the new system. Chen et al. believed this finding was a result of introducing the process flow and progress markers within the user interface. The subjective results were in line with the task performance data. Operators preferred the mock-up system as they were able to enter specific geographical locations and perform searches by entering specific criteria into the database system. There was a strong preference for the new proposed system in terms of: easy to learn, ease of use, intuitiveness, speed and accuracy of retrieval, helpfulness, level of comfort, and overall its effectiveness. The *integration* of information sources was considered important for improving operator performance in the London Underground control rooms (Luff and Heath, 2001).

The study by Chen et al. (2005) examined task performance issues within a real-world context, and like the study London Underground study by Luff and Heath (2001), both studies examine one specific task: locating a scene. As both these studies examines one specific task within one specific control room setting, the findings are limited and therefore do not provide a broader understanding of the task performance issues within control rooms where CCTV and other related technology is heavily used. For example, the interaction between different control room stakeholders was not examined, nor was there a discussion of the interplay between users when using communication devices such as radio and telephone.

In contrast to the study by Chen et al. (2005), McCarthy et al. (1997) carried out a 12-month ethnographic study to *compare two different systems currently being used by operators* within two ambulance controls rooms. The field research involved field observations and open-ended interviews with operators at two different ambulance control rooms. The main objectives of the field study were to: 1) understand how operators performed a specific task (*locating a scene* of emergency in order to dispatch an ambulance); 2) compare the effectiveness of tasks performed by operators at each control room; and 3) use these findings to assess the potential usefulness of the technology used at one control room if it was to be used at another control room (where technology is not heavily used). The control rooms which were evaluated included:

- Ambulance Control Room 1 ('ACC1'):
    - Located within an urban environment.
    - Operators were supported by a 'high technology system.'
    - Tasks involved: receiving calls, entering jobs electronically into a system, locating scenes using a database system, passing the information to a gazetteer (a reference tool for information about places and place names), which identifies the possible location, confirms the location to the caller, and then passes to the dispatch team electronically.

- Ambulance Control Room 1 ('ACC2'):
    - Located within a rural environment.
    - Operators were supported by a 'low technology system' (a manual system to locate a scene).
    - Operators are given access to a database of telephone numbers of public telephone boxes, local general practitioners, and other important landmarks and locations.
    - Tasks involved: receiving calls, locating a scene of emergency using a paper map of the area, and then writing the job details down on note paper.

The analysis of the field research revealed that the most critical activity operators performed was identifying a scene of emergency. This task was described as highly complex since postal addresses were difficult to identify (for e.g., some houses were not numbered, some street names clashed etc). McCarthy et al. described the nature and context of this scene locating task at each of the ambulance control rooms in detail.

Similar to the study conducted by Chen et al. (2005), field research was carried out to identify the problems operators experienced when locating a scene within a control room environment. McCarthy et al. found that task performance was severely affected as a result of ineffective communication between public callers (those requesting an ambulance) and control room operators (those receiving calls from public callers to dispatch an ambulance). Communication was ineffective for a number of reasons: 1) pubic callers had strong regional accents; 2) public callers provided operators with insufficient location descriptions; and 3) operators lacked experience in locating scenes of emergency. McCarthy et al. suggested that some technologies, such as the gazetteer and an electronic map could compensate for operators' lack of direct access to the knowledge on the local geographical areas. This particular control room provided detailed insights into the nature of ambulance control room work; however the study findings are largely task descriptive, and does not provide a detailed discussion on the implication of these problems on operator performance and system effectiveness.

To conclude, there have been a number of field studies within control rooms which examine the nature and context of operator tasks. These studies were carried out with the aim of improving system design, apart from the security control room study by Gill et al. (2005) which examined the overall effectiveness in terms of crime reduction. In review of these studies, there is a common theme: all involved operators performing what would be widely known (by operators and other researchers) a 'reactive task' whereby the operator is involved in locating a scene using CCTV cameras, video technology, and maps based artefacts.

Reactive tasks are time critical tasks and the ability to perform these tasks effectively and efficiently depends on two factors: 1) how well an operator's work-system is set-up and configured and 2) the effectiveness of communication (verbal and technical) between operators and other users. The transport control room studies (Chen et al., 2005 and Luff and Heath, 2001) did not examine the role and effectiveness of technology in much detail across a wide range of control rooms. Instead, field research was limited to 1-2 control rooms to examine a specific task with the aim of improving system design. Furthermore, secondary tasks such as administrative paper work and making copies of CCTV evidence were not mentioned in these studies.

Based on the previous control room studies, a number of important research questions were identified for further investigation in Study 1 (see Chapter 4):

1. What tasks are performed by CCTV operators where CCTV and other associated technologies are used within a control room environment?
2. Do CCTV operators have a good understanding, knowledge, and experience in their tasks and the surveillance areas?
3. What factors do CCTV operators use to aid their situational awareness?
4. How do CCTV operators communicate and collaborate with other CCTV control room stakeholders, and is it effective?
5. Are operators provided with integrated (e.g., a geographical information system linked to CCTV cameras) or segregated tools?

It is clear that, a better understanding is needed on the nature and context of CCTV, what tasks are performed by operators and what problems they encounter, and it affect on other stakeholders and the technical parts of the system – thus the socio-technical impact. This understanding is important so that it can then be used to improve the effectiveness of CCTV system design, configuration, and use.

2.4 Control Room Guidance

The guidance on physical control room design is important in improving the CCTV operator's health, safety, comfort, and performance. Currently, the BS EN ISO 11064 (ISO, 2004) – 'The Ergonomic Design of Control Centres Standard' is the only guidance available on the design of control centres. The standard is based on ergonomic principles and offers guidance on the physical aspects of control rooms, such as workstation arrangements, control room and workstation layout, use of displays and controls, and maintenance. Overall, much of the ergonomics guidance for workplace environments apply to office environments, however the standard does consider workstation design and human factors affecting operator performance within control rooms. For instance, the standard recognises that an operator is likely to adopt a typical viewing distance to the screen between 750-1000 mm. In contrast to typical office tasks, the typical viewing distance would be much closer: 500-600 mm. The reason there is a difference in monitor viewing distance in an office environment and a control rooms is because in control rooms, a typical array of equipment on an operator's workstation is likely to be 2-5 large screens which are viewed in parallel. The size of this array and the requirement to have an overview of all of the screens requires a 750-1000 mm distance (Wood, 2001). The standard also points out that if a viewing distance of a visual display increases, it will have a direct impact on font selection so that characters subtend the recommended angle of visual arc at the eye (minimum 15 minutes of arc). The ergonomics requirements for the use of off-workstation and shared displays is also given in the form of ideal viewing angles and direct sightlines.

Despite the wide range of ergonomic guidance on display and equipment use within control rooms, separate guidance is need to provide CCTV practitioners and users on the design configuration and set-up of CCTV-specific technology, such as cameras, recording quality etc. This type of guidance is needed as it considers ergonomics in two ways: by taking into account both the: 1) physical and 2) cognitive factors associated with control room and CCTV design, both of which have a notion on task design. The distinction between physical and cognitive ergonomics is next described.

Physical ergonomics is concerned with the anatomical, anthropometric,[2] biomechanical,[3] and physiological characteristics of human users, thus relates to a user's physical activity at work. In the context of a security control room, the following aspects of an operator's physical workplace should be considered during the design process: the operator's working posture, repetitive actions, and the layout of their personal workstation. These factors should be considered in the control room design process in order to achieve high task performance, minimise discomfort, and promote a good health and safety culture. As control rooms are shared work environments, the design of an operator's physical environment should also be considered, for example, required levels of lighting, temperature, space and the minimisation of noise to ensure that operators achieve high performance and experience minimum

[2] Study of human body measurements for the purpose of designing a user's physical environment based on their body size and characteristics.

[3] Study of a user's body mechanics for the purpose of considering the physical design of tools and systems that require physical exertion.

distractions, discomfort and job stress. The BS EN ISO 11064 (ISO, 2004) addresses these aspects of physical ergonomics for the design of control rooms.

Cognitive ergonomics is concerned with the human mental processes (human cognition) which relate to performance within the workplace. Human cognition includes attention, memory, audio/visual perception, decision making, and motor response. These aspects of human cognition can influence the way in which tasks are performed and how decisions are made, as well as many other interactions that take place between users and the system. There is little understanding about the cognitive factors surrounding CCTV video (i.e., impact of low video quality on visual perception, decision making and motor responses).

Hollnagel and Woods (1983) argue that human-machine systems should be analysed, conceived, and designed as cognitive systems which produces 'intelligent action:' its behaviour is goal-oriented and the user interacts with it using heuristic knowledge. A number of cognitive models of human operators have been developed in the process control industries to help researchers understand how operators interact with systems in monitoring and control tasks (Bridger, 1995). These tasks within these settings are described as *cognitive control tasks*, as the operators deal with facts and rules - unlike CCTV operators who respond to targets of interest in the perceptual mode of interaction (Bridger, 1995). Further research is needed to examine the context of CCTV security, to discover *how* operators think and perform when making observational decisions when responding to targets and events from CCTV video, as well as how they communicate and collaborate with other CCTV stakeholders.

2.5 Chapter Summary

In this chapter, a critical literature review of previous control room studies is given. The control room studies are reviewed under two sections: security and non-security control room studies. In addition to this, a review of the current control room standards and a background to control room ergonomics is given.

Previous studies conducted within control rooms examine the nature and context of operator tasks – many of which examine just one particular task: locating scenes using CCTV video and other types of technologies. One of the main findings in the security control room study (Gill et al, 2005) was that 'CCTV is not effective' in lowering crime and the public fear of crime. Although this study did not specifically focus on the technical failures of CCTV technology, Gill et al. identified a number of technical problems which affected operator performance when operators observed and reviewed CCTV video footage. Firstly, operators were expected to monitor a vast number of CCTV cameras and video monitors at one time. Secondly, when CCTV video is used for post-incident investigation work by the police, the recorded video at a majority of control rooms was very low. In addition, it was found that in most cases, the deployment of a CCTV system is decided based on the funding available and partial advice from security consultants. Gill's evaluation was conducted during the turning point of CCTV technology. Since then, there have been a number of important developments, which have changed the way in which CCTV is being used. There is now a broad range of CCTV tools and systems to support the security tasks of a diverse collection of CCTV owners and users. The impact of these changes within a security control room environment requires an evaluation to understand what technology is being used and how effective they are for operator tasks.

In two of the non-security control room studies (Luff and Heath, 2001 and McCarthy et al., 1997), the failure to maintain the CCTV camera work environment and equipment (e.g. video monitors) reduced operators' situational awareness when identifying scenes on CCTV. Furthermore, audio communication between operators and individuals outside of the control room was found to be ineffective. Chen et al. (2005) demonstrates that, field research data collected within a control room environment can be used to inform the design of new systems. However, structured field research is needed to understand the overall context of CCTV, its effectiveness, and the wider implications of poor performance.

This review has shown that there is currently no HCI description or understanding of CCTV operator tasks and this is needed as a basis to improve operator performance and the design of technology used within the control room to support operator tasks. This research is presented in Chapter 4 and the findings were used to develop best-practice guidance on CCTV system design, configuration, and use (see TEC-VIS in Chapter 10).

Chapter 3
Field Research: Methodology

"Failure to take account of context can lead to failure of large scale systems." **(McCarthy et al, 1997).**

In order to understand how users interact with CCTV, it is essential to observe them in context and talk to them to identify the key problems associated with the technology they use to accomplish their tasks (Holtzblatt and Jones, 1993). It is important to take context into account in developing an understanding of work in the field, as tasks are not only shaped by the tools and systems used, as noted by Carroll (1990), but by a wide range of social, organisational and environmental factors. This chapter details the two main field methods used in field research for assessing how operators perform their tasks in their workplace and what problems operators experience: 1) observations and 2) interviews. These field inquiry methods are used to develop a detailed understanding of work processes, failures, and technology. This chapter identifies the appropriate methodology for the CCTV control room field study (Study 1, see Chapter 4) conducted for this thesis.

3.1 Introduction

In the discipline of HCI, there has been an increased awareness of the need for the design process to be driven by context research. This was considered important in the design of systems to understand users, how their work is performed and the external influences surrounding their work activities. When design is led by context research, system designers and developers are provided with a rich and deep understanding of the system's users and their work activities.

Context research is achieved when a researcher goes into a user's place of work, to identify the processes and problems within a socio-technical system. Field work in HCI research is conducted specifically to gather user requirements to identify the tasks which users perform with the system (Norman, 1983). This data is gathered through observations and direct questioning with users which is used to form a task model to inform the design of a system. In addition to identifying the users' task model, and challenges in their work - technical failures can also be identified. This information is then used to creatively design and test a new system to overcome the challenges and improve task performance effectiveness and the overall system design (ISO, 1997; ISO, 1999).

In the context of a security control room environment, it is important to study the different interactions between the social (i.e., different CCTV stakeholders) and technical systems (e.g., CCTV controllers, cameras, maps and intelligent interfaces, etc.) to understand what tasks are being performed by CCTV operators, how they are performed, and the problems that affect task performance. By studying these interactions through context research it is possible to examine how well the overall security system is operating in terms of its functional capabilities. Furthermore, it is possible to identify overall whether the system is fit for its intended purpose.

3.2 Interviews

Rich qualitative and quantitative data can be elicited from users through interviews. This research method is used in HCI research to gather users' insights, opinions, and attitudes towards a system they interact with and about their workplace practices. Interviewing is flexible as the interviewer can change the questions asked during the sessions to gather specific information from particular users. Interviews are also participatory, as they require both the interviewer and the user to participate in an interactive conversation about their work. There are three main interviewing techniques used in HCI research (Neilson, 1993):

1. Unstructured interviews
2. Semi-structured interviews
3. Structured interviews

3.2.1 Unstructured Interviews

Unstructured interviewing is typically used during the early stages of a system evaluation to explore general issues, and therefore to understand context at the outset of a long-term study. The interview dialogue between the researcher and user is based on informal conversations. Prior to the interview, the interviewer prepares a number of topics they wish to learn about the user and their work. There can be some focus on the subject matter, however, the interviewer does not necessarily ask specific questions; instead, broad topics are presented for open discussion. The aim of unstructured interviews is to gather as much background information as possible about the user's current experiences and expectations about the: work they perform, environment, systems used, and the various people with whom they interact with. The information gathered from unstructured interviews can be very rich, therefore time consuming to analyse. The main practical limitation with this technique is that the researcher may unintentionally ask leading questions which may result in losing focus of the research objectives.

3.2.2 Semi-structured and Structured Interviews

Semi-structured interviewing involves the researcher using a focused questioning design, which is guided by a script that has some degree of flexibility. Specific issues of interest to the researcher can be explored in more depth by asking the interviewee to elaborate and provide additional information. The semi-structured elements of this interviewing technique can increase the reliability of research data allowing for quick and easy analysis (Cooligan, 1999).

Structured interviews on the other hand are designed with a specific and predetermined agenda, whereby set questions are used to guide and direct the interviews. Structured questions are designed to elicit short and specific responses from users so that quantitative data can be easily gathered. This is a particularly useful technique when there is very little time and access available to interview users. This method is also useful for comparing data from several work environments of the same type. For example, Gill et al. (2005) achieved this in a security control room study to compare data between 14

CCTV control rooms. In this study, quantitative data was gathered from the structured interviews with control room staff, such as: hours of work, type of technology used for recording CCTV video, number of CCTV cameras, and crime rates.

Depending on the style of questioning the interviewer takes, there may be problems associated with conducting structured interviews with users at work. For instance, interviewees may perceive the 'structured' nature of the interview as intimidating and too formal. This is likely if users are expected to discuss the social aspects of their work. On the other hand, there are benefits to using a structured interviewing style. Firstly, structured interviews are tightly scripted, therefore the data is easy to collect and analyse. Secondly, a structured interview can be repeated at a later date using the same interview protocol and data can be compared over time.

For all types of interviewing, the length can be relatively short (typically lasting 20 minutes to one hour) and conducted within the users' workplace; however, a noise-free environment is essential for capturing as much detail as possible. For security research, it is important that interviewees be put at ease at the start of the session to encourage open discussion. It is also fundamental that background information about the workplace setting, the interviewees and the context of the work is gathered in advance of the interview sessions. Interview questions and topics should be detailed in a logical order. Furthermore, leading questions that make assumptions, as well as long and complex questions should be avoided. Forming bias opinions during interviewing (e.g., stereotyping genders) can happen quite easily, and should also be avoided.

Within security work environments (such as police and CCTV control rooms), responses to interviews questions may be limited as the nature of the work is highly sensitive and personal information is frequently discussed. Therefore, interviewing security personnel could be challenging for researchers, as access to some places and information can be difficult to obtain.

3.3 Observations

Observations have become an increasingly popular field method in HCI research. Naturalistic observation is traditionally used by anthropologists and psychologists to study people and their behaviours in their natural habitats. In HCI research, observations are used to extract meaning from users at work when performing their everyday tasks. The aim of this type of observation is to gain an understanding into *how work is performed*, *how tools and systems* are used, and the problems with task performance. The main distinction between naturalistic observation and other forms of observation in the field is that the events must be observed and recorded unobtrusively; thus, the researcher must make every effort to observe users without intervention.

To gain a close and intimate familiarity with individuals interacting within a natural setting, participant observation can be used. Participant observation is one of the most common approaches to data collection used in field research, since it *"…provides answers to the contextual questions that cannot be answered by interview alone"* (Morse, 1994, p 31). Participant observation is a naturalistic method of observation typically used for research conducted over a long period of time (typically ranging from six months to a number of years) and is a useful method for studying interactions taking place within a particular stakeholder group. Due to the complex nature of CCTV control room environments, it is difficult for a researcher to observe events and extract meaning from users effectively over long periods of time. Furthermore, participant observation would be difficult to conduct within a security work environment, as the observer (if inexperienced in security) would need to learn the language security professionals speak and act as naturally as possible without posing any risks to the workers or the organisation.

3.3.1 Ethnography

Ethnography is an observational technique that is well established in sociology research. It is used to uncover the social organisation of activities (Sharp et al., 2007). Traditionally, ethnographic studies conducted in sociology research have been conducted from a particular theoretical viewpoint and for the purpose of contributing to theory (Paay, 2007). However, ethnography is now regarded as a common approach to HCI research and design. This practice has been motivated by a growing need to design for complex real-world situations. More specifically, ethnography has been used in HCI as a field research technique to elicit requirements for the design of new products and systems across a wide range of work environments.

In the past, ethnography research has been used for longitudinal single case study research, typically involving participative observations within one particular setting and context over a period of six months or more. Longitudinal ethnography has been used in studies examining several different types of control room environments. For example, air traffic (Bentley et al., 1992 and Twidale et al., 1994); underground transport (Luff and Heath, 2001); ambulance services (McCarthy et al., 1997; Blandford and Furniss, 2005) and a financial trading floor (Heath, Jirotka, Luff and Hindmarsh, 1993).

The key issue with conducting longitudinal ethnography research is the practicality of carrying out observations within certain work environments:

1. Gaining access to a workplace over a long period of time may not be possible in some situations, as the presence of a researcher in the workplace may appear to be too invasive. This is particularly an issue when observing users in security work environments.

2. Conducting field observations can be resource intensive. To gain as much insight as possible for longitudinal observations, more than one researcher is needed to observe and take notes.

3. Conducting observations over a very long period of time may not be practical for the researcher or for the workplace owner/manager.

As a result of the practical resource and time constraints, various modern ethnographic techniques have emerged. Hughes, King, Rodden and Anderson (1994) describe several different applications of ethnography which can be used as a field research method to improve system design:

1. Concurrent ethnography: The design is influenced by ongoing ethnography, which is carried out in parallel to the development of the system.
2. 'Quick and dirty' ethnography: Shorter ethnographic studies are conducted when time is limited. Data from these observations can be used to inform designers with a general background of the workplace and the activities being studied to help design new or improved systems.
3. Evaluative ethnography: Ethnographic studies are carried out to verify or validate a set of pre-formulated design concepts and ideas.
4. Pre-examination of previous studies: Previous studies are re-examined to inform initial design thoughts and concepts.

As has been discussed, conducting research in the field is time and resource consuming, therefore costly. Millen (2000) argues that very often researchers are confronted with the demand of spending time in the field carrying out the research while, *"…matching the pace of ever-quickening product development cycles"* (p 285). In light of this view, Millen introduced a new form of ethnography: rapid ethnography, which aims to work around these time pressures. This form of ethnography is given a different term to the 'quick and dirty' ethnographic technique, yet both these techniques are described as the same and regarded as a means to*"…fill the gap between short design cycles and the long, complex nature of ethnographic research"* (Ramachandran, Kam, Chui, Canny and Frankel, 2007, p 2).

The main criticism in attempting to cut time and the methodological processes in gathering field data is that there is a major risk that the researcher will lose sight of observations that may be important to the study objectives. Furthermore, time-saving techniques may not work for some research projects where the work environment is large and complex (physically large and where complex technologies and processes are employed), and also consists of many stakeholder groups. If more time is made available for field research, there are indeed more opportunities for the researcher to observe events and experience the changes within the work environment (O'Reilly, 2005). Adequate time is needed to carry out field research to build an understanding of large and complex socio-technical system in a holistic manner, to gain a detailed understanding of the users, their tasks and processes, stakeholder relationships, and well as the problems users experience during task performance. Typical large and complex work environments include: public CCTV control rooms, large nuclear power plant control rooms, a major airport security control room etc.

By modifying traditional ethnographic research in this way, field research data is gathered more effectively and efficiently. The data can then be used to support the design of systems. To achieve this however, new and emerging ethnographic methods need to follow a *structured methodology* to ensure that specific objectives of the research are met and the results can be applied to solve real-world problems in the workplace.

Contextual inquiry (Beyer and Holtzblatt, 1998) is an adaptation of several field research techniques taken from disciplines such as anthropology, sociology, and psychology. The technique provides concepts to guide data gathering and the analysis of field research which considers the various contextual factors surrounding the user's work. This research method was used for the field research detailed in Chapter 4, but with the addition of a structured observation checklist (see Appendix C) to help focus the observations. The use of a structured checklist was a useful means of identifying and presenting the results in a meaningful way. The following sections describe the contextual inquiry method in more detail. Also discussed, is the justification in using this research method and the specific methodological approach taken for the field study in this thesis.

3.3.2 Contextual Inquiry

The various contexts in which a given system will be used should be considered at the very earliest stages of system specification and design; and this can be accomplished through contextual inquiry (Beyer and Holtzblatt, 1998). The principles and practices of this technique were developed commercially at Digital in 1986 by Hugh Beyer and Karen Holtzblatt. The technique was developed based on building system models while keeping the users' intentions, concepts, and workflows in mind.

Being aware of the various contextual factors of a system or product is important to the development process in order to determine whether the system meets the needs of its intended users. This participation is achieved by allowing users to articulate their current work practices, tasks, challenges and experiences. The essence of this inquiry is the use of direct observations and semi-structured interviewing with users whilst they perform their tasks at work. Understanding users at work is represented through models of current work practice, descriptions of their work, and task performance failures (related to the user's task, skills, processes, work environment, technology used, and communication and collaboration with other stakeholders).

3.3.2.1 The Concepts

This method does not follow a collection of steps for gathering and interpreting user information, but a set of concepts to guide the process of data gathering and interpretation. These concepts are described by Holtzblatt and Jones (1993) using three core principles:

1. Context: For the design of a new product or system, it is important to consider the various contextual factors which will affect a user's performance when interacting with the system (see Table 3.1). In the discipline of ergonomics, standards such as the BS EN ISO 9241-11 and ISO 13407 (ISO, 1997; ISO, 1999) recognise the role of context within usability and user centred design, and define usability in relation to context as *"…the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use"* (ISO, 1999, p 2). The standards define the process of understanding and specifying the context of use as one of the main stages of a human-centred design process (Maguire, 2001a). In fact, the usability of a product depends on its context of use, and products should be designed for specific contexts (Maguire, 2001b).

The actual process used to understand the contextual factors for a system through contextual inquiry involves a researcher going out into the field to talk to users whilst they perform their tasks. Beyer and Holtzblatt (1998) found that if users are interviewed informally about their work, they will describe their experiences in abstractions and summaries, rather than in rich detail. By observing and interviewing users whilst they work, contextual value is added to the research, allowing for richer descriptions. Another benefit of interviewing users as they perform their tasks is the ease of recalling previous experiences when and where they are relevant.

2. Partnership: It is also important to form a connection with the users who are being observed at work. To build an accurate picture of users' tasks, observations should be active (rather than passive), by developing a dialogue with users. The principle of partnership is recognised through dialogue so that designers can become aware of a user's experience of work and tool use (Holtzblatt and Jones, 1993).

4. Focus: By focusing on the users, it is possible to create and direct an understanding of the situation, environment and the users' tasks. The purpose of this focus is to allow the researcher to ignore the minor and unimportant aspects of the environment they observe (e.g., organisational politics, personal problems outside of work etc.). By fixing the focus of the inquiry, the researcher will be able to interpret observable facts that relate to the users' work and their working environment. Having focus in the field inquiry also allows the researcher to focus on gathering data related to the research objectives.

3.3.2.2 The Process

The contextual inquiry process involves the identification of the users and the workplace being studied. Once this is established, a framework for the contextual inquiry should be developed which defines the structure of the field visits. Within this framework, several facts need to be established through an *introduction*, whereby the researcher tells the site owner, manager, and users who they are and why they are there to assure confidentiality and anonymity (also seeking permission for recording). Users, of course, will not be provided with an explanation of the background of the research objectives, but will be told that the observer is interested in learning about their work, their experiences, and how they use their tools and systems. It is also important to inform users how long the researcher will be present. The contextual interviews should then begin.

The researcher should start by asking the user to provide an overview of their work, the tools they use, and their opinions on their work processes, procedures, and specific tasks. During the observations, the researcher should try to understand every user action and goal through active observation and inquiry (using a semi-structured set of questions).

As people perform their work without thinking about the specific processes, they are likely to talk about their work in abstract. It is the researcher's job during the contextual inquiry activity to form an understanding of users at work and build a rich picture of their experiences and actions. This requires the researcher to gather concrete data from the field to form task models (a detailed representation of the tasks users perform). One approach developed for the research conducted in Chapter 4 was the SEE technique – which aims to gauge more detail from users when they describe their work or a particular issue:

1. **S**pecify
2. **E**laborate
3. **E**xemplify

In the context of a CCTV control room field study, an operator may say, *"... here, we tend to do this."* The researcher should then inquire further to understand the context and the meaning behind this statement by responding:

> *"That's interesting, what do you mean when you say here? Do you mean at your company, or within your team? What specifically do you tend to do? Tell me what you mean. I think I understand. Can you think of any examples of when that has happened?"*

Also, when the user is interacting with the system, the researcher should make every effort to understand the tool (e.g. what it is, why it is used, who else uses it, and the problems with using the tool etc.) and if possible take photos and video during user interaction. Table 3.1 provides an example of the types of inquiry questions and the different contextual factors that should be considered in field research. Once these questions have been answered, any assumptions related to the focus of the research should be verified with users based on the interview responses and observations made.

Table 3.1: Example Questions and Contextual Factors

Example Questions	Contextual Factors
Who would use it?	- User ▪ Profile and roles of users ▪ Skills, knowledge and experience ▪ Training and qualifications
What task are you performing **What** tools are you using?	- Task ▪ Goal, output, and physical and cognitive demands - Equipment ▪ What is being used, where is this located and duration used
When would you use the system?	- Task: Frequency, duration, trigger, dependencies, risks of error, health and safety, information and resources needed, criticality of task and procedures.
Where would you carry out these tasks?	- Environment: ▪ Social environment: location of system and performance of tasks, impact of use with other users, assistance, interruptions, communication. ▪ Physical environment: temperature, visual conditions, noise, space, layout of workspace and workplace, user's posture.
Why are you doing that?	To understand other contextual factors related to the work users are involved in such as social, psychological, technical, economic and political factors.

The basic framework of a contextual inquiry consists of an introduction, contextual interviews, and the validation and invalidation of assumptions. Finally, the field visit should end with a wrap-up, which involves a summary of what was learnt from the observations and interview responses. The wrap-up allows for the clarification of uncertainties or issues that were not raised during the inquiry. The sessions should then be closed by thanking staff for their time and providing feedback. Once the field visits are complete, the workplace should be revisited to verify and clarify the findings. Following this data gathering exercise, a short discussion should be conducted with the site owner or manager about the results, and the researcher should provide a summary of the major understandings.

3.3.2.3 The Analysis

The analysis of contextual inquiry data, according to Holtzblatt and Jones (1993) takes place both during the contextual interviews with users and after the field study is complete. This involves the researcher co-interpreting users' actions and experiences. On-site field research analysis is however not a suitable method when studying several work environments of the same type (e.g., CCTV control rooms), as there may be many different types of findings discovered from each visit. Furthermore, conducting an analysis and forming conclusions on-site is a risky approach as observations and user feedback may inadvertently be misinterpreted. This is likely to happen when researching a vast and complex work environment.

The data gathered through a contextual inquiry is mainly qualitative, but may also include quantitative data. Holtzblatt and Jones (1993) suggest that it is useful to interpret data by recording understandings. For the field research conducted in Chapter 4, this can be achieved methodologically [list adapted from Beyer and Holtzblatt's (1998) principles]:

- Describing the various CCTV stakeholders and their tasks.
- Describing the operators' work activities for the benefit of conducting a task analysis.
- Compiling a list and detailed description of the various artefacts used during task performance.
- Compiling a list and detailed description of the problems and failures associated with task performance when using artefacts (task performance issues).
- Compiling a list and detailed description of the problems and failures operators experience when communicating and collaborating with other CCTV stakeholders (stakeholder conflicts).

Data gathered from a field study can amount to large amounts of note paper and if each finding is written on a post-it note, there would be hundreds. These post-it notes can be used to form an Affinity diagram (described as the 'wall method' by Beyer and Holtzblatt, 1998) to visualise common themes and issues (Figure 3.1). The process of grouping field notes involves selecting an arbitrary piece of information and deciding what it means and what it relates to by placing it into an unlabelled group on a white board or wall space. As the groupings emerge, individual items (the post-its) move between groups until they fit into defined categories. It is not until the items are placed within groups that they are appropriately labelled. Once the groups are defined and appropriately labelled, the items can then be grouped hierarchically. Creating an Affinity diagram not only allows vast amounts of data to be organised, but also, *"represents an effective means of communicating the emerging system vision. This information can be used in requirements documents, decision-making processes, and to focus design meetings"* (Holtzblatt and Jones, 1993, p 204).

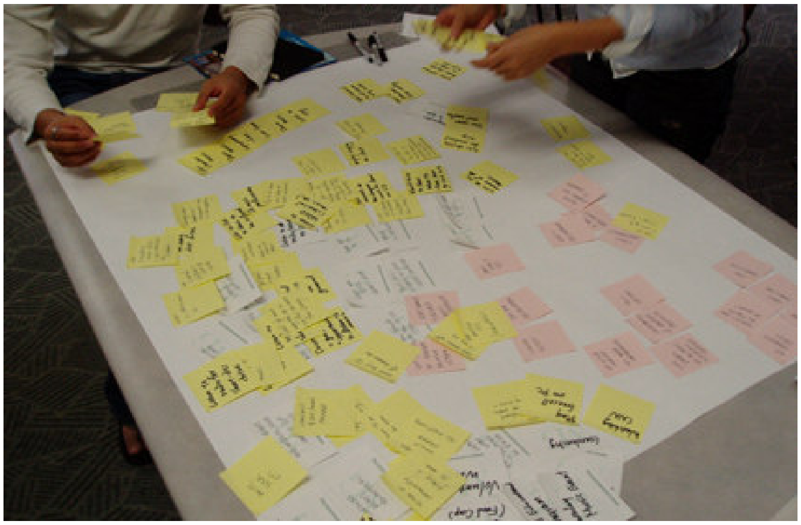


Figure 3.1: An example of an Affinity diagram.[4]

3.4 Summary of Field Research Methods

In summary, there are two main field research methods used in HCI research: interviews and observations. These methods can be used together to gather rich data about users who interact with a system. Ethnography is a field research method which allows researchers to elicit rich qualitative data through naturalistic observations and interviews to gain a deeper understanding of work performed by users in a real-world setting. Ethnography is a method of inquiry as it answers not only the 'what' questions, but also addresses the 'why' questions (e.g. *"Why are tasks performed in this way?"* and *"Why do users communicate in that way?"* and so on).

Ethnography has been used in prior research for single case studies to understand the social and work interactions of users within specific work domains (e.g., offices, financial institutions, transport control rooms). Qualitative data gathered through an ethnographic study allows a researcher to understand how users interact with other individuals within the socio-technical system and the various technologies when carrying out their tasks. Data gathered from an ethnographic study can be complex and also quite verbose, which means that the analysis will be difficult when the purpose of an ethnographic study is to design a new system. Furthermore, a great deal of time and resources are needed to process and analyse ethnographic data, making the task of translating field data into tangible design recommendations or guidance difficult. This issue was identified by Sharp et al. (2007), who found that *"…many developers are unsure how to integrate ethnographic evaluation into development cycles"* (p 332).

In the light of these limitations, other variations of ethnography have been put forward such as rapid ethnography to reduce the time required to conduct field research. In addition to these newer variations of ethnography, contextual inquiry can be used which is a collection of field methods which follows a

[4] www.zoefolio.com/2006_recollect/img/affinity.jpg

structured approach to data collection and analysis. This technique was developed for HCI research, to inform system design; however, it is also used in research with the aim of improving an existing design or a set of ideas. The following summarise the main differences between ethnography and contextual inquiry:

- Ethnography is a sociological field research method which is typically used for *longitudinal research* (six months to several years) to understand the nature of work users carry out, people interactions, system failures and the work environment for the elicitation of user requirements for system design.
- Contextual inquiry is a collection of field research methods based on three concepts: 1) context; 2) partnership; and 3) focus, and is typically used for field research carried out over a *shorter period of time* (ranging from a few hours to a week). Contextual inquiry is used in HCI research to understand the context of systems - thus the overall socio-technical system.

3.5 A Critique of Methodologies Used in Previous Control Room Studies

There have been many field studies carried out in control rooms and control centres. These studies are critically reviewed in Chapter 2. The methodological strengths and limitations of each of these studies are summarised in Table 3.2 and critically discussed in this section.

Table 3.2: Summary of Methodologies Used in Previous Control Room Studies

Control Room Domain	Study Type	Field Research Methods Used	Methodological Strengths & weakness(+/-)[5]
CCTV control room (Gill et al., 2005; Gill and Spriggs, 2005)	Evaluation	- Quantitative data collection - Semi-structured interviews	+ Uses interview and observation guides. + Data triangulation. - No action based recommendations provided. - Methodology excludes discussion of data processing and analysis.
Transport control room (Heath and Luff,2001)	Ethnography	- Observations (naturalistic)	+ Detailed findings revealed about CCTV tasks performed within the London Underground. - No details given on field work data collection method or analysis.
Traffic management control centre (Chen et al., 2005)	Evaluation and Design	- In-situ observations - Interviews - Questionnaires - User scenarios	+ Data triangulation. + UCD and contextual research used to propose new design to improve task performance. - No detailed analysis of operator tasks given.
Ambulance control room (McCarthy et al.,1997)	Ethnography	- Observations - Semi-structured interviews - Conversation analysis	+ Detailed analysis of operator tasks given. - Data too rich and unstructured to be applied for systems designers.

In the study by Gill et al. (2005), three research methods were used to gather field data at several CCTV control rooms:

1. Quantitative data gathering: An incident template was used to record details for all surveillance incidents detected by CCTV operators on camera in the control room (after actively monitoring a target for more than one minute).

[5] + Methodological strengths.
- Methodological limitations.

2. Qualitative interviews: At least one operator from each control room was interviewed using a semi-structured questionnaire guide. The objectives of the interviews were to examine operational behaviour, operators' attitudes to CCTV, the criteria used for surveillance, working relationships between other CCTV stakeholders, and operators' impressions about their job.
3. Qualitative observations: Operators were observed to understand how they performed their work, whether they worked effectively with other CCTV stakeholders, whether any standards and procedures were followed, what training was received, and how they viewed their relationship with the police. The observations were facilitated by the use of an observation guide. The observations were used to also verify the responses from the interviews with operators.

Interview and observation guides are an effective way of structuring field work activities, particularly when conducting research within workplace environments where physical access and time is restricted. Another strength identified with Gill's research methodology was that it allowed for data triangulation: by the use of interview and observation data. Data triangulation can strengthen both the validity and quality of the data (Patton, 1990) and takes a combined account of operators' actions and attitudes whilst performing tasks within a real-world setting. The main limitation with the methodology was the lack of description given on *how* the field research data was analysed. Furthermore, the results discussion did not provide actionable recommendations to improve the way in which CCTV is used within security control rooms.

Similar to Gill et al. (2005), Luff and Heath (2001) also failed to provide specific details on the methodology and field work analysis in their study at the London Underground control rooms. It was reported that field analysis was conducted through observations and video recordings to examine the awareness and monitoring of events on CCTV within platforms (station supervisor tasks). Luff and Heath (2001) claimed that, *"an orientation formed by ethnomethodology and interaction analysis"* (p 28) was undertaken. However, it was unclear how the field research was conducted and how the field data was processed. The findings from the field work were very detailed, and although some recommendations were put forward to improve system design (see a review of this study in Chapter 2, Section 2.3), the results are still too difficult for designers to apply for the design of systems to improve operator task effectiveness.

The only control room study which has gathered field research data with the intention of designing a new user interface system was the traffic control centre study by Chen et al. (2005). The aim of the research was to improve operator performance when they performed a surveillance task (searching and locating outdoor scenes). Chen et al. followed a User Centered Design (UCD) approach to understand the current processes and systems operators used to manage incidents. The field research involved conducting in-situ observations and interviews with operators to gather insights into their current working practices. The findings were then used to devise appropriate use scenarios. The field data was also used to formulate a written questionnaire to collect more refined feedback on specific aspects of the user interface which operators' used. Similar to the CCTV effectiveness study by Gill et al. (2005), more than one field method was used to verify subjective comments.

In the ambulance control room study by McCarthy et al. (1997), the importance of understanding context was considered valuable for the design and evaluation of systems used within ambulance control rooms. A 12-month field study was carried out to examine the organisation of work in two ambulance control centres. Observations, conversation analysis, and open-ended interviews were conducted with ambulance operators. From the findings, detailed descriptions were provided on a highly critical activity: locating an ambulance scene. McCarthy et al. extensively discusses the wider implications of the field research for understanding operator tasks - in context, for the design and organisation of systems. Despite this, no practical recommendations were provided to help designers improve system design to increase operators' effectiveness in their surveillance tasks.

In summary, all of the control room studies (see Table 3.2) fail to describe *how* field data was collected and analysed in detail. Furthermore, these studies - particularly the ethnographic studies (Luff and Heath, 2001 and McCarthy et al., 1997) do not provide a discussion on *how* the findings can be used to improve system design. This confirms the view by Sharp et al. (2007) that, field research studies are (and still not) designed to be integrated into product development cycles. The main reason why field research cannot be easily applied to system design is because the procedure for gathering and analysing data lacks structure. Without structuring field research data, a number of difficulties may arise:

1. Important information may be missed. There is a risk that the observation and interview responses that are relevant and important to the study objectives may be overlooked.
2. Study aims and objectives may alter without the researcher being aware. Field research involves gathering rich data about activities in the workplace. Without the use of a structured framework to gather this data, there is a risk that too much information will be gathered, which may lead to the researcher having too much data to analyse.
3. Data analysis will be very difficult if too much data is gathered, as stated in 2) the analysis will be time consuming.

Contextual inquiry is a useful requirements capture and analysis technique which can overcome these difficulties described. This technique involves gathering data by following a set of concepts (context, partnership, and focus), placing strong emphasis on the use of structured methods for gathering and analysing the field data gathered. In this thesis, contextual inquiry was chosen to understand the context and use of CCTV in the field and to specifically identify the factors which reduced the effectiveness of CCTV (see Chapter 4). Although contextual inquiry is not an evaluation method, the process of gathering rich and detailed user-based feedback is achieved through naturalistic observations and semi-structured interviews whilst users carry out their tasks in their place of work.

Paper surveys were trialled in the early stages of Study 1; however the responses were not very detailed. The data from the surveys was not analysed as the response rate was low. Data gathered through user-based reports (surveys and interview methods) alone were not considered as reliable methods in identifying the wide range of problems which reduced the effectiveness of CCTV.

In addition to user-based reports, other observation-based methods could have been used in Study 1 such as action research and grounded theory. Action research involves the researcher actively getting involved with the research material, as opposed to being an observer. The results from action research provide relevant information grounded in practical action as well as informing theory. Qualitative data and analysis techniques are used to form the basis of interpretive idiographic research. Grounded theory is a theory-building qualitative method, which is a particularly useful method for investigating factors and issues that are previously unknown to the researcher. This method involves the researcher forming an area of study and allowing theory to emerge from the data. Both action research and grounded theory involve qualitative data collection and analysis with the aim of building theory. Whilst these methods are commonly used in field research, they were not considered for the control room field study for a number of reasons. Firstly, there was a requirement to collect quantitative data (e.g. number of CCTV cameras, monitors, faulty equipment, time spent on tasks, and so on) in addition to qualitative data as part of the context exercise. Secondly, the study objectives for Study 1 (see Section 4.1) were developed as a result of previous research findings on CCTV effectiveness (Gill et al., 2005). Thirdly, rather than building on theory, one of the objectives of Study 1 was to use the research findings to develop a conceptual framework which included a set of best-practice recommendations on CCTV design. For these reasons, contextual inquiry was chosen to gain a detailed understanding of how CCTV is used, how CCTV stakeholders interact and then identify the factors which interfered with task performance.

The contextual inquiry technique was considered useful for the field research since it involves the use of several methods (observations and interviews) and approaches which allows for data triangulation. It was necessary to develop a task-based contextual inquiry approach in order to build a practical framework to improve task performance (Task-Effective Video In Security: TEC-VIS). This involved modifying the contextual inquiry technique by making use of a structured observation checklist (see Appendix C) like the one used by Gill et al. (2005). This checklist was designed to help focus and gather specific data on operator tasks during the observations. This checklist was also used to guide the interviewing (semi-structured interviews) with control room managers and CCTV operators during task performance. The checklist was mirrored with the research objectives for Study 1.

A flexible, semi-structured interviewing technique was chosen over other interviewing techniques, to gather as much insight and experiences from operators as possible. To achieve this reliably, the ‘SEE’ technique was applied. Data gathered was analysed by grouping the notes into hierarchical categories on a large wall space and then identifying themes (performance issues). This method involved creating an Affinity diagram (see Figure 3.1). The creation of an Affinity diagram is a methodological technique for organising a large amount of rich data to aid data analysis. The complete Affinity diagram was used to draw a number of practical recommendations to improve the design configuration and use of CCTV systems. These recommendations are presented in the research contributions in this thesis (see the TEC-VIS framework in Chapter 10).

3.6 Chapter Summary

This chapter has provided an overview of the main field research methods and types of analysis used in HCI research. This review has been provided purposely to identify the appropriate methodology for use in the CCTV control room study carried out for the research of this thesis (see Chapter 4). Contextual inquiry was identified as the most appropriate methodology for Study 1, since it provides useful qualitative and quantitative data for this thesis. Contextual inquiry, an ethnographic process used in HCI research (Beyer and Holtzblatt, 1998) is often used in system design when the design and development of a new system will be short. Contextual inquiry provides a logical and structured approach for gathering and analysing data for research that is carried out under shorter time spans. The research technique also follows a number of concepts important for conducting field work: context, focus, and partnership. The technique was slightly altered by making use of a structured observation checklist which mirrored the study objectives. Furthermore, the 'SEE' technique was used to draw meaningful data from operators which involved asking CCTV operators to: 1) Specify; 2) Elaborate; and 3) Exemplify when self-reporting issues about their work.

To date, contextual inquiry has not been used as a field method to examine task performance problems within security control room environments. The next chapter details a field study in which contextual inquiry was used for understanding the context of CCTV and performance issues at 14 different CCTV control rooms.

Chapter 4

Study 1: Control Room Field Study

"You don't get us watching TV anymore!" **(Anonymous CCTV Operator, 2006)**

A review of prior control room studies has revealed that there has been very little research conducted in the field of HCI which examine the task performance issues associated with technology used for security tasks. Prior studies have uncovered some task performance issues within control room environments (Gill and Spriggs, 2005; Gill et al., 2005; Luff and Heath, 2001; McCarthy et al., 1997) however, the research so far conducted has not involved a systematic examination of security tasks or provide best-practice recommendations to improve the design of CCTV and control room performance.

In this chapter, an extensive eight-month field study is detailed, which was carried out at 14 CCTV control rooms. The field study was conducted to gain an understanding of the context in which CCTV is used, identify the tasks performed by CCTV operators and other CCTV stakeholders, and examine the factors that reduce the effectiveness of CCTV. The chapter finishes with a discussion of the findings and then reviews the study in relation previous control room research.

4.1 Study Objectives

A field study was conducted to investigate the effectiveness of current CCTV technology and its impact on operator tasks, security processes, and stakeholders. The cognitive factors that can be used to assess CCTV operators' performance in their tasks include measuring their mental workload (i.e., how much information they absorb when dealing with radio communication) and their vigilance (i.e., how they decide whether to react to incidents observed on video monitors). The purpose of understanding an operator's cognitive functions in their work is to determine whether they are able to understand and digest the information available to them, make correct decisions, communicate, and execute their tasks. Examining the cognitive aspects of a CCTV operator's work has been largely disregarded in the HCI and ergonomics research domains. The outcome of such an evaluation can help in the design of security tasks, the control room environment, and the systems to allow for the proper allocation of tasks across the system. This study addresses research goal 1 (see Chapter 1, Section 1.3.1), with the following objectives:

1. Gain an understanding of the tasks and contexts for which CCTV is used within security control room environments.
2. Identify the social and technical constraints of operator tasks and conflicts between CCTV stakeholders inside and outside the control room.
3. Evaluate the effectiveness of CCTV and other technology used by CCTV control room operators to support their security tasks.

4. Use the findings to create a set of best-practice recommendations for security practitioners and owners to improve the effectiveness of CCTV use in control room environments.

4.2 Control Rooms Visited

The individual field visits were carried out over an eight-month period from September 2005 to April 2006. 14 CCTV control rooms were visited, with each visit lasting 4–6 hours. Contextual inquiry was used to structure the observations and interviews. An effort was made to visit control rooms having a maximum number of operators on shift, in order to gather as much insight in a single visit as possible. While this was not a strict criterion, it was a valuable way to assess task performance issues when there were staff shortages. The visits were arranged through the support of the CCTV User Group,[6] a private membership community for CCTV stakeholders. Managers who were involved in the research took part voluntarily, and participated following an email request. A majority of the control rooms visited were based in Greater London. Table 4.1 summarises the different CCTV control rooms visited in this study.

Table 4.1: Summary of CCTV Control Rooms Visited

CONTROL ROOMS VISITED	TYPE OF CONTROL ROOM	GEOGRAPHICAL LOCATION
4 police control rooms	1 = Heathrow airport 3 = Public Surveillance: CCTV control rooms	Greater London: - 1 west of London (Heathrow airport) - 2 north of London - 1 south of London
10 public/private managed CCTV control rooms	10 = Public Surveillance: CCTV control rooms	- All in London

From the pool of interested participants, the 14 CCTV control rooms were chosen which were similar to one another (common users, tasks, and technology). Variety was also important to consider the varying problems. Thus, control rooms which were from different boroughs, different geographical locations, and those using new and old technology were included to assess variety of performance issues. The characteristics of the control rooms were identified during the initial stages of research when organising the participants for the field study. The control rooms in London were chosen as they were the busiest and most convenient for travel purposes, so that a greater level of interaction could be observed (see Table 4.2). It should be noted that since 14 control rooms were available for this study, the research findings are representative of current location based surveillance tasks.

[6] http://www.cctvusergroup.com

Police control rooms operate similarly to CCTV control rooms, the main difference being that police control room operators are trained police staff as well as trained CCTV operators. Police control room operators are able to deal with incidents directly, as they have the authority and access to police officers and other necessary individuals. There are very few police CCTV control rooms in the UK; this is mainly because there are not enough police officers trained to work in CCTV control room environments. The level of funding for CCTV security each control room received varied. Some are funded by the Home Office CCTV schemes and some rely on public funding and other government resources.

Table 4.2: Control Rooms Visited and Managers' Security Goals

Control Room	Security Goals for CCTV	Technology	Camera to Operator Ratio
1	Traffic enforcement and surveillance.	Analogue 50% Digital 50%	160:5
2	Monitor incidents and crime.	Analogue 100%	110:4
3	Prevent crime and protect the public.	Digital 100%	90:3
4	Review images on behalf of police.	Analogue 30% Digital 70%	111:6
5	Provide council support and city surveillance.	Analogue 30% Digital 70%	200:4
6	Deter and detect crime, assist in identification, arrest and prosecution of offenders, and reduce fear of crime.	Analogue 100%	141:3
7	Catch criminals, track lost and stolen vehicles, traffic enforcement, and provide emergency resources for residents.	Digital 100%	110:3
8	Deter and detect crime, support borough in emergencies, and support police operations.	Digital 100%	87:3
9	Prevent crime and catch criminals.	Digital 100%	80:1
10	Work with adult community service and community section to provide safety and support using CCTV.	Digital 100%	120:1
11	Crime and disorder reduction, interception and arrest of offenders, provide elderly support using call centre and alarm support services.	Digital 100%	96:3
12	Public safety, supporting police officers and operators in operations and incidents.	Analogue 100%	108:2
13	Prevention and detection of crime, traffic management, public re-assurance, provision of evidence for civil and local proceedings.	Digital 100%	65:2
14	Handle incidents reported by Heathrow Airport Limited and others, and to provide and coordinate security.	Analogue 100%	4:3

Two of the control room managers (control rooms 7 and 14) requested a recommendation report following the observation visit. Control room 7 is managed by staff working for a north London local authority. A report was specifically required following a visit, as this control room was undergoing a major upgrade in technology (analogue to digital) and the manager wanted feedback on problems prior to the upgrade work. This report was provided to the control room and a letter of receipt was received (see Appendix A).

Control room 14 is a police control room (command and control room) which is based at Heathrow airport's terminal 1. This is the only control room for which permission was obtained to use its name in this thesis. A chief inspector from the Metropolitan Police constabulary approached the thesis author in April 2005 requesting an evaluation report detailing the problems within the control room and also requested recommendations. This control room was also due to upgrade (from an analogue to a digital networked system), as part of a wider upgrade programme to improve the airport's security technology. The focus of the observations involved assessing operators' workstation layout (e.g., equipment placement and positioning), identifying how technology was being used (e.g., radio and other audio tools), and how well operators performed their tasks. The findings and recommendations following the field visit to this particular control room were presented to the Metropolitan Police as well as to the technical project team. A letter in receipt of this report was received by the Metropolitan Police project leader (see Appendix B).

4.3 Method and Procedure

Contextual inquiry (Beyer and Holtzblatt, 1998) was used for the field visits to identify:

1. the security goals for CCTV for the different CCTV control rooms;
2. the security tasks operators carry out;
3. how CCTV tools and other associated technologies are used by operators; and
4. the interactions between operators and other CCTV stakeholders working in and outside the CCTV control rooms.

Two field research methods were used to carry out the contextual inquiry: 1) structured interviews with control room managers and semi-structured interviews with CCTV operators and 2) structured observations with operators whilst performing their tasks. As mentioned in Chapter 3: Section 3.4, questionnaires and surveys were used to gather information from each of the control rooms, however very few responses were returned and so interviews and observations were used.

Table 4.3 details an interview checklist which was developed in for the interview with managers and operators. These methods were used to gather rich and detailed information about users in the field. Probing questions were developed for use with direct observations to ensure that the interviews were conducted as effectively as possible. The combined use of these methods was necessary to understand the context of CCTV, what tasks are performed by CCTV operators and how they perform their tasks using different tools and systems.

Table 4.3: Interview Checklist Used in This Research Study

PRE-INTERVIEW: INTRODUCTION
▪ Who is the interviewer? ▪ What institution are they from? ▪ Provide a background on the research topic. ▪ Who do they want to interview and why? ▪ For what will the information be used? ▪ How long will the interview take?
INTERVIEW DATA GATHERING CHECKLIST
▪ How will the personal and sensitive data be protected? ▪ Seek informed consent on publishing the data (if applicable). ▪ How will the data be analysed and used for future work? ▪ Seek permission for recording audio during the interview sessions (if applicable).
INTERVIEW CHECKLIST
▪ Conduct the interviews in an environment in which others will not be distracted. ▪ If interviewees reveal any out of the norm comments, avoid giving any strong emotional responses, as this may influence the interviewee's willingness to be open in later questions. ▪ Ask interviewees to respond to questions in an open and honest manner and assure interviewees that their responses will be kept safe from other personnel. ▪ Inform interviewees that they can interrupt at any time to ask for clarification when unsure of any question.
POST-INTERVIEW CHECKLIST
▪ Summarise the interview theme and sum up the main issues. ▪ Confirm the main findings with the manager/operator to ensure the findings are accurate. ▪ Seek clarification of any comments made that may appear difficult to analyse later. ▪ Once the interview has been concluded, debrief the user about the study's objectives in more detail than given in the introduction. ▪ Re-affirm that the information gathered from the interview will be kept anonymous and confidential.

Prior to each visit, control room managers were provided with the objectives of the field visit as well as the planned structure and details of the visit. The plan included: a guided tour of the control room and other facilities, an interview with the manager, and 4–6 hours of continuous observations with operators in the control room.

Control room managers were all asked the following questions:

1. How many operators work at the control room in one shift?
2. What shifts do operators work?
3. What type of crime is observed at your control room?
4. What is your role at the control room?
5. What are your goals for CCTV at your control room?
6. How many cameras do operators use to monitor the surveillance areas?
7. What type of technology is used in your control room to store and transmit CCTV video (digital, analogue, or both)?

These questions were chosen based on the objectives for this study, which were formulated purposely to address research goal 1. The observations, the specifics of the control room technology, operator tasks and usability problems, were recorded on note paper by one researcher (thesis author). Observation notes were made by following an observation checklist (see Appendix C), which includes a number of questions related to an operator’s behaviour and the technology they use. This checklist was used to prompt the observations. This checklist was also prepared by mirroring the study objectives, so the contextual observations could be focused on the operators’ actions, behaviours and attitudes whilst performing tasks and interacting with other CCTV stakeholders.

4.4 Results

4.4.1 Analysis

It was assumed that the responses provided by CCTV control room managers and operators were open, accurate and honest. There were, however, some instances in which this was not the case, as detailed later. Following each visit, the field notes were organised by creating an Affinity diagram which involved the following:

1. Single observations related to the study objectives were summarised on post-it notes.
2. The individual observation notes were specified in greater detail and then linked together to form a set of larger components.
3. Components identifying similar findings were then grouped.
4. The observations which were grouped in components were labelled and placed into hierarchical order by significance in relation to how they reduce the effectiveness of CCTV.

The following groupings (themes) emerged following the analysis of the field research data:

1. CCTV stakeholders, their roles and goals for CCTV
2. Technology used within CCTV control rooms and its functions
3. Task performance issues when operators use technology
4. Stakeholder conflicts between CCTV users

4.4.2 Findings

4.4.2.1 Identification of CCTV Stakeholders

The first theme involved the identification of the different CCTV stakeholders who work within the CCTV control room system and their roles (see Table 4.4). Starting with the stakeholders at the bottom of Table 4.4, CCTV operators liaised with local authority staff to provide support on housing security, safety and maintenance issues using CCTV and telecommunication tools (mainly telephone). When operators reacted to crime on-screen, they communicated with local businesses, such as the high street shops, clubs and bars using radios (these stakeholders are described as 'participant CCTV radio users' in the findings). Operators also collaborated with police officers and inspectors to gather up-to-date information about crime occurrences within their dedicated surveillance areas via face-to-face meetings, email, or by fax communication. Police officers frequently visited CCTV control rooms (local to their police station) to provide news and photographs on suspects. The level of communication and collaboration between CCTV operators and police control room operators was high in comparison to other participant CCTV radio users. CCTV control room operators only initiated radio calls to police operators when they are very certain an incident was about to take place. Once contact was made, operators shared their camera views with police operators to help them locate the scene or target for on-street police support.

Each CCTV control room manager's goals for CCTV were identified from face-to-face structured interviews (see Table 4.2). Although some of these goals were quite general, it was clear that each of the manager shared a common high-level security goal: use CCTV to detect and prevent crime in public places. However, the specific security goals for the control rooms varied from one control room to another, and depended largely on the surveillance area, funding available, and type and level of crime in the borough being served.

Table 4.4: CCTV Stakeholders and Their Roles

	Stakeholder Roles
Control Room Manager	Ensure that all CCTV operators and team leaders operate all equipment and cameras to comply with the CCTV code of practice for the control room.
	Operate radio communication equipment and any other equipment.
	Liaise with police to coordinate a response to detected incidents.
	Ensure communication between departments and public is achieved.
	Compile statistical reports for management.
Control Room Supervisors	Operate all equipment and cameras in line with the CCTV control room policies and the CCTV code of practice.
	Liaise operationally with police to coordinate a response to detected incidents, including the use of shared monitor views.
	Record all events and actions taken in a clear, legible, and accurate written format to record these onto other digital media.
	Manage the production of evidence by CCTV operators from the initial telephone call or visit by an applicant through to the completion of the statement and bagging of the evidence.
Operators	** All Control Room Supervisor Roles **
	1. Proactive surveillance: } See Figure 4.5 for a detailed 2. Reactive surveillance: } breakdown of each task. 3. CCTV video review and administration: record all events and actions taken in a clear, legible, and accurate written format to record these onto other digital media.
Police Officers/Inspectors	Use police radios to contact CCTV control room operators for incident support using CCTV cameras.
Police Control Room Operators	Use of shop and pub watch radios to contact CCTV control room operators for reporting suspicious and actual incidents on-site.
Participant CCTV Radio Users	Communicate council maintenance and public safety problems to CCTV control room operators.
Local Authority Staff	Use police radios to contact CCTV control room operators for incident support using CCTV cameras.

4.4.2.2 Manager Interview Responses

The interview responses from managers are summarised in both Table 4.2 and Table 4.5.

Table 4.5: Summary of Control Room Managers' Interview Responses

Question	CCTV MANAGERS' RESPONSES
1. How many operators work at the control room in one shift?	(on average): 5 CCTV operators (day shift) 3 CCTV operators (night shift)
2. What shifts do operators work?	All of the control rooms operate live surveillance for 24-hours and operator shift patterns generally consisting of four 12-hour day shifts, a two day break, and then four 12-hour night shifts.
3. What type of crime is observed at your control room?	The type of crime observed in London was found much broader, including a higher number of violent and drug-related crimes: Gun crime, violence, theft, cash machine robberies, shoplifting, graffiti, car crime, underground ticket touting, burglary, fly tipping, breaking traffic rules, street robberies, drug dealing and drug abuse. Outside London, the most prevalent crime included property theft, shoplifting, damage, burglary, anti-social behaviour, and drug dealing.
4. What is your role at the control room?	See Table 4.2
5. What are your goals for CCTV at your control room?	See Table 4.2
6. How many cameras do operators use to monitor the surveillance areas?	See Table 4.2
7. What type of technology is used in your control room to store and transmit CCTV video (digital, analogue, or both)?	See Table 4.2

4.4.2.3 Artefacts Used to Support CCTV Tasks

Artefact is a term often used by sociologists and HCI researchers to describe an item used in the workplace. An artefact can be low-tech (paper) or high-tech (technology). Table 4.6 provides a summary of technologies used across the different CCTV control rooms visited. The specific technological artefacts used by CCTV operators are shown in Table 4.7.

Table 4.6: Summary of Main Technology Artefacts Across Control Rooms

No. of control rooms	Artefacts
10	Analogue radio
4	Analogue CCTV
4	Digital radio
7	Digital CCTV
4	Upgrading from analogue to digital
4	Geographical mapping and information system[7] (see example in Figure 4.1)
2	Automatic Number Plate Recognition (ANPR) system

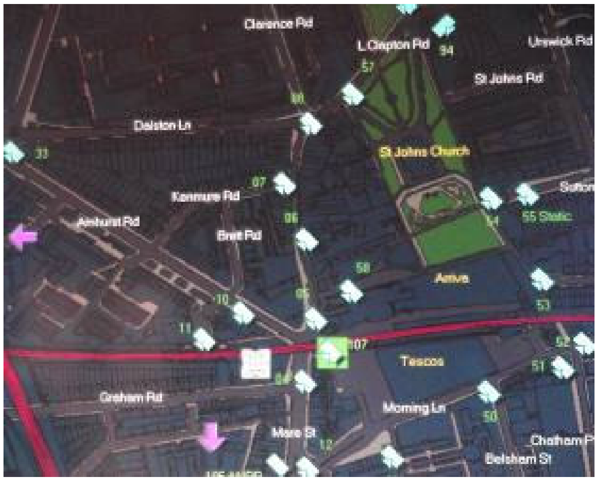


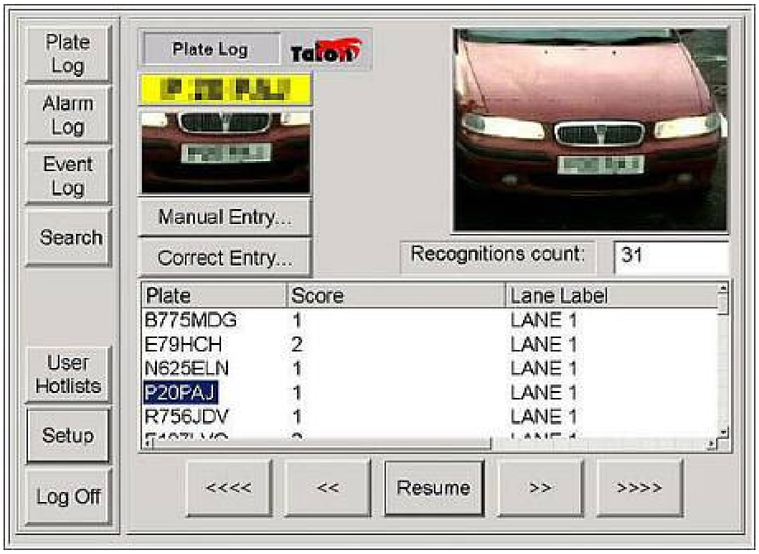
Figure 4.1: Geographical mapping and information system.

[7] A tool which allows operators to locate relevant CCTV cameras for instant views using an electronic map on which cameras are overlaid, or a database which can be searched for the CCTV cameras. In cases when this system is not used, operators locate a scene by locating the camera ID on a paper list and looking up the street name on a geographical map (paper or on a PC), and then selecting the correct camera in the system.

Table 4.7: Artefacts Used to Support CCTV Operators in Their Tasks

Artefacts	No. used per operator	Function of Artefact
Monitor wall	5–15	Proactively monitor camera activity.
CCTV spot monitor	1–5	View CCTV directly on a video monitor facing the operator for close-up inspection of events.
Camera joystick/pan, tilt and zoom (PTZ)	1	Adjust the view of an interesting scene on a camera to gain better visuals.
Camera controller: keyboard and touch screen interface	1	Enter camera numbers and adjust picture quality settings to gain better visuals of scene.
Personal computer (with Internet connection)	1–3	Access e-mails and Internet/Intranet, and preview of post-recorded CCTV video footage. Alter imaging properties of live camera view to improve visuals.
Geographical mapping and information system	1 (only present at 4/14 control rooms)	Access geographical maps of surveillance areas, which show the locations of all CCTV cameras and other relevant mapping data.
Automated surveillance system	1–2 (only present at 2/14 control rooms)	Use of ANPR, body and facial recognition systems with CCTV cameras for intelligent support in locating criminals.
Evidence reviewing PC and video monitors	1–2	Replay of CCTV video footage for verification by operators/police.
VCR/CD/DVD recording device	1–3	Create copies of CCTV video and images for the police for investigations.
Pub/shop watch radio	1–3	Make and receive contact with local businesses to support and detect incidents and targets.
Police CAD radio	1–5	Relay information on crime between CCTV operators, police control room and street police officers.
Telephone	1–3	Receive incoming calls from public and council departments on any security related issues. Internal control room communication.
Incident and handover logging book	1	Log crime and road traffic incidents for record keeping and for information sharing purposes.

In the control rooms in which an ANPR system is used, operators are assisted in tracking lost and stolen vehicles passing ANPR cameras. The tool automatically alerts the operator with an audio signal and instantly displays: an image of the suspect car, the vehicle registration number (VRN) plate, the alert status, and the system's electronic match of the VRN (see Figure 4.2). The system can be configured to alert operators for different situations. Once an alert is flagged by the system, the operator's task is to observe the ANPR monitor, compare the CCTV image showing the VRN and identify whether the computer match is correct. If the VRN does not match, the operator deactivates the alert by pressing the 'mismatch button' on-screen and continues with their proactive and reactive surveillance tasks. If the VRN is a correct match, the operator accepts the alert, informs the police control room for immediate support. Whilst the operator is in radio/telephone contact with the police they attempt to locate the vehicle from camera to camera, until ground police attend the situation. The ANPR system identifies vehicles registered under the Driver and Vehicle Licensing Agency (DVLA) and the Police National Computer (PNC) databases.

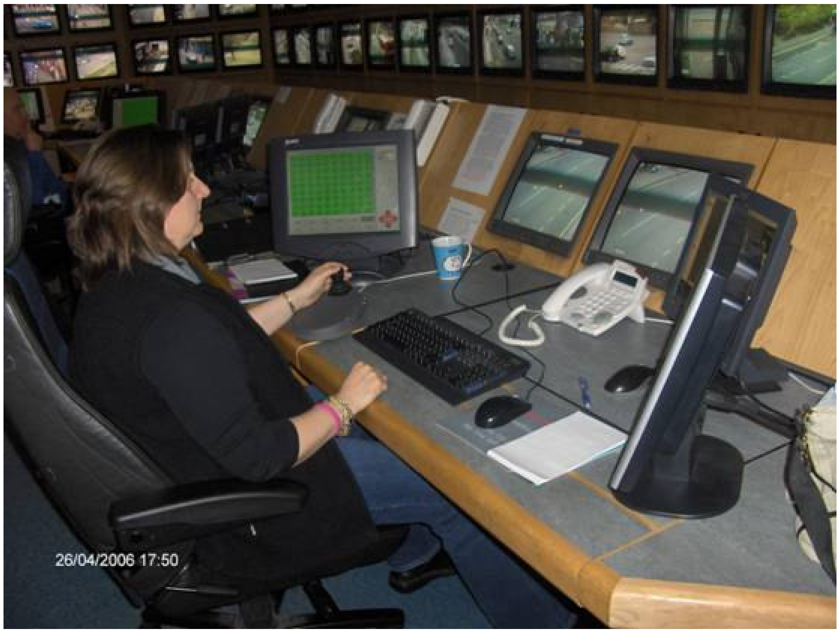


Recognitions count: 31

Plate	Score	Lane Label
B775MDG	1	LANE 1
E79HCH	2	LANE 1
N625ELN	1	LANE 1
P20PAJ	1	LANE 1
R756JDV	1	LANE 1

Figure 4.2: ANPR system used in control rooms 7 and 4.

The typical CCTV operator's work environment set-up is shown in Figures 4.3 and 4.4. The operators in both images are watching their spot monitors in a proactive surveillance task. In Figure 4.4, the operator is using a radio headset to manage radio calls. Both operators and their managers at these control rooms gave their permission to be photographed and have the images published.



26/04/2006 17:50

Figure 4.3: Typical CCTV operator's work environment set-up #1.

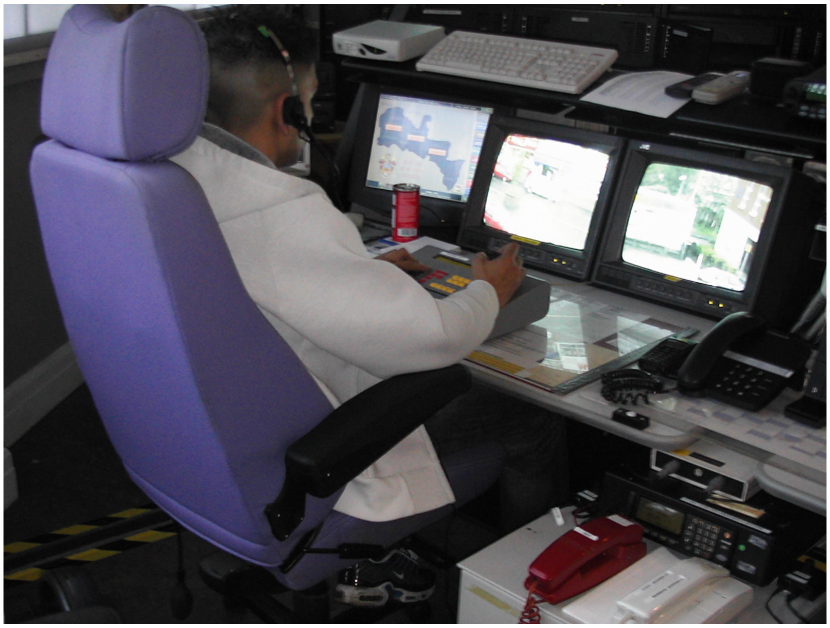


Figure 4.4: Typical CCTV operator's work environment set-up #2.

4.4.2.4 CCTV Operator Tasks

In order to determine what tasks were being performed and how they were being performed, operators were observed whilst performing their tasks. They were also probed about how and why they carry out their activities. Task modelling could have been used to determine what tasks were being performed, however this technique was not used as observations and probing operators during significant events captured more meaningful data. From the data, it was discovered that CCTV control room operators performed three core tasks (task analysis detailed for each of these tasks, see Figure 4.5, 4.6, and 4.7):

1. Proactive surveillance
2. Reactive surveillance
3. CCTV video review and administration

These tasks are described in Table 4.8. The average time spent on each of these tasks in a typical shift and the types of artefacts used are also detailed.

Table 4.8: Operator Tasks, Percentage of Time Spent on Each, and Artefacts Used

Operator Task	Time Allocated (approx. % of a 12-hour shift)	Artefacts Used
1. Proactive surveillance	Day shift: 60%	Observing video on spot monitors: PTZ controls (1) Geographical mapping and information system (1) Paper map (1) Spot monitor (desk facing monitor) (1–2)
	OOH[8] shift: 30%	Observing video on several monitors: Monitors aligned against wall (10–15)
2. Reactive surveillance	Day shift: 30%	Police CAD[9] Radio (1) Pub/shop watch radio (1–2) Personal Computer with Internet (1) Telephone (1–3) PTZ controls (1) CCTV spot monitor (1–3) Monitor banks (10 --40) Intelligent CCTV system (0–1) Paper/electronic map (0–2)
	OOH shift: 65%	
3. CCTV video review/admin	Day shift: 10%	Evidence review PC (1–2) VHS tapes, DVDs & CDs Incident/handover log book (1) Tape/CD/DVD labels Evidence bags
	OOH shift: 5%	

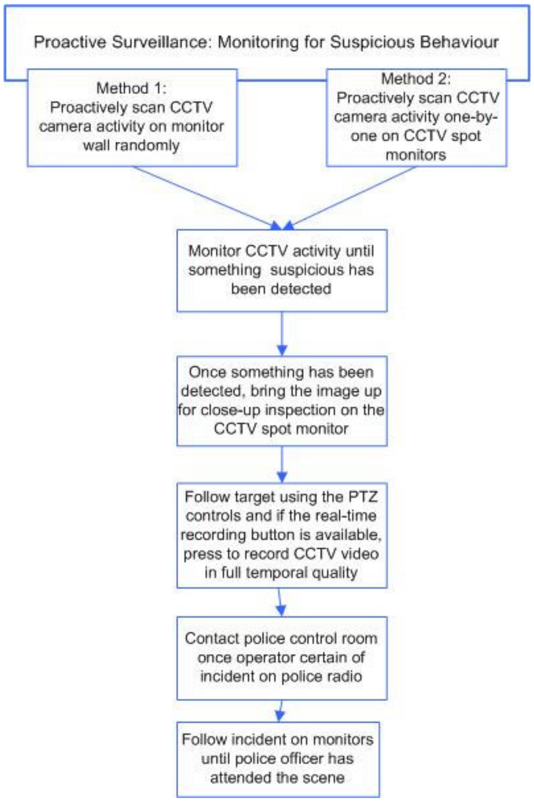
[8] OOH = Out Of Hours: weekend and night shifts.
[9] CAD = Computer Aided Dispatch is a radio system used by UK police operators and CCTV control room staff.

Proactive surveillance

Proactive surveillance involves CCTV operators *detecting* suspicious behaviour, *recognising* and *identifying* suspicious targets by scanning CCTV activity on a monitor wall or on 1-2 spot monitors. Proactive surveillance is usually carried out during what operators refer to as the 'grave-yard shift' (between 4–7 a.m.) and all day Sunday.

As observed, this task is carried out by operators using one of two methods. Figure 4.5 details a task analysis for each of these proactive surveillance methods adopted by operators:

1. Proactive random scanning of CCTV camera activity on monitor wall
2. Proactive scanning of CCTV camera activity on each individual CCTV spot monitor



Proactive Surveillance: Monitoring for Suspicious Behaviour
Method 1: Proactively scan CCTV camera activity on monitor wall randomly
Method 2: Proactively scan CCTV camera activity one-by-one on CCTV spot monitors
Monitor CCTV activity until something suspicious has been detected
Once something has been detected, bring the image up for close-up inspection on the CCTV spot monitor
Follow target using the PTZ controls and if the real-time recording button is available, press to record CCTV video in full temporal quality
Contact police control room once operator certain of incident on police radio
Follow incident on monitors until police officer has attended the scene

Figure 4.5: Proactive surveillance task analysis.

In this study, **method 1** was commonly adopted by the less experienced CCTV operators, who found it easier to scan CCTV video using the monitor wall, as they were less familiar with and confident in using the different systems available to them at their workstations. It was also found that operators new to the job and surveillance areas struggled to memorise camera numbers and locations, which prevented them from using a geographical mapping and information system (at the control rooms where these were used).

Method 2 was commonly adopted by experienced CCTV operators. Experienced operators felt comfortable using a geographical mapping and information system when proactively monitoring CCTV. This method was found to be the most efficient and effective method for proactive surveillance. This method was used successfully only by experienced operators who were: 1) familiar with the surveillance areas; 2) experienced in the task and 3) experienced and confident in using the PTZ controls, since PTZ controls a geographical mapping and information system.

Each operator scanned on average 10–15 monitors along the monitor wall using PTZ controls to focus particular scenes of interest. Operators (particularly trained operators) made every attempt to capture good evidence of incidents and targets. This was achieved by adjusting the angle of a camera, using the real-time button to maximise the frame-rate, focusing cameras on a target's face, etc. During radio communication between CCTV operators and police staff, the NATO phonetic language was used to describe and clarify names (e.g., 'A' for alpha, 'B' for beta, 'C' for Charlie) as well as police identity codes.[10] When operators contacted the police control room once they detected an incident (during proactive surveillance), they would describe the event, location, and describe the targets involved.

Operators did not adopt any particular scanning strategy with either method. Operators were asked during proactive surveillance: *"What scanning patterns/methods do you use to scan CCTV video activity using the monitor wall?"* All operators replied that it was pretty much random. They were further probed and asked: *"Where do you choose to scan first and why? How certain are you when something suspicious is happening?"* Most operators said that they knew where to look and this was simply based on intuition. *"It was like sixth sense, and I don't know where I should be looking as anything could happen at any time.... I can just tell if something is going on there even though no one tells me."* At another control room, another operator commented, *"…we don't really use any particular method to detect suspicious movements or people."*

When operators performed proactive surveillance, the time and place at which events were being monitored were crucial factors to the task. For instance, operators said that before monitoring scenes, they first consider the day of the week, the time of day, and any special events planned on that day (e.g., major sporting events, religious events, and festivals). These factors help operators improve their situation awareness. During the task, operators also looked out for targets with revealing, unusual body language and negative emotions (e.g., distress, panic, anger) to help detect suspicious activities. At one control room, an operator pointed out a middle-aged, well dressed man on their spot monitor who had extremely pale skin and was dragging one side of his body as he walked. The operator identified this particular target as a drug addict: *"Now…he's definitely a druggie."* This particular operator was familiar with the appearance of drug users from previous experiences working in CCTV and security.

[10] The UK police use the following identity codes: IC1 = White European; IC2 = Dark European; IC3 = Afro-Caribbean; IC4 = Asian; IC5 = Oriental; and IC6 = Arab.

In addition to using 'CCTV operator intuition,' temporal factors relevant to CCTV activity, and visual cues used to detect suspects, operators took into account crime patterns (previously observed on CCTV in the control room or as a result of existing knowledge from other CCTV stakeholders). This observation was noted following comments made by several operators:

- *"Criminals don't get out of bed until about lunchtime."*
- *"Shoplifting in the town is likely to happen after 3:30 when school kids finish from school or during the Easter and summer holidays."*
- *"I know this telephone box, it's a classic spot for crack smokers who can hide in there, and we can still tell what they're doing."*
- *"In the high drug problem areas, I have seen patterns of behaviour so I know what to look out for and at what time… like really smart business men come to [town name] first thing on a Friday to collect their drugs ready for the weekend."*
- *"I know that druggies look really pale and worn down, and even if they are dressed well I can tell that they are drug abusers by the way they walk, they drag on one side of their body when they walk…. it's so obvious!"*
- *"I think it's like sixth sense. I am doing something over here and looking at the monitors here (operator points at screens on the far left corner) … and then I look out the corner of my eye (operator looks to the top right of his side) and I know something is happening over there even though I can't see from here."*
- *"A lot of the time we suspect people when they are standing in small groups in nearby alleyways and car parks. A lot of the time we figure this out as we have caught gangs dealing drugs or discussing shoplifting in the past…. it's very much directed by what we have seen and acted on in the past."*
- *"We have photographs of previous convicts which the police supply us with. The photographs of the criminals who are wanted for serious crimes like bank robberies and knife attacks …we put their pictures on the wall. We occasionally look in this book if we suspect we saw someone in the CCTV."* (Other photos of criminals are kept in a folder. The operator displayed the folder and flicked through several pages of criminals ~150).

Operators identified key crime patterns which aided their proactive surveillance, and suggested that they base their intuition on pre-existing knowledge and experiences in crime patterns (types, time of day, etc.) and the characteristics of criminals. The identification of crime patterns has several implications for both the research and developments in criminology and intelligent CCTV design. For instance, these patterns can be used to inform the design of intelligent CCTV systems. Furthermore, the identification of these patterns are a starting point for developing a taxonomy of crime patterns and behaviours which can be used to deliver better training for CCTV observers.

In most control rooms, operators can share a camera display with police control room operators when an incident is detected during proactive surveillance that requires immediate police support. This sharing facility is activated by the operator by pressing a 'share' button at their workstation. Once the camera is shared, only the CCTV operator is able to control the camera movement using their PTZ controls (except at one police control room).

In five of the 14 control rooms, proactive surveillance also involved operators carrying out traffic surveillance and enforcement. This task involved operators detecting drivers committing traffic contraventions, such as parking on a double yellow line, parking on a red route in central London and driving/parking in a bus lane during peak hours. If the operator observed a contravention on their monitors, the details are noted in a logbook and processed manually by passing the information and CCTV video evidence to the traffic enforcement department. In one control room, this task was supported by the use of a road accident recovery organisation database. When operators were unsure of a contravention, they accessed detailed information about the vehicle by entering the VRN into a database on their PC. Figure 4.6 provides a breakdown of this traffic enforcement task.

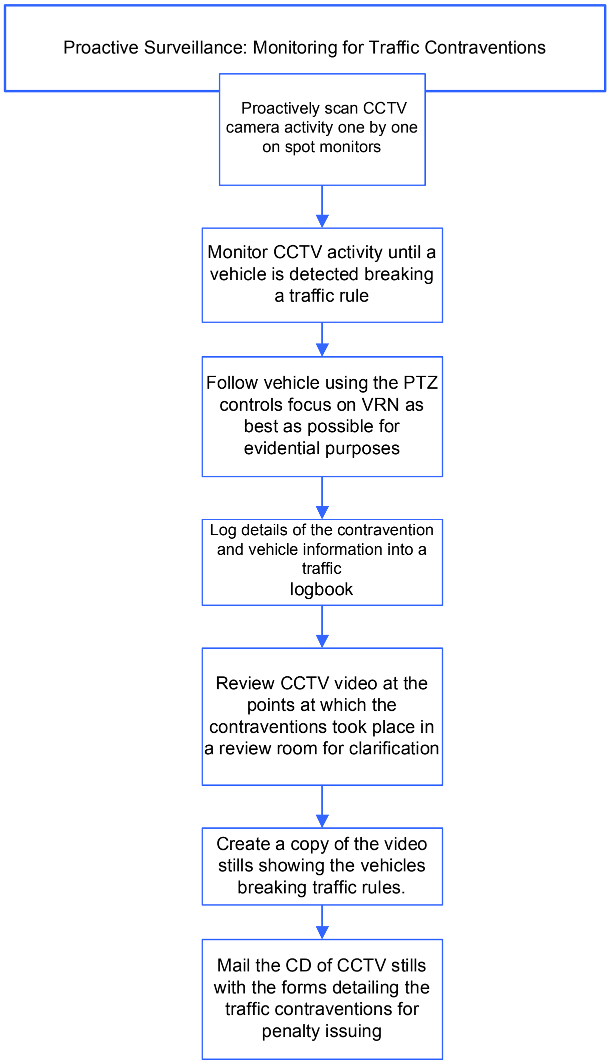

Proactive Surveillance: Monitoring for Traffic Contraventions
Proactively scan CCTV camera activity one by one on spot monitors
Monitor CCTV activity until a vehicle is detected breaking a traffic rule
Follow vehicle using the PTZ controls focus on VRN as best as possible for evidential purposes
Log details of the contravention and vehicle information into a traffic logbook
Review CCTV video at the points at which the contraventions took place in a review room for clarification
Create a copy of the video stills showing the vehicles breaking traffic rules.
Mail the CD of CCTV stills with the forms detailing the traffic contraventions for penalty issuing

Figure 4.6: A task analysis for a traffic enforcement task.

Reactive surveillance

Reactive surveillance involves operators reacting to information they receive through radio, telephone, email, fax, and face-to-face communications. They can also react to information from an intelligent CCTV system (i.e., ANPR system). Once information is provided about an incident, operators instantly react and locate the scenes by either entering the camera number into the computer system or by inspecting the CCTV activity on the monitor banks. When operators receive calls from participant radio users (e.g., pubs, clubs and shops), the callers first identify themselves on the radio by giving their business name and location in the form of two letters, and the control room's camera number (e.g., 'Mike Sierra six' was the radio caller's ID at control room 3 for camera number 6 at the Marks and Spencer store). Figure 4.7 details a breakdown of the reactive surveillance task performed by CCTV operators.

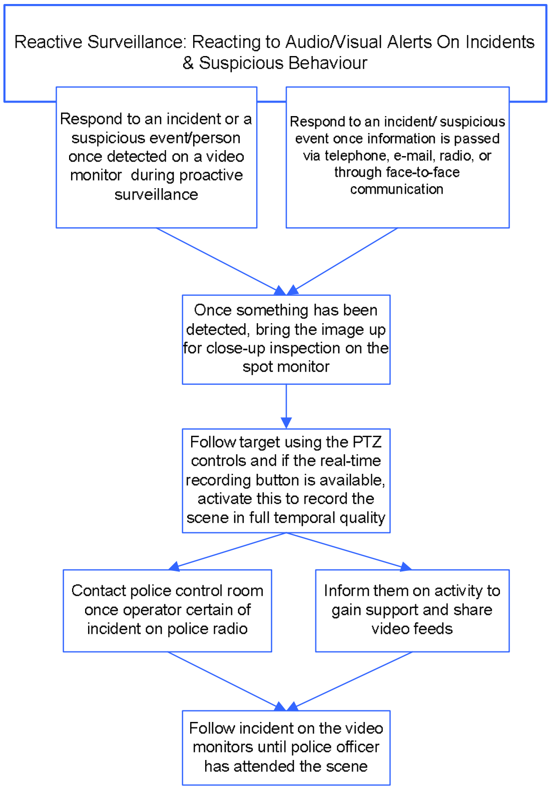

Reactive Surveillance: Reacting to Audio/Visual Alerts On Incidents & Suspicious Behaviour
Respond to an incident or a suspicious event/person once detected on a video monitor during proactive surveillance
Respond to an incident/ suspicious event once information is passed via telephone, e-mail, radio, or through face-to-face communication
Once something has been detected, bring the image up for close-up inspection on the spot monitor
Follow target using the PTZ controls and if the real-time recording button is available, activate this to record the scene in full temporal quality
Contact police control room once operator certain of incident on police radio
Inform them on activity to gain support and share video feeds
Follow incident on the video monitors until police officer has attended the scene

Figure 4.7: A task analysis for a reactive surveillance task.

Telephones are rarely used by CCTV control room operators when performing reactive surveillance. On a few occasions, operators receive telephone calls from the public related to neighbourhood issues, but telephone calls are mainly used for internal communications between control room staff. In five of the 14 control rooms in London, an emergency telephone service was in use for operators to deal with emergency calls from the public and council departments. At these control rooms, operators dealt with issues such as property fire incidents, water leaks, and noise disturbances. This task was carried out in parallel to other tasks (proactive, reactive and CCTV video review and administration tasks), and at times operators found that the incoming calls were distracting. In two of the 14 control rooms, ANPR technology was used to support operators performing reactive surveillance.

CCTV video review and administration

These tasks are performed when operators are not performing surveillance or attending to radio calls and involved the following:

1. Logging all incidents observed on-screen into a database and/or logbook.
2. Preparing working copies of CCTV video footage for the police and traffic enforcement departments to aid in investigations and also to provide evidence to the court.
3. Labelling and bagging of CCTV video evidence for the police.

These tasks were not observed during the field visits, as they were performed outside of the control room in a video analysis room where access was not permitted to visitors. Operators enjoyed these administrative tasks least, as it took them away from watching the CCTV monitors. Operators described administrative tasks as dull, boring, mundane, and lonely. One operator commented, *"…doing any admin work is boring and not as interesting as working with CCTV … also you don't get to chat to anyone in that room, as you're on your own – in silence."*

4.4.2.5 Task Performance Issues

Having identified the different CCTV stakeholders, their roles and tasks (study objective 1), the next phase of the analysis involved the identification of the social and technical constraints associated with operator tasks, as well as conflicts between CCTV stakeholders (study objective 2). This investigation was necessary to determine why CCTV is ineffective, and more specifically, what problems operators experienced with their tasks and at what cost. The aim of this investigation was to determine how the technology, operator tasks, and processes could be improved (study objective 3). The task performance issues were identified through interviews and observations with operators whilst performing tasks. The task performance issues were identified by referring to the factors detailed in the observation checklist used to structure the note taking (operator's tasks, workstation set-up, artefact usage, situation awareness, and processing of video footage). These issues were grouped into categories which shared common themes. The analysis was achieved by creating an Affinity diagram (see Chapter 3, Section 3.3.2.3). Each major theme was then rated, based on how significantly it impacted the effectiveness of CCTV, and hierarchically organised based on these ratings. Figure 4.8 illustrates the effectiveness scale used to rate each performance issue:

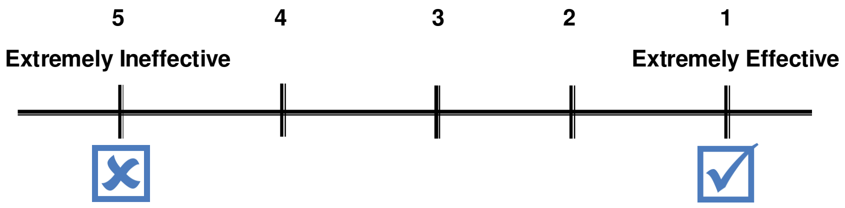


Figure 4.8: CCTV Effectiveness 5-point Likert rating scale.

As well as identifying problems associated with the performance of operator tasks, it was also possible to identify the potential (as well as actual) conflicts between CCTV stakeholders, which are a result of the performance issues. These issues are discussed in full in the following section.

From the analysis, five key areas were identified as contributing to the ineffectiveness of CCTV (a total of 17 issues):

1. Poor camera positioning
2. Ineffective workstation set-up
3. Difficulty in searching/locating scenes
4. Very low-quality CCTV video recordings
5. Ineffective stakeholder communications

These issues are summarised in Figure 4.9, which also indicates the number of control rooms affected by each issue and their CCTV effectiveness ratings. The graph illustrates that the biggest issues were the use of very low-quality CCTV video recordings, poor camera positioning and ineffective stakeholder communications. The specific issues (17) identified under each of these five factors which affected task performance are detailed in Table 4.8. Each of these factors are described in more detail in the following section.

Table 4.8. Summary of Factors Affecting Task Performance and CCTV effectiveness

Factor Affecting Task Performance	No. of Control Rooms Affected	Effectiveness of CCTV Rating
Poor Camera Positioning	8	4
Ineffective Workstation Set-up	4	3
Difficulty in Searching/Locating Scenes	5	3
Very Low-Quality CCTV video recordings	9	5
Ineffective Stakeholder Communication	8	4

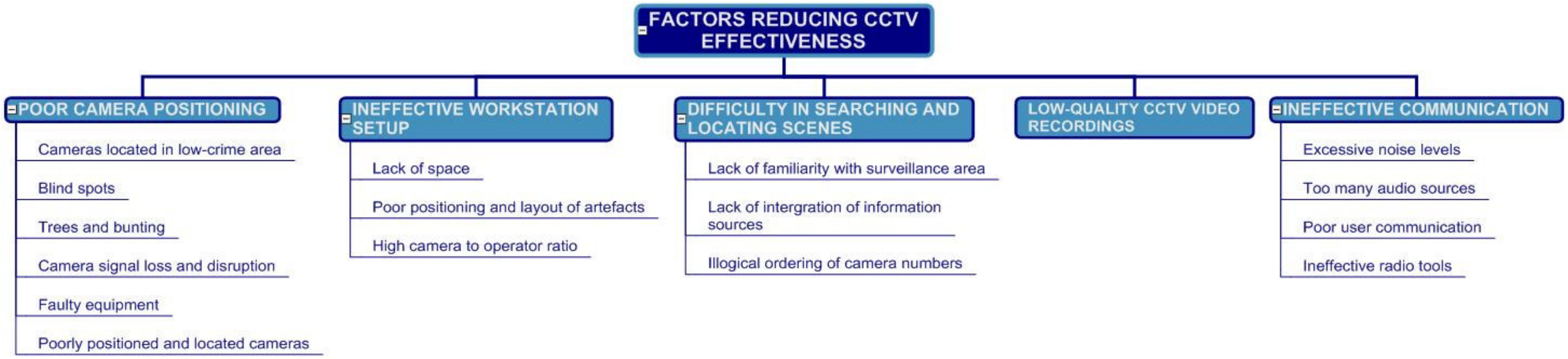

FACTORS REDUCING CCTV EFFECTIVENESS
POOR CAMERA POSITIONING
Cameras located in low-crime area
Blind spots
Trees and bunting
Camera signal loss and disruption
Faulty equipment
Poorly positioned and located cameras
INEFFECTIVE WORKSTATION SETUP
Lack of space
Poor positioning and layout of artefacts
High camera to operator ratio
DIFFICULTY IN SEARCHING AND LOCATING SCENES
Lack of familiarity with surveillance area
Lack of intergration of information sources
Illogical ordering of camera numbers
LOW-QUALITY CCTV VIDEO RECORDINGS
INEFFECTIVE COMMUNICATION
Excessive noise levels
Too many audio sources
Poor user communication
Ineffective radio tools

Figure 4.10: Breakdown of performance issues that reduce CCTV effectiveness.

Poor camera positioning

Cameras located in low-crime areas: Operators complained that fixed CCTV cameras (immovable camera) were left in 'useless locations' where no crime occurred and where little activity took place (e.g., no people or cars). Operators in several of the control rooms found that many of the cameras were originally installed in crime hotspots (where crime *was* very high), and the purposes and functions of these cameras had not been reviewed. At control room 10, operators said that some of the fixed CCTV cameras were installed over ten years ago and no assessments had been made to relocate them to locations where they are now needed. Operators felt very frustrated, as there were too many cameras on the system which were of no use to them. An operator from control room 6 commented on this issue: *"Nine out of 10 times, nothing happens on these cameras. This is quite annoying, as I could think of plenty of places to relocate them."* This problem was identified in six of the 14 control rooms; where CCTV cameras had been left in their original positions without review for over eight years. Over time, operators identified these cameras and developed a tendency to ignore them when performing proactive surveillance.

Blind spots: While some cameras are located in low-crime areas, many of the control rooms had installed cameras in busy, crime-prone areas (e.g., high streets, markets, park entrances, taxi ranks, train and bus stations). Despite this, many operators complained about the poor positioning of cameras, which resulted in blind spots where parts of the camera scene were invisible. Operators explained that criminals are aware of some of these blind spots and 'hide' from the cameras (e.g., suspects were seen crouching beneath a camera during an observation at one control room). Operators were also aware of this, as they were told by street police officers that suspected criminals were hiding beneath cameras when dealing drugs. This was also identified by Luff and Heath (200) in their London Underground control room study. At control room 5, one operator said some people are aware that they are being watched, as the 'shoebox' style cameras give away the direction and angle of view. As a result, operators prefer to perform their surveillance tasks using 'dome' style cameras, where the camera is concealed in a semicircular black dome: *"People don't know where to stand under dome cameras as the angle of the camera is hidden."*

At control room 13, operators complained that they were not informed by management about housing and building developments taking place within camera surveillance areas they were required to monitor. The building developments involved the erection of scaffolding material and a lot of traffic from construction vehicles, which blocked some camera scenes where surveillance was necessary. Operators at these control rooms (two of the 14) said that if they were informed on the building development projects within the surveillance areas, they could keep an 'eye out' for any problems, such as building site trespassers and children playing near work areas. Furthermore, operators said that if there was better communication between town developers and control room management, control room managers could ensure that CCTV cameras are not obstructed by building scaffolding and work vehicles.

Trees and bunting: CCTV control rooms located within greener areas also caused problems for operators locating cameras scenes. There are many cameras which are physically obstructed by overgrown trees and bunting. For instance, at control room 13 (which is located in the heart of a national forest) an experienced operator commented about camera obstructions: *"The pain of life we have with natural trees, Christmas trees, bunting, and carnival decorations!"* Seasonal decorations, especially Christmas lighting was a huge problem for operators trying to scan camera views in close-up detail on their spot monitors, as the Christmas lights over-illuminated the camera lenses and altered the its functionality. Bunting, such as carnival banners and decorations also affected camera settings, causing cameras to self-focus to nearer objects. Managers asked operators to ignore cameras that are obstructed until the trees are cut down and the decorations are removed; however, operators said that overgrown trees were never pruned or cut down. One manager commented on this particular issue: *"It was down to one government department not talking to another and this was because the planning Section within the council would not allow trees to be cut down because of their environmental policy."*

Operators stressed that it is important not to lose sight of any camera views, particularly in areas where there is a high level of activity and crime, for example within town centres. At control room 10, the council did not authorise their staff to maintain trees and foliage in the town centre, due to strict environmental policies. The control room manager said that he was left with no choice but to install an additional CCTV camera near a camera that had been occluded by trees to ensure that operators had access to these scenes. In this particular instance, it is clear that when the environment of the system changed, no changes were made by control room managers to alter the system, which meant that the original goals set out for the system could not be achieved and the effectiveness of the CCTV was reduced.

Camera signal loss and disruption: Poor weather conditions (heavy rain and wind) disrupted the transmission of analogue video signals. This was identified as a problem at six control rooms. During temporary signal loss or disruption, operators showed signs of disappointment, distress, and frustration. When signal loss occurred over a very long period of time (e.g., two hours), some operators were relieved and used the opportunity to take a break from their workstations. Operators reported that in general, camera maintenance and repairs took too long and they felt helpless when it happened.

The delivery of CCTV video to control room monitors was very poor at two of the control rooms as a result of poor weather conditions (microwave radio transmission was used to stream video). On two occasions, operators at two different control rooms were observed struggling to locate targets and vehicles due to signal loss and disruption. Signal loss also occurred at the control rooms which used digital technology; camera scenes would freeze and in some cases, scenes would disappear altogether until the system was reset by the control room manager. On these occasions, control room operators blamed the incompetence of installation and maintenance engineers, as well as their managers for not recognising the problems.

Faulty equipment: At four of the control rooms, more than five cameras were faulty. Operators said that their managers spent less time and money on repairs and more on equipment upgrades. A shift supervisor said that some CCTV cameras were installed over 12 years ago and none of these cameras were properly maintained or even repaired. At control rooms 9 and 14, three CCTV cameras could not be displayed on the monitor wall, as they were damaged by vandals and needed repairing/replacing. At control room 7, 12 CCTV cameras were owned by a local housing association and were monitored by CCTV operators. These cameras produced markedly low-quality video compared with the CCTV control room cameras. The control room manager was questioned about these cameras and said that, *"it was out of [his] hands,"* as he did not own these CCTV cameras.

In addition to faulty CCTV cameras, there were many other artefacts which failed to work, reducing operators' effectiveness in their surveillance tasks. At control room 10, the video displayed on three monitors were out of focus, and operators suspected this was possibly due to the older cameras having a faulty iris. At another control room, there were several cameras whose movements were *"…too slow to catch the crime."* When operators attempted to move these slower cameras using the PTZ camera controls, the scene appeared to shake and move very slowly. On a number of occasions operators showed signs of frustration when attempting to change the direction and angle of the cameras. It was clear that the mechanical workings of the older cameras were worn out and needed replacing altogether.

At control room 7, three different intelligent CCTV systems were in use: 1) an ANPR system; 2) a person tracking system and 3) an object tracking system. The two tracking systems were not fully functional, and thus, could not be used by operators. They had been introduced on a trial basis, but as the control room manager explained, the tracking systems did not work in trial mode as the system manufacturers had provided an incorrect version of the software.

Poorly positioned and located cameras*:* Operators on previous shifts would leave movable cameras in 'useless' positions (e.g., facing the sky and road instead of road bends, junctions, and building entrances/exits). Some camera controls were also left in full zoom mode or out of focus. Operators suspected that these settings were not corrected, because other operators were lazy, careless, or bored (and hence, played with camera controls during quiet periods). One operator commented that *"operators should leave cameras trained so that they are looking at something decent."* Another operator agreed, saying that, *"If operators do their jobs properly and leave the cameras in their active positions it's fine then, cos … I'm at my disposal then once I got all the cameras in front of me."*

Ineffective workstation set-up

Lack of space*:* Four of the 14 control rooms were in the process of upgrading their technology in a number of ways: adding more CCTV cameras (and video monitors), upgrading the network and storage facilities from analogue to digital and integrating intelligent CCTV systems, such as person and body tracking systems to the control rooms. Upgrade work had led to constant change in the operators' work environments, which in some cases caused work performance difficulties. For instance, equipment was regularly moved to different positions (see next issue) and old systems were stored in the control room even though they were no longer used. As the focus of this study was to understand the context of CCTV, a comprehensive ergonomic workstation assessment was not completed. Nevertheless, it was discovered that seven operators experienced physical discomfort when performing CCTV tasks at their workstations due to a lack of space.

Poor positioning and layout of artefacts: At six of the 14 control rooms, operators reported that they were unhappy with the physical set-up of their workstations. The cause of this was mainly linked to changes made during and after upgrade work. One operator at a police control room was five months pregnant and found it difficult to reach items on the front of her workstation (radio consoles and telephones). She instead concentrated on proactive surveillance, whilst receiving support from her colleague to physically respond to incoming radio and telephone calls. At control room 8, an operator complained that her PTZ controller (joystick) wire was too short and as a result used the controls with her left hand, which she stated was unnatural and uncomfortable after a long period (see Figure 4.11).

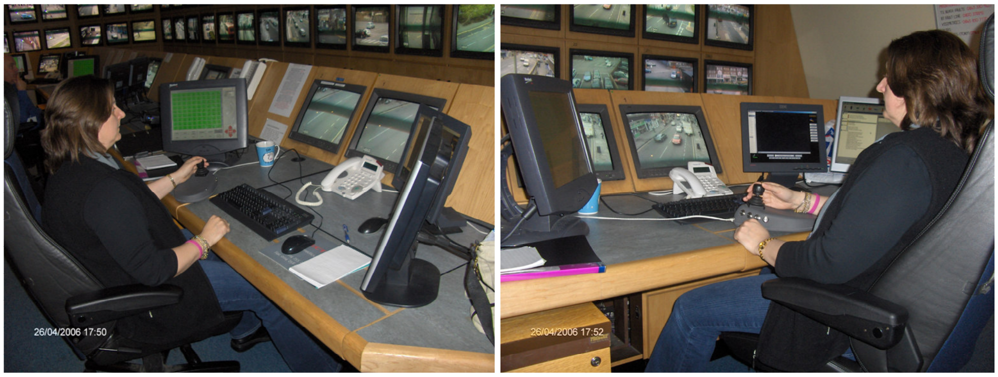


Figure 4.11: PTZ joystick used in #1 original position and #2 preferred position.

At control room 7, the ANPR monitor was not very well positioned for the operator to visualise alerts and vehicle registration numbers. The image in Figure 4.12 shows that the ANPR monitor was positioned at almost standing height on the wall too far away from operators. During alerts, operators were seen physically stretching and straining their eyes to gain a better view of the information being displayed. The manger explained that this was *"just a temporary set-up."*

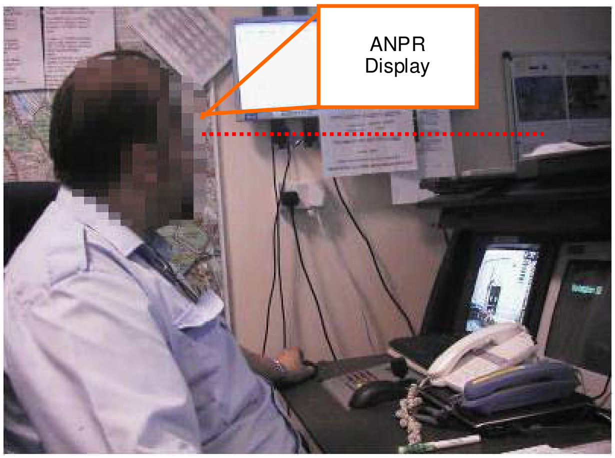


Figure 4.12: ANPR monitor poorly positioned.

High camera to operator ratio: The number of CCTV cameras used for surveillance depended upon the size of the surveillance area and the level of funding available for additional cameras. The larger the area of surveillance, the larger the number of cameras displayed in the control room. At the control rooms visited, not every CCTV camera was displayed on every single video monitor; instead, a set number of cameras were displayed on approximately 60 monitors, side-by-side along the monitor wall. Some camera displays were multiplexed (i.e., four video outputs displayed on one monitor) and some were displayed on video monitors using an autocycling/switching mode (an alternative method of presenting the output of CCTV cameras at specific time intervals). Operators complained that there were far too many cameras to monitor and their managers were adding more and more CCTV cameras to the system. One operator commented that, *"the increase in the number of CCTV cameras at one control room was overwhelming, particularly in the last six months."* If operators have a large number of cameras, more time is needed to search through them during proactive and reactive surveillance tasks. Furthermore, having access to a large number of cameras increases the operator's memory load when recalling camera IDs and locations during a reactive surveillance task.

In the control rooms where CCTV was being monitored on behalf of a housing association, operators struggled to see scenes displayed on multiplexed monitors. If it is assumed that each operator is responsible for scanning an equal number of CCTV cameras along the monitor wall, the camera to operator ratios (as shown in Table 4.2) indicate that operators have too many cameras to monitor at one time, but this is only the case when performing proactive surveillance using method 1 (see Section 4.4.2.4). Managers explained that operators do not usually monitor CCTV activity passively. Instead, most use their spot monitors to inspect CCTV video close-up. They regard the CCTV overload issue as a temporary problem which associated with staff shortages (e.g., staff off sick or on leave).

Difficulty in searching and locating scenes

Lack of familiarity with surveillance areas: For both proactive and reactive surveillance tasks, operators maintained their situation awareness with the use of geographical maps (paper maps and electronic maps) of the surveillance areas. Operators who were new to the job and the surveillance environment struggled to locate scenes in unfamiliar locations - one reason being that, they did not live in the area. This reduced their efficiency in responding to incidents.

Lack of integration of information sources: At four of the 14 control rooms, operators used a geographical and information system that allowed them to view a camera scene instantly by touching an area of the screen displaying the surveillance map. This was also accomplished by entering the camera ID or street name into the camera database which was linked to CCTV video outputs and a map. However, some systems did not offer an integrated link to view video. An operator commented: *"This system is not that usable to be fair. I would have preferred if we had a database where we can enter street names or camera numbers. Once the details such as the postcode or street name are entered, a map should display the regions showing our control room camera numbers."* Instead the system only allowed the operator to select an area on the map and pan and zoom to a desirable level showing camera locations. At the control rooms in which only paper maps and camera lists were available (ten of the 14 control rooms), operators were encouraged to memorise the camera locations, which is likely to cause difficulties (e.g., memory recall) for the inexperienced and older CCTV operators.

Illogical ordering of cameras: Another factor which reduced the ease of locating scenes was the illogical placement of CCTV cameras on the list and in the database. In a majority of the control rooms visited, cameras were organised by installation date and not grouped by geographical location. Furthermore, where geographical information and mapping systems were in use, any new cameras added to the system were placed at the end of the camera list. This created an illogically organised list of cameras which resulted in difficulties for operators when they were required to locate the right camera at the right time. One operator commented: *"The new cameras were good, but hard to find and not placed in any proper order which is annoying at times."*

Low-quality CCTV video recordings

Operators performed proactive and reactive surveillance (observational) tasks by scanning CCTV video in real-time and did not use post-event recorded video for any of their observational tasks. Operators did, however, process recorded CCTV video (e.g., prepare working copies of CCTV video for the police) to help police investigate a crime. Operators said that they were responsible for data retrieval, and not video analyses. At times, they found this task to be *"very time consuming and tricky,"* because the video quality was never as good as what they (and the police) expected. Across the control rooms visited for this study, operators felt that the CCTV video recordings were very low and unusable for police use.

When control room managers were asked what they thought about the quality of the CCTV video recordings, 11 managers responded positively and believed their systems produced videos at acceptable quality. The remaining managers were less certain. Nine operators said that the quality of the CCTV recordings were too low. It was only possible to identify the recording frame rates at 11 of the 14 control rooms (see Table 4.10). At the remaining control rooms, both managers and operators had very little knowledge of CCTV video quality. One operator at control room 4 commented on this issue: *"The police are never able to make out incidents at night, such as nightclub fighting and shop burglaries, because of things like lighting and camera quality - the recordings are just not up to scratch."* Recording (and streaming) CCTV video at very low frame rates reduces the effectiveness of CCTV, as the police will not be able to use the recordings to investigate crime or use it as evidence in court. An overview to temporal video compression and the implications of lowering video frame rate for a detection task are presented in Chapter 5.

Following discussions with managers, it was evident that they had very little knowledge and guidance on how to configure their CCTV systems to ensure that the CCTV recordings were suitable for operator tasks. Managers said that they decide on the video quality parameters based on the number of days, and cameras required for 24-hour recording (which is usually 28–31 days). The type of digital CCTV system deployed was chosen by the manager with the help of a security consultant, and the purchase was constrained by a strict budget. This was also the case in the CCTV control room study conducted by Gill et al. (2005).

Table 4.9: Chosen CCTV Recording Frame Rate at 11 Control Rooms

No. of Control Rooms	Frame Rate
4	1 fps
3	5 fps
1	6 fps
1	8 fps
2	12.5 fps

The following comments were made by three control room managers:

1. Control room 2: CCTV video was recorded at 6 fps and employed M-JPEG compression which was set to record video at high quality.

   "Small quantities are fine. However, when bulk downloads are required for counterterrorist police, the download takes hours and the process is long and cumbersome."

2. Control room 4: CCTV video was recorded at 5 fps and operators used the 'real-time button' to increase the recording frame rate from 5 fps to 25 fps when an incident was detected. The system used M-JPEG compression, which was recorded at medium quality.

   "I'd say the video quality is good, but only when the real-time button is pressed. So it's good when we know an incident is taking place. Operators either see something happening or they're informed by the police control room on the radios. This, we find, is the best compromise, as we wouldn't be able to increase our budget any further if we bumped up the video quality – especially as we've got over a hundred cameras tied to our system."

3. Control room 11: CCTV video was recorded at 12.5 fps, and employed MPEG-4 compression which recorded video of low to medium quality.

   "The quality is alright at times. I mean we record at a high frame rate so it's pretty much smooth, so we know exactly what's happening. I think the only issue is that…often the police have had problems establishing the identities of ASBOs and other low life criminals as the pixel count in the videos we record are quite low. But this isn't much of a problem in the day just at night when you can't see much anyway."

Out of the 14 control rooms studied, three used a real-time recording facility. The real-time feature is popular for recording CCTV video at full temporal quality. As one operator noted, *"when an incident is taking place, like a fight or a shoplifting incident, we can hit the real-time button in the control room which raises the recording frame rate from 2 fps to maximum quality [25 fps]."* The only issue with this facility is that operators must notice an incident and then immediately press the real-time button. Those incidents happening on cameras not directly visible to the operator are recorded at a low frame rate. The effect of lowering CCTV video frame rate on a detection task is investigated empirically in Study 2 (see Chapter 9).

Ineffective Communication

Excessive noise levels: In the busiest control rooms, there were many incoming radio and telephone contacts; as a consequence, the noise level was high. In five control rooms, it was noted that operators struggled to hear information over the radios and telephones. For instance, at control room 14 (the police control room at Heathrow airport), CCTV operators complained about the noise. An intercom system was used at the control room for operators to contact security personnel. One operator complained: *"…that thing is so high pitched it's piercing! But I suppose at least we know we can hear it alright."* Despite this, operators felt that *"…nothing much could be done to change it as management had other issues to resolve."*

Too many audio sources: At the control rooms based in central London, operators were required to process radio and telephone calls continuously, particularly during day shifts. Operators (at all control rooms) stated that all radio communication is recorded onto tape so that any discrepancies can be verified if necessary. Operators new to the job were observed making notes when talking on the phone and radio and experienced operators reacted to incidents whilst on the call. It was clear that there were too many radio channels assigned to operators performing reactive surveillance tasks. For example, at control room 11, two operators working side by side were expected to respond to seven different communication devices, all of which were positioned on top of a filing cabinet (see Figure 4.12). One of the operators commented about this set-up: *"We are overwhelmed with the number of phones and radios here. At times they go off all at once and we can't tell which one is going off, as the tones all sound the same."*

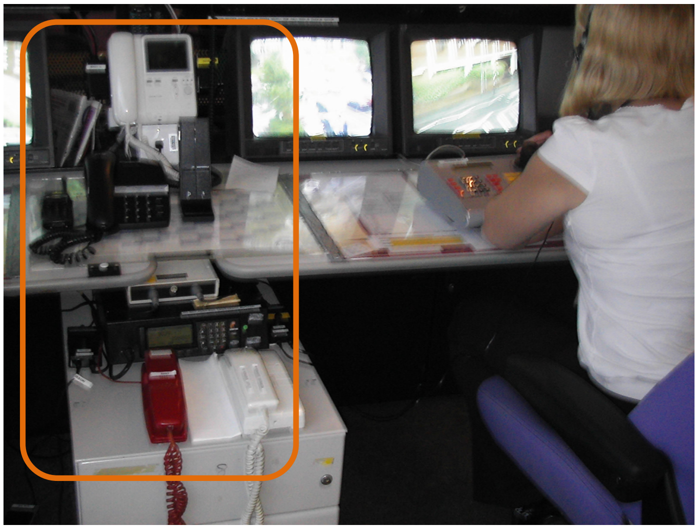


Figure 4.12: Several communication tools placed in the same location.

In control room 9, there was only one operator working on a shift and he was expected to manage eight different communication devices by himself. The operator was required to respond to calls from the police radio control room, council, the public, as well as monitor 80 CCTV cameras on his own. During the observations, the operator was asked how well he thought he managed the different audio devices, and replied: *"Well I try to do everything as well as I should, but it's like I'm an octopus with all this equipment."*

Poor communication: CCTV operators and police staff used the NATO phonetic language to describe and clarify names and spellings. Police identity codes were also used to describe the ethnicity of targets. The use of the standardised languages and codes worked very well even in the noisiest control rooms. However, not all radio users were familiar with the codes, which made communication between CCTV operators and participant radio users less effective. At control room 10, operators found that, *"…some shop managers give beautiful descriptions, but others are very vague, and they don't tell us where the target is, which is the bit of information we need!"* Furthermore, shop radio users, especially in central London were difficult to understand, because English was not their first language and they had strong accents. Operators tried their best to understand callers by asking them to slow down and repeat themselves. This was observed on a number of occasions during the visits. Poor communication reduces rapport between CCTV operators and retail participant radio users. To compound the problem further, operators also found it difficult to build a good relationship with shop radio user, since shop staff turnover was high which meant they were in contact with different people every time they received a call.

The effectiveness of radio communication was also reduced (at five of the 14 control rooms) as shop staff were not able to provide incident information to operators clearly. Operators reported that shop staff usually panicked, spoke too fast, and gave too much unnecessary information. At other times they gave too little information or provided important information in the wrong order. An operator at control room 7 commented that, *"shop radio users are incompetent, as they shout over the radio and blabber. All we need to know is who they are, the location of the incident and clear information about the incident."* Consequently, operators misunderstand shop radio callers and at times are led to believe that the incident is more (or less) serious than it actually is.

Faulty radio tools: In six of the 13 control rooms, communication between CCTV operators and other radio users was ineffective as a result of faulty radio equipment. The delivery of the audio from analogue radio systems was affected by poor weather conditions (similarly to the delivery of analogue video). In control room 10, three operators using (analogue) radios were observed asking shop staff to repeat themselves due to signal transmission problems. At another control room, operators reported hearing a 'crackling' sound each time they spoke over the radio. In the control rooms using digital radios, some channels were not used because they were faulty or misconfigured. *"They would sound like they were talking through a goldfish bowl because the radios were not encrypted and when it was encrypted it knocked the phones out somehow."*

4.5 Discussion

As a result of the contextual inquiry studies conducted at 14 different CCTV control rooms, 17 task performance issues were identified, all of which impacted the effectiveness of CCTV. In this section, the implications of these findings described in Section 4.4.2.5 are discussed in relation to the operators' task performance and the impact each had on different CCTV control room stakeholders.

4.5.1 Discussion of Findings – Stakeholder Implications

Poor camera positioning

Cameras located in low-crime areas

During the observations, there were many instances in which operators were unable to locate scenes quickly enough following reactive calls. This resulted in delays in police response to incidents, as well as confusion when operators were in radio contact with police control room operators and street police officers. Operators felt frustrated and embarrassed when they lost track of people. The older CCTV cameras remained installed and accessible to operators, and the new CCTV cameras were illogically placed within the system. This made it difficult for operators to locate scenes quickly. As the effectiveness of CCTV cameras were not assessed on a regular basis, operators were unable to support radio users during reactive surveillance effectively. This issue affected the relationship between operators and police staff.

Blind spots

The presence of blind spots will reduce the operator's confidence in what they see as they will have a limited view of the scene. They are also likely to commit errors when identifying targets or detecting events. A stakeholder conflict was observed between a control room manager and a superstore manager at one control room. A large superstore had recently been constructed at a town centre, which unintentionally created blind spots for two CCTV cameras. The superstore manager refused to provide funding to reinstall two new cameras. After six months and a series of meetings between council members and the control room manager, the council authorised funding for two additional cameras to be installed near the two obstructed cameras. In this particular situation, the stakeholder conflict existed between the control room manager and the superstore manager who refused to provide installation funding for two CCTV cameras.

Trees and bunting

Trees and bunting altered the camera functionality and views which made it difficult for operators to discern the identities of targets and interpret events during the performance of proactive and reactive surveillance tasks. Problems associated with maintaining camera environments created stakeholder conflicts between control room managers and the council, as evidenced in the scenario presented earlier detailing how one control room manager was forced to deal with the problem of overgrown trees obstructing cameras by installing new cameras.

Camera signal loss and disruption

When operators were asked about the problems they experienced during signal loss and disruption, they explained how it reduced their ability to respond to crime when in contact with external agencies (e.g., police control room operators, shop staff, council staff and the public). Operators also claimed that radio users who contacted the control room to report incidents had no patience and did not take the time to understand their technical difficulties. This conflict occurs between operators and radio users, and leads to miscommunication, frustration and delayed responses.

Faulty Equipment

When operators experienced problems locating and sharing cameras with the police control room operators, they felt frustrated and embarrassed. Some operators were also afraid that police operators would not understand their difficulties: *"The police operators expect us to provide them access to the cameras they want instantly and it's not always that easy."* A number of operators claimed that police operators were not always considerate towards them when they struggled to provide them with a camera view. This was perhaps due to a lack of awareness. *"If they knew, they probably wouldn't understand."* Thus, a conflict occurred between CCTV operators and police operators as a result of faulty equipment.

Poorly positioned and located cameras

Operators found it very frustrating when the CCTV cameras were installed in a 'useless' positions and locations. At one control room, an operator said that, *"…one camera linked to the control room was pointing away from an estate where the cheapest drugs could be bought."* This operator was aware of the crime hotpots from previous surveillance work with the police. Cameras installed in poor locations will increase the time required to locate scene of interest as these camera will take up space within the camera list, reducing the operator's efficiency in locating scenes of interest. Similar to the issues described above, the stakeholders most affected by this issue were the CCTV operators and participant radio users who performed reactive surveillance tasks. If CCTV operators cannot perform their tasks effectively due to poor configuration of their equipment, they will not be able to support, or be supported by, police staff and other radio users.

Ineffective workstation set-up

Lack of space

Although a lack of space does not directly impact other CCTV stakeholders, it appeared to contribute to internal conflicts between operators and their managers. For instance, there appeared to be a poor level of communication between operators and their managers when dealing with issues related to workstation set-up. Operators said that they avoided reporting these types of problems as they felt it was insignificant compared to the problems they experienced with technology. They were also reluctant to mention these issues due to the bureaucracy associated with equipment removal and replacement. Furthermore, there was no process in place at a majority of the control rooms to report issues related to an operator's work environment.

Poor positioning and layout of artefacts

See the stakeholder implications discussed in the preceding Section.

High camera to operator ratio

Three of the 14 control room managers admitted that their operators were somewhat overloaded with CCTV video. One manager was defensive and said, *"It's just a short term issue and is due to the problems with staff retention and sickness, rather than the number of cameras."* The remaining 11 managers did not think that operators were overloaded with the number of cameras being displayed on the monitor walls, *"as operators do not usually monitor all of them in one go."* Despite these findings, many operators complained about the number of CCTV cameras: *"…our manager just keeps on adding more and more cameras to the system and forgets that we have one pair of eyes!"* This problem created conflicts between operators and other radio users when the operators were required to locate scenes under time pressure. The longer the camera lists, the less efficiently operators are able to locate scenes, thus reducing their response times when dealing with queries from CCTV stakeholders external to the control room.

Difficulty in searching/locating scenes

Lack of familiarity with surveillance areas

At two control rooms, experienced operators drew maps of the surveillance areas, crime hotspots and camera numbers to help operators new to the job/control room with their surveillance tasks. These operators were asked how remembered the ID and locations of up to 80 cameras, and the responses were similar amongst operators:*"…it's like revising for an exam…we remember the ones we use the most and we are tested on locations and camera numbers from time to time by our supervisors."* Operators had a tendency to remember cameras which were used most frequently. Certain cameras were also regularly monitored as they were recalled easily. Experienced operators had favourite cameras which were based on the image quality, angle, and type of camera. The main problem with focusing on specific cameras is that operators are likely to miss incidents which may be captured on other cameras (which are ignored). Although crime hotspots were identified by operators, not all operators took these into account. It was clear that training was needed to help less experienced operators familiarise themselves with the surveillance areas and take into account of crime statistics.

When operators communicated with the police operators using radio, police operators were at times impatient with the less experienced CCTV operators who were unable to immediately locate the scene. These operators felt pressured and wished that police operators would understand the nature of their work and be more patient.

Lack of integration of information sources

Analogous to the stakeholder implications discussed for the issue above, at the control rooms where systems were not integrated, radio communication was ineffective between CCTV operators and participant radio users. This was because operators were not able to locate scenes quickly enough to identify targets and interpret the events being observed. This was also found to be a problem when operators used the ANPR system. Incident response was more efficient and communication between radio users improved when the camera controls (keyboard and touch-screen interface) were directly linked to video outputs.

Illogical ordering of cameras

This issue mainly affected the two-way communication which took place over radios between operators and the following users:

1. Council staff who report local community problems: Operator response was slow, as they were unable to locate the correct cameras.
2. Police control room operators: There was a delay in operator response when they tracked targets and cars from one camera to another. On a few occasions, operators were seen displaying cameras from unrelated locations within the surveillance area, as the ordering was confusing.

As a result, these radio users received limited support from control room operators, resulting in police delays to serious incidents. This may lead police operators to rely less on CCTV operators for remote incident support and more on police officers for immediate street support.

Low-quality CCTV video recordings

Operator task performance was not directly affected by this problem as they interacted with real-time video for their security observation tasks, and not recorded CCTV video. Despite this, operators were expected to capture CCTV video (using PTZ controls) so that the police could use them for investigations and evidence. At times, the technology was too old, broken, and lighting was poor to achieve effective video capturing. Stakeholders most affected by the use of low-quality CCTV video are the police. One police officer who was visiting a control room said that he was very pleased with the quality of CCTV from that particular control room, as it was often recorded in full frame rate. This only applied to the video recorded once the operators activated the real-time facility (which is not available in all control rooms). *"When this real-time recording facility is not used, and video is recorded before [at least an hour before] an incident took place, the CCTV video is more or less useless."*

Two crime reduction police officers at other control rooms painted a slightly different picture about the quality of digital CCTV video:

> Officer 1: *"It varies greatly across the different systems, but the demand is for high quality."*
>
> Officer 2: *"Nationally (and internationally) they remain poor. We continue to reject 80% of all recorded images associated with a criminal offence as being worthless for identification purposes. Evidentially, far less than 20% are useful. In the investigative process, no accurate figures are available and depending on the seriousness of the crime (homicide and terrorism for example) all recorded images will be used; although I estimate well over half are of no value whatsoever – purely due to poor quality."*

These sentiments were echoed by a forensic image analyst in an email conversation shortly after the field visits were complete: *"We do not produce CCTV, but we use the products supplied by police, solicitors, etc. In the majority of cases it is not good enough to provide positive identification. At best, most are capable of recognition only"* (Oxlee, 2004).

'Positive identification' means that there is no doubt in the analyst's mind that Person A is one and the same as Person B. This is normally only achieved if there is a unique identification mark (such as a scar or mole) or if the similarities are numerous and overwhelming. 'Recognition' on the other hand means there are similarities to the extent that someone who knows a person well or has studied imagery of that person in detail believes it possible that the subjects under comparison are one and the same. 'Recognition' might have some probative value, but normally only as support to other evidence. 'Identification' on the other hand might stand as proof in its own right.

If CCTV video is recorded at very low quality (spatial and temporal) it will greatly affect the effectiveness of the CCTV system. Specifically, it will affect the police's tasks when investigating incidents (e.g., identifying targets and establishing what occurred). This specific issue was taken further and investigated with users (see Study 2: Chapter 8 and Study 3: Chapter 9). These studies have been conducted to: 1) understand how untrained participants perform a face identification task and an event detection task; 2) examine the effect of lowering spatial and temporal video quality on a face identification and scene detection task and 3) determine the most effective video quality level for these tasks to improve the effectiveness of CCTV when used for both real-time and post-event security observation tasks.

Ineffective communication

Excessive noise levels

Excessive noise within the control room distracted operators and at times confused operators when relaying or receiving information over the radio and phones. High frequency alarm tones from the ANPR and intercom systems annoyed operators too. It is clear that high noise levels affect operators' moods whilst at work, which reduces their patience in dealing with incidents. The knock-on effect of excessive noise is the ineffectiveness of communication with other radio users, which in the long term has the potential to affect relationships between operators and other stakeholders with whom they are in regular contact.

Too many audio sources

The use of numerous audio sources (phones, radios and alerts from other systems) created a noisy atmosphere within the control room which: 1) distracted operators when concentrating on tasks; 2) delayed call responses; 3) made it difficult to differentiate different levels of priority calls and 3) caused information overload. Operators blamed their managers for allowing too many businesses to participate in the CCTV radio scheme, which resulted in a high volume of radio calls. Most managers stated that operators' workloads varied over the day and night, and when call volumes were high, they themselves *"would provide additional support if operators need it, but this rarely happened."* Despite staff resource issues, the radio channel numbers remained fixed. The stakeholder conflicts for this issue was between CCTV operators and their managers, as well as the operators and external agencies, such as the police control room operators, the police, council staff and the public.

Ineffective user communication

Communication was most effective between CCTV operators and police control room operators, and the least between operators and shop radio users. This difference was apparent as both CCTV and police staff used a standardised language and codes (e.g. NATO phonetic alphabet and police identity codes) when giving information over the radio. If the operator cannot elicit the right information within the right span of time, operator will be likely to miss incident and misinterpret reports. These user-to-user communication issues affected all radio and telephone users, particularly shop radio users. These issues have led to operators building poor relationships with shop radio users. Operators are very negative towards radio users and perceive them to be incompetent and unprofessional.

Faulty radio tools

Having old, unreliable and untested radio systems is a serious problem as it reduces the effectiveness of communications between operators and stakeholders (police staff, council, and shop/pub staff). When performing reactive surveillance tasks, operators received incoming calls from these radio users who guides them to the right cameras to capture criminals and crime scenes. Communication is crucial for this task, as much of the crime at the central London control rooms was detected through reactive surveillance. Poor communication led to the development of poor rapport between operators and radio users, particularly with shop staff for whom comprehension was poor due to accents and the poor delivery of information.

4.5.2 Review of Field Study in Relation to Previous Research

The field study detailed in this chapter used a task-based contextual inquiry - a requirements capture technique which focused on examining operators' tasks. This technique was used to understand the context and use of CCTV, and fundamentally the factors which interfered with operator performance when using CCTV and other associated technologies in the field. This research followed a socio-technical approach in addressing research goal 1 (see Section 1.4.1) and involved carrying out naturalistic observations and interviews with operators at 14 CCTV control rooms. This field technique was altered so that the questions and observations focused on the tasks operators carried out in order to identify the factors affected task performance when using CCTV technology (see Section 3.5). In addition to following a task-based contextual inquiry, a structured observation checklist (see Appendix C) was used to systematically gather observations relevant to the research goals for this study. In addition, a structured interview checklist (see Table 4.3) was used to identify each of the CCTV managers' security goals, stakeholders' roles, operators' tasks, artefacts used, and the performance issues. This technique has not been applied in previous field research in CCTV.

Semi-structured interviews and observations with operators revealed a number of factors which reduced the effectiveness of CCTV, specifically those reducing operators' performance in their tasks and their effect on the security goals for the system, have been identified. These issues were discovered in context using the 'SEE' technique as described in Chapter 3, Section 3.3.2.2. In contrast to previous control room studies (see the literature review in Chapter 3), the study presented in this chapter describes in detail the methods and techniques used in the field. This involved describing the contextual inquiry process, how the data was analysed. One key benefit in describing the methodology and analysis in detail is so that other researchers can understand the study and replicate it.

The findings revealed that operators were not only confronted with too many cameras and monitors, a finding also identified by Gill et al. (2005), but they were also overloaded with audio information from radio, telephones, and other systems. Some of the performance issues identified in this study have been noted in previous control room studies. For instance, Gill et al. (2005) found that both CCTV operators and control room managers had very limited knowledge of digital CCTV systems, with many control rooms recording video at very low frame rates. Furthermore, the systems were purchased following partial security consultant advice. Gill et al. found that some of the CCTV cameras were poorly suited to low light conditions, which reduce the effectiveness of CCTV recordings.

In the London Underground study conducted by Luff and Heath (2001), operators were also confronted with low-quality CCTV video (real-time video), due to poor installation and configuration. Furthermore, several cameras were poorly positioned, resulting in blind spots, which prevent operators from observing the entirety of the surveillance scene. If CCTV operators are required to view hundreds of cameras, their jobs should be made as easy as possible; the information should allow them to interpret scenes quickly and accurately.

In addition to these findings, the ineffectiveness of communication between operators and radio callers was discovered in McCarthy's ambulance control room study (McCarthy et al., 1997). However, a wider range of communication failures were revealed in the study detailed in this chapter: such as the use of faulty radio tools, background noise, poor placement of radio tools, excessive radio contact, etc. Additional factors which reduced effective communication have been identified in this study as a result of the wide range of control rooms examined, unlike McCarthy's study in which only two different ambulance control rooms were studied. It is clear that the design, layout, and process of communication have been overlooked by those who design and implement security systems and control room environments. Communication is a crucial part of an operator's work activities and CCTV owners need guidance on how communication can be improved and maintained at a high standard to avoid operator errors and delays in performing proactive and reactive surveillance tasks.

A number of issues related to camera accessibility and operator communication have been identified in previous control room studies. However, no other study has provided transferable lessons for good practice to improve the effectiveness of control room work performance and the effectiveness of CCTV. The purpose of this research has been to identify key CCTV tasks and the various technologies used to perform these tasks. The performance issues have been identified to: 1) isolate problems associated with CCTV video quality for further empirical research and 2) to use these findings to form a framework which provides best-practice recommendations for security practitioners and owners in order to improve the effectiveness of CCTV.

4.5.3 Field Study – Contributions and Limitations

A number of research contributions have been made as a result of this contextual inquiry study:

A task-oriented contextual inquiry methodology:

Following a comprehensive review of the field research methods used in HCI research, contextual inquiry was chosen as the most suitable methodology for this study. This method was chosen as it takes into account the user's actions and attitudes within a real-world setting during task performance. The approach was modified to allow the data to be used for best-practice recommendations, and this was achieved by focusing the observations in-line with the study objectives, localising issues, and rating them by severity. The use of a structured framework was applied successfully in this study as a wide range of task performance issues were identified within a short period of time. This methodology, therefore, followed a task-oriented approach as the evaluation focused on the users, their tasks and the system in context. Previous control room studies have focused observations and data collection to single study environments (one or two control rooms), which limited the scope of the research findings for control rooms operating under specific contexts.

Good access to security control rooms for inquiry:

The field observations and interviews carried out at the 14 CCTV control rooms were completed successfully with the support and cooperation of local authority and private security control room managers. The visits were organised over many months prior to the field study start date, which involved several emails and lengthy telephone conversations with various stakeholders (control room managers, supervisors, consultants, and police staff). This field study is considered timely research. Four months prior to commencing this study, in July 2006, terrorists plotted suicide bomb attacks on central London. Although there was some delay at the start of the study, gaining access to CCTV control rooms in London was thankfully straightforward, despite the heightened security.

Key limitations of this study include:

Difficulty gaining initial support from control rooms for participation in field study:

There were very few email responses to the study announcement (sent by the CCTV user group chairman). After eight weeks, the response rate was still very low (10%), which was not considered acceptable for the research. This low response rate delayed the start of the study by two months. The study was later publicised following a presentation at a CCTV user group conference to 500 security stakeholders which increased its exposure. Following the presentation, over 30 CCTV control room managers demonstrated interest. Over 20 CCTV control room managers were contacted individually by telephone to organise the visits. It was necessary to build rapport with managers prior to the visits to build confidence in the research. Despite this effort, there were some control room where very little data could be gathered from managers and operators.

Duration of filed observations limited to a maximum six hours:

The field observations were carried out at most control rooms for approximately 4-6 hours. Some control room managers were not physically available to assist for the entire day and some managers did not want the visits to last too long, as they felt the presence of a visitor could disrupt operators at work. If it had been possible to conduct the observations over a longer period of time (i.e., eight hours during the day and eight hours at night), there would have been a greater opportunity to uncover additional performance issue details (or more issues).

Information was not always disclosed:

Some control room managers and operators were hesitant when discussing negative aspects of their work practices, procedures and, tools. This was expected. Even though managers were informed that their identities would be kept anonymous they were extremely conscious of their comments and appeared worried at times about the issues being made public. Managers were worried their responses would affect their control room's reputation and that they would receive publicity. CCTV operators on the other hand, were worried about disclosing information about other staff and reporting problems in case management found out and they lost their jobs. This was evident from the initial comments operators made before describing problems they experienced in the control room:

> *"This is between you and me…yeah…."*
>
> *"You won't tell my manager will you?"*
>
> *"This is off the records, but can I just say that…"*
>
> *"Don't write this down, but…"*

Personal staff issues operators spoke about were not recorded as they were not relevant to the study objectives. The names and specific locations of the control rooms visited were kept confidential and anonymous when describing the issues in this thesis and related publications. There was however one exception: the police control room at Heathrow airport. A chief inspector from the Metropolitan Police constabulary approached the thesis author in April 2005 requesting an evaluation report detailing the problems within the control room and also requested recommendations. Due to time constraints during this control room evaluation it was not possible to talk to each operator (observed and interviewed) and describe the study objectives (see Section 4.1). In contrast, all operators at the other 13 control rooms were informed of the study objectives. Although this approach maybe perceived as unethical, there was no difference in the way operators interacted and responded during the task observations. Furthermore, operators at the control room were aware of the upgrade work and informed by their supervisors that the visits were related to improving their work environment and overall work.

4.6 Conclusions

The environment of a CCTV system undergoes many changes throughout its lifecycle as a result of the different social, technical, and environmental developments surrounding the system. The CCTV operator’s job is often overlooked by those who design and deploy control room systems. The factors which reduce the effectiveness of CCTV not only affect operators’ task performance, but also affect other stakeholders since communication and collaboration with others is an integral and important aspect of their work. From the field study findings, it is apparent that control room managers are spending money and time upgrading and expanding their control rooms as the goals for CCTV diversify and increase over time. Managers were also very optimistic about introducing new technology into their control rooms to support operators (e.g., more CCTV cameras, and digital and intelligent CCTV systems).

In order to achieve effective CCTV performance, it is crucial that the design management of CCTV systems be carefully considered. Anything that can improve the effectiveness of CCTV and the operator’s job is valuable. The technology used in CCTV security control rooms, if configured and used properly, has the potential to provide tremendous support for CCTV operator security tasks. If, however, the technology is poorly designed, the operator will experience great difficulty in using the system and it will be a wasted investment. All equipment that is deployed in a CCTV security control room should fulfil a purpose and be an essential part of the system and not be deployed ‘for the sake of it.’ To make security tasks and operations much more effective there is a need to improve: 1) the design of the technology; 2) the operator’s workspace and workflow processes; 3) the set-up of the surveillance areas outside the control room; 4) the administration required for equipment repair and purchase and 5) equipment maintenance (inside and outside the control room).

The findings of this study have been translated into a number of best-practice recommendations to improve the effectiveness of CCTV systems in three ways: 1) design; 2) configuration; 3) system usage. These recommendations are detailed in the TEC-VIS framework which is presented in Chapter 10.

PART 2: CCTV Video Quality

Chapter 5: CCTV Video Quality Background

Part 2 of this thesis addresses research goals 2 and 3, which focus on the effectiveness of CCTV video quality for security observation tasks. In this chapter, Part 2 begins by reviewing the background of CCTV video quality. Chapter 6 then presents a critical review of the literature detailing previous video quality studies, and methods used to assess the quality of video used for security tasks are presented in Chapter 7. Chapters 8 and 9 then detail the CCTV video quality studies conducted for this thesis.

Video quality was specifically chosen for the empirical research portion of this thesis, as the use of low-quality CCTV video was identified in Study 1 as having the most significant impact on the effectiveness of CCTV. Gill et al. (2005) also found that a number of control rooms were recording CCTV video at very low frame rates. Gill found that some control room systems were recording CCTV video at frame rates as low as ¼ fps; thus, the video was 'virtually useless' for police investigations.

Why is CCTV video being recorded at such low quality levels? Gill et al. (2005) found that the CCTV systems in the CCTV control rooms investigated were being deployed following advice from consultants and decided upon based on tight CCTV budgets. This was also the case in the CCTV control rooms examined in Study 1. It was evident that CCTV owners and manager lacked the knowledge necessary to properly configure the CCTV recording devices, because digital technology is new and evolving, and there is very little tested guidance available for these systems.

Before beginning a discussion on CCTV video quality and its impact on task performance, this chapter will first provide an overview of: 1) CCTV technology; 2) its uses in security and surveillance applications; 3) the main differences between analogue and digital CCTV systems and 4) video compression and its effect on task performance. A critical review of the current guidelines and recommendations for CCTV video quality is also presented and the chapter will conclude with a discussion of the key challenges for various CCTV stakeholders using low-quality CCTV video.

5.1 Introduction

Recent advances in computing technologies and video compression, together with the availability of faster networks, have increased the flexibility of digital CCTV. This increased flexibility has allowed digital CCTV systems to be applied to a wide range of applications, which serve a number of different security goals. Research is needed to understand how different CCTV stakeholders are affected by these changes.

5.1.1 Moving from Analogue to Digital CCTV Systems

First-generation CCTV systems are based on analogue technology, which records video from a number of surveillance cameras directly onto videotape. At the time of their development, analogue CCTV systems were perceived to be easy-to-use and affordable; however, many shortcomings have been revealed over the past 20 years:

1. Recording video on tape is inefficient. Unlike digital video, it is time consuming to track, transfer, search, and copy analogue video. Inefficient access to surveillance video slows real-time and post-event criminal investigations substantially.
2. Tape-based systems require constant human intervention. They require tapes to be replaced regularly for recording purposes. This process is time consuming and susceptible to human error since tapes can easily be overwritten.
3. Continual reuse of videotapes can wear down and degrade the video quality over time. The use of low-quality video will reduce a user's performance on security tasks, such as the detection of events and identification of targets.
4. Analogue video captures images of much lower resolutions (typically 240–340 TV lines). In contrast, digital video is capable of recording up to 400 TV lines. Low-resolution video contains less detail, making it difficult for human observers to detect and identify objects and people.
5. Analogue systems cannot be used for remote surveillance tasks, in which camera scenes are monitored from long distances through the use of a network (see Super Cops, 2006).

Digital CCTV systems were purposely introduced to overcome the shortcomings of analogue systems. In theory, the move from analogue to digital CCTV should eliminate all of these problems. However, digital CCTV systems can be configured by system owners, and there is anecdotal evidence that CCTV owners are tempted to reduce costs by reducing video quality to accommodate low budgets (Cohen, 2004). There is an inevitable trade-off between video quality and cost: uncompressed, high-resolution, high-frame-rate video requires more bandwidth and disk space. The trend has been to spend money on increasing the number of cameras, rather than increasing bandwidth and storage space. Digital CCTV requires considerable processing, storage, and data transmission capabilities, and these are expensive. There is an additional problem, in that CCTV owners do not have the knowledge necessary to properly configure digital CCTV systems. This issue was identified in Study 1 and by Gill et al. (2005), and the problem was echoed by a CCTV professional: *"…the problem we see is often digital video simply does not live up to the promises on the box and few managers have the technical knowledge to really get to grips with the technology"* (Fry, 2005).

When the video quality is reduced, the file size is also reduced. This can be accomplished by altering one of the following:

1. Image resolution: Most CCTV systems record digital video at CIF resolution (352x288), but offer configuration options to lower the resolution even further (e.g., QCIF resolution: 176x144). The lower the resolution, the less detail preserved; thus, performance of a security task such as face identification becomes harder.

2. Video compression: Video that is excessively compressed will lose a lot of detail and detail is removed by the encoder. Depending on the type of video CODEC used and the extent to which it is applied, video compression can introduce distortions and artefacts to the video.

3. Frame rate: In the CCTV industry, video that is recorded or streamed at more than 15 fps is considered to have a high frame rate. Low-frame-rate video (1–5 fps) is perceived to be 'jerky' and 'choppy.' Low frame rate video is difficult to follow, as frames are discarded by the encoder. The lower the frame rate, the higher the chance in missing an event.

More expensive CCTV systems offer greater flexibility in recording and streaming video (i.e., more than two video resolutions); less expensive systems provide limited settings for video recording. There are many different types of digital CCTV systems, and all have different operating requirements and constraints. There are a number of factors that can affect the quality of CCTV video and the observer's performance in accomplishing security observation tasks. The various variables which were considered as influential factors for an observer's response to targets and events from CCTV video (both recorded and real-time CCTV video) are listed in Table 5.1. A number of these variables were derived from the literature reviewed in Part 2 of this thesis as well as from personal communication with a number of CCTV stakeholders.

Table 5.1: Factors Affecting CCTV Video Quality and Observer Performance

Target	Technical	Environmental	Owner
Gender (Section 8.4.4) Race (Section 8.4.3) Ethnicity (Section 8.4.3) Gait (Section 5.2.2) Posture (Section 5.2.2) Clothing & accessories Skin tone (Section 8.4.3) Hair style (Section 8.4.5) Hair colour (Section 8.4.5) Facial features, e.g. scars, moles, birth marks (Section 6.2) Emotional state Way in which objects are carried Who person is with	Camera height Camera resolution Camera position Age of equipment Camera connections Type of CCD chip Type of recorder Recorder (storage) Data capacity Network capability (Section 6.2) Network reliability Video CODEC (Section 6.2) Compression level (Section 6.2) VHS tape quality (Section 10.2.4)	Physical obstructions Weather conditions Lighting capture area Accidental damage Vandalism Business of scene Special events	Lack of knowledge about technology Technophobic user Equipment not maintained regularly Low budget for system/network Poor advice from installers and sales
			User
			Training Experience in task Level of skill Motivation Familiarity with system

The CCTV operator's task typically involves the detection and identification of suspicious events and targets in real-time by observing scenes on several video monitors within a control room. The police and forensic investigators' task involves analysing recorded post-event CCTV data following an incident. As part of the investigation procedure, an investigator's job is to try and use any part of a CCTV video as evidence in court. This is difficult at times, because recorded CCTV video is not always of sufficient quality to identify events or suspects. This issue has posed problems for criminal investigations in the past (Bromby, 2002).

Changes in CCTV technology have led to an increasing number of systems being deployed by public and commercial sector organisations to achieve an increasing number of security tasks. As a result, there is now a wider range of CCTV users, with varying skills and experience. Table 5.2 provides a number of examples in which digital CCTV technology is currently being utilised for different applications by different groups of CCTV users. As noted in Chapter 1, some CCTV owners are now even recruiting untrained members of the public to assist with security and surveillance tasks, for e.g.:

1. In the US, web users view live CCTV video via the Internet to monitor the Texas-Mexico border for illegal crossings and alert the authorities (Web Users to Patrol, 2006).
2. In the UK, residents of a housing project view digital CCTV images from their television sets by subscribing to a community safety channel and alert the police by telephone if they witness unusual events or suspicious individuals (Rights Group, 2006).

Table 5.2: Examples of Digital CCTV Applications and CCTV Users

APPLICATION	DESCRIPTION	TYPICAL CCTV USERS	CURRENT EXAMPLE
Digital CCTV system	Locally store CCTV video on a PC or DVR hard-disk and view on-site	Private: shops, hotels, clubs, bars, etc. Public: local authorities and police	Local authority public street CCTV surveillance
Networked CCTV	Locally store and view CCTV images from remote location(s) using network	Private: shops, hotels, clubs, bars, etc. Public: local authorities and police authorities	Athens Olympics (2004)
Mobile network	View surveillance video, which is transmitted via a mobile network, on a hand-held mobile device Mobile IP cameras can also be used to view CCTV at remote locations	Public: police authorities and intelligence groups	UK Police
Digital image enhancement	Digital video of poor quality can be enhanced using several image enhancement techniques	Public and private: digital forensic practitioners and UK Home Office	Kalagate (1989)
Intelligent	DVRs and cameras can be linked to intelligent automated systems, such as people/object/vehicle tracking & recognition systems (e.g., to aid in spotting suicide attempts, overcrowding, etc.)	Private: transport companies Public: local authorities and UK police	Automatic Number Plate Recognition Intelligent 'Nice Systems'
Digital video identification parade systems	CCTV stills can be used to create identification parades replacing line-ups	Police authorities	VIPER

These examples illustrate how CCTV detection and identification tasks are now being performed by untrained participants, and that this is a growing trend. These schemes have been regarded as ‘controversial’ and privacy advocates fear that untrained participants may potentially form bias judgments, abuse the use of CCTV and spy on their neighbours (Web Users to Patrol, 2006). The empirical research conducted for this thesis examines how these CCTV users are able to perform a face identification and detection task, and assesses the performance and effectiveness of these tasks given the image quality that these systems currently deliver. The following section presents an overview of digital and Internet Protocol based CCTV systems. Video compression is then discussed with regard to its impact on a user’s perception of image quality, data storage, and transmission.

5.2 Digital CCTV Video

Digital CCTV video is recorded onto a computer or a Digital Video Recorder (DVR) hard disk, so that post-events can be investigated by the police and forensic experts. Digital video can also be transferred over the Internet for the purpose of real-time monitoring. Digital systems can be configured to record video at different quality levels; the choice of quality depends on: 1) the number of days continuous recording is required; 2) the number of cameras set-up to record CCTV video and 3) the hard disk and network access speed available. The quality level should also considered depending on the security observation tasks being carried out with the video, as different tasks have different video quality requirements. The level of quality is to be chosen by the CCTV system owner. This is typically achieved by setting a level on the system on an unlabelled numerical scale (typically ranging from 1–100). CCTV system manufacturers describe these levels by stating the number of days of recording that can be achieved and the video frame rates.

Video is efficiently stored through compression. In general, CCTV video security and surveillance systems utilise three types of compression formats: 1) MPEG; 2) Wavelet and 3) M-JPEG (Robertson and Monro, 1997). These CODECs share similar principles and methods, yet produce different video quality results. The objective of compressing CCTV video is to provide cost benefits for the system owner by:

1. minimising the use of hard-disk storage space in order to achieve greater data retention, and
2. facilitating the sending and receiving of CCTV video over a network without a loss of data or delay in transfer, thereby reducing bandwidth requirements.

Video can be compressed in two ways: spatially and temporally. An overview of each of these compression methods are provided in the following sections along with a discussion of the effect on perceived video quality and the implications for tasks performed by CCTV users. These sections provide a background for the two empirical research experiments presented in the second half of this thesis (see Chapters 8 and 9).

5.2.1 Spatial Video Compression

Spatial video compression is achieved using a video compression device (an encoder), which is found within the hardware of the surveillance system that compresses video. The encoder receives video images, either directly from the lens of a standalone system or via the IP network in a networked digital video recorder (DVR) system, and then converts these video signals from analogue to digital. The digital video is then compressed using a hardware or software compression algorithm. Changes take place within the video, which result in a reduction in video file size. As a result, the total recording time and data transfer rate of CCTV video is greatly reduced.

Video compression algorithms work by replacing pixel-related information with more compact descriptions and are categorised into 'lossless' and 'lossy' compression techniques. Lossless compression is employed when data must be decompressed to its exact state before compression. Lossy compression, on the other hand, attempts to eliminate redundant or unnecessary information contained in the video signal. Lossy CODECs identify information that may go unnoticed and eliminate it to reduce the size of the video. Hence, lossy compression works on the assumption that the data does not have to be perfectly restored and discards data in order to achieve the data transfer over the Internet at low bit rates.[11] It is important to note that this technique introduces visible distortions and artefacts to the video during the compression process. The more the video is compressed, the more distortions and artefacts will appear in the video, thus reducing the effectiveness of CCTV video for security observations tasks. The errors in the video file will vary and depend on the picture content, type of compression CODEC, and the processing power of the system. For this reason, it is difficult for many users to configure digital CCTV systems, particularly users who are use to analogue CCTV systems in which data is not compressed and is stored on physical VHS cassettes.

Video compression is measured as a ratio of the amount of data entering the compression system. A compression ratio of 1:1 indicates no compression; 10:1 indicates that there is 10 times less data after compression, and so on. There is a wide range of video CODECs available, many of which are standardised (e.g., MPEG-4); however, some are proprietary (e.g., Wavelet). The development of video CODEC standards have been driven by the multimedia mass market, rather than the CCTV market, and CCTV owners are not aware of the differences between the CODECs in terms of performance and quality (Cohen, 2006). Despite the complexities associated with digital video compression, CCTV owners should be provided with a basic explanation of the CODECs and the impact of increasing video compression when using CCTV video for security observation tasks.

5.2.1.1 MPEG Compression

The International Standards Organisation (ISO) set-up the Moving Picture Experts Group (MPEG) in 1988 to form standards on audio and video compression. MPEG is a four-part compression standard

[11] Bit rate is the number of bits transmitted.

defining the coding methods and techniques to reduce the amount of redundant information contained in video. Each standard addresses different standard bit rates. MPEG-1 was designed for video production and presentation with a bit rate of up to 1.5 Mbps, MPEG-2 and MPEG-3 were designed for 1.5 to 15 Mbps. MPEG-4 is a standard designed for video and web compression video. It is a very popular video CODEC used in CCTV security and surveillance systems. This type of video compression employs three different compression processes, as summarised in Table 5.3.

Table 5.3: An Overview of Compression Processes for MPEG

Compression Form	Type of Compression	Compression Process
Temporal Redundancy	Inter-frame P-Frame	▪ Compression is applied across frames. ▪ Some information between frames is discarded and only the differences between the frames are kept.
Spatial Redundancy	Intra-frame I-Frame	▪ Compression is applied to the reference frame (single I frame). ▪ Any identical pixels in the I frame are discarded.
Statistical Redundancy	Prediction/Motion Compensation	▪ The encoder predicts the changes that occur between frames. ▪ Line and field synch codes and long codes are replaced with shorter codes to reduce data contained in the video.

Encoded pictures are classified into I frames (intra-coded pictures: the reference frame), B frames (predicted pictures), and P frames (bi-directional pictures). An MPEG frame structure is described as a group of pictures. I frames are transmitted every 12th frame to serve as the detailed reference frame, P frames are interleaved between I and B frames, and B frames are made up from interpolated information from adjacent I and P frames. The areas defined (I, B and P frames) are called macro blocks. Macro blocks are made up of 8 x 8 blocks of pixels. After compression takes place, each block is independently run through a discrete cosine transform (DCT) algorithm. This is a technique for converting a signal into elementary frequency components. Following this, a process called ‘quantisation’ is performed, which involves significant data reduction.

5.2.1.2 Wavelet Compression

Wavelet compression is a relatively new method of intra-frame compression. Wavelet CCTV systems are not as popular as MPEG CCTV systems, as Wavelet compression technology is still fairly new and has not established itself as a standard. Rather than breaking a video frame into blocks (like MPEG compression), Wavelet compression works by analysing entire video frames. The components of the video signal are split into a range of frequency bands (typically 42 bands). The higher frequencies, which are invisible to the human eye, are discarded, and the remaining frequency bands are then subjected to spatial compression. Further compression is then applied to the remaining data. Each video frame is processed individually and checked three times for optimal compression. For this reason, the video has much greater storage requirements than MPEG-4 video, at the same bit rates.

5.2.1.3 M-JPEG Compression

M-JPEG stands for Motion-Joint Photographic Experts Group. M-JPEG, like MPEG, is an industry standard compression format, but employs intra-frame compression like Wavelet. M-JPEG CODECs work by compressing each frame separately, and only the differences within the frame are stored. Moderate compression ratios are achieved with M-JPEG video compression (i.e., 15:1 to 25:1). Compared with MPEG, M-JPEG compression requires considerably more bandwidth and storage space. Table 5.5 summarises the main differences between these three common compression schemes in the context of CCTV video recording equipment.

Table 5.4: A Comparison of MPEG, Wavelet, and M-JPEG Compression Schemes

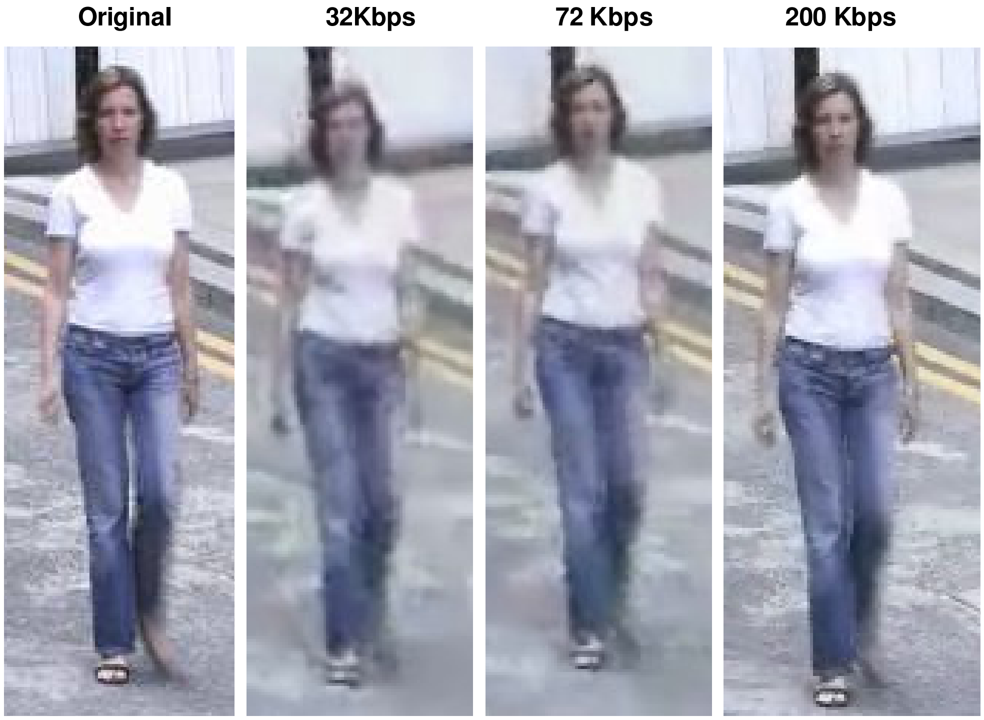
	CCTV Compression CODEC		
	MPEG	**WAVELET**	**M-JPEG**
Compression Method	Inter-frame	Intra-frame	Intra-frame
Compression Ratio	High 30:1 to 100:1	Moderate 15:1 to 25:1	Moderate 15:1 to 25:1
Appearance	Blocky artefacts	Soft blurred edges	Blocky artefacts
Data Storage Cost	Low	High	High

Different software applications offer different encoding methods and techniques; thus, video quality differs for each. The effect of video compression on final picture quality is illustrated in Figures 5.1 and 5.2. The images in these figures were produced by recording video of a woman walking towards a camera. The video was then encoded at three different video quality levels by altering the compression bit rate into the following levels: 32 Kbps, 72 Kbps, and 200 Kbps using an MPEG-4 and Wavelet video CODEC. A single frame is presented for each of these conditions to illustrate the effect of MPEG-4 and Wavelet compression on the perceived quality of the image. MPEG-4 and Wavelet video compression were specifically chosen, as these CODECs are commonly used in digital CCTV systems. These CODECs were also used to assess task performance in Studies 2 and 3 (see Chapters 8 and 9).

The images in Figures 5.1 and 5.2 illustrate that when video is compressed excessively using either MPEG-4 or Wavelet video CODECs, the finer details within the scene (i.e., target's eyes, nose, and mouth) are imperceptible. Table 5.6 shows that MPEG-4 video compression has a greater impact on reducing image file size than Wavelet video compression. The results are likely to differ if different types of video CODEC (and different versions) are used. The level of compression needed also depends upon the amount of change within the scene.

Table 5.5: Impact of MPEG-4 and Wavelet Compression on Image File Size

Encoding Bit Rate (Kbps)		Compressed File Size (Kb)
Original image file size		4, 200
MPEG-4	32	55
	52	57
	200	70
WAVELET	32	66
	52	75
	200	83


Original

32Kbps

72 Kbps

200 Kbps

Figure 5.1: The effect of MPEG-4 video compression on perceived image quality.

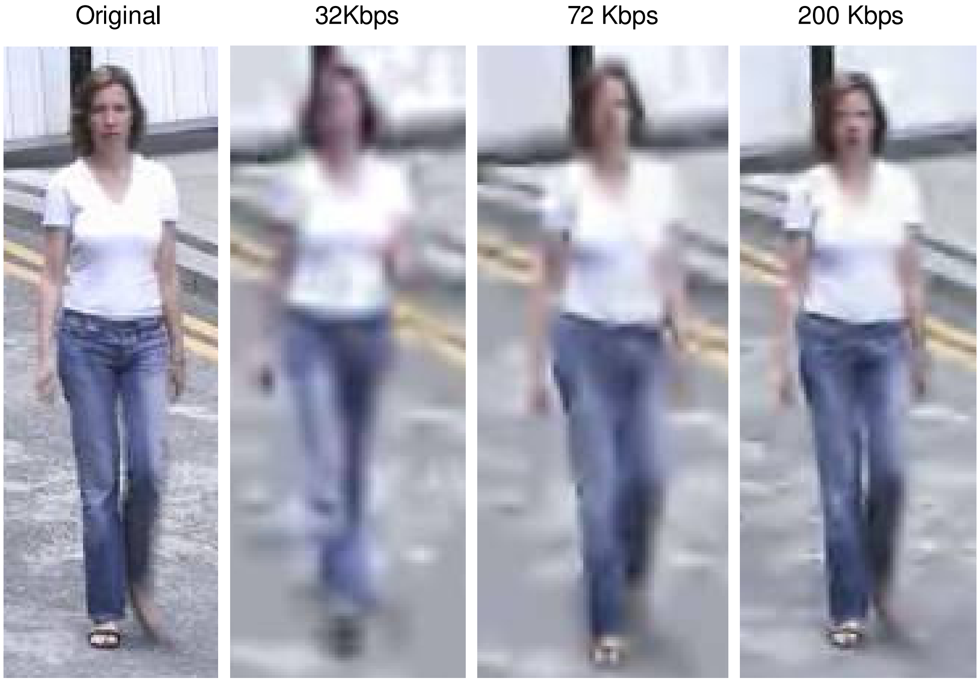


Figure 5.2: The effect of Wavelet video compression on perceived image quality.

There are several negative consequences associated with producing low-quality CCTV video:

- Police's task in investigating crime will be difficult (Bromby, 2002).
- Task of identifying and recognising perpetrators, particularly unknown individuals is surprisingly error prone (Klatzky and Forrest, 1984; Bruce et al, 1999).
- CCTV video evidence will not be admissible in court.

5.2.2 Temporal Video Compression

Video is temporally compressed when the frames within the video are reduced to: 1) increase the time available for recordings and (2) free up additional hard-disk space for further recordings. The frame rate chosen depends upon how much hard-disk space is available. For networked CCTV systems (IP CCTV systems), the frame rate is altered depending upon how much bandwidth is available for data transfer. The less storage space available and the lower the Internet speed, the less video that can be stored and delivered over the network.

For Phase Alternating Line (PAL)[12] CCTV security systems, frames are displayed at a rate of 25 frames per second (fps). Each frame consists of two interlaced fields (even and odd fields), giving a field rate of 50 Hz. In general, low frame rates (1-5 fps) produce jerky images when objects and people are moving. Video streamed at high frame rates and video recorded at high frame rates (12+ fps),

[12] PAL is a dominant television and video standard used in Europe.

appear to be smooth flowing. High frame rates generally produce better quality video, because the video contains more information about the scene:

- Video is perceived by humans to be ‘smooth’ at frame rates between 12–25 fps. Barber and Laws (1994) found that the lip movements of talking heads and its corresponding audio synchrony is maximised at 12 fps.
- Video recorded or transmitted over the network at very low frame rates (i.e., 1–5 fps) will appear to be jerky/choppy and this effect will increase with increased motion.
- At 1 fps, only one image is stored or transmitted per second. At this rate, very little information will be captured and fast moving actions are unlikely to be noticed if viewed in real time.

5.2.3 Video Resolution

Video resolution is the degree of detail contained in video, which is defined in X by Y values. The resolution is stated as the number of picture elements of the video displayed in horizontal and vertical rows. Most CCTV systems record video at CIF resolution 352x288 and *“this is mainly because of the cost savings to all CCTV stakeholders”* (Cohen, 2006). In addition, CCTV manufacturers have found it much more economical to manufacture digital CCTV systems which record video at CIF resolution (Aldrige, 2007). In comparison to 4CIF resolution, only a quarter of the data storage handling capacity would be required, making it a cheaper option for CCTV owners. Aldrige noted that CCTV suppliers also make profits by selling these cheaper systems and *“…the market exists for CIF CCTV systems as very few customers have a performance specification that can tell the difference between CIF and 4CIF.”*

Some digital video recorders are designed to allow CCTV owners to choose up to three video resolutions with which to record video. Despite this, a majority of digital CCTV systems can only record and transmit video at a set video resolution (usually CIF). There are no guidelines or recommendations on the most effective video resolution for the different CCTV observation tasks (e.g., face recognition and identification, detection of events, etc.), and not all security observation tasks require CCTV video at the same level of detail or resolution. For instance, a monitoring task that requires an observer to monitor overcrowding outside an underground station will not require high-resolution video to perform this task. However, if the observer must identify suspicious people, they will require access to high-resolution video to discern the facial features of a target in order to confidently say that the individual is a suspect.

Previous research by Bruce et al. (1999) identified that observers performed poorly when attempting to recognise faces from low-resolution analogue CCTV video, and performance was even worse when recognising unfamiliar targets. In another face matching experiment, Bachmann (1991) found that the identification of faces was affected very little when the number of pixels in an image was reduced from 74 pixels to 18 pixels. These studies (see full review in Chapter 6) identified that there are task performance issues associated with low-quality CCTV video and images used for face recognition and identification tasks. However, no studies to date have provided objective video quality recommendations to improve the effectiveness of CCTV video when used for observation tasks.

5.2.4 IP CCTV Systems

Digital CCTV video can easily be transferred over the Internet for the purpose of remote monitoring, enabling CCTV users to gain rapid access to surveillance video in order to react to incidents and also to archive video at other locations. In Europe, there are many large-scale networked CCTV systems (e.g., Brussels and Luton Airport, World Cup in Stuttgart, Athens 2004 Olympics, Dutch high-security prison, Amsterdam central station, Nederlandse Spoorwegen, and Dutch National rail network). There are many other small-scale and permanent IP CCTV systems being used throughout the UK.

IP CCTV systems work by digitising analogue signals and then compressing the data using a video encoder or video server. The encoding process occurs within the DVR device or in an IP camera. A standard web browser allows the user to connect to the CCTV system, which is configured to receive live video over the web, from anywhere in the world. The diagram presented in Figure 5.3 was created to illustrate the basic set-up of a networked CCTV system.

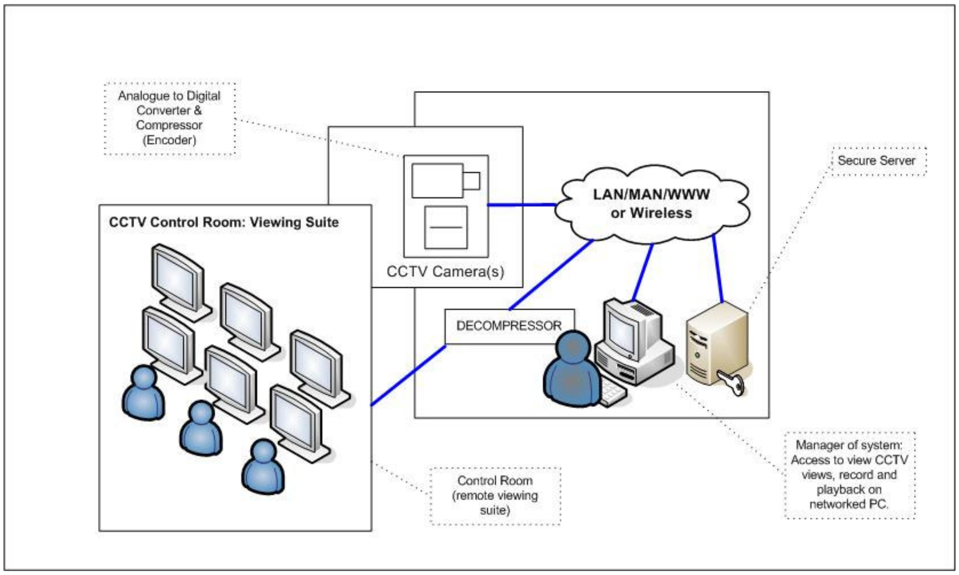

Analogue to Digital Converter & Compressor (Encoder)
Secure Server
LAN/MAN/WWW or Wireless
CCTV Control Room: Viewing Suite
CCTV Camera(s)
DECOMPRESSOR
Control Room (remote viewing suite)
Manager of system: Access to view CCTV views, record and playback on networked PC.

Figure 5.3: The basic building blocks of a networked CCTV system.

There are a number of networked CCTV systems that have been set-up for untrained CCTV users (new and emerging CCTV users), with which CCTV video can be monitored on the Internet to support the police in crime detection (e.g., see Web Users to Patrol, 2006 and Rights Group, 2006). In addition to these schemes, 'nanny cams' have become popular amongst modern householders in the UK to monitor their homes and businesses. These nanny cams are often sold as off-the-shelf packages and come with cheap cameras which offer limited video quality. Within the camera market, there are also many sophisticated higher-end home CCTV systems in use. One particular homeowner invested in a £20,000 home CCTV security system that worked successfully in a major burglar arrest (Leydon, 2006).

Home CCTV systems are becoming increasingly popular amongst householders as they are cheap and easy to set-up. There is, however, a two-fold problem with the use of these systems: 1) householders are not trained or experienced CCTV users; therefore, they are likely to make errors in security tasks (e.g., misjudge a situation or incorrectly identify an innocent person for a criminal) and 2) low-cost CCTV systems produce lower quality video (in real-time or recorded), which further reduces the observer's effectiveness in security tasks.

5.2.5 Streaming CCTV Video – The Effect on Quality

When CCTV video is streamed over a network, it is subject to degradation at two levels when the video is processed: encoding and transmission. Once the video is compressed, it is then transmitted in a series of packets over the network. During transmission, video delivery may be disrupted by either end-to-end delay or packet loss. These two transmission artefacts can occur as a result of the following factors:

1. Processing power of the machine at the client and/or server end of the network
2. Encoder type and quality
3. Internet speed (bandwidth) allocated for video distribution
4. Level of concurrent network usage

Trained and experienced CCTV users, such as the police and forensic experts, use IP CCTV systems for running special operations to solve crimes. For the remote monitoring of special operations, CCTV video is received from several IP cameras and observed at a control room. The images are viewed by police control room operators and used to provide support and intelligence information to street police staff. IP CCTV systems are also used by private businesses for monitoring their staff and customer activity on the weekends and in the evenings. If a low bandwidth is chosen for data transfer, the CCTV video will be streamed at low bit rates. At low bit rates, video is compressed for efficient data transfer, lowering the video quality. The effects of video compression are a reduction in CCTV observers' vigilance when observing video scenes and increased risk of observational errors.

Many IP digital CCTV systems offer a choice of bit rates (from as low as 32 Kbps to 1 Mbps) for remote video surveillance. There are however no guidelines or recommendations available for CCTV owners on the bandwidth requirements for the effective performance of remote security observation tasks. The configuration of a digital CCTV security system should be based on an assessment of the goals for the CCTV system, the user's skills and experience, the task requirements, as well as the capabilities of the system. As no two CCTV systems are the same, it is impossible to apply one set of configurations to all observation tasks and systems. The requirements for multimedia applications can be generalised quite simply; however, digital CCTV video systems are complex compared to analogue systems. There are a wide range of factors that can alter video quality and overall performance of a digital CCTV system.

5.3 Guidelines and Recommendations for CCTV

At present, there are two types of guidance available which offer information and advice on video quality and recording aspects of digital CCTV systems:

1. The Data Protection Act (DPA): CCTV Code of Practice (2008) and the Home Office Guidelines (Cohen, 2006).
2. Expert guidance – CCTV consultants and special working interest groups.

5.3.1 Policy Recommendations

Mead (1998) states that, *"video recordings which are used for eyewitness identification purposes can be of no value to the court if the recording is of such poor quality"* (p 1). Legally, a judge can only accept CCTV video evidence that is authentic and sufficiently intelligible. In the interest of CCTV owners, victims and prosecutors, it is vital that recorded CCTV video be used reliably as evidence in court, especially when CCTV is the only piece of evidence available for a case (although this is rare in criminal prosecution cases).

The Data Protection Act (DPA): CCTV Code of Practice (2008) is the current piece of UK legislation that deals with video quality, privacy and data protection issues. The purpose of the legislation is to ensure that images produced by CCTV systems are legally admissible in court. Within the Quality of Data section of the act, principles 3, 4 and 5 state that, *"images produced by a system must be as <u>clear as possible</u> to ensure that <u>they are effective as possible</u> for the purposes for which they are intended for."* The legislation has been regarded by control room operators as vague statements and difficult to interpret" (Gill et al., 2005). 'Effective as possible' and 'clear as possible' are vague and not easily translated into good practice. The DPA was published during the time when only analogue CCTV systems were in use, and therefore the guidelines refer primarily to these systems. There is some reference to digital and intelligent CCTV systems; however, automatic facial recognition systems are only mentioned briefly.

Since the DPA was published, CCTV and video technology have undergone many changes. Firstly, many analogue CCTV systems have been replaced by digital and IP CCTV systems. Secondly, the CCTV market has expanded significantly as a result of technological developments in digital CCTV. These systems now meet a number of different needs and support CCTV users who have different security requirements. Thus, the legislation on CCTV is out of date and does not cover a wide range of users, systems and CCTV applications. CCTV users are therefore left to decide their own design configurations without guidance.

The Home Office have published guidelines defining minimum performance standards for different CCTV observational tasks where CCTV systems are used (Aldrige, 1994; Cohen et al. 2007). The Home Office and the ACPO[13] have recognised that the guidance provided within the early version of the Operational Requirements (Aldrige, 1994) *"does not adequately address the recommended quality of recordings, or how to test the systems to ensure that the recorded images were fit for purpose,"* as they were mainly intended for real-time CCTV monitoring tasks and published before digital CCTV systems were in use (Gerrard et al., 2007, p 12). These guidelines are described and critically discussed in Chapter 7, Section 7.5.

The most recent version of the UK Home Office Operational Requirements (Cohen et al*.*, 2007) was published during the development of this thesis and provides a number of recommendations on the video quality requirements for achieving effective CCTV. These recommendations are referred to in many publications, including British Standards and Information Commissioner Officer (ICO) advice. Despite this, the recommendations are not specific, and therefore difficult to apply in practice.

To illustrate:

- **Recommendation 1:** If the target speed is high or the scene is complex, CCTV video should be recorded at a high frame rate [high frame rate is defined as 5+ fps].

The highest frame rate for PAL (and VHS systems) is 25 fps. It is unclear why a frame rate 5x less than the maximum was considered to be a high frame rate. There is no evidence that this recommendation was put forward following experimental tests with users.

- **Recommendation 2:** When a target is slow, the frame rate could be reduced to optimise storage.

This particular recommendation is informative, as it gives the view that low frame rates can be used where targets move slowly. The recommendation is, however, too vague when configuring the frame rate to suit their data storage needs. For instance, CCTV owners will be uncertain how *slow* a target should be, and therefore, how much the video frame rate should be reduced. Instead, owner-specific examples should be provided where a given frame rate is recommended for specific video scenarios (e.g., covert versus overt crime).

The Home Office and ACPO published short, concise (two page) guidelines on digital CCTV systems in which CCTV images are used by UK police authorities for investigative purposes (Home Office, 2005). The guidelines were prepared in attempt to address the difficulties the police face as a result of the widespread migration of CCTV systems from analogue to digital. The guidelines provide recommendations on video and image quality, storage, export and playback. These recommendations encourage users to define: 1) what they want to achieve from each camera, and 2) the specific activities required for monitoring. With reference to video quality, the supporting notes within the guidelines

[13] Association of Chief Police Officers.

state that, *"there are no definitive performance criteria for video to be legally admissible and it is up to the courts whether the pictures are acceptable or not as identified earlier."* Although CCTV owners have some access to up-to-date guidance on digital CCTV (Cohen et al., 2007; Home Office, 2005), these guidelines are not being widely used amongst small organisations and individuals (Gerrard et al., 2007). Those who are aware of them are not adhering to them (Police Guidelines, 2007).

5.3.2 Expert Recommendations

The Scientific Working Group for Imaging Technology (SWGIT) proposed recommendations on CCTV video quality for a specific set of CCTV users: forensic imaging practitioners (SWGIT, 2004). These recommendations apply specifically to digital CCTV systems set-up to record CCTV video within banks and shops:

- **Recommendation 1:** CCTV systems must capture and record at least one complete field per camera per second [this is equivalent to half a frame per second]. If a rate is selected below this value it may result in inadequate temporal coverage of events in the scene.
- **Recommendation 2:** The recording spatial resolution for a digital video recorder should be at a minimum resolution of 640x480, although the SWGIT strongly encourages higher resolution video to be recorded wherever possible.
- **Recommendation 3:** For recording sequences of interest, lossless compression should be employed. [This type of compression involves a class of data compression algorithms which allow the original data to be reconstructed from the compressed data]. It is strongly recommended that: 'the lowest possible amount of compression must be used for recording CCTV.'

There are three criticisms with these recommendations:

1. Within the SWGIT (2004) report, there is no evidence to suggest that these recommendations were developed as a result of task-based evaluations with CCTV users performing different security observation tasks, such as face identification and crime detection.
2. These recommendations were specified for two very different environments (banks and shops). The goals for CCTV in these two environments will be different; therefore, the video quality requirements for CCTV will be different for each environment.
3. The recommendations are broad. No video quality requirements are given for specific security observation tasks (e.g., identification, recognition, detection, and monitoring).

Security systems architect Morton (2004) identified four main causes for low CCTV image quality:

1. Recording systems produce low-resolution images
2. Video is excessively compressed
3. Inadequate compression technology
4. Video recorded at very low frame rates

Based on interviews with security personnel,[14] Morton put forward a number of recommendations for three different types of digital CCTV users in the USA (see Table 5.7) carrying out the following tasks:

1. Investigation: A task carried out by security professionals who examine recorded video.
2. Video monitoring: A task in which the observer monitors events from real-time scenes, typically on a video monitor (quad-display).
3. First response: A task carried out by the police and forensic teams who access critical up-to-date security video on a mobile device. The information is used to prepare for situations and also to obtain adequate resources to save victims in an emergency.

Table 5.7: Video Quality Requirements (Morton, 2004)

CCTV Users	Spatial Resolution	Temporal Resolution	Compression	Data Rate
1. Investigators	640x480	30 fps	Minimal	High
2. Video monitors	320x240	5 fps	Medium	Medium
3. First responders	160x120	10 fps	High	Low

There are two problems with Morton's recommendations:

1. Although the video quality requirements are provided for different CCTV users, the requirements for different types of CCTV tasks are not given.
2. The recommended video compression levels and data rates are not specific enough; therefore, they are likely to be open to interpretation.

Like the SWGIT recommendations, the recommendations offered by Morton were put forward based on expert knowledge, rather than through CCTV user assessment – and "*the only effective means of ensuring that a system is usable is to periodically assess the system and test user responses*" (Fléchias, 2005, p 40).

[14] The exact number of interviews conducted was not reported.

In conclusion, the current guidance and recommendations available on digital video quality requirements for CCTV systems contradict one another, are too vague and are unreliable, as they have been put together through informal research and craft knowledge, rather than scientific evaluations with users. CCTV owners need access to knowledge and specific guidance on video quality requirements for different security observation tasks. This guidance will ensure that low-cost digital CCTV systems will record CCTV video that can be used for real-time tasks performed by human observers and also by automated CCTV systems. The guidance will also ensure that recorded CCTV video will be usable for police investigations and will be acceptable as evidence in criminal prosecuting courts.

5.4 Chapter Summary

This chapter has provided an overview of digital and networked CCTV systems, the process of video compression (MPEG and Wavelet) and the effect of compression on perceived image quality. The degree to which video quality is lowered through spatial and temporal compression depends upon the trade-offs CCTV owners make between the cost of video storage/delivery and video quality. A number of different variables can alter a CCTV video system performance; these variables are linked to the user, system and security environment. As there are so many variables that can alter CCTV video quality, it is impossible to recommend one set of parameters to be applied to all types of security observation tasks. The recommendations currently available to CCTV users for configuring and measuring the performance of CCTV systems are not based on experimental tests with real-world CCTV users. In addition, they contradict one another, and are too vague. Furthermore, much of the Home Office guidance on video quality and performance testing applies to analogue CCTV systems – as these guidelines were developed in the pre-digital CCTV era. The next chapter provides a critical review of the previous experiments which have examined the impact of video quality on the performance of various tasks. The review serves as a background literature review for the two video quality experiments conducted in Chapters 8 and 9.

Chapter 6

Video Quality – Task Performance Experiments

There have been a number of studies conducted in HCI which have investigated the effect of low-quality video on a user's performance for a number of different tasks, such as distributed learning, lie detection, video conferencing, and gaming. These studies were designed to establish minimum video quality requirements with the aim of improving the efficiency of storing video locally and streaming it over a network. The literature review in this chapter was performed to provide background for the two CCTV video quality studies carried out for this thesis. These studies can be found in Chapters 8 and 9.

This literature review demonstrates that there have been surprisingly few studies conducted which investigate the impact of lowering video quality on a user's task performance and perceptions of quality for CCTV applications. Such research is needed to understand how CCTV users of varying skill levels are able to perform security observation tasks with low-quality digital CCTV video. This research is needed to establish minimum video quality requirements to ensure that the tasks are being performed effectively and efficiently, and also to identify training needs. Evaluations with users will provide reliable, valid and applicable guidance for CCTV owners wishing to deploy a digital CCTV system.

6.1 Introduction

CCTV video systems *"are developed to be used either in real-time to prevent disasters or crimes, and/or to extract knowledge for a-posteriori investigation"* (Cucchiara, 2005, p 4). Traditionally, CCTV video recordings have been used by trained experts, such as control room operators, the police and forensic investigators. A CCTV operator's job involves carrying out four main observation tasks: monitoring, detection, recognition and identification of suspicious events and targets (Aldridge, 1994) as proactively and reactively (see Chapter 4, Section 4.4.2.4). These observation tasks are performed in real time by observing CCTV activity on a monitor bank or on 1-2 spot monitors in a control room. The police and forensic investigators' task is principally investigative and involves a detailed analysis of post-recorded CCTV video footage. As part of the investigation procedure, an investigator's job is to try and use any part of the CCTV video for evidence to prosecute criminals in court. This task has been particularly challenging for these CCTV users, as recorded CCTV video is not always of sufficient quality to identify events or suspects (Bromby, 2002). This issue was identified as a major task performance issue in Study 1 (see Chapter 4, Section 4.4.2.5), in which a number of police officers regarded the CCTV recordings at several CCTV control rooms to be 'virtually useless.' As has been mentioned, this issue was previously identified by Gill et al. (2005). In addition to the problems associated with the use of low-quality video for post-event tasks, the use of low-quality digital CCTV video is equally problematic for real-time tasks in which the CCTV user observes live surveillance video which is streamed across a low-bandwidth network.

In the previous chapter, it was noted that the policy and expert guidance currently available on CCTV video quality are: 1) contradictory; 2) insufficient and 3) untested. The UK Home Office have recognised the need to develop and provide guidance for CCTV owners on video quality, storage and transfer of digital CCTV to improve the effectiveness of CCTV systems (Gerrard et al., 2007), and there have been a number of video quality studies conducted with users to assess task performance when video quality is compromised for cost; yet, *"no evaluation or benchmarking studies have attempted to assess image quality for the purpose of commercial and law enforcement applications"* (Sirohey, Wilson, and Chellapa, 1995, p 705). Thus, research is needed to develop a set of CCTV video quality requirements to ensure that a wide range of CCTV users (of varying skill levels, experience and working environments) are able to achieve high performance in their security tasks. This research is only possible by conducting evaluations with users to assess the impact of video quality on task performance using task performance measures. In the following sections, the previous video quality studies most relevant to CCTV tasks are detailed. These studies conclude with a discussion of the video quality guidance needed for CCTV tasks.

6.2 Effect of Lowering Video Resolution and Increasing Video Compression

The most important video quality parameter for achieving high levels of accuracy in an identification task for which an observer is required to identify a specific object or human target is video resolution, as this is what provides the detail within a video scene. As discussed in Chapter 5, video which has been recorded or streamed at low resolutions will show less detail than video recorded at high resolutions. By default, most digital CCTV systems record video at CIF resolution (352x288); however, many can be set to lower resolution formats (QCIF: 176x144) (Cohen, 2006). There are serious consequences associated with recording CCTV at very low resolution rates when the video is to be used for critical tasks, such as real-time suspect identification or face recognition during a post-event analysis:

1. Low-resolution CCTV video will contain fewer pixels and therefore will hold less detail, which will make it very difficult for an observer to identify or recognise an individual. Typically, scars, moles, birthmarks and other fine facial details will be indiscernible.
2. The use of low-resolution CCTV video will create significant challenges for police and forensic experts examining the identities of unknown suspects during criminal investigations, particularly when the experts are under time pressure.
3. If CCTV video has been recorded at a very low resolutions (spatial and/or temporal rate), it is unlikely that it will reach a criminal prosecution court as supporting evidence.

The impact of lowering spatial and temporal video quality on final image quality and the performance of security observation tasks is discussed in greater detail in Chapter 5.

Eyewitnesses are not experienced in processing and analysing CCTV video. An eyewitnesses' role during a criminal investigation procedure is to identify a suspect from CCTV images (usually 1–2

CCTV images). The identification process typically involves the eyewitness identifying a known target from episodic memory (i.e., a memory of events, times, places, associated emotions and other conception-based knowledge in relation to an experience). Jurors treat eyewitness identification as compelling evidence in both civil and criminal trials. Despite this, the reliability of an eyewitness statement has been questioned on three scientific grounds (Green, 2004):

1. Poor visibility conditions (such as lighting)
2. Humans are generally poor at face recognition (unless the target is known)
3. Procedures used to obtain identification may be biased

In addition to these factors, eyewitness identification errors are to be expected if the CCTV video or image being used for the identification task is low in spatial quality, due to the low level of detail provided. CCTV video is now being used by untrained CCTV users to support the police authorities in real-time crime detection within local communities through the use of remote surveillance systems (see Chapter 5, Section 5.2.4 and 5.2.5). It is unclear how these users perform security observation tasks. For instance, which features within a scene do observers rely upon when identifying targets? Furthermore, it is not evident that these users are capable of supporting the police for real-time crime detection. This is an important issue which requires empirical research, particularly with low-quality video, as a large number of CCTV systems are either low-cost or configured without guidance on video quality. These research questions are empirically examined in Study 2 (see Chapter 8).

Psychologists have previously investigated performance issues associated with face recognition accomplished by human observers (Bruce et al., 1999; Burton et al., 1999; Henderson et al., 2001). These studies were carried out to assess performance and identify ways to improve eyewitness reliability. Burton et al. (1999) conducted a face recognition experiment and compared task performance between trained and untrained CCTV users. Both trained (police officers) and untrained participants were required to observe 10 video clips (lasting three-seconds) showing an unknown target walking through a building entrance. Participants were then shown a 2x10 array of high-quality A4 colour photographs of faces (all of whom were unfamiliar to the observer). Participants were told that half of the faces within the array were targets who appeared in the video clips and the other half were look-alikes. The recognition task required participants to choose one target within the photograph array whom they felt was present in the video clips, and then rate their confidence in their response using a 7-point Likert scale. The results showed that recognition performance for both untrained and trained participants was very poor. These findings demonstrated that regardless of skill level and experience in face recognition, people are simply poor at recognising unfamiliar faces. However, the study did not detail the specific factors which affected recognition performance. Thus, it was not clear whether participants were shown analogue or digital video in the task, and at what video quality level. The CCTV video quality was described by Burton et al. (1999) as: *"…poor, though tolerable for a low-cost system"* (p 243).

Henderson et al. (2001) also examined observers' performance in a face recognition task using a mock-up of a crime scene which was filmed using a low-cost video surveillance system. A bank raid scene was filmed over a 39-second period within a high street bank using three CCTV cameras. A number of CCTV images were captured showing three targets (actors who played the roles of bank robbers) from different angles in different poses. These video sequences were shown to participants in less than four-seconds. They were then immediately shown a 2x10 photograph array. Participants struggled to recognise robbers' faces due to the CCTV images being too poor in quality when compared to a set of high-quality photographs. Like the photograph array used in Burton et al.'s (1999) study, 50% of the photographs were distractor faces, which resembled the robbers who appeared in the bank robbery scene. These distractor faces were chosen by asking the actors to choose faces which they thought were similar to their own faces from a selection of actors' faces in a catalogue. As in Burton et al.'s study, the type of video used and the quality level were not stated in the study literature, and this is because the study examined the effect of target familiarity, rather than the relationship between video quality and face recognition performance.

Both of these face recognition experiments (Burton et al., 1999; Henderson et al., 2001) required participants to recognise targets from memory after previewing short video clips. This method was chosen to assess eyewitness reliability and compare task performance between trained CCTV users and untrained participants within a laboratory setting. In addition to being unfamiliar with the targets, there were a number of other reasons why eyewitnesses demonstrated poor recognition performance.

Green (2004) discusses some of the issues that may contribute to memory failure during eyewitness identification:

1. Low resolution – the face in the video may not be as clear in the observer's memory as the one viewed.
2. Observer bias – eyewitnesses often have a tendency to construct memories, so that missing information is supplied from expectations and/or biases from other memories, or from an external source such as TV, newspaper, other witnesses, or the police.
3. Systematic perceptual distortions in memory – for example, small objects appear larger and large objects appear smaller; also, certain colours and shapes are remembered as brighter and bolder.

Bruce et al. (1999) chose to assess face recognition performance with participants using an item verification method, rather than memory recall. This procedure involved using real-world CCTV video taken from an analogue CCTV system which recorded video at 'relatively low quality.' Two different groups of participants were recruited, students, and police officers unfamiliar with the targets shown in the CCTV. Participants were shown a series of three-second video clips (black and white) of a target walking towards a camera. Two different sets of video clips were presented to each group; one half contained targets with whom participants were familiar, and the other contained unfamiliar targets. The task required participants to match a target in the CCTV video with a target from a high-quality photograph (an array of 20 different faces was used for each video clip). In contrast to the previous memory recall experiments, participants in this study were given the opportunity to compare the video clip with the photos for as long as necessary using the replay button on the video recorder. After a target was matched, they were asked to rate their confidence in their decision using a 10-point Likert scale.

The results showed that participants who were familiar with targets performed better (average ratings of 6.5) than when they were not familiar with the targets, despite the low quality of the CCTV video. When participants were unfamiliar with the targets in the CCTV video their performance level dropped (average ratings of 4.7). It was therefore concluded that the participants performed well when they were familiar with the targets because the CCTV images contained sufficient low-spatial frequency information which observers used to match the stored visual appearances of the faces from high-quality photographs. This low-spatial information helped participants fill the missing bits of information in the low quality images from the high quality image stills (Bruce et al., 1999). However, this is not possible when the observer is not familiar with the target appearing in the CCTV video clip – a typical scenario for identification and recognition tasks performed by CCTV operators, police and other untrained CCTV users. These findings are important for the development of eyewitness tasks; however, further research is needed to identify the digital video quality requirements for identifying unfamiliar targets using CCTV images.

The effect of lowering image resolution on face recognition performance was examined by Bachmann (1991). Participants in this experiment were asked to match six different non-quantised[15] face images with a collection of computer-quantised face images. The quantisation conditions employed were: 15, 18, 21, 24, 32, 44 and 74 pixels per face. The participants' task performance (correct and incorrect matches) and response times were measured as participants identified a large number of faces (1152). The results revealed that the identification of faces was affected very little when spatial quantisation (pixellation) was reduced from 74 pixels to 18 pixels. Bachmann concluded that images of faces with a total of 18 pixels or fewer are unacceptable for face recognition. Consistent with Bachmann's results, Costen, Parker and Craw (1994) reported similar findings. In this study, participants were able to match original non-quantised face images with their corresponding images that were quantised between 8–16 cycles[16] per face (equivalent to 16–32 pixels per face). The quantisation conditions used were: 42, 23, 12, and 9 pixels per face. The results of these studies by Bachmann and Costen are not reliable, for two reasons. Firstly, a very small sample of participants took part in the experiments; Bachmann used four participants and Costen recruited eight participants to complete the task. A small sample size will mean that the experimental design will have less power, thus reducing the validity of the results when being applied to the population. Secondly, participants in Bachmann's study were expected to identify an unrealistic number of face images (1152). If participants are provided with too many stimuli, they will

[15] Quantisation in the context of image processing is a lossy compression technique achieved by compressing a range of values to a single quantum value (see Chapter 5, Section 5.3.1).
[16] One cycle is equivalent to two pixels.

eventually become bored and fatigued. This effect is likely to flaw the experiment's results. Although these two studies provide the spatial requirements for CCTV images (lowered through spatial video compression), further task-oriented experimentation is required using realistic task scenarios and users.

Prior psychological research into face recognition showed that face recognition performance is affected by video images and the observer's familiarity with the target in the CCTV images. Previous research has also identified that face recognition performance is affected by spatial frequency (Bachmann, 1991; Costen et al., 1994). These studies were conducted during the peak production of analogue video systems. Further research is needed to establish the minimum spatial and temporal resolution requirements for CCTV observation tasks, such as face identification and detection.

6.3 Effect of Lowering Video Frame Rates

The choice of video frame rate is an important factor in the recording and streaming of live CCTV video, as it can affect the performance of security observation tasks, particularly tasks which involve depicting movement and specific actions. Low-frame-rate video (i.e., <= 5 fps) contains at most 5 of 25 images of the scene being stored/streamed in a second. By lowering the video frame rate, potentially important frames are discarded, which may result in the unintended removal of vital evidence from a scene. Card, Moran and Newell (1986) stated that the frame rate needed for closely related images nearer together in time than the cycle time of perceptual processing must be identified as the same object. Card calculated that 10 fps is needed for an animated image on a video display must be refreshed to give the illusion of movement.

However, CCTV system users commonly record video in time-lapse mode at rates between 5–8 fps, and it has been observed that rates as low as 1 fps are being used by many CCTV users (Cohen et al., 2007; Gill et al., 2005).

The frame rate requirements for recording and streaming CCTV video are likely to be different for the four main security observation tasks (identification, recognition, detection and monitoring). For instance, for a task which requires an observer to detect an incident from CCTV video, the frame rate level is more influential than video resolution since the ability to follow events from one frame to another is essential in comparison to resolving detail of a person's face. The observer's performance and effectiveness will be reduced as a result of several environmental and technical factors (e.g., noise distractions, dirty camera lens, camera occlusions, and so on); however, the one variable which has the potential to reduce the observer's performance significantly is the frame-rate of CCTV video (recorded or real-time). The use of low-frame rate CCTV video for security observation tasks, such as detection, affects more than one CCTV stakeholder:

- CCTV users: When low-frame-rate video (i.e., 1–5 fps) is delivered over a network it does not show events in a clear and smooth flow, as many frames within the video sequence are discarded during data transmission. Low-frame-rate video may inadvertently lead to the user completely missing an incident or losing track of a criminal they previously detected.

- Police: It is very difficult for police to establish the identities and movements of targets when reviewing post-event recorded video footage which has been recorded at a low frame rate and low resolution. This difficulty reduces the confidence police have in the video for interpreting criminal events, and will slow investigations. This is a serious issue if there is no evidence other than the CCTV video available. The difficulties faced by police when confronted with low-quality CCTV video were identified by Gill et al. (2005) and Bromby (2002), as well Study 1 of this thesis.
- Criminal Prosecution Service (CPS): CCTV video which is recorded at very low frame rates will contain fewer frames, thus less information. If the frames showing the vital crime scene are discarded by the system, the CCTV video footage cannot be used as evidence for prosecuting criminals in court (Oxlee, 2004).

There have been just two experiments that have examined task performance issues associated with CCTV video used in a detection task. The first study investigated whether antisocial or criminal behaviour could be predicted by novice and expert human observers (Troscianko et al., 2004). This research was carried out with the aim of identifying predictive factors to support human operators in detecting crime observed in real-time with public CCTV cameras. Two groups of observers were recruited for the study: 'experts' (50 professional CCTV control room operators), and 'novices' (50 university students). Participants were required to observe 18 CCTV scenes and judge whether something 'bad' had occurred in the moments following an event from the video clip. Participants were then asked to give a rating which reflected how bad the scene was using a 7-point Likert scale. Troscianko et al. (2004) found that there was very little difference in detection performance between the experts and novice observers, suggesting that predictive cues are *"…automatic, low level and not strongly modified by experience in the task"* (p 96). The authors did not state the quality level of the CCTV video used for the study. This study is unique in that it examines human observer performance with real CCTV video footage, and demonstrates that experience with CCTV is not entirely essential for a detection task. Therefore, this research supports the notion that untrained observers are capable of performing detection tasks - such as the two schemes described in Web Users to Patrol (2006) and Rights Group (2006). This may not necessarily be true when using low-quality CCTV video for tasks.

The second detection experiment examined the impact of using low-frame-rate video on a crime detection task (van Voorthuijsen, et al. 2005). In this particular study, a number of crime events were staged in public by actors and police officers outside a railway station and used as CCTV video for the detection test. The scenes included bag and cell phone snatches, threatening and harassing incidents, as well as some incidents showing missing and wanted persons. The detection test was carried out with two groups of expert CCTV users: 22 police academy students and 16 CCTV operators. The users' performance in detection was measured using the Surveillance and Monitoring Assessment Exercise (SAMAE), a CCTV training software tool, with the following frame rate levels: 25 fps, 2 fps and ¼ fps. The results showed that for both participant groups, their ability to detect incidents decreased as CCTV video frame rates decreased.

To date, this is the only study which examines the cost versus quality dilemma for a real-world security task. Despite this, the study presents only preliminary findings; thus, no significance levels are stated. There are two major weaknesses with the study design:

1. The performance in the detection task was not measured using standard HCI task performance measures. Instead, a CCTV training software tool was used to record and analyse performance. The method and measures used by this tool were not described in the literature, making the results difficult to apply.
2. The reasoning behind the frame rate conditions (25 fps, 2 fps, and ¼ fps) was not stated in the literature. These frame rate levels do not cover the typical frame rates applied by CCTV owners (Cohen et al., 2007). Thus, these conditions do not allow a threshold frame rate to be identified, which means that the results from the study are impossible to apply in practice.

In the wider field of human centered multimedia research, there have been a number of studies which have assessed the impact of lowering video frame rate on task performance. These studies are reviewed in the remainder of this section, as they are relevant to the CCTV video quality research conducted in the second half of this thesis. They all involve the assessment of performance with users actively engaged with video lowered in quality through temporal and spatial compression. The findings from these studies were used to form the hypotheses in empirical studies 2 and 3 (Chapters 8, Section 8.3 and Chapter 9, Section 9.3).

In a lie detection study conducted by Horn, Karasik and Olsen (2002), participants were asked to judge whether a person was lying in an interview observed on playback video. Performance was measured using a 6-point Likert truthfulness scale with video played back under the following video quality conditions: 320x240 and 29.97 fps; 106x80 and 10 fps; 106x80 and 5 fps; 53x40 and 10 fps; and 53x40 and 5 fps. The results showed that a slight reduction in image resolution impaired lie detection accuracy; however, performance was the worst when video was played back at 5 fps. At this frame rate, behaviours such as shoulder shrugs, eyebrow movements, and eye blinks were difficult to interpret.

In a video conference study, Wilson and Sasse (2000) conducted an experiment to establish the effect of frame rate reduction on an observer's subjective and objective perceptions when viewing an interview of a video conference call. Participants viewed university admission interviews, which took place over the Internet. The interviews were conducted between a student (the participant) and an admissions tutor, and were purposely scripted to mimic the interactions typical of a university interview situation. The interviews were conducted using an IP video conferencing tool and participants viewed video that changed in frame rate every 5 minutes as follows: starting at a baseline of 16 fps, increasing to 25 fps, decreasing to 5 fps, and then increasing again to 25 fps. The impact of changing the video frame rate was assessed by recording the participants' subjective responses through a post-experiment questionnaire. Subjective data was gathered from participants by collecting their perceptions of the video quality using a Likert scale. User cost was also measured by recording the participants' physiological responses, such as their heart rates, skin conductance and blood volume pulses, as the video frame rate was altered. The results showed that when the video interviews were observed at 5 fps, participants' heart rates and skin conductance rose, and as a consequence their blood volume pulse declined. The change in the participants' physiology indicated that they were more stressed when video was observed at a low temporal quality level. However, the subjective ratings were not in line with user cost, since the participants' perceived video quality ratings did not correlate with their physiological responses. 84% of the participants did not notice a change in frame rate even though physiological responses indicated they did. This study illustrates the importance of using both subjective and objective methods for assessing the impact of video quality. The subjective participant responses provide an indication of what the users perceive when observing low-quality video. These responses varied between users and depend largely on their previous experience with low-quality video and their expectations for video quality when performing a particular video task. User cost was not considered to be a suitable measure for the video quality experiments conducted for this thesis, as the conscious effort users expend is less important than measuring their performance in CCTV tasks.

In another video communication study, naïve American sign-language users demonstrated a high level of performance in learning and recognising signs using video played back at frame rates as low as 5 fps (Johnson and Caird, 1996). Whilst this low frame rate was considered acceptable, the performance is likely to differ with other types of tasks, such as face identification with CCTV video, since users in this study may have relied more on the interpretation of gross hand movements than subtle facial expressions and movements.

The effectiveness of low-frame-rate video has also been examined with gaming tasks. In one particular study by Swartz and Wallace (1993), participants were asked to fly a vehicle in a simulator that displayed animated video at different frame rates. Task performance deteriorated at lower frame rates (i.e., 7.5 fps, 4 fps, and 2 fps). The results showed that at 4 fps video quality was high enough to achieve acceptable performance for the gaming task. In contrast to these findings, a frame rate as low as 3 fps was found to be unplayable for a shooter video game task (Claypool, Claypool and Damaa, 2006). In this study, task performance was measured by recording the number of times a user killed an animated target (a 'bot'). Performance in the gaming task was compared with the game displayed under two different video resolutions 320x240 and 640x480. Claypool et al. found that task performance was *"remarkably insensitive to different resolutions for a range of game conditions"* (p 2) and concluded that different resolutions did not have a significant impact on task performance in comparison to video frame rate.

During the process of video streaming, video is digitised and compressed. In the digitisation process, video is streamed at low bit rates (using low bandwidths) and the high level of video compression that takes place introduces distortions to the video and removes potentially important frames from the scene. A number of CCTV owners are introducing real-time remote surveillance tasks as 'add ons' to their existing digital CCTV systems as a flexible way to monitor and detect events from a remote location in real-time mode. The effect of streaming CCTV video using low bandwidths (i.e., 32 Kbps or less) has serious implications when used for real-time critical CCTV observation tasks. To date, there have been no studies conducted in HCI to investigate the task limitations associated with using low-bit-rate video for security observation tasks performed by human viewers.

Within the computer science domain, there has been only one study conducted to examine the effect of streaming low resolution CCTV video at low bandwidths for computer-based tracking and identification tasks (Korshunov and Ooi, 2006). This study examined the scalability issues for large-scale distributed video surveillance systems (typically streaming CCTV video to 1000 CCTV cameras), in order to identify a trade-off between task accuracy and video quality. In the first experiment, face detection was assessed using an open-source face detection algorithm (Viola-Jones method) with ground truth (manually tagged CCTV images). In this assessment, face images were compressed using JPEG ranging in quality from 1 to 100. The detection test was completed with a total of 507 face images, and the detection index (average detections) and false alarm index (average false alarms) were recorded. The results showed that the average performance of face detection failed to improve with CCTV images compressed at JPEG quality beyond 20. However, detection performance declined with CCTV images below a quality level of 7. A conservative image quality threshold rate of 20 was proposed for the effective and economical distribution of CCTV video for automated detection. At this quality rate, it was suggested that a CCTV image with a file size of '15.8 Kb,' which translated to a reduction of bandwidth by 29 times, was acceptable.

This image quality recommendation for an automated face detection task, although conservative and based on tests with a large number of ground truth CCTV images, will be difficult to apply in practice for a number of reasons:

1. The CCTV stimuli used for the detection test does not accurately represent the quality characteristic of distributed CCTV video, since video which is distributed over a network undergoes video compression. In this study, the face *images* were compressed using varying JPEG image quality, rather than video compression.
2. The study does not state the minimum resolution required for effective automatic face detection when images have been compressed.
3. An unlabelled image quality scale was used to alter quality (1 to 100). This technique is useful for experimental purposes; however, these quality rates are difficult to apply in practice. For distributed CCTV systems, bit rate (Kbps) is a better measure of quality, as it is an objective means of varying video quality. Video bit rate is also a familiar quality parameter for users in the context of video compression, storage and network.

Following these tests, Korshunov and Ooi (2006) validated this threshold CCTV image quality level through experimentation to determine whether the rate was effective using M-JPEG compressed video. The validation of the threshold quality level was confirmed. As there were several methodological limitations to this study, specifically with regards to the production of unrepresentative CCTV stimuli for the detection tests, the proposed critical image quality level of 20 is not reliable or valid, and therefore, cannot be applied to real-world CCTV deployments.

6.4 Chapter Summary

In this chapter a critical review of the task performance experiments is given. The psychology research has demonstrated that both trained and untrained CCTV users are generally poor at identifying unfamiliar faces from poor quality analogue CCTV video and images. These experiments mainly assessed task performance with observers who were familiar with faces from CCTV images when high-quality photographs were available for direct comparison. Although these experiments demonstrated that low-quality CCTV video significantly affected face recognition performance with unknown faces, the specific video quality parameters were not measured as the aim of these studies were to examine the task performance differences when the observer was familiar and unfamiliar with the faces shown in the CCTV images. To understand whether digital video compression affects face identification performance and what impact excessive video compression has on the identification of unknown faces, one key question is raised: how much should the user compress digital CCTV video before it starts to affect the user's ability to identify faces from CCTV images? In addition to this, how do untrained CCTV users go about identifying unknown faces from CCTV images? What cues are used? What makes the task difficult?

The frame rate studies reviewed all investigate the relationship between video frame rate and a user's ability to perform tasks with the video. It is clear that the frame rate required for a given task is dependent upon the perceptual demands of the task, as well as the content being shown to the observer. For example, for person-to-person communication conducted over the Internet, 5 fps was considered unsuitable (Wilson and Sasse, 2000). In the gaming study conducted by Claypool et al. (2006), task performance deteriorated with a decrease in frame rate even when the game was played back at a frame rate of 7 fps. These results suggest that higher frame rates are needed for tasks with video content including people (e.g., Horn et al., 2002; Wilson and Sasse, 2000) as opposed to animated targets (e.g., Swartz and Wallace, 1993; Claypool et al., 2006).

It is unclear at this stage how much video frame rate can be lowered while ensuring observers can effectively detect crime on CCTV video. The lie detection task (Horn et al., 2002) is perhaps the most similar to a CCTV detection task, and may provide some indication of the temporal video quality requirements. Both of these tasks involve a user detecting suspicious actions and targets on video. It may seem that the detection of facial movements and body language would require higher temporal resolution, thus higher quality; however, the indicators of some crimes can be equally subtle and hard to detect (e.g., the movements of a skilled pick-pocket, or the actions surrounding a discrete drug deal carried out in a public space). Detecting these types of subtle actions in low-frame-rate video can prove to be challenging for some CCTV users, particularly untrained and inexperienced users performing CCTV security observation tasks. The lie detection study raises an important question for the CCTV video quality research undertaken for this thesis: If micro-facial movements (such as eye blinks) and body language (such as shoulder shrugs) are difficult to detect in video played at 5 fps, is it also difficult to detect subtle actions associated with CCTV video at this rate? The questions to both the identification and detection tasks with CCTV video are tackled empirically in Study 2 (Chapter 8) and Study 3 (Chapter 9).

Chapter 7

Impact of Video Quality: Methodology

The aim of this chapter is to provide a critical review of the different objective and subjective methods used in HCI research to evaluate the effect of lowering video quality on user tasks. A critical review is also presented on the methods proposed by the UK Home Office (Cohen et al., 2007) for evaluating the performance of a CCTV system and CCTV video quality. The material presented in this chapter provides the basis for the selection of the methodology used to conduct the experiments detailed in Chapters 8 and 9.

7.1 Introduction

HCI and multimedia researchers have attempted to assess the impact of low-quality video on user tasks for two main reasons: 1) to conserve hardware storage space and bandwidth and to, 2) improve the design of multimedia applications, so they are effective and usable. A number of different objective and subjective evaluation methods have been used. The type of evaluation method employed depends upon several factors, such as:

1. context in which the task is being performed – spatial and temporal factors,
2. characteristics of users – skills, knowledge and experience,
3. mental exertion required to perform the task – easy/difficult, life threatening etc.,
4. and information and measurements to be gathered through the evaluation.

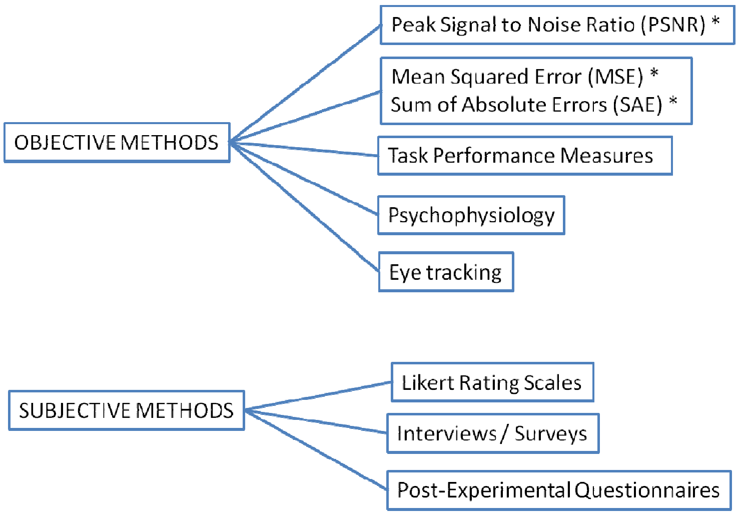
The purpose of an evaluation of a task-based video application is generally to assess performance of user tasks and overall system performance. Whitefield, Wilson and Dowell (1991) defined an evaluation as, *"…an assessment of the conformity between a system's actual performance and its desired performance"* (p 66).

The evaluation criteria for CCTV video can be objective, subjective or both:

1. Objective:
   - The impact of video artefacts and distortions on the overall performance of the system are predictively measured.
   - The users' performance is measured through the use of classic HCI task performance measures (correct detections, errors, etc.), and user costs (psychophysiological responses) as video quality is lowered.

2. Subjective:
   - The user's perceptions are measured as video quality is lowered using Likert rating scales. The users' perceptions relate to how they feel about the quality.

In early studies, evaluations were conducted to assess the impact of video quality on users utilising a number of interactive applications (e.g., e-learning, desktop video-conferencing, gaming and mobile TV). The most common objective and subjective methods used in HCI research for evaluating user interfaces and the impact of lowering video quality on user tasks are shown in Figure 7.1. Each method is described in more detail in the following two sections.


OBJECTIVE METHODS
Peak Signal to Noise Ratio (PSNR) *
Mean Squared Error (MSE) *
Sum of Absolute Errors (SAE) *
Task Performance Measures
Psychophysiology
Eye tracking
SUBJECTIVE METHODS
Likert Rating Scales
Interviews / Surveys
Post-Experimental Questionnaires

*Predictive measures (not actual measures).

Figure 7.1: Objective and subjective methods used for assessing video quality.

7.2 Objective Methods for Evaluating Video Quality

7.2.1 Predictive Measures

Objective evaluation methods involve an assessor measuring performance numerically. Typically, the assessor tallies the number of errors a user has made on a series of tasks using a video application which is degraded to different quality levels. Video quality can also be measured without involving the user in the evaluation process (quantitatively). Two of the mathematical methods traditionally used for measuring quality include the peak signal-to-noise response (PSNR) and mean squared error (MSE)/sum of absolute errors (SAE). These are pixel-error metrics commonly used by the signal and image processing community in the research domain of computer science for measuring video quality at the pixel level. Although these methods of assessment are easy to compute, they suffer a major drawback in that they are extremely poor at predicting subjective quality judgments (Girod, 1993). Miras (2004) agrees, suggesting that the metrics are relatively easy to compute, *"…but cannot match the discriminations in quality in which humans can and cannot see that are less or more annoying"* (p 56).

This was demonstrated by Wang, Speranza, Vincent, Martin and Blanchfield (2004) who used PSNR and SAE measures to assess the impact of the following encoding parameters: spatial resolution, frame rate and quantization,[17] with different video scenes. Based on the results of the experiment, Wang et al. (2004) stated that *"PSNR and SAE do not adequately reflect perceived video quality when changes in spatial resolution and frame rate are involved, and are therefore not adequate for assessing quality (in a multi-dimensional rate control scheme)"* (p 8). There is a fundamental problem with mapping PSNR data and subjective methods, since PSNR data does not take into account of temporal video quality changes (e.g., a drop in the bandwidth during a peak in Internet traffic). This limitation was identified in a mobile TV study conducted by Knoche, McCarthy, and Sasse (2008) in which PSNR measures were used to gain a rough estimate of video quality content. In this study, Knoche et al. examined the effect of broadcast TV shots on a mobile device at reduced video resolution levels. Users were asked to state whether the video clips were acceptable or not, as the quality varied in resolution. The PSNR values provided no indication of the degradation for extremely long shots of different sizes and low resolution. PSNR and other predictive objective methods of video quality assessment also do not take into account the contextual factors related to the video application and the task.

7.2.2 Actual Measures

Task performance measures are popular amongst usability researchers and practitioners who rely on them to provide objective indicators of task effectiveness and efficiency. Task performance measures are not suitable for passive tasks; they are only suitable for evaluation of user performance for tasks that require user interaction with an application. Monitoring real-time CCTV video is a good example of an active task for which task performance can be measured.

The tasks for the evaluation can either be contrived in a laboratory or real tasks carried out in the field. Where task performance measures are used for assessing a specific video application, the choice of tasks, the number of tasks and the duration of each should be carefully selected so that they represent how the application is actually used. Tasks should also be designed so that they do not mentally overload the user's cognition during the evaluation. Too many tasks or tasks that are too complex will lead to a cognitive overload, which may affect the user's responses. Similarly, if the tasks are too easy or unrealistic, the results will not be valid.

For tasks that require the user to discriminate between objects in a video scene, signal detection theory (SDT) techniques can be used to measure performance. SDT allows an assessor to determine how well a user is able to distinguish a signal from noise, and to identify the threshold in signal detection. SDT assumes that the user does not passively receive information, but actively makes difficult judgements under conditions of uncertainty. The types of measures that can be determined based on a

[17] Quantisation in the context of image processing is a lossy compression technique achieved by compressing a range of values into a single quantum value.

psychophysical experiment (using SDT) include the four task performance measures shown in Figure 7.2.

		SIGNAL	
		YES	NO
RESPONSE	YES	HIT	FALSE ALARM
	NO	MISS	CORRECT REJECTION

Figure 7.2: 2x2 matrix of signal detection measures.

The hit rate (percentage of correct responses) and the false alarm rate (the percentage of incorrect responses) are used to determine the threshold levels of performance in terms of target detection sensitivity. The hit and false alarm rates are calculated for each type of stimuli condition and provide a sensitivity measure (*d'* prime/index: discriminability). A perfectly sensitive participant will have a hit rate of 1 and a false alarm rate of 0, and therefore, will be described as an observer who is very good at discriminating noise from signal targets with a very high *d'* value. A participant with a hit rate equal to the false alarm rate will be described as a completely insensitive observer and the *d'* value will be low. The distribution of hits and false alarm pairs for each condition can be plotted graphically to show the observer's performance in a detection task. This plot indicates the sensitivity between the signal and the noise stimuli. The hit and false alarm values can be plotted on a Receiver Operating Characteristics (ROC) curve. A ROC curve shows the trade-off between sensitivity and specificity for the targets identified by the observer. For every false alarm rate value, the plot shows the hit rate that will be obtained to yield a particular sensitivity level. If performance is set at chance (d'=0), the ROC curve is a major diagonal line, called the chance line. As the observer becomes increasingly sensitive to the stimuli, the ROC curve shifts towards the upper left corner of the vertical axis (hit rate axis).

SDT measures have been used in video experiments, for example, experiments designed to evaluate a human observer's ability to recognise faces using computer algorithms (Schwaninger, Lobmaier and Collishaw, 2002). Subjective methods can also be combined with task performance measures to provide a holistic analysis of a user's performance and their perceptions of video quality.

Another popular objective method used in HCI and usability research studies is psychophysiology. Psychophysiology is the branch of physiology that deals with the relationship between physiological processes and thoughts, emotions and behaviour. Wilson and Sasse (2000, 2004) conducted a number of experiments and found that biological signals such as heart rate (HR), blood volume pulse (BVP) and galvanic skin response (GSR) are key indicators of stress and arousal when users are confronted with low-frame-rate video. Bouch, Sasse and Wilson (2001) also found that if users are presented with low-quality video and audio when using a video conferencing system, they *"expend extra effort* *perceptually without being consciously aware"* (p 3). As users increase their efforts to decode low-quality media, this effort becomes apparent in their physiological responses.

Wilson and Sasse (2004) noted that a user's HR is a valuable indicator of the body's overall activity level. During stressed or aroused states, the user needs to increase blood flow to the working muscles in order to act upon the stressful situation and to ensure the body is prepared for a 'fight or flight' response. Thus, if the user's HR increases, it indicates that they are feeling stressed or anxious. HR and BVP can be recorded using the same sensor, which is placed on a user's fingertip. The sensor records infrared reflections, which are shone on the skin's surface. The BVP waveform exhibits the characteristic periodicity of the beating of the heart, and each beat represents the blood forcing itself through the blood vessels. Under stress, the user has a lower BVP value at their extremities than in their working muscles. SC is another psychophysiological measure of stress. It is an indicative measure of the user's arousal level (also known as galvanic skin response). Two small silver chloride electrodes are placed on the user's index and second finger. A small, imperceptible voltage is applied to the electrodes and the conductance is measured between the two sensors. SC increases when the individual is startled by novel or interesting stimuli and when anxiety is experienced.

Although collecting physiological response measurements from users is easy, in practice, processing and inferring conclusions from objective data is perhaps the biggest challenge for researchers. For instance, physical signs cannot be used to positively determine human behavioural traits, such as stress, strain, boredom, frustration and so on. Instead, these signs should be considered indicators of the user's affective state and not their actual affective state.

Eye tracking is a method used for capturing a user's attention to a visual interface containing text, imagery and/or video content. Eye tracking is an objective technique in which fine temporal-grain data is collected and used to infer various higher cognitive states; many of which do not reach consciousness and are therefore not verbalisable. The interpretation of eye tracking data provides insight into different cognitive states, such as difficulty with visual processing, attention, stress and reading skills (Ikehara and Crosby, 2005). In the 1970s, eye tracking became a very popular technique in HCI research. More recently, it has been used to identify usability issues of web interfaces (e.g., Ellis et al., 1998; Cowen, 2001).

There are a number of eye tracking tools that exist today, ranging from head-mounted trackers to remote, portable eye trackers. Tracking of the human eye is now achieved using a low-power; infrared light-emitting diode (LED) located in the centre of an eye tracker camera lens illuminating one or both of the observer's eyes. The LED generates a small, but very bright reflection on the surface of the observer's cornea, which creates a 'bright-pupil' effect, and light is reflected off the retina. The x-y coordinates of these reflections are continuously registered by the eye tracker, and can therefore be used to generate images of the observed areas. Images, such as heat maps and scan paths, can be manually created from the raw data, or software can be used. In addition to visual data, an eye tracker produces raw spatial and temporal data which provide information about the user's duration of gaze (fixations), fast conjugate changes of eye positions between fixations (saccades) and pupil diameter changes, as well as other information. These images and information produce an overview of the observer's regions of interest. Figure 7.3 illustrates a typical set-up for an eye tracking study and summarises the processes involved in viewing eye-tracking data. This diagram was created for illustrative purposes.

1. LED light from eye tracker illuminates participant's eyes

2. LED creates a 'bright-pupil' effect, reflecting light off the retina

3. Bright pupil reflection enables eye tracker to record the participant's X-Y screen coordinates.

4. The assessor is able to observe the participant's eye movements and facial expressions on the observation monitor.

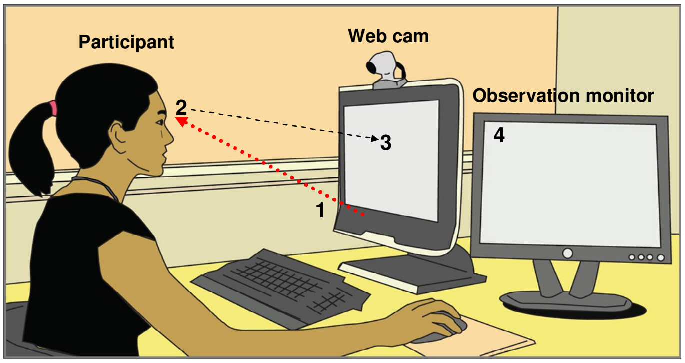


Figure 7.3: An overview of the set-up and process involved in an eye-tracker study.

Eye trackers are becoming increasingly easy to set-up and operate; however, large amounts of data can be produced following an eye-tracking evaluation, which can be time and resource consuming to process during post-analysis.

Although research shows that the point of fixation usually coincides with the point of interest (Just and Carpenter, 1980), humans are also capable of processing information in their peripheral vision, which is usually three to four degrees off the centre area of fixation. Human peripheral vision is of much lower resolution than foveal vision, which means that observers may only have a rough idea of the events happening outside of the fixation point. Users naturally use their peripheral vision to determine what is and what is not interesting. For example, when an operator is tracking a target using CCTV on a video monitor, the movements detected in the operator's peripheral vision will be registered and only revisited if these movements are interesting. Furthermore, the user is likely to revisit areas of the scene where movement attracts attention, especially if these areas are not natural parts of the scene. If the movements or actions detected within the user's periphery are suspicious, they are also likely to be revisited. Peripheral vision is therefore used to define targets for subsequent eye movements or saccades. The use of interviews, think-aloud protocols, and task performance measures can be combined with eye tracking to facilitate an in-depth analysis.

There have been a number of eye-tracking studies carried out in HCI research, which have examined eye movement patterns with static images; however, there have been very few studies conducted with video stimuli. One particular eye-tracking study evaluated a video-based task (Muir, Richardson and Hamilton, 2005). In this study, eye tracking was used to identify the characteristic eye movements of users whilst observing sign-language video, which was lowered in spatial quality. This study showed that when video quality was lowered, the observer's gaze was focused on finer facial movements, rather than on large, gross hand movements, where a majority of the signing took place. These findings were consistent with the findings of a sign language study conducted by Agrafiotis et al. (2003).

Eye tracking can be applied in HCI research and can be a very useful technique for evaluating the effectiveness of CCTV video when used by human observers for security observation tasks. It can be used to understand where users look when scanning CCTV video on a monitor wall within a CCTV control room to determine eye movement behaviour under poor visibility. Eye tracking was used successfully for a study conducted during the research for this thesis. This study examined the difference between scanning patterns of trained CCTV operators and untrained participants performing a detection task for a railway crossing. This study is not presented in this thesis, due to confidentiality reasons (i.e., the research was conducted in collaboration with a research consultancy and a commercial organisation).

7.3 Subjective Methods for Evaluating Video Quality

In order to establish the impact of low-quality video on a user's perceptions, it is useful to measure their perceptions through verbal accounts, questionnaires and Likert rating scales. There have been a number of studies conducted which illustrate the importance of gathering user ratings in addition to objective methods (see Knoche et al., 2008; Veltman and Gaillard, 1998). Likert rating scales can be used to measure a user's perceived video quality response. They are usually labelled using discrete points, with each point indicating a different level of subjective opinion of the test material. The data gathered from ratings are objective when an average value for each point on the scale across users is calculated. Likert rating scales are commonly used in many types of attitude research surveys and are easy for the user to complete, although they have been criticised for being biased: *"bold statements usually produce positive bias responses"* (Hill, Brierley and MacDougall, 1999, p 76). The results from Likert rating scales are also unreliable, as users are forced to give a single rating of the stimuli under timed conditions. In addition, when uneven numbered Likert scales (e.g., 5-point and 7-point) are used, the user is likely to give a middle, rating when unsure what rating to give. Overall, subjective bias can result in skewed and unreliable data, which is undesirable for video quality user assessments.

The ITU-T[18] image quality and impairment scales are widely recognised by telecommunication researchers and commonly referred to as 'MOS[19]' scales. The ITU Recommendation BT.500-1 (2002) was originally proposed by the ITU-T as a set of recommendations for assessors to use in the subjective evaluation of the performance of television systems. The user's perceived quality for audio and video can be measured using two different image quality and impairment scales (see Table 7.1).

[18] International Telecommunication Union.
[19] MOS stands for Mean Opinion Scores.

Table 7.1: ITU-T Quality and Impairment Scales

Quality Scale	Impairment Scale	Score
Excellent	Imperceptible	5
Good	Perceptible but not annoying	4
Fair	Slightly annoying	3
Poor	Annoying	2
Bad	Very annoying	1

In the recommendations, two methods are given for assessing multimedia quality with users:

1. Double stimulus continuous quality scale (DSCQS)
2. Single stimulus continuous quality evaluation (SSQCE)

The method applied depends upon the context of the evaluation. DSCQS is a method used for evaluating new systems or the effects of network transmission paths on video quality. Typically, the researcher asks the user to assess pairs of pictures or video clips (lasting approximately 10–12 seconds) from the same source; one is degraded and the other is an original. The user rates the quality after each presentation using a quality scale. At the end of the assessment, the mean scores for each test condition are calculated to obtain the MOS. SSQCE, on the other hand, is used for assessing picture quality when the impairments fluctuate widely within the scene. In this test, users are asked to assess the overall picture quality of each presentation using a quality scale.

Despite the popularity of image impairment and quality MOS scales, they have been disregarded by many researchers for a number of reasons:

1. Giving subjective ratings and verbal reports may interfere with a user's task.
2. Giving subjective ratings may affect the user's performance in the task (Wilson and Descamps, 1996; Wilson and Sasse, 2004; Bouch and Sasse, 1999).
3. The scales are not internationally equivalent or designed to assess changes in video quality over time (Watson and Sasse, 1998; McCarthy, Sasse and Miras, 2004).
4. Degradations that are not consciously noticeable cannot be measured using ITU scales, as participants are unaware that these degradations exist (Knoche and de Meer, 1999).
5. Gradual differences in ratings may not be detected due to individual differences in judgement (Knoche and de Meer, 1999).

When evaluating video quality for network-based video applications, it is essential that the video test stimuli be realistic and representative of the application being assessed. The ITU-T recommends that test video sequences be presented to observers for approximately 10–12 seconds. However, such short exposures to video degradations mean that the user is likely to forget the artefacts present in the video. Therefore, task performance measures are more suitable, but only if the experiment requires the user to perform tasks.

Following a critique of the ITU MOS methodology, Knoche and de Meer (1999) proposed a task-oriented approach for assessing video quality with users, which considers context. This approach involves conducting standard task performance tests based on a set of scenarios, taking into account the context of the application under assessment. This method allows the researcher to compare and reproduce the results in future experiments. In addition to task performance measures, subjective data can also be gathered from users (interviews during and after the evaluation). Users can be asked post-experimental questions, such as: "At what point did you feel the quality was insufficient?" and "What aspect of the task did you feel you performed poorly?" This task-oriented approach was applied in both the video quality experiments carried out in this thesis (see Chapter 8 and 9).

7.4 A Combined Assessment for Evaluating Video Quality

Shackel (1981) describes a usability evaluation framework based on three levels of criteria: dimension, performance and attitude, which can be applied for assessing systems and products. The three criteria are categorised into three different assessment methods: analytical, objective and subjective. For each method, several types of measurement techniques can be applied by the assessor (see Table 7.2).

Table 7.2: Evaluation Framework Developed by Shackel (1981)

Criteria	Method	Measure Type
Dimension	Analytical	▪ Physical ▪ Anthropometric
Performance	Objective	▪ Physiological ▪ Operational ▪ Experimental ▪ Functional
Attitude	Subjective	▪ Psychological

Shackel argued that it is better to choose three dimensions for an evaluation than one single dimension. This view was also supported by John and Marks (1997), who compared several HCI evaluation methods and suggested that "*not one method is suitable for assessment and also the use of all evaluation methods are of limited value*" (p 16). Dix, Finlay, Abowd and Beale (2003) are in agreement and recommend that, ideally both objective and subjective approaches should be used. Combining objective methods with subjective data collection methods is a concept also favoured by other researchers who have examined the impact of media video quality on users (Bouch et al., 2001; Wilson and Sasse, 2004).

7.5 Methods for Assessing CCTV Video Quality

The purpose of assessing the performance of a CCTV system is to identify whether the technical configurations of the system, such as video quality (e.g., frame rate, video compression and resolution) as well as the camera set-up (e.g., location, angle, focus etc.) are effectively established to ensure that the goals for the CCTV are being met. This type of performance assessment is not often achieved, as CCTV owners have limited access to guidance on the design, configuration, and use of digital CCTV systems. Where guidelines are available, CCTV owners have little awareness of their existence and their importance in achieving effectiveness (Police Guidelines, 2007). Consequently, there are many digital CCTV systems that are not fit for purpose. The current performance guidance, which outlines requirements and methods for assessing CCTV video quality, are reviewed in the following sections.

7.5.1 Home Office Operational Requirements

The Operational Requirements Manual (Aldridge, 1994), a Home Office Scientific Development Branch (HOSDB) report published over a decade ago, was one of the first sets of guidelines to provide non-technical, informative and methodological procedures for use by a wide range of analogue CCTV owners. The manual outlines the key factors that can affect the performance of a CCTV system and provides a checklist for CCTV owners to complete in determining their operational system requirements. The methods provided can be used for existing and proposed CCTV security systems. The manual also provides a checklist to identify problems, which covers issues related to picture quality and content factors. The checklist is generally used to describe an area of surveillance interest in relation to a marked CCTV site plan. It offers the user checkpoints that can be followed for testing the performance of a system. The manual was re-issued in 1995 for the assessment of perimeter surveillance systems (Aldridge and Gilbert, 1995). Since this manual was published, a number of CCTV owners in the UK have referred to it for guidance.

Over the past few decades, digital technology has revolutionised the CCTV market and digital and networked CCTV systems have become increasingly widespread throughout the UK. The Home Office has recognised this technological change and in the past 10 years they have started to develop guidelines for second and third generation CCTV systems and an updated version of the Operational Requirements Manual was published in 2007 (Cohen et al., 2007). The new version addresses some aspects of digital and IP video quality; however, due to the complexity of these systems, considerably more experimental research is needed to provide CCTV owners with a set of guidelines for effective design configurations for digital CCTV systems.

7.5.2 Rotakin Standard Test Target and Image Height Guidelines

Aldridge and Gilbert (1995) developed a series of CCTV performance tests that can be used with a standard test target to evaluate analogue CCTV systems once an intruder alarm is triggered within a secure environment. These performance guidelines relate to the four security observation tasks defined by Aldridge (1989) and are based on a 625-line CCIR[20] standard video system. In conjunction with these guidelines, a test target, called the 'Rotakin', is used as a measurement tool to help with the configuration of a video security system (see Figure 7.4). The Rotakin target is a matt black, flat panel cut out in the shape of a male (1.6m x 0.4m wide). The Rotakin panel bears high contrast resolution bars and a resolution wedge chart. The markings on the target represent the resolution in TV lines per picture height when the target fills the vertical picture. These resolution bars are used to calibrate the target from top to bottom.

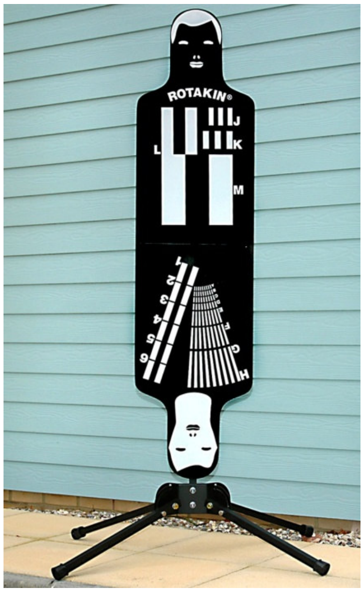


Figure 7.4: Rotakin ® standard test target.

In the most recent Home Office Operational Requirements Manual, Cohen et al. (2007) suggest that the appropriate image size of an on-screen target should be aimed towards meeting particular requirements, rather than towards meeting a minimum standard for image quality. For a monitoring task, the CCTV camera should be configured so the targets occupy no less than 5% of the screen; for a detection task, targets should occupy no less than 10%; for the recognition of a target, no less than 50%; and no less than 120% for identification. Figure 7 illustrates the image heights in Rotakin sizes for these four CCTV observation tasks.

[20] Consultative Committee for International Radio.

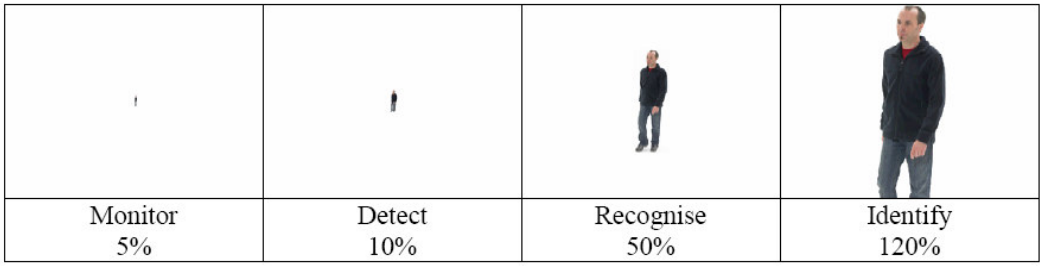


Figure 7.5: Home Office recommended image heights (Cohen et al., 2007).

In summary, the Rotakin standard test target may be used to assess the following:

1. Task area identification – areas to be viewed should be clearly identified within a plan and described if necessary.
2. Image height – size of the images produced on the CCTV monitors.
3. Anticipated minimum user response time necessary for the system to be effective.

The image height guidelines shown in Figure 7.5 (Aldrige, 1994) were originally proposed for analogue systems; thus, they cannot be applied to digital CCTV systems, as the recording process is different. Digital CCTV video, unlike analogue CCTV video, undergoes video compression during data storage and transmission over a network, which cause the quality of the video to be reduced. This reduction in quality means that the image height guidelines cannot be applied to digital CCTV systems that have been set-up to record video for spot-event investigative tasks, or for tasks performed in real-time over the Internet.

7.6 Chapter Summary

In this chapter, the different HCI methods used for evaluating video quality were reviewed, as were the specific methods used for assessing CCTV video quality with observers. This review demonstrated that there are currently no established methods in HCI for evaluating digital and networked CCTV.

The Operational Requirements Manual was developed when analogue CCTV systems were in a period of peak production and use. The current version of the manual (Cohen et al., 2007) does not yet include a framework of guidance on digital and IP CCTV systems. Specific guidance is needed for those who use digital systems. This guidance must consider the impact of both spatial and temporal video compression, and can only be developed through empirical evaluations with users.

Task performance measures are the most suitable measures for evaluating CCTV video quality, as task performance is the most important measure of CCTV effectiveness. If high task performance cannot be achieved, the system will be ineffective and not fit for purpose. In addition to task performance, it is also important to identify the impact of video quality on an observer's confidence and perceptions of video quality. Subjective ratings and opinions from users can provide an indication of the level at which video quality affects a user's judgement. Combining objective and subjective data will strengthen the validity of the results. A task-oriented approach (Knoche and de Meer, 1999) is a useful approach, as the evaluation takes into consideration the real-world measures and variables that may affect an observer's performance in a video task. This approach also provides CCTV owners with objective indications of video quality in the form of performance, a threshold criterion that can easily be applied to digital CCTV systems used in the real-world.

Chapter 8

Study 2: Face Identification with Compressed CCTV Images

In this chapter, a study into the effectiveness of digital CCTV for a security observation task: face identification is detailed. The study relates to the literature review on digital CCTV, which was presented in Chapters 5, 6, and 7.

8.1 Introduction

In Study 1 (Chapter 4), the use of very low-quality CCTV video for police tasks was identified as one of the biggest issues contributing to reduced CCTV effectiveness. On a number of occasions during the field visits, the police were unable to discern the identities of targets from post-event (recorded) CCTV video during their investigations. This was also an issue found in a number of CCTV control rooms in a previous study (Gill et al., 2005), and has been noted as problematic for prosecutors (Bromby, 2002; Mead, 1998).

The study reported in this chapter investigates the effect of lowering CCTV video quality – specifically spatial video compression on a face identification task. As digital CCTV systems are being used by a wider range of users – not just traditional CCTV users such as the police and security staff, but also untrained users (see Web Users to Patrol, 2006; Rights Group, 2006), there is a need to: 1) assess how these users identify targets from CCTV video; 2) determine the reliability of using untrained CCTV users for identifying faces from low-cost digital CCTV systems and 3) determine what video quality level is the most effective for this task.

8.2 Research Motivations

This study was motivated by the need to identify objective and actionable video quality requirements for a face identification task. The video quality requirements for this specific task are needed through evaluations with users to help support CCTV practitioners and owners design, configure and use CCTV systems as effectively as possible. Much of the existing guidance on CCTV video quality is vague, inconsistent and developed through craft-knowledge, rather than assessing tasks with real-world tasks users through scientific experimentation. The identification of a minimum digital video quality level will ensure that CCTV owners will be recording and streaming CCTV video which is usable, task effective and efficient (for data storage and streaming purposes).

8.3 Research Goals and Hypotheses

The research goals for this experiment are as follows:

1. Develop an understanding of how untrained CCTV users perform a face identification task (where the targets are unfamiliar) with low-quality digital CCTV images, typically taken from video compressed by a digital CCTV system (recorded video).

2. Establish which video compression CODEC (MPEG-4 or Wavelet) and level is the most effective when used for identifying faces, and how much video should be compressed.
3. Determine the minimum bit rate required for a face identification task and incorporate this requirement into the best-practice framework (see Chapter 10).

The two hypotheses addressed by this experiment are as follows:

- H1: As the quality of CCTV video is lowered through spatial compression, the observer's ability to identify faces from CCTV images will decrease.

This hypothesis is the main focus of the study and was defined based on the current practices with digital CCTV systems. CCTV owners are attempting to cut storage and transmission costs to save money running their CCTV systems and as a result compromising video quality over cost (Cohen, 2004 and Gerrard et al., 2007).

- Hypothesis 2 (H2): As the quality of CCTV video is lowered through spatial compression, identification performance will be higher with MPEG-4 compressed CCTV images than Wavelet.

An additional interest at this stage of the research was to establish empirically whether face identification performance will be better or worse with CCTV video encoded using an MPEG-4 or Wavelet video CODEC. These two particular video CODECs were considered in this study as they are two of the most common video CODECs used in commercial CCTV equipment (Robertson and Monro, 1997 and Walker, 2005). No previous work has been done to compare the effects of these two CODECs on video quality and task performance for security applications; therefore, a comparison was made in this study.

8.3.1 Participants

The face identification experiment was conducted with 80 paid participants at the Department of Computer Science, University College London. The participant sample comprised of untrained CCTV users: 22-46 years of age (mean = 28.42 years, SD = 5.24 years), of which 47 were female. As this experiment assessed identification performance using targets unfamiliar to participants, those volunteers who appeared in the CCTV images were not recruited in the study. Participants for the identification task were recruited from an opportunity sample and included a wide range of ethnicities (see Table 8.1).

Table 8.1: Ethnic Groups Chosen for the Targets in the Identification Test

Identity Code[21]	Ethnicity Type	Targets in CCTV	Participant Recruited
IC1	White European	8 males / 8 females	45
IC2	Dark European	0	0
IC3	Afro Caribbean	8 males / 8 females	6
IC4	Indian Asian	8 males / 8 females	19
IC5	Oriental	8 males / 8 females	10
IC6	Arab	0	0

8.3.2 Materials and Data Collection

As there are no standardised CCTV video data sets available for assessing CCTV video containing targets with human observers, mock-up CCTV video was recorded to create the CCTV images for this identification experiment. This involved filming 64 video clips of an individual walking towards the video camera. For each clip, the individual held one of 64 different face masks. This procedure simulated a real target walking towards a digital video camera. A single image of the target at the same point was taken from each video clip and used for the 64 CCTV images for the face identification test. The walking mask method was adopted, instead of filming 64 individuals walking one-by-one towards a video camera, principally to focus participants on the face of the target during the task. Although this walking mask method loses 3d information from targets' faces, this method was used to keep targets' characteristics, such as their height, physique, and gait constant to allow for an accurate measure of identification performance.

The first stage in preparing the CCTV images involved creating the 64 face masks. This was achieved by taking photos of 64 individuals (age range: 20-40 years old). Sixteen targets were taken from each of the four different ethnic category groups (see Table 8.1). These ethnic groups were chosen to provide CCTV images containing people from a wide ethnic mix. Each person was photographed in a full-face pose and asked to give a neutral facial expression within controlled lighting conditions using a 2-mega pixel digital camera. Each face image was cropped so that only the face was visible and the background was removed. The images were resized so that each face printed to the equivalent size of an average human head and printed onto high-quality matt finish colour paper. Matt paper was used to avoid light reflections during filming.

[21] These ethnicities are based on the UK police Identity Code (IC) categories. ICs are used by police officers and staff when describing the ethnicity of an individual through official communication channels.

Each face mask was attached to a clear plastic strip and held in front of the individual. Video was recorded in an indoor, well lit squash gymnasium. This environment was considered suitable for the recordings, as indoor sports halls provide evenly distributed lighting from floor to ceiling. Figure 8.1 and Figure 8.2 illustrate the recording set-up with measurements in elevation and plan view.

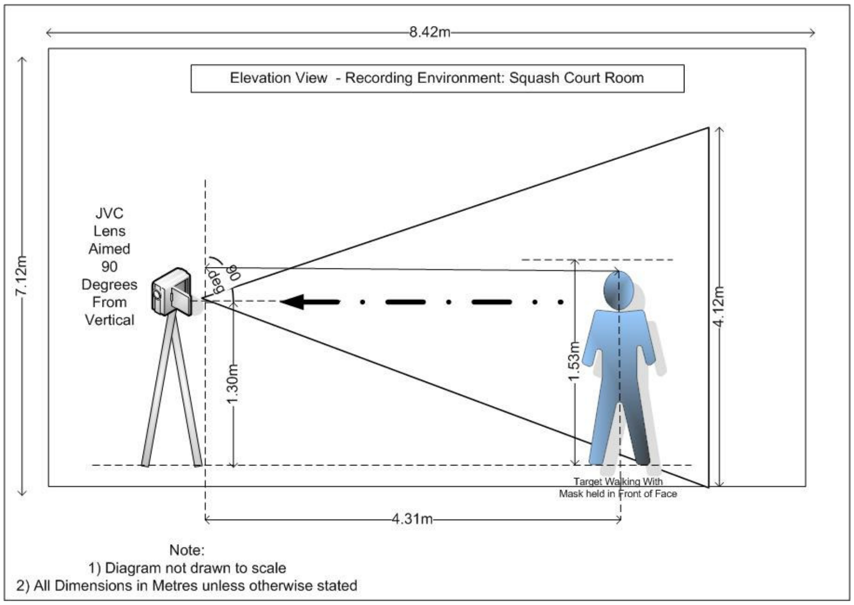

8.42m
Elevation View - Recording Environment: Squash Court Room
7.12m
JVC Lens Aimed 90 Degrees From Vertical
90 deg
1.30m
1.53m
4.12m
Target Walking With Mask held in Front of Face
4.31m
Note:
1) Diagram not drawn to scale
2) All Dimensions in Metres unless otherwise stated

Figure 8.1: An elevation view of the CCTV video recording set-up.

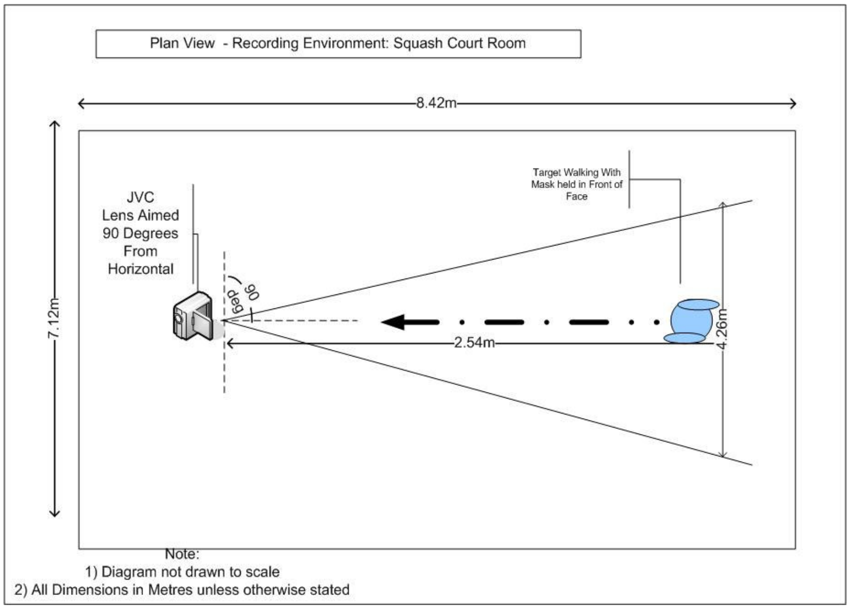

Plan View - Recording Environment: Squash Court Room
8.42m
7.12m
JVC Lens Aimed 90 Degrees From Horizontal
90 deg
Target Walking With Mask held in Front of Face
2.54m
4.26m
Note:
1) Diagram not drawn to scale
2) All Dimensions in Metres unless otherwise stated

Figure 8.2: A plan view of the CCTV video recording set-up.

The next stage required taking an additional 32 photographs of (new) individuals under the same location and lighting conditions using the same video recording equipment. These photographs were used for the target look-alikes and the individuals shared the same ethnicity, but intentionally bore no close resemblance to the matching targets' faces. These look-alike targets were included in the test so that it would be possible to detect any identification errors when the degraded CCTV images were shown across the different video quality conditions and levels. Half of the photographs (32 out of 64) were taken of individuals who looked like half of the targets who appeared in the CCTV images (signal absent condition) and the remaining 32 were exact matches of the targets appearing in the CCTV images (signal present condition). These photographs were printed onto high quality matt photo paper (5.5x5 cm) and affixed to white paper to create the photo book. The photo book was used for the identification task, in which each target in the book was compared with a target on-screen.

The CCTV images (within the test and the images within the photo book) were arranged in random order to ensure that the video quality and CODEC conditions were observed in no repeated order. Figure 8.3 depicts examples of the types of target used for the face masks with their corresponding look-alike faces. In this figure, correct and look-alike targets have been taken from each ethnic group for illustration.

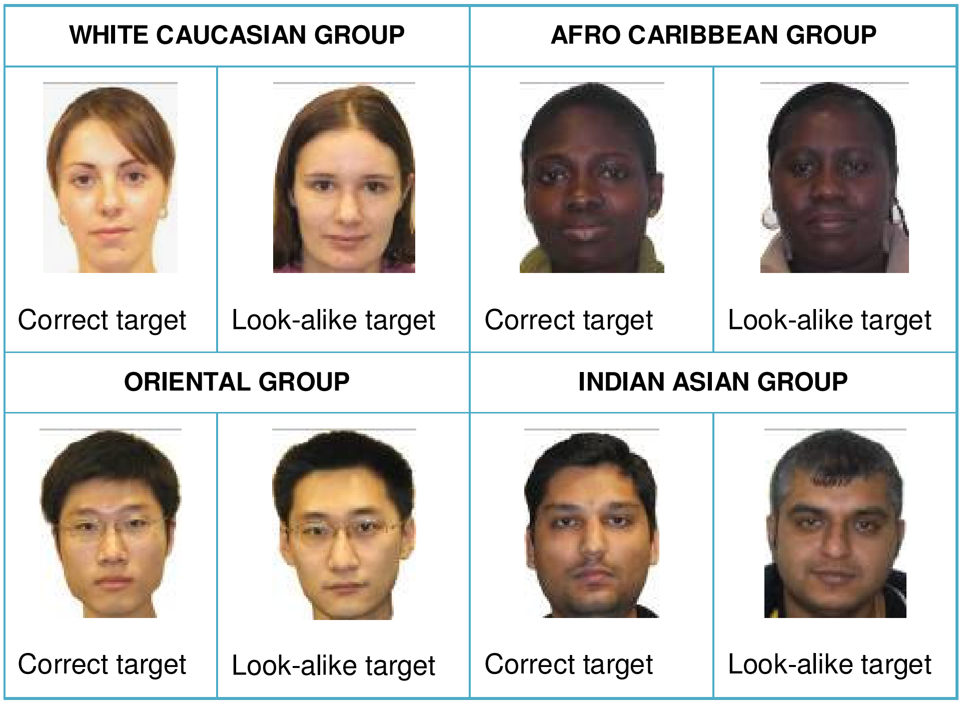


Figure 8.3: Examples of face masks used and corresponding look-alike targets.[22]

The suitability of each look-alike target (for each ethnic group) was validated by asking ten volunteers (untrained in CCTV tasks) to complete the identification test (with all of the 64 CCTV images in full quality) using the photo book. The results showed that three of the look-alike targets were very difficult to identify and this was because their facial features were very similar to their corresponding targets

[22] Permission to print images was granted by all participants via model-release consent forms.

from the CCTV images. These look-alike targets were replaced and the identification test was repeated with another ten volunteers. Performance in the second test was successful.

Table 8.2 lists the variables which were controlled in this experiment. It was necessary to control these variables so that the independent variable, video quality, could be altered effectively and the dependent variable, task performance and confidence, could be measured reliably.

Table 8.2: Description of Recording Variables Controlled

Variable Chosen for Control	How Variable was Controlled
Recording lighting	The recordings were made under uniform lighting conditions. This was achieved by recording the video in a squash gymnasium where the lighting levels were even.
Recording background	Any change in the background of a video clip (still or moving) alters the way in which the video is encoded and therefore perceived by the observer. All recordings were made in the same location with the same set-up on the same day.
Physical target appearance	The same person was used for all of the recordings, wearing the same plain dark clothing. The person's gait and posture was also maintained as best as possible on each recording.
Target walking speed	The walking speed of the target was controlled as best as possible. For each recording, the target was instructed to walk towards the video camera at an average walking speed of 0.8m/s.
Target expression and pose	The face masks were taken of individuals with a neutral facial expression in a full-face pose under the same lighting conditions and location.
Camera height	The height of the video camera recording the targets was kept at an eye-level height of 1.30m. The height was deliberately not matched with typical CCTV camera heights, as this would introduce another variable into the experiment.

The original video scenes, which were recorded in the squash gymnasium, were lowered in quality through digital video compression. This was achieved by using the following video CODECs:

1. MPEG-4 video CODEC (Microsoft Windows Version 3), using the Microsoft Windows Video Encoder 9 Series software.

2. WAVC[23] video CODEC (one type of Wavelet video CODEC), using Capture Professional V6.02 software.

Digital video conversion was necessary so that the video clips could be compressed into the different video quality levels (32, 52, 72, and 92 Kbps). Following video compression, all of the video clips were lowered in resolution from 720x576 to CIF resolution (352x288) - a common format used for recording CCTV video. One frame (image) from the video was selected from each of the 64 video clips and used for the test presentation. This image showed the target head to thigh with the face at approximately 17x 27 pixels. This level of zoom was chosen to match the corresponding 120% Rotakin target size criteria required for an identification task (Cohen et al., 2007). An example of an encoded CCTV still used in the experiment with its look-alike target image (MPEG-4, 52 Kbps condition) is shown in Figure 8.4.

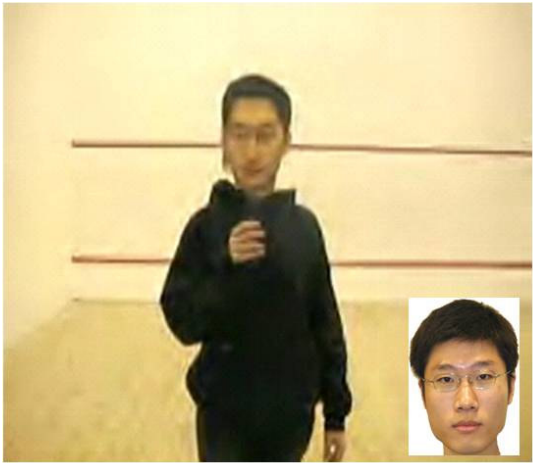


Figure 8.4: An example of a degraded CCTV image and the look-alike photograph.

8.3.3 Independent Variables

The independent variables and levels chosen for this study were:

1. Video CODEC: MPEG-4 and Wavelet
2. Video quality: 32, 52, 72, and 92 Kbps

Video bit rate was chosen to lower video quality as it is an objective and meaningful measure of CCTV quality, particularly for systems set-up for remote surveillance. It is also an easy, accurate and replicable method for altering video quality for experimental purposes. The lowest rate of 32 Kbps was chosen as many digital video recorders offer this rate as the minimum quality level for streaming video.

[23] WAVC stands for Wavelet Approximated Video Compression. This video CODEC is available at: http://www.cwaip.nus.edu.sg/demo/wavc.htm.

For example, the Indigo Vision VideoBridge Networked video recorder offers CCTV owners network configurable rates ranging from 32 Kbps to 1 Mbps for remote transmission and the equivalent for recording video locally. This rate has also been used in previous research on video quality. The highest video quality level was 92 Kbps. Although this is a relatively low video quality rate for streaming video, it is a common bit rate (or equivalent quality rate for recording video) applied by CCTV owners (Cohen, 2004).

8.3.4 Dependent Variables

Two measures were recorded from each participant:

1. **Task performance**: Performance in the task was measured using a binary point system. A score of 1 = correct answer (possible 50% true positives and 50% true negatives) is given under the two following circumstances:

   - True positive = the participant believed that the target in the stimuli (CCTV image) was the same as the corresponding target in the photo book and this was true.

   - True negative = the participant believed that the target in the stimuli was not the same as the corresponding target in the photo book and this was true.

   A score of 0 = incorrect answer (possible 50% false positives and 50% false negatives) is given under the two following circumstances:

   - False positive = the participant believed that the target in the stimuli (CCTV image) was the same as the corresponding target in the photo book and this was not true (it was a look-alike target).
   - False negative = the participant believed that the target in the stimuli (CCTV image) was not the same as the corresponding target in the photo book and this was not true (it was a look-alike target).

Using these scores, it was possible to measure the number of hits (correct matches) and false alarms (type I errors) to identify performance in relation to the stimulus sensitivity. This analysis was taken from the yes/no procedure, a signal detection technique (Green and Swets, 1976).

2. **Confidence in identification:** After each image was matched, participants were asked to select a rating which best reflected their confidence in their response using the following 5-point Likert scale (see Table 8.3).

Table 8.3: Identification Confidence Scale

Likert Rating	Rating Label
5	That is definitely the same person
4	That is the same person
3	That could be the same person
2	That is not the same person
1	That is definitely not the same person

8.3.5 Design

A 2x4 within-subjects experimental design was adopted. In order to balance the face images (for target and ethnicity), the CCTV images were randomly ordered. Each face within the presentation was a different person, and as all 80 participants saw the 64 faces it was necessary to randomise the video quality conditions fairly to avoid face bias (e.g., some faces may be easier to identify if they are highly distinctive, and more so if they are assigned a high-quality video condition). To overcome possible face bias, each of the 64 video clips were encoded into the video quality x video CODEC conditions. This required the production of eight separate counterbalanced tests consisting of 64 faces in each test, with each face encoded differently from another. Counterbalancing was necessary to rotate the conditions across participants and conditions.[24] The age of the targets were not balanced, therefore this factor was not randomised as it was not part of the study objectives. Target ethnicity and gender were both equally balanced and randomised in the presentation. Table 8.4 provides the set-up and criteria for the identification test.

[24] 64 possible combinations arose from 2 factors = MPEG-4 + Wavelet = 2 x 4 levels = 32, 52, 72 and 92 Kbps = 8 x 8 order combinations = 64.

Table 8.4: The Criteria and Set-up Used for the Identification Study

Scenario set-up:	Untrained CCTV Users
Location of CCTV video recording:	Indoors
CCTV images created to simulate:	Post-event recorded—playback mode and low-bandwidth real-time mode
Recorded to show:	People walking towards CCTV camera (no specific activity)
Task:	Face identification when good quality face image is available for comparison
Field of view:	Horizontal
Features used for task:	Face and hair only
Maximum distance CCTV video captured:	4.31m
Shutter speed:	1/50
Video resolution:	CIF: 352x 288
Presentation form of CCTV:	CCTV still
Average target pixels per stimuli:	~446 pixels

8.3.6 Procedure

On arrival, participants were first briefed about the experiment and asked to read and sign the study instructions and give their informed consent. They were then tested for visual acuity using the standardised Snellen eye chart test and an Ishihara colour blindness test[25] (see Appendices D and E). One-by-one, participants were seated in an adjustable chair facing a 21-inch video monitor positioned at a recommended eye-to-screen distance of 25 inches (62 cm). The viewing area of the monitor was set at approximately 15° below horizontal eye level. Figure 8.6 shows the set-up of the experiment with a participant performing the face identification task.

[25] The Ishihara test was used as a colour perception test and an indication of a person's ability to distinguish detail in colour (i.e., detect differences between ethnicities). The Snellen test was used to determine those participants with poor visual abilities. Both the Snellen and Ishihara tests are currently used by the Royal National Institute of the Blind (RNIB) and government institutions to classify people with visual impairments (acuity and colour perception).

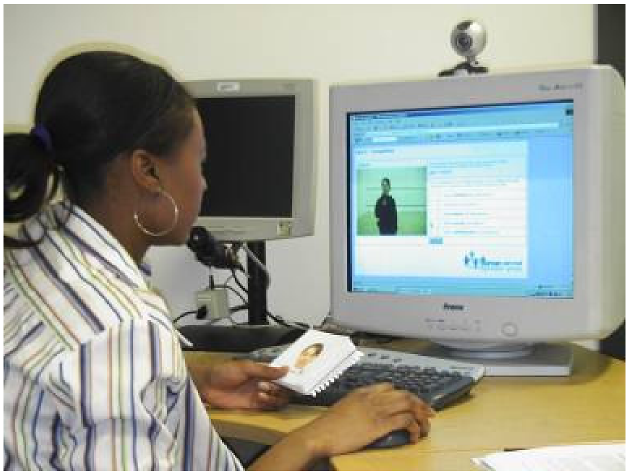


Figure 8.5: Participant performing identification task.

Participants began by familiarising themselves with the task through practice with four CCTV stills (two stills were taken from each condition). For each CCTV image, participants were asked whether the target in the image matched the corresponding target in the photo book. Only a 'yes' or 'no' response was accepted. They were then asked to give a confidence rating for each response. Participants were told that they were under no time pressure to complete the task – a time limit would have potentially affected their performance in the task. However, participants were told to spend roughly '30 seconds' on each CCTV image. Upon task completion, participants were asked to complete a post-experiment questionnaire (see Appendix F). The questionnaire included the following questions:

1. Rank the difficulty in the task by target ethnicity. Please explain your rankings.
2. Which was harder to identify: males, females or either and why?
3. What features did you use to identify the faces?
4. Overall, did you find the task difficult?

8.4 Results

8.4.1 Task Performance and False Alarms

Performance in the task was calculated into hits (total true positives and false negatives) and false alarms (false positives) across the video quality conditions and levels. The results are shown in Figure 8.6 and Figure 8.7[26].

To summarise, the task performance results in Figure 8.6 show that:

- As MPEG-4 video quality increased from the 32 to 92 Kbps, the average number of hits (correctly identified targets) increased by 11 targets (~20%).

[26] The error bars in both graphs represent the standard error.

- As Wavelet video quality increased from 32 to 92 Kbps, the average number of hits remained steady from 52 to 92 Kbps, but increased from 32 to 52 Kbps by 5 targets (~8%).

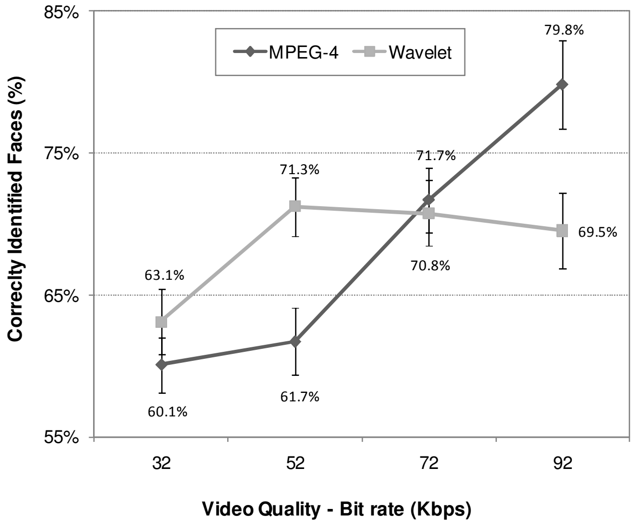


Figure 8.6: Identification performance with an increase in video quality.

The false alarm results in Figure 8.7 show that:

- As MPEG-4 CCTV video quality increased, the number of false alarms on average decreased by ~18%. This decrease was equivalent to 11 targets.
- However, as Wavelet video quality increased, the number of false alarms increased on average by ~9%. This increase was equivalent to 6 targets.

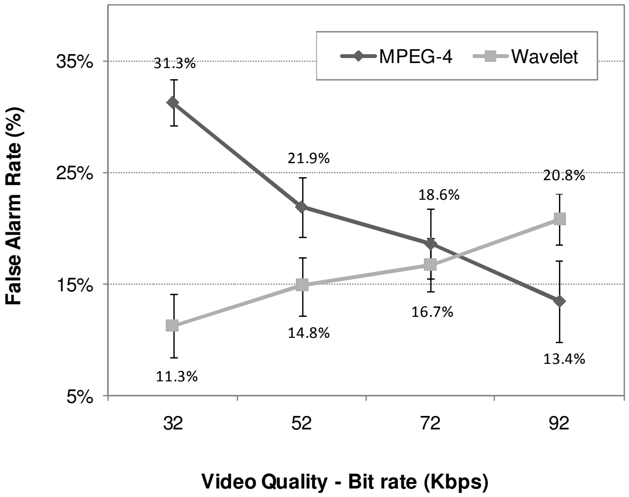


Figure 8.7: False alarm rate with an increase in video quality.

The analysis of H1 was conducted using a repeated measures analysis of variance (ANOVA) test. Shapiro-Wilk W (test of sample normalility) and Levene's test for homogeneity of variance test indicated that the data was normally distributed and exhibited homogeneity of variance. A two-way ANOVA was considered suitable, as the dependent variable measures (task performance and confidence ratings) were repeated across participants independently. There is no non-parametric equivalent test.

The ANOVA test showed that:

- A within-subjects factor had a significant effect on face identification performance [F (3, 237) = 16.50, $p < 0.001$]. Therefore, as CCTV image quality increased, the average number of correct identifications also increased.
- There was no significant improvement in face identification performance with one video CODEC over another, [F (1, 79) = 0.004, $p = 0.9530$].
- There was a significant interaction effect between video bit rate and video CODEC, [F (3, 237) = 5.37, $p < 0.001$].

To examine the ANOVA results further, a pair-wise comparison between the video bit rate conditions were made. Paired t-tests were calculated between the video bit rate levels for each video CODEC. To maintain the overall Type I error rate (α) across all comparisons at 0.05, a Bonferroni-corrected significance level for a two-tailed test of 0.008[27] was used. Table 8.5 provides a summary of the pair-wise comparisons, the p values and the significance for each. The test showed that:

- For MPEG-4 encoded video, there was no significant increase in identification performance as image quality increased from 32 to 52 Kbps.
- There was, however, a significant increase in the number of correct identifications between MPEG-4 video quality levels: 52 Kbps and 72 Kbps [t (79) = 3.33, p = 0.001], and between 72 Kbps and 92 Kbps [t (79) = 2.81, p = 0.006].
- The differences in the average number of correct identifications between the Wavelet conditions were *not significant.*
- The pair-wise comparisons between the conditions revealed that the significant ANOVA result previously found with video bit rate and the interaction effect was true.
- The comparisons, however, revealed another finding: face identification performance significantly improved beyond 52 Kbps (with MPEG-4 encoded CCTV); this improvement was not found at all with Wavelet.

[27] To establish the significant difference between conditions, 6 pair wise comparisons was required (0.05/6 = p = 0.008).

Table 8.5: Pair-wise Comparison Results for MPEG-4 and Wavelet Conditions

Video Condition	Comparisons Between Levels	Pair-wise comparison result	Significance
MPEG-4 CODEC	32-52 Kbps	*t* (79) = .305, *p* = .761	Not significant
	52–72 Kbps	***t* (79) = 3.33, *p* = .001**	**Significant**
	72–92 Kbps	***t* (79) = 2.81, *p* = .006**	**Significant**
Wavelet CODEC	32–52 Kbps	*t* (79) = 2.45, *p* = .016	Not significant
	52–72 Kbps	*t* (79) = .207, *p* = .836	Not significant
	72–92 Kbps	*t* (79) = .508, *p* = 613	Not significant

An ANOVA with video bit rate as a within subjects factor showed that there was a significant difference with false alarms between the bit rate levels [*F* (3, 237) = 3.97, $p<0.001$], but there was no significant difference when comparing false alarm rates between the two video CODECs. There was however, a significant interaction effect between video bit rate and video CODEC, [*F* (3,237) = 6.59, $p<0.001$].

To establish the actual significance of the changes with false alarms, paired t-tests between the video bit rate conditions and video CODECs were also made. The analysis showed that as MPEG-4 video quality increased from 32 to 52 Kbps, the false alarms significantly decreased [t (79) = 3.41, $p<0.008$]. There was, however no significance decrease in the false alarms when video quality increased beyond 52 Kbps (MPEG-4). There was no significant change in the false alarms as Wavelet quality decreased.

8.4.1.1 Signal Detection Analysis

The study design adopted a psychophysical methodology based on signal detection theory (SDT) to determine how sensitive participants were when asked to identify a real target (signal present condition) from a look-alike target (signal absent condition). The number of hits (correct matches) and false positives (false alarms) were converted into standardised *Z* scores[28] and subtracted from one another to give the discriminability index (*d'*):

$$d' = z_{\text{hits}} - z_{\text{false alarms}}$$

[28] A Z-score quantifies the original score in terms of the number of standard deviations that the score is from the mean of the distribution. Z score is calculated as the average hit rate – mean score / standard deviation of score.

d' is a statistical measure used in signal detection theory and is a measure of the difference between the hit and false alarm rate. The *d*' values were calculated so that it was possible to statistically determine how well participants were able to discriminate between a signal stimulus (correct face) and a noise stimulus (look-alike face). The larger the *d*' value, the better the observer was at making this discrimination. The *d'* values for the bit rate and video CODEC quality conditions in Figure 8.8 show that as MPEG-4 CCTV image quality increases, face discrimination also increased. For Wavelet images, the sensitivity between correct and look-alike faces showed the opposite effect; whereby, as the quality of Wavelet CCTV images increased, face discrimination decreased, but only slightly as quality increased from 32 Kbps to 92 Kbps. Measuring the observer's ability to discriminate a correct target from a look-alike target in low-quality CCTV images is important, as it determines the reliability of CCTV images lowered to given quality levels (and video CODEC type) for a face identification task.

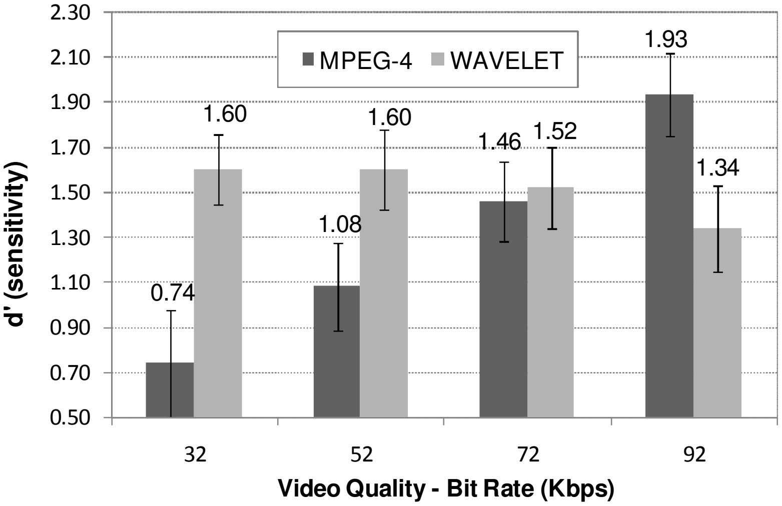


Figure 8.8: *d'* values for MPEG-4 and Wavelet video quality conditions.

An ANOVA test for discrimination values (*d'*) across conditions and levels showed that there was:

- a significant difference in *d'* as MPEG-4 video quality increased [F (3, 237) = 7.54, $p <$ 0.001],
- no significant change in *d'* between MPEG-4 and Wavelet CCTV as video quality increased, and
- no significant change in *d'* between the Wavelet video quality conditions.

The locus of false alarm and hit pairs for the two video CODEC and video bit rate conditions was plotted as ROC curves (see Figure 8.9 and Figure 8.10). The plots show ROC curves for each of the conditions (MPEG-4 and Wavelet) in the experiment. These were calculated assuming that both the underlying 'signal' and 'noise' distributions were Gaussian distributed with equal variance.

In Figure 8.9, the curves show that as MPEG-4 video quality increases, the observer's sensitivity in discriminating between a correct and look-alike target from CCTV images increased. At 32 Kbps, the ROC curve furthest away from the upper left corner indicates that the CCTV images at this video quality level is unreliable for the face identification task.

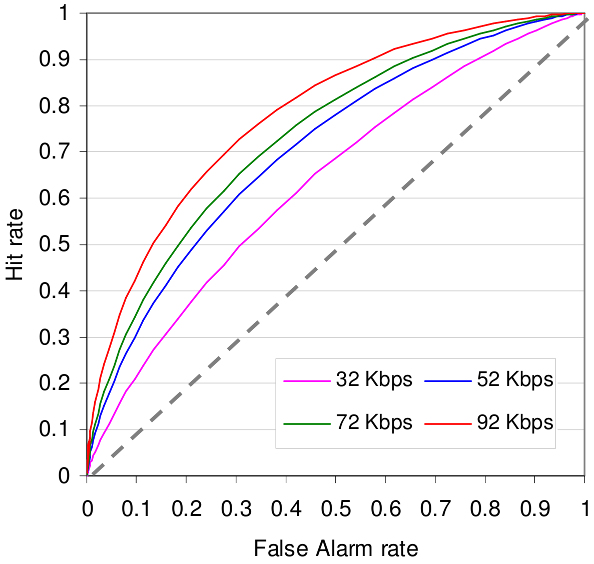


Figure 8.9: ROC graph for MPEG-4 conditions.

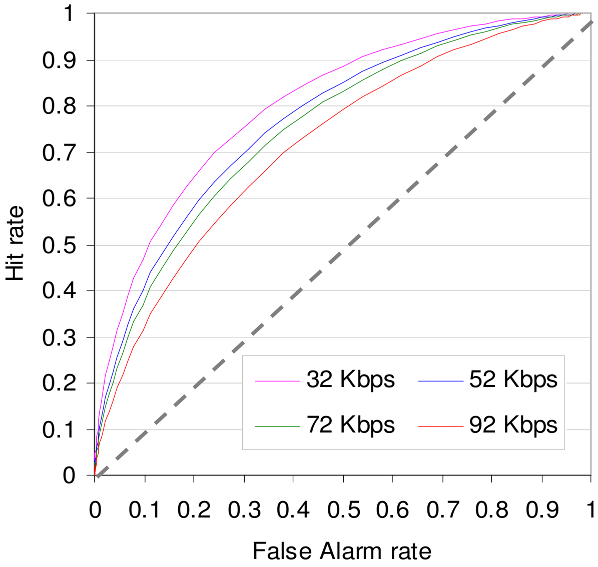


Figure 8.10: ROC graph for Wavelet conditions.

The ROC curves in Figure 8.9 and Figure 8.10 illustrate the finding that participants did not perform very well when identifying targets from Wavelet CCTV images compared with MPEG-4 CCTV images. Task performance increased from 32 to 52 Kbps, and then remained steady as quality increased 52 to 92 Kbps. The average number of false alarms unexpectedly increased as Wavelet video quality increased. This finding is also illustrated in the ROC curves which show the reverse discrimination curves for the MPEG-4 and Wavelet conditions. These results, therefore, suggest that observers (untrained) will have great difficulty in identifying unfamiliar faces from very low-quality CCTV images – particularly from Wavelet CCTV images. These results shed light in the way in which real-world CCTV observation tasks are carried out and highlight the need to update existing policy recommendations on CCTV systems and the use of digital CCTV for security tasks. The performance results also highlight the dangers in recruiting untrained CCTV users to perform security tasks that carry high risks if errors are committed (e.g., wrongly identifying an innocent target as a criminal). Particularly when CCTV video quality is seriously compromised (e.g., compressed to 52 Kbps or less).

8.4.2 Confidence in Identification

The confidence ratings averaged around 2.81 across all MPEG-4 conditions and 2.79 across all Wavelet conditions; therefore, there was very little difference between samples to warrant further statistical analysis. After completing the task, some participants commented that overall they were **very unsure of their performance** and felt they had no choice but to give a rating of 3 for almost all responses.

8.4.3 Performance and Subjective Responses: Ethnicity

Following task completion, participants were asked to complete a post-experiment questionnaire. The first question (1a) asked participants to rank the target ethnicities in order of identification difficulty. A ranking of 1 = very easy to identify and a ranking of 4 = very difficult to identify. The average rankings per ethnicity group across participants were calculated and these values are given in Table 8.6. The rankings show that:

- A ranking of 1 was given (easiest target ethnicity group) for White-Caucasian targets (38%) and Oriental targets (30%).
- A ranking of 3 was given (second most difficult target ethnic group) for Indian-Asian targets (49%).
- A ranking of 4 was given (most difficult target ethnicity group) for Afro-Caribbean targets (66%).

Table 8.6: Average Difficulty Rankings Across Ethnicity Groups

Ranking	Target Ethnicity Groups			
	Afro Caribbean	Oriental	Indian Asian	White Caucasian
1 (Very Easy)	5%	30%	8%	38%
2	4%	25%	17%	34%
3	6%	21%	49%	4%
4 (Very Difficult)	66%	4%	6%	4%

A Friedman test on the ranking data showed that the rankings between target ethnicity groups were significantly different from each other [$\chi2$ (3) = 31.41, $p < 0.001$].

Question 1b asked participants to explain their choices in rankings in question 1a. Comments from this question were analysed firstly by identifying categories and then marking the frequency with which comments fell into each category (see Table 8.7). From the analysis, two categories were identified:

1. Afro-Caribbean and Indian-Asian faces were very difficult to identify (73%).
2. White-Caucasian and Oriental faces were very easy to identify (63%).

The responses to question 1b also revealed that 62% of the participants found it easier to identify faces that belonged to the same ethnicity group as their own; some of the comments which supported this finding include:

- *"…darkness of the skin makes it more difficult to see traits, and because I am white, I might also be more familiar with white faces."*
- *"As I am of a white origin I am use to seeing same face colour through school/family experiences."*
- *"I am Oriental, and I guess I can associate the facial features and pick up little differences better with my own ethnic group."*
- *"I think I recognise my own ethnic group, Oriental a bit better."*
- *"I probably found Oriental faces easier to spot (compared to White-Caucasian) because I come from an Oriental community!"*

Table 8.7: Subjective Comments on Difficulty in Identifying Ethnically Different Targets

Participant's Comments	
CATEGORY 1	*"In the blurred images – the facial features of Afro-Caribbean faces were very difficult to distinguish."*
	"Afro-Caribbean faces looked like silhouettes with no obvious features; this was true in some cases with Indian-Asian faces too."
	"Darker skin tones tend to become black-brown blurs on the CCTV images."
	"It was difficult to recognise the distinctiveness of each of the black faces on screen."
	"Darker faces were impossible as there is no light reflected off their faces like other faces."
	"Couldn't really see the Afro-Caribbean faces properly. Dark skinned and unclear pictures make it difficult to identify dark-skinned people on screen."
CATEGORY 2	*"…the oriental faces seemed to have better defined features, face shape, etc."*
	"It seemed that the White-Caucasian faces had more distinguished features."
	"The features of the Oriental faces seemed to be more distinct, especially the noses and length between eyes."
	"…you could see the features of the lighter coloured skin better. It was easier to see the different shading on their face."
	"The oriental eyes and features seem to be more expressive making it easier to read."
	"I probably found Oriental faces easier to spot (compared to White-Caucasian) because I come from an Oriental community!"

To establish how well participants were able to discriminate targets belonging to different ethnicities, *d'* was calculated for each target ethnicity group. Figure 8.11 shows that participants were better at discriminating between real targets and look-alike targets in the following order of ethnicity:

1. Orientals
2. Indian-Asians
3. White-Caucasians
4. Afro-Caribbeans

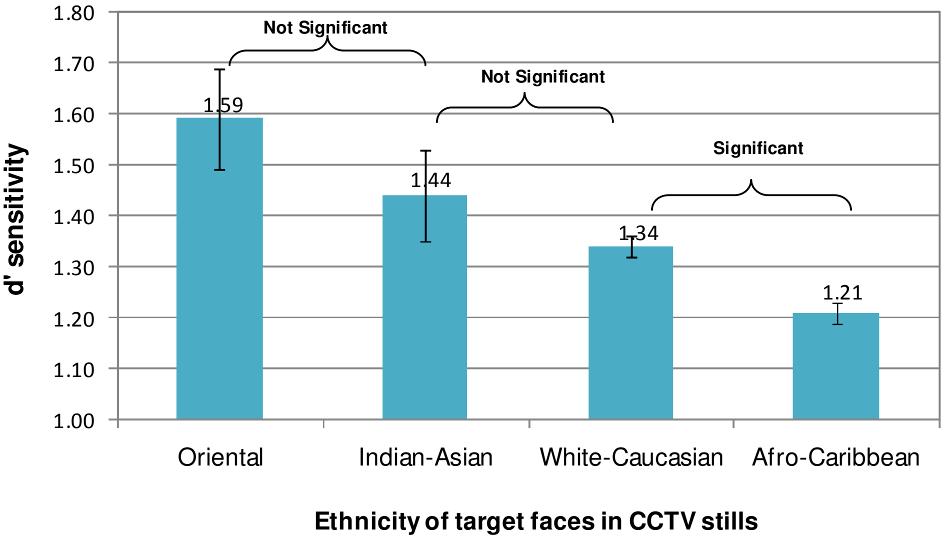


Figure 8.11: *d'* performance in identifying faces per target ethnic group.

To establish whether the *d'* in Figure 8.11 values were significantly different between the target ethnic groups; six paired *t*-tests between the target ethnicity groups were completed using a 0.008 Bonferroni-corrected significance level for a two-tailed test. The tests showed that *d'* was significantly different between:

- Oriental and Afro-Caribbean targets [*t* (79) = 8.60, *p* < 0.001].
- Indian-Asian and Afro-Caribbean targets [*t* (79) = 8.00, *p* < 0.001].
- Oriental targets and White-Caucasian targets [*t* (79) = 2.87, *p* < 0.001].
- There was no significant difference in *d'* between Oriental and Indian-Asian, and between the Indian Asian and White Caucasians.

No analysis was made on the hits and false alarms for participant ethnicity (as the participant's ethnicity was not controlled in this study). Overall, the significant results for target ethnicity showed that participants do perceive differences between different target ethnicities, and the ability to identify targets from digital CCTV (taken from a mixed ethnic population) is affected as video quality is lowered through video compression. Furthermore, the results show that some ethnic groups are harder to identify than others.

8.4.4 Performance and Subjective Responses: Gender

In the post-experiment questionnaire, question (2) asked participants whether they found male or females easier to identify (or neither). The responses showed that:

- Females are easy to identify (44%).
- Males are easy to identify (31%).

- Neither male nor female are easy to identify (25%).

The common reasons given in relation to target gender are given in Table 8.8. The subjective comments made by participants with regard to task difficulty by target gender were examined, and it is clear that hair and head shape were commonly being used to identify targets from the CCTV images.

Table 8.8: Reasons for Identifying Targets Across Gender Type

	REASONS FOR TARGET CHOICE
FEMALES	*"Females usually have long **hair** so they're easier to recognise. Females also have feminine characteristics which also help in recognition."* *"Female **hairstyles** help and the shape of their faces."* *"Women with long **hair** are easier to distinguish. There were also some women with earrings who were easier to recognise."* *"For females, the differences created by changing **hairstyles** introduced some doubts in identification. But, it made it easier for me to rely on other features like face shape and ears."*
MALES	*"Maybe easier to identify a man with **facial hair**, if he was bald or has a particular head shape."* *"Males are easy to identify because of their **hairline** or jaw line, which tends to be less pronounced in females."* *"Males, because of various **head shapes**."* *"Men were easier to spot if you went on the basis of their **head shapes**."* *"It was sometimes easier to identify the men if you went on the basis of their **moustache**, the parting of their **hair** and **hair length**."*

An analysis was conducted to determine whether there was an actual significant difference in target gender preference. A 2x2 Chi-Square test revealed a significant relationship (very close to $p = 0.05$) between target gender preference and participant gender [$\chi 2$ (2, N=60) = 5.99, $p = 0.041$]. Performance in identification per target type between participant gender groups was very close:

- Female participants correctly identified 22.3% male targets and 21.7% female targets.
- Male participants correctly identified 22.3% male targets and 21.47% female targets.

8.4.5 Identification Cues

Question 3 asked participants to state the specific features they used to help them identify the targets from the CCTV images. The frequency with which feature was used in the task is summarised in Table 8.9. The responses show that 14 different features were used. Subjectively, participants were mainly using internal face features: eyes, lips/mouth, nose, eyebrows, and face shape. Despite this, the most common non-face feature used to aid identification was hair. Hair was frequently mentioned when participants were asked about gender difficulty (see Table 8.8).

Table 8.9: Frequency of Features Used for Face Identification Task

	Feature	Frequency (%)
Internal Face Feature	**Eyes**	**51**
	Mouth/Lips	**44**
	Nose	**34**
	Eyebrows	**28**
	Face shape	**20**
	Facial hair	14
	Distance between eyes	8
External Face Feature	**Hair**	**53**
	Skin colour	4
	Hair line	3
	Glasses	3
	Jewellery	3

The final post-experiment question (question 4) asked participants to state whether they found the task easy or difficult. Overall, a majority of participants found the task difficult (94%).

Participants remarked on the task difficulty:

- *“Made me aware how easy it is to be wrong when trying to identify someone from a CCTV. There is a lot of scope for error.”*
- *“I wasn't very confident about any of my decisions. It was a really difficult task for me.”*
- *“In the vast majority of pictures, if I was a police officer, I would definitely hesitate to say for certain that I had correctly identified someone.”*
- *“The resolution of CCTV images should definitely be improved to catch the guilty.”*

8.5 Discussion

8.5.1 Objective and Subjective Findings – Implications for CCTV Tasks

The study results partially support H1: As the quality of CCTV video is lowered through spatial compression, the observer’s ability to identify faces from CCTV images will decrease. The results showed that as MPEG-4 CCTV video quality increased from 52 to 92 Kbps, the number of correct detections increased significantly. The number of false alarms also significantly increased, but not beyond 52 Kbps. This result was not found with Wavelet CCTV video. Thus, H2 was accepted, to some extent: As the quality of CCTV video is lowered through spatial compression, identification performance will be higher with MPEG-4 compressed CCTV images than Wavelet. This predictive hypothesis was formulated and tested objectively through the use of task performance measures following anecdotal claims by CCTV video and imaging experts that, *"Wavelet video encoded at the same bit rate as MPEG-4 video is perceived as better quality"* (Walker and Cohen, 2006). Although a significant result was not found with Wavelet CCTV, further experimentation is needed with other types of Wavelet CODECs, since there is more than one type of Wavelet CODEC. Furthermore, Wavelet CODECs have not been established as a standard. These results, particularly in relation to the Wavelet conditions, imply that CCTV practitioners and owners should be cautious when configuring their CCTV recording and streaming systems to very low-quality levels.

In addition to the impact of video quality, the results revealed another factor which affects an observer's ability to identify targets from low-quality CCTV: target ethnicity. Consistent with previous research (Malpass and Kravitz, 1969; Rehnman and Herlitz, 2006; and Kassin, Tubb, Hosch and Memon, 2001), participants performed better when identifying targets who shared the same ethnicity as their own. This is known as the 'other-race effect' whereby people are better at recognising faces of their own ethnicity as they are more familiar with them through everyday social interactions (Elliot, Willis and Goldstein, 1973). This finding reinforces that the other-race effect still exists, an effect which has been previously identified in a number psychology experiments (e.g., Sheperd, Deregowski and Ellis, 1974; Chiroro and Valentine, 1995; Chance and Goldstein, 1996).

O'Toole, Jiang, Roark and Abdi (2006) explained that darker skinned faces are difficult to process as there are *"large scale structural and reflectance differences that exist between different faces of different races"* (p 299). In this study, an analysis of the objective and subjective participant responses confirmed this finding. Afro-Caribbean targets were indeed the most difficult ethnicity to identify in comparison to other ethnicities. Despite this finding, there were some inconsistencies between the performance and subjective responses for the difficulty in identifying targets of certain ethnicities from degraded CCTV images. For instance, participants ranked White-Caucasian targets the easiest to identify and Afro-Caribbeans the most difficult, yet performance in the task was the highest with Oriental targets. This result may have been due to participants being too confident when identifying lighter skinned targets from CCTV images (White-Caucasian and Orientals), as a result of their perceived difficulty. Furthermore, Indian-Asian targets were ranked as the second most difficult to identify, yet participants were the *least* sensitive when discriminating between White-Caucasian and Afro-Caribbean targets (see Figure 8.12) and not Indian-Asians. It is likely that this over-confidence led to a larger number of false alarm errors when identifying White-Caucasians at the lower video quality conditions.

In terms of the features used for target identification, hair was the most reported feature used. Hair is a highly unreliable feature for a face identification task (particularly with low-quality CCTV video), as it is often used to disguise one's identity which can misguide the observer's judgement quite easily (Foley and Foley, 1998). Hair (both head and facial hair) is rarely used by expert CCTV users. Therefore, non-expert CCTV users should not use hair as a primary cue when identifying unknown targets from CCTV video and images - regardless of the level of video quality. The most reported internal face features used included eyes, mouth/lips, and nose. These features were also identified as important features by other researchers when observers processed face information (e.g. see Rimell, Keval, Mansfield and Hands, 2005 and Laughery, Alexander and Lane, 1971).

Amongst adults, there is an inner face advantage for the recognition of familiar over unfamiliar faces (Ellis, Shepherd, and Davies, 1979; Young, Hay, McWeeny, Flude, and Ellis 1985). However, prior eye-movement studies have found that fixations predominantly fall within the inner face region in face recognition and matching tasks, irrespective of familiarity (Stacey, Walker, and Underwood, 2005; Walker-Smith, Gale, and Findlay, 1977). These eye-tracking studies confirm the subjective findings in this study, however further research is needed to objectively determine the observers' area of interest when identifying faces unknown to them from low-quality CCTV video.

In this face identification study, there were many female participants (*n* = 46) in comparison to male participants (*n*=34). Overall, participants perceived female targets to be the easiest to identify from CCTV images. This finding was consistent to a previous face processing study (Lewin and Herlitz, 2002). The objective findings however did not reveal a significant difference between task performance for male and female participants. If an equal number of female and male participants were recruited, it would have been possible to compare performance by target gender and participant gender. The results could have then been used to highlight the (potential) increased risk in recruiting female only or male only untrained CCTV users for security observation tasks using low-quality CCT video.

In review of these findings, it can be concluded that in addition to low-quality CCTV, target ethnicity is the main factor which can affect the observer's perceptual process when identifying unknown targets. Although previous research in psychology has already established that: 1) humans are poor at identifying unknown targets from video and images (Burton et al, 1999; Bruce et al, 1999; Henderson et al, 2001) and 2) people are better at identifying faces which share the same ethnicity as their own, these findings highlight the importance of utilising high-quality and usable CCTV video/ images for an identification task particularly when used by untrained CCTV users. These findings are particularly relevant to digital and networked CCTV systems deployed within ethnically diverse populations. In addition, greater attention is needed when configuring the video quality of a CCTV system which will be used by untrained CCTV users. To aid this process, it is recommended that CCTV owners follow the recommendations on CCTV video quality which is presented in the TEC-VIS framework (see Chapter 10, Section 10.2.4).

8.5.2 Review of Study 2 in Relation to Previous Research

Prior to this study, Aldridge (1994) stated in the UK Home Office Operational Requirements that, *"there are no definitive performance criteria for video to be legally admissible and it is up to the courts whether the pictures are acceptable or not as identified earlier."* This suggests that it is simply up to the criminal prosecution service to decide whether CCTV evidence brought to the courtroom is good enough to be used as evidence supporting criminal cases. If so, then what about CCTV video and images used for detecting and investigating crime? As discussed in Chapter 5, existing policy recommendations and expert guidelines on CCTV video quality are insufficient, as they mainly apply to analogue CCTV systems (Aldridge, 1994; Data Protection Act, 1998). Current guidance which addresses the video quality and storage issues for digital CCTV systems are too vague and are inconsistent to one another (see Morton, 2004; Cohen, 2007; Data Protection Act, 1998).

To develop objective and easy to apply CCTV video quality recommendations for CCTV observation tasks such as face identification, an empirical study was conducted with 80 participants. This study specifically examined how capable and reliable untrained CCTV users are in performing a face identification task. These users were specifically chosen, as there are an increasing number of untrained CCTV users being recruited to perform CCTV tasks. Furthermore, by assessing these types of users it is possible to recommend a conservative video quality rate which can be applied by all types of CCTV users. This study was conducted using task-oriented measures to identify the minimum video should be compressed (spatially) and also to establish the most effective video CODEC, MPEG-4 or Wavelet, for a face identification task.

In addition to providing video quality guidance, this research study builds upon previous face recognition research (Burton et al., 1999; Bruce et al., 1999; Henderson et al, 2001) and examines the practical constraints in using digital CCTV video for a common security observation task, face identification. The research conducted in this chapter is an important contribution to the field of human centered security and CCTV. It provides a paradigm for future CCTV experiments with users, and it provides CCTV owners with a recommendation on video quality requirements for a face identification task. This recommendation is useful when configuring a digital or networked CCTV system, and is detailed within the TEC-VIS best-practice framework in Chapter 10.

8.6 Conclusions

Study 2 investigated the research goal 2 (see Chapter 1, Section 1.4.2). This experiment empirically investigated the effect of lowering CCTV video quality on task performance. This investigation was conducted as the use of low-quality digital CCTV caused difficulties in identification tasks for both CCTV control room operators and police staff in Study 1 (see Chapter 4).

8.6.1 Substantive Conclusions

1. MPEG-4 CCTV systems are suitable for recording and streaming CCTV video when used for a face identification task using real-time and post-event recorded CCTV video.
2. MPEG-4 CCTV systems should be configured at a bit rate of 52 Kbps or above for a face identification task involving the identification of unknown targets.
3. Wavelet video CODECs are not yet a standard; thus, further testing is required with different types of Wavelet CODECs to determine the most effective video compression level needed for a face identification task.
4. Observers are better at identifying targets who share the same ethnicity as their own. This target bias is known as the ‘other-face effect’ which has been previously found by

psychologists. The findings from Study 2 highlight that CCTV owners should be aware of potential target ethnicity bias when recruiting untrained CCTV users, particularly when using a low-cost digital CCTV system.

5. Following on from point 4, CCTV users untrained in face identification performed worst when identifying dark-skinned targets from low-quality CCTV images. The UK population is becoming ethnically diverse; thus, CCTV owners should ensure that their CCTV system camera environment is optimally configured so that darker skinned targets are visible when captured on CCTV video. This can be achieved by positioning CCTV cameras in well lit open spaces and/or providing additional lighting.

6. CCTV system owners who recruit untrained CCTV users should consider providing training and education on the relevant security observation tasks to ensure that observers do not commit errors, unnecessary bias or use unreliable cues when identifying criminals.

8.6.2 Methodological Conclusions

A number of strengths were revealed in this study:

1. Every effort was made to control the many different variables in this study (video camera settings, target and environmental factors). These variables were controlled purposely to create a standardised set of CCTV images in order to focus the evaluation on target identification performance when video quality was lowered. Furthermore, this study promotes an easy to replicate methodology.

2. As well as controlling the recording environment variables, the individuals appearing in the CCTV images were taken from four different ethnicity groups (and an equal number of females and males). Obtaining a cross-section of demographics was necessary to assess face identification performance with targets who were representative of people captured on everyday CCTV video. The targets chosen for the CCTV images, therefore, strengthened the external validity and reliability of the research findings.

3. A large number of participants were recruited for the experiment (*n*=80), which means that the data obtained from both the objective and subjective participant responses have high statistical power.

There were also a number of limitations in this study:

1. An analysis of the performance data across the 64 targets revealed that two targets were ‘too easy to identify’ regardless of video quality. These targets were visually distinctive and atypical from the other targets. For example, one easy to identify target was an Indian-Asian male target wore a black turban (100% performance was achieved with this target). Another easy to identify target was an Oriental female who had very large artificial looking lips. It is likely that these two targets were easy to identify since they were atypical. Light, Kayra-Stuart and Hollander (1979) identified this type of face bias and noted that faces which are rated as typical are not recognised as well as faces rated as atypical.

2. Different versions of MPEG-4 and Wavelet video CODECs exist and some versions will produce different video quality results. It was not possible to assess face identification performance with different versions, as there are several, and the number of conditions would be too high to counterbalance and would be too great for users to perform. There is of course an opportunity in the future to repeat Study 2 using different video CODECs and versions to determine the effect of each type on face identification performance and confidence.

3. The walking mask method used for the production of the CCTV stimuli was chosen to standardise the CCTV images. Whilst the walking mask method gives the impression that the CCTV video images were not real, it was the best compromise given the difficulties associated with varying views and expressions. It was assumed that the exposure and printing of the masks produced a representative gray scale with the correct gamma and illumination. Shadow detail from the CCTV images may have potentially influenced face identification, but this is not certain without further evaluation. As the face masks were taken within a well lit environment, a high level of detail was retained in the high-resolution face masks; thus, the appearance of the faces as seen by the camera would not have changed significantly during the walking mask phase. Of course, the level of detail may not have been captured optimally with darker skinned targets. Nevertheless, the walking mask method was considered suitable to test the differences in face identification performance between target ethnicities, as this has not been previously investigated with CCTV images.

Gaining access to realistic CCTV video stimuli can be very difficult, particularly if high-quality uncompressed video is required. To ensure that the data holds high ecological validity, the targets recorded for the video scenes must realistically represent real-world CCTV. Rather than selecting an opportunity sample, typically consisting of White-Caucasian middle aged males, the participant sample for CCTV targets should contain a wide and equal range of ethnicities and ages (and equal gender). For those researchers who wish to assess the effectiveness of CCTV video quality when used for security observation tasks, it is recommended that the data used for the test be high quality and the content be representative of real-world CCTV imagery. Also, a reliable video CODEC should be used for altering video quality (easily manipulated and replicated by other researchers). Variables such as environmental lighting, recording height and angle should also be controlled as best as possible, without factoring in motion blur, which is a very common issue when recording video using a digital camera recorder.

In this study, task-oriented measures (Knoche and de Meer, 1999) were used to measure performance. Classic task performance measures, such as the correct/incorrect identifications and false alarms were straightforward to measure and easy for reporting task performance. It is therefore an ideal method for assessing the effectiveness of video quality with human observers. As well as using task performance measures, a task-oriented method was used to increase the ecological validity of the study even further to take into account of the context, users and the application. By adopting this task-oriented approach, requirements for the task can be appropriately applied to real-world tasks.

Chapter 9

Study 3: Detecting Events with Low-Frame-Rate CCTV Video

In the previous chapter, Study 2 was presented in which the impact of spatial video compression on a face identification task was investigated. Detection tasks are increasingly being performed by untrained CCTV users and there is currently no guidance or training available for them. In this chapter, Study 3 is detailed. This study investigates the effect of *temporal* video compression on an observer's ability to detect events from CCTV video. This study examines how untrained CCTV users perform a detection task under four low temporal video quality conditions and the effect it has on their confidence.

9.1 Introduction

Traditionally, CCTV video and images are used by expert-CCTV users, such as CCTV operators and investigators. CCTV operators carry out their tasks in real-time, while CCTV investigators are involved in the investigation of crime and analysis of post-event recorded CCTV video and imagery. As part of the investigation procedure, an investigator's job is to interpret the CCTV video and images as best they can and present them as supporting evidence in court proceedings for criminal prosecution. Low-quality CCTV video can result in errors. This real-world problem was identified in Study 1, the control room field study (see Chapter 4). This issue was also raised over half a decade ago by Bromby (2002) who claimed that recorded CCTV video is not always of sufficient quality and this has posed problems for criminal investigations in the past.

With the recent developments in digital and networking CCTV systems, CCTV owners (both public and private) are now recruiting untrained and inexperienced CCTV users to carry out security observation tasks over the Internet. There are two problems in recruiting untrained CCTV users. Firstly, these users are not trained in vigilance tasks and lack experience in using security video; thus, there is the risk they may misinterpret events. Secondly, video is delivered over the Internet and not necessarily at broadcast quality, due to the high costs associated with fast data networking speeds. CCTV video streamed across a finite bandwidth may be excessively compressed, which will lead to delivery of low-quality CCTV video for viewing on the observer's video monitor at their remote location. Using low-quality CCTV video for security observation tasks, particularly for detection tasks, is highly undesirable as events may go undetected (Cohen, 2006). Untrained CCTV users are also likely to make biased judgements without understanding the task and the impact of their responses to incidents.

When surveillance video is streamed over a bandwidth-constrained network or played back locally (and recorded at very low frame rates: 1–5 fps), the resulting video is difficult to process. For example, at 1 fps, only 1 out of a maximum of 25 frames is stored per second. At this rate, it is hard to interpret gross movements, such as dropping objects on the floor, large hand signals, and moving cars. It is also difficult to detect fine movements, such as facial expressions (Horn et al., 2002). Events are also difficult to interpret when the speed varies; the faster the action, the more information lost within a single frame.

Within the human centered multimedia research domain, there have been a number of studies conducted which have investigated the impact of low-quality video across a wide range of tasks. These studies are reviewed in Chapter 6. Surprisingly, there have been no studies conducted to examine the performance issues associated with CCTV video applications. There is a need to assess the impact of lowering CCTV video frame rate on task performance to establish recommendations and guidelines that will help CCTV users configure their systems to produce usable CCTV video recordings. An empirical experiment was designed to investigate the effect of lowering the frame rate of CCTV video on detection performance, a key task performed by a wide range of CCTV users.

9.2 Research Motivations

This study was motivated by the same factors identified in Study 2 (see Chapter 8, Section 8.2).

9.3 Research Goals and Hypotheses

The research goals for this experiment are as follows:

1. Assess the impact of lowering the frame rate of CCTV video on a security task performed by untrained CCTV users.
2. Determine the minimum frame rate required for a detection task and incorporate this requirement into the best-practice framework (see Chapter 10).

The two hypotheses addressed by the experiment are as follows:

- Hypothesis 1 (H1): As the frame rate of CCTV video is lowered, the observer's ability to detect events will increase.
- Hypothesis 2 (H2): As the frame rate of CCTV video is lowered, the observer's confidence in the task and their perceived video quality ratings will decrease.

9.3.1 Participants

The detection experiment was conducted with 80 paid participants at the Department of Computer Science, University College London. The participant sample comprised of untrained CCTV users: 23-38 years of age (mean = 29 years, SD = 7.12 years), of which 49 were female. Each participant was paid £5 for participating in the study.

The actors who appeared in the CCTV video footage were not permitted to take part in the detection test. Untrained participants were purposely recruited for the detection experiment to provide an understanding of how reliable untrained users are in detecting events from low-quality CCTV video; a task typically performed with real-time CCTV video or recorded video (eye witness scenario).

The evaluation was conducted to identify the threshold performance level at different frame rates. By assessing untrained participants, it was possible to define a conservative threshold frame rate for effective detection performance. The rate is conservative, as the performance results are based on tests with untrained CCTV users, which means that the rate can be applied to a wide range of CCTV systems used by CCTV users with varying skills.

9.3.2 Materials and Data Collection

As there were no standardised CCTV video data sets available for assessing CCTV video with human observers, mock CCTV video scenes were recorded for the detection experiment. Eight crime scenes and eight non-crime scenes were filmed using a mini-digital video camera recorder. Each scene was filmed for approximately 10 seconds. The duration of the video clips was restricted to 10-12 seconds for one main reason: to ensure that each participant was able to complete the study within a 30-40 minute period and no more. The crime scenes included various theft incidents and drug dealing activities. The non-crime scenes included everyday activities, such as people chatting, reading text messages and picking up keys from the floor. For each scene, three different scenario descriptions were presented to participants and only one scenario was correct. These scenarios are provided in Appendix G.

In filming the mock scenarios, every effort was made to record a variety of crime and non-crime scenes, in addition to capturing actors belonging to different ethnic backgrounds and genders. However, these factors were not factored, as they were in Study 2, since there was no intention to compare the observer's ability to detect events by target gender, ethnicity or scenario type. These variables were not part of the objectives for this study. All of the recordings were made on the university's grounds for the purpose of safety and privacy in filming. The scenarios chosen for the video recordings were deliberately kept short, simple, realistic and representative of everyday non-crime and crime scenes. A group of eight volunteers (untrained actors; 50% male) took part in the filming. The actors were not selected using any particular demographic criteria and were required to sign a consent form allowing their data to be used in the detection experiment and for publications. For each video clip, actors were briefed on the scenes. The set-up variables and characteristics of the video stimuli used in the tests are shown in Table 9.1.

Table 9.1: Recording Set-up for CCTV Stimuli

Scenario set-up for picture users:	Untrained CCTV users
Location of CCTV video recordings:	Outdoors
Timing of CCTV event:	Post-event recorded and real-time mode
Recorded to show:	People interacting with one another/people committing crimes
Purpose of CCTV set-up:	Detection of events on CCTV video
Field of view:	Horizontal
Features used for task:	Movement and actions of individuals (no vehicles)
Distance at which CCTV video captured:	Varied depending on scene
Shutter speed:	1/150
Video resolution (playback):	352 x 288 (CIF)
Presentation form of CCTV data:	Video; length 10–15 seconds

Prior to filming the scenarios, the video camera was set using the following parameters:

- **Automatic Gain Control (AGC): ON**

As most CCTV cameras are designed with this feature turned on (by default), the camera used in the filming has the AGC function was activated. AGC increases the amount of amplification under low-light conditions to bring the video signal up to its minimum required level. AGC also minimises noise and picture attenuation.

- **Shutter Speed: 1/50**

Shutter speed, like AGC, is a feature used to control the amount of light entering the camera. A slow shutter speed is generally used under low-light conditions, allowing a longer length of time for light to enter the camera sensor. At shutter speeds above 1/40, motion blur is not visible. When preparing the video recordings, a shutter speed of 1/50 was chosen to avoid unnecessary motion blur.

Once the scenes were recorded, each recording was edited in segment length and spatial video quality for the different frame rates. The following video quality parameters were chosen:

- **Video resolution** was reduced from its original resolution of 720x576 to 352x288 (CIF resolution), since this is the typical recording and streaming resolution used in most low-end digital CCTV systems.
- **Spatial quality** was reduced, through a process called encoding, at a bit rate of 2000 Kbps using the Microsoft MPEG-4 Video CODEC (Version 2). The CODEC and compression rate were chosen as these are commonly used in many digital CCTV systems when storage is

limited (Cohen et al., 2007). No frames were discarded during video compression. As in Study 2, bit rate was chosen to lower video quality, since it is an easy, replicable and objective method for reducing video quality.

- **Frame quality** was then reduced to the following frame rates: 1 fps, 5 fps, 8 fps, and 12 fps. The lowest level (1 fps) was chosen, as this is what most systems offer as the lowest rate; 5 fps was chosen, as this was the threshold frame rate identified in the video conference and lie detection studies conducted by Wilson and Sasse (2000) and Horn et al. (2002), respectively. Both of these studies involved a user performing an active task processing video containing people. For this reason, this CCTV detection study was purposely designed to examine whether or not 5 fps is suitable for a CCTV detection task. Frame rates 8 and 12 were chosen, as these are the next two equal levels on the frame rate scale (more or less). Also, 12 fps was chosen for the highest video quality condition, as this is the point at which video is generally perceived by humans to flow smoothly. In addition to these justifications, these levels were chosen because they have been identified as the most commonly used recording frame rates (Cohen et al., 2007). Figure 9.1 illustrates the steps taken to prepare the CCTV video stimuli for the detection test.

To validate the CCTV video scenarios, all 32 video clips were assessed in full frame rate by 32 volunteers who were untrained in CCTV tasks. The trial was completed to ascertain task difficulty. An average performance of 98% was obtained for the detection test. Participants reported that they occasionally pressed the wrong key, which led to errors in the task. The video was further validated by two video and imaging experts from the Home Office Scientific Development Branch who work in crime and security video analysis. Both experts rated all of the video scenes as *'realistic and representative of typical CCTV video'* (Cohen, 2006); thus, they were accepted for the CCTV video stimuli for the final detection test. Figure 9.2 shows three images taken from the crime scenario video clips and three taken from scenarios in which no crime takes place.

9.3.3 Independent Variable

The independent variable chosen for this study was video frame rate. This variable was assessed under four frame-rate levels: 1, 5, 8 and 12 fps, as justified in Section 9.3.2.

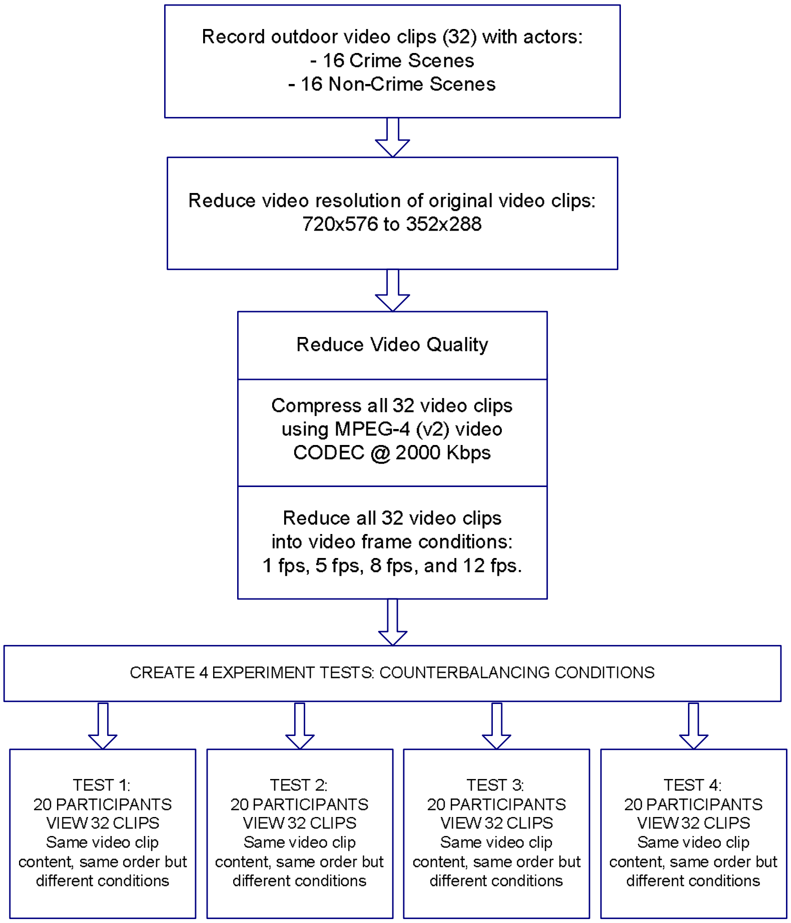
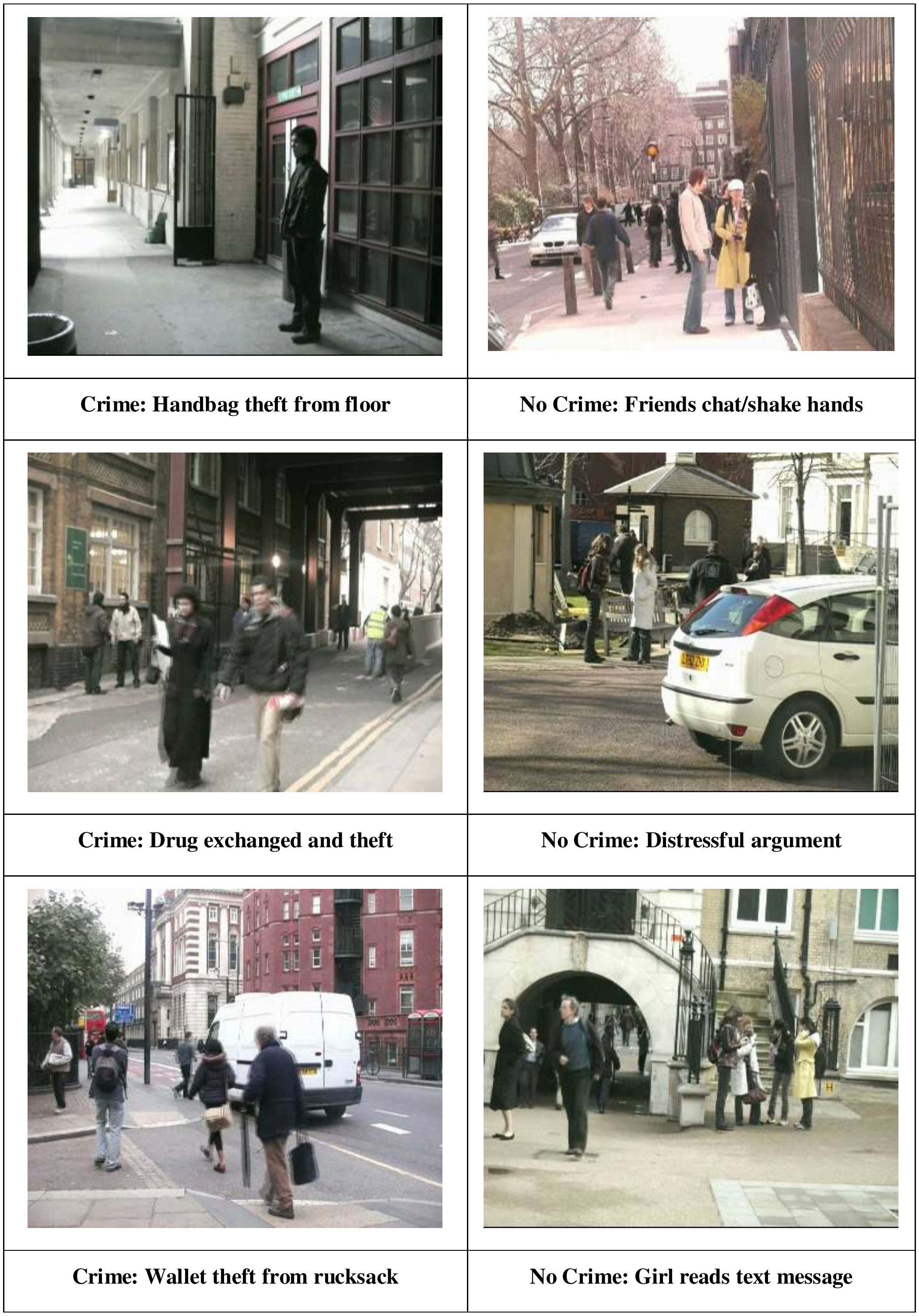


Figure 9.1: Video recording and encoding processes for CCTV stimulus.



Crime: Handbag theft from floor

No Crime: Friends chat/shake hands

Crime: Drug exchanged and theft

No Crime: Distressful argument

Crime: Wallet theft from rucksack

No Crime: Girl reads text message

Figure 9.2: CCTV stills taken from non-crime and crime scenario video clips.

9.3.4 Dependent Variables

Three measures were recorded from each participant:

1. **Task performance**: This is a measure of the participant's ability to correctly detect scenes. Task-oriented measures (Knoche and de Meer, 1999) were applied in this study, whereby each correctly detected scenario was scored 1, and 0 for an incorrect detection. These task performance measures were the same as those used in Study 2, since they take into account the context of the task (detecting events in real-time or with post-event CCTV video), the user performing the task (human observer) and the application being used (digital CCTV system). The reason for using task-oriented measures was so that a real-world task scenario could be replicated within a laboratory setting.

2. **Task confidence**: Participants were required to rate how confident they were in giving their response to each video clip observed. Their confidence in the task was measured by gathering confidence ratings using a 5-point Likert scale (1 – not very confident and 5 – very confident).

3. **Perceived video quality:** Participants were also required to rate the perceived video quality for each video clip observed. This rating was made using a 5-point Likert scale (1 – very bad and 5 – very good).

9.3.5 Design

A 1 x 4 within-subjects experimental design was adopted. The video frame rates were counterbalanced appropriately to avoid order and practice effects; this was achieved by not showing participants video clips with the same frame rate consecutively. To overcome video content bias, each of the 32 video clips were encoded into all four frame rates, which created four counterbalanced tests (consisting of 32 video clips each). The video scenarios were all encoded differently. The four separate tests were prepared to rotate the studied conditions[29] across participants. The ordering of the conditions for each of the four experimental tests was achieved by applying a 4x4 Graeco-Latin square design. The detection task was designed in this way to determine detection performance for the different frame rates.

Once the stimulus was prepared, the video clips were inserted into the study's program, which played as soon as the participant was ready to begin. After each video clip preview, the participant was required to choose the scenario which matched the video scene. Participants were specifically asked to pick one likely scenario from three, instead of responding to an open-ended prompt. Participants were provided with multiple choices (x3) so the performance in the detection task could be measured objectively and a threshold rate could be established. If subjective responses were gathered from participants for each scenario, the analysis and scoring of each response would have been too difficult

[29] 1 factor = Video frame rate and 4 levels = 1, 5, 8 and 12 fps = 1 x 4 = 4.

to determine detection performance across the frame rates. Furthermore, scoring qualitative responses is open to interpretation and subject to manual error.

9.3.6 Procedure

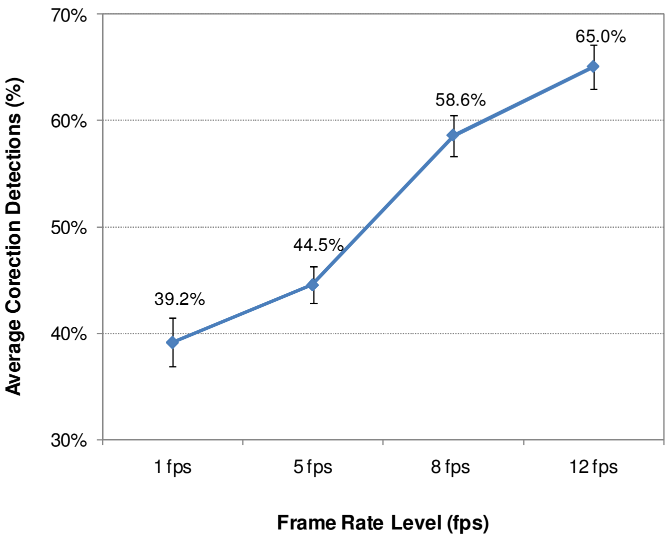
On arrival, participants were verbally briefed about the experiment and asked to read and sign the detection study's instructions. Two sample video clips were then shown for practice, each containing three scenarios. The first example was a non-crime video clip, and the second was a crime clip. Participants were seated in an adjustable chair facing a 21-inch PC monitor positioned at the recommended eye-to-screen distance of 25 inches. The viewing area of the monitor was set at approximately 15° below eye level. Once participants clearly understood the instructions, they were asked to press 'enter' on the PC to commence the detection test. Once the video clip played, participants were given the opportunity to view the video clip again; they were not permitted to view the clip after the second viewing. Participants were limited to 30-minutes to complete the entire detection test. In this period, participants were asked to read the instructions, practice the test with two video clip scenes, and then complete the test with 32 video scenes. After each clip was viewed and participants had read three short scenarios and were required to pick one scenario that best described the video clip they saw. Participants were then asked to rate their confidence in their response and the perceived video quality using two different scales (see Section 9.3.3). Once the test was complete, participants were asked two post-experiment questions:

1. Did you find the task difficult? If a participant responded 'yes', they were asked question 2.
2. What aspect of the task did you feel affected your performance?

9.4 Results

9.4.1 Task Performance

The total correct detections were calculated, averaged across participants and then categorised by frame rate. Figure 9.3 shows that the average number of correct detections increased by almost a quarter as the video frame rate increased from 1 to 12 fps.



N.B. The error bars in this graph represent the standard error.

Figure 9.3: Average task performance across frame rates.

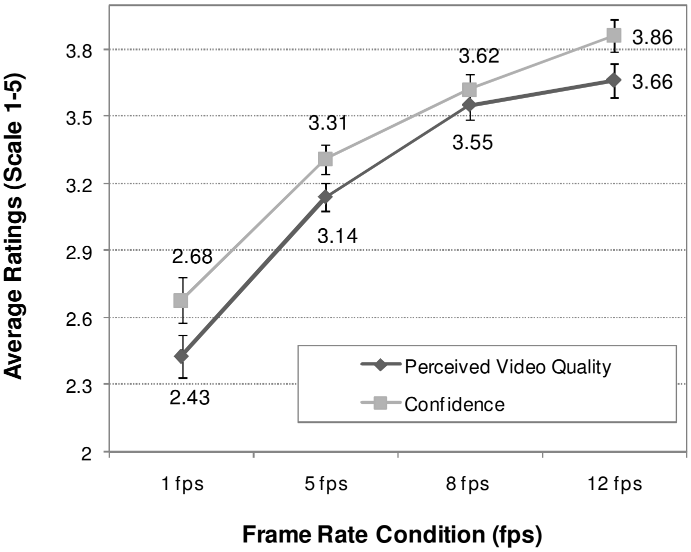
The analysis of H1 and H2 was conducted using a non-parametric Friedman test, a statistical test used to compare observations gathered on a single group of participants to detect differences in three or more groups of related samples. A non-parametric test was used to analyse the results, as the task performance data was not normally distributed and the data within the x-axis (in Figure 9.3) was ordinal. Unlike the parametric repeated measures of an ANOVA or paired *t*-test, a non-parametric Friedman test makes no assumptions about the distribution of the data (e.g., normality); thus, this test was chosen to provide a cautious analysis.

Prior to this analysis, Spearman rank-order correlations were performed for each unique pair-wise combination of the three dependent variables measured in this study (task performance, task confidence and perceived video quality). This correlation test was necessary to identify whether or not these three variables would be treated as related measures during data analysis to ensure that the data would not be over-tested. No strong correlations were found between the dependent variables using the Spearman rank-order correlation test; therefore, each measure was tested independently using the non-parametric Friedman test.

A Friedman rank test showed that an observer's ability to detect events on CCTV video increased significantly as video frame rate increased [$\chi 2$ (3) = 63.4, $p < 0.001$]. A follow-up test was carried out to determine the actual differences in performance between the video frame rates. The results showed a significant difference in detection performance between frame rate levels 1 fps and 5 fps, 5 fps and 8 fps ($p<0.0001$), but not between 8 fps and 12 fps ($p = 0.294$).

9.4.2 Ratings

The task confidence and perceived video quality ratings were averaged across participants and for each frame rate to determine the average ratings. Figure 9.4 shows that as video frame rate increases from 1 to 12 fps, both confidence and the perceived video quality ratings rose. At 8 fps, the ratings were starting to reach a plateau. These subjective findings support the insignificant result in detection performance between 8 and 12 fps.



N.B. The error bars in this graph represent the standard error.

Figure 9.4: Perceived video quality and confidence ratings across frame rates.

A Friedman test showed that the confidence ratings were significant as CCTV frame rate increased [χ2 (3) = 63.4, p<0.001] and comparisons also revealed significant differences between all the frame rate levels. The confidence ratings were also found significant [χ2 (3) = 109.8, p <0.001] however, individual comparisons revealed no significance between the perceived video quality ratings from 8 to 12 fps (p =0.818).

9.5 Discussion

9.5.1 Subjective and Objective Findings – Implications for CCTV Tasks

The detection experiment presented in this chapter investigated the effect of temporal video compression on a CCTV detection task performed by untrained human observers. The results from this study are in line with both of the study hypotheses:

Hypothesis 1 (H1) was accepted: "As the frame rate of CCTV video is increased, an observer's ability to detect events will increase." Although the average performance for the detection task increased as the video frame rate increased, this increase was only significant when the frame rate was increased from 1 fps to 8 fps - this finding was unexpected. Although there was a significant difference found between 1-8 fps, this study furthers the work by Horn (2002) – in that detection performance is actually worst at 1 fps and not 5 fps. Furthermore, the study findings reveal that effective detection with CCTV video is achieved at a threshold frame rate of 8 fps. Further evaluations would be needed to verify this claim for sure. These objective task performance results imply that CCTV owners can save a third of their costs in data storage and network streaming, if they apply the 8 fps (instead of 12 fps) recommendation.[30]

Hypothesis 2 (H2) was also accepted. As the frame rate of CCTV video is increased, an observer's task confidence ratings and their perceived video quality ratings will also increase. Both the average task confidence and perceived video quality ratings increased as video frame rate increased from 1 to 12 fps. These ratings were significantly different across the frame rates. However, there was one exception: the average perceived video quality ratings between 8 and 12 fps were not significant. This is possibly because observers were not able to tell the difference in video quality (perceptually) between 8 and 12 fps, and this is why task performance did not improve beyond 8 fps.

Participants were asked questions after they completed the detection test purposely to verify the objective task performance results. The method of combining objective task performance and subjective perception data was also used in the face identification experiment (see Study 2 in Chapter 8). The responses revealed that 72 participants (90%) found the task difficult. Participants' comments reveal that much of the task difficulty was a result of the participants noticing a perceptual difference between very low frame rates and very high frame rates when observing CCTV video:

- *"Some clips were quite hard actually, especially the slow motion video clips. Everything was slow, but I think there were some bits missing in the video. It felt like the video was not long enough for me to take everything in before the question came up."*
- *"I'd say there were about three different quality levels, the lowest quality one was the hardest as these were choppy … bits of information were missing definitely."*
- *"I'd say I got a few wrong but only when the video was really low in quality, probably 70% of the video clips were easy and it was obvious what was going on. Some of it was just impossible because it was all flickery and slow."*
- *"Some events happened and I didn't even realise, I could tell I missed bits even after the second watch. I realised something had happened after I read the scenarios really. The ones which were the worst in quality went like a flash."*

[30] 12 fps - 8fps = 4 frames/100 = 33% saving.

- *"It was much harder when the video kept stopping and starting … I couldn't really focus on the actions in these clips."*
- *"Some clips went really quick and I couldn't see what happened."*

From these comments, it was clear that the task was increasingly difficult for participants when observing video at very low frame rates (e.g., 1–5 fps). Task difficulty was reflected in the task performance results, for which the average number of detections from CCTV video reached no higher than 44.5% (see Figure 9.3). Task difficulty was also confirmed with the decrease in task confidence as the video frame rate was lowered (see Figure 9.4).

In reviewing the objective and subjective results, it is evident that if CCTV video is increasingly compressed, it will have a negative impact on the performance of a detection task with CCTV. Furthermore, detecting events on low-frame-rate video will reduce the observer's confidence in the task and their perceptions of video quality. This study examined the effect of one type of video compression, temporal, on detection performance. It would have also been useful to assess the combined effects of spatial and temporal video compression on CCTV detection to determine whether performance, confidence and user perceptions decrease even further.

As with Study 2, the research findings from this study provide objective video quality recommendations based on evaluations with a large sample of novice CCTV users. It would have been valuable to compare the detection performance of these users with expert CCTV observers. However, there were difficulties in recruiting expert users to participate. Even so, previous research in CCTV has identified that there is very little difference in performance between these groups of users when detecting the *on-set* of a crime incident (Troscianko et al., 2004). Additionally, untrained participants were purposely assessed, as these users are now being recruited as part of crime reduction schemes (see Web Users to Patrol, 2006 and Rights Group, 2006). Thus, these untrained participants were chosen in order to understand how reliable these users really when performing security observation tasks, particularly when digital video quality is low. For this study there was an additional advantage in recruiting untrained CCTV users, in that a conservative CCTV frame rate could be recommended for all types of CCTV users of varying skills and experience.

CCTV video tasks are more error prone than other video tasks, such as video conferencing and gaming, due to the wide variations in content and video quality. As discussed in Chapter 5 (see Table 5.1), there are many variables which can influence an observer's response to events and targets on CCTV video, including frame rate, environmental lighting to the camera capture area, camera angle and height, the target's actions, clothing, etc. It is therefore important that CCTV owners treat the minimum 8 fps level for real-time and post-event detection tasks as a recommended level rather, rather than a strict guideline.

9.5.2 Review of Study 3 in Relation to Previous Research

Horn et al. (2002) found that 5 fps was not acceptable for detecting lies on video. The same was found to be true for CCTV detection tasks; however, at 8 fps detection performance reached a threshold. This finding supports the rule-of-thumb figure proposed by Card et al. (1986): whereby frame rates must be at least 10 frames per second.

Prior to this study, there had been only one other study conducted to examine the relationship between frame rate and detection performance (van Voorthuijsen et al., 2005). The results of this prior study were not valid, as the experimental methodology was weak for a number of reasons. Firstly, participants' performance was not assessed under a wide range of frame rate conditions. Secondly, the frame rates chosen to assess detection performance did not consider the typical frame rates used by CCTV owners when recording or streaming security video. Thirdly, CCTV operators were recruited to assess detection performance. Recruiting these users was inappropriate, since in the UK CCTV operators do not typically observe events from low-frame-rate video. Although van Voorthuijsen et al. (2005) observed that lowering CCTV video frame rate affected the user's ability to detect events; no recommendation was put forward (i.e. a threshold frame rate) and this was possibly because the experiment was a pilot study and only preliminary findings were reported by the authors.

The study detailed in this chapter is the first to examine CCTV task performance by applying a task-oriented approach, which brings context into the evaluations by:

1. using untrained CCTV users. New and emerging CCTV users are increasingly being used to detect suspicious events on the Internet and on low-cost CCTV systems where network and storage resources are scarce;
2. assessing a common CCTV observation task (Aldridge, 1994) performed by both experienced and inexperienced CCTV users; and
3. using typical video frame rates applied by CCTV owners: 1, 5, 8 and 12 fps (Cohen et al., 2007). There was no need to assess performance at 25 fps (maximum frame rate), since video is perceived by humans to be smooth in appearance above 12 fps.

9.6 Conclusions

Study 3 investigated research goal 3 in this thesis (see Section Chapter 1, Section 1.4.3) and addressed one of the most significant issue with CCTV, which was identified in Study 1 (see Chapter 4).

9.6.1 Substantive Conclusions

An MPEG-4 CCTV system (at CIF resolution) should be configured so that video used for an event detection tasks performed by human observers of varying skill and experience is recorded and streamed at a minimum video frame rate of 8 fps. By assessing users untrained in CCTV tasks, it was possible to put forward a conservative frame rate recommendation. This rate can be applied to all types of CCTV systems and used by CCTV users of varying skill and experience.

This finding is a valuable contribution as it allows CCTV owners and practitioners to ensure that their CCTV systems will record and streams usable and effective CCTV video for a detection task. There are serious implications associated with using low-quality (temporal) CCTV video for security observation tasks. For instance, at low frame rates (i.e., 1–5 fps), potentially important frames will be discarded by the system, depending upon the speed of the activity within the scene. This means that crime and/or pre-crime activities will not be captured on the video recordings, nor be observed in real-time by an observer. Without having access to scenes vital to a crime situation, the observer's confidence will be low, negatively affect their vigilance, and directly impact their performance in the task. Also, when an observer detects crime using recorded CCTV video, an observer will be unable to piece together evidence to prosecute potential suspects.

This threshold frame rate for detection is an important finding for HCI researchers, particularly for those involved in the research and design of intelligent CCTV systems. The research in this study demonstrates that if a human observer is unable to correctly detect an event on CCTV video played back below 8 fps, it is unlikely that an intelligent system operating remotely will be capable of detecting events below 8 fps.

These subjective findings are also key findings in this study as they illustrate the importance of providing all types of CCTV users with training and education on general CCTV security and CCTV tasks, as well as necessary vigilance skills. In addition to training, CCTV users should be informed of the consequences of task failures to provide them with an understanding of CCTV and its uses in security and surveillance. With this type of task support, and the 8 fps recommendation, CCTV users will improve their confidence and performance in this type of task.

9.6.2 Methodological Conclusions

A number of strengths were revealed in this study:

1. Evaluating CCTV video quality can be very challenging as there are no standardised video data sets available for testing task performance. The CCTV data sets that are available have been produced for testing intelligent detection systems (such as the UK Home Office image library for intelligent detection systems) and not CCTV systems used by human observers for security tasks. As a result, the CCTV stimulus used in this study was pre-recorded of which included typical CCTV scenarios for the detection experiment. A number of parameters were controlled to keep the quality and content of the video clips unique from one another and representative of real-world CCTV scenarios.

2. A large sample of participants were recruited for the experiment (n = 80), which meant that the data has high statistical power, unlike the small samples used in the study conducted by van Voorthuijsen et al. (2005) of just 22 and 16 participants each.

3. Task-oriented measures (Knoche and de Meer, 1999) were applied for the evaluation of the CCTV detection task. By using task-focused measures (e.g., classic HCI measures), it was possible to analyse and report the results objectively. The objective (Likert ratings) and subjective (post-experimental questionnaire) responses were also combined in the analysis to triangulate the results to strengthen the reliability of the data. Furthermore, the detection task considered an existing CCTV scenario using commonly applied frame rates (Cohen et al., 2007). A task-oriented approach is extremely valuable for the multimedia research community, as it provides a human-centred approach for examining the limitations of resource-constrained multimedia systems (not just CCTV systems). This approach takes into account context when designing video task-based experiments with users, thereby adding ecological validity to the research methodology and results. The experimental paradigm used in both Study 2 and Study 3 should be followed by other HCI researchers examining video quality in user tasks to identify where the biggest gains in performance can be achieved, and then weigh the cost of higher quality against the cost of a drop in performance.

There were also a number of limitations in this study:

1. A few of the actors recruited for the CCTV filming struggled to act out some of the scenarios. This meant that the filming took longer than anticipated. Some of the scenarios were filmed several times until the actions were satisfactory. Despite making several recordings of the same scenes, some were acted differently than others, because there was a difference in the acting styles of the volunteers.

2. The duration of each of the events recorded on video was limited (10-12 seconds) for the purpose of managing participant task time to 30-40 minutes. This is a limitation of the study since some events (crime and no crime) can last longer than 12 seconds, and some events include pre-cursor activities leading up to the main event. Although the results provide a 8 fps recommendation, this recommendation is based on a system which has captured live events lasting 10-12 seconds. It must be noted that this length of time is a realistic length (and common) for events such as pick-pocketing, and the non crime events captured for this study. The TEC-VIS framework (Chapter 10, Section 4) emphasises that CCTV owners *must test their system performance* in context. Thus, the results from this detection study provide a recommendation for detecting events from video clips lasting 10-12 seconds.

3. The speed of events could not controlled (e.g., stealing an object from someone) during filming affected the way in which the video recorder compressed the video. This was unavoidable as each scenario was acted out at different speeds and this varied because of the different skill and ability of the actors recruited for the study.

4. Careful consideration was needed in choosing the recording environment to prepare the CCTV stimuli, as the business and complexity of the scenes also had the potential to affect the observer's perceptions, and possibly their performance in the task. The activity

within the background of the scene and the content may have also affected the encoding behaviour. This was controlled as best as possible by filming the video at the same period of time on the university's grounds.

5. As in Study 2 (see Chapter 8), a digital camera recorder was used to prepare the CCTV video clips. Digital CCTV video is recorded in the real world using CCTV cameras and is stored or streamed via a digital video recorder. It was not possible to install a CCTV system to record the CCTV stimuli for this study as this was expensive and had practical implications for the research. Despite this, the camera was configured to closely match a low-cost CCTV video recording system. For future CCTV video quality evaluations, the set-up of the recording environment should be matched as best as possible to existing CCTV system configurations.

To take this research further, it would be a valuable exercise to examine observer performance for a detection task in real-time across a network, and also compare performance for different CCTV crime scenes (i.e., theft, violence, deceptive crime, drug abuse etc). Advanced and intelligent CCTV systems can be designed to record and stream video where only movement is detected and record/stream video containing only key reference frames ('I' frames). The use of these advanced features will save users time and cost in operating a digital networked CCTV system.

Chapter 10

TEC-VIS: A Best-Practice Framework for CCTV

The results from the research carried out for this thesis (see Chapters 4, 8, and 9), were used to develop a best-practice framework for the effective design, configuration, and use of digital CCTV, which is entitled Task Effective CCTV Video in Security (TEC-VIS). This framework has been designed for CCTV security owners and practitioners and provides them with a substantive set of guidelines on digital CCTV. The framework has been validated through an external peer review conducted by three experts who work in the fields of HCI, ergonomics and security. As a result of the critical review, a number of amendments have been made to the original version of the framework, which can be found in Appendix H.

In this chapter, the TEC-VIS framework and each of its eight phases are then described. Following these descriptions, a CCTV deployment scenario is provided. This scenario illustrates how the TEC-VIS framework can be applied.

10.1 Introduction

A security observation environment involves human observers performing security tasks with the support of a range of tools and systems. The size of the environment can vary, and it can be either a control room or a security room. The human observer is typically trained and experienced in performing tasks such as monitoring, controlling, identifying and recognising incidents, objects and human targets. The different types of tools and systems the CCTV user may interact with include access control systems, radios, telephones, video monitors, digital video recorders, and intelligent CCTV systems (e.g., automatic detection and face recognition systems).

When planning the design of a CCTV security system, it is crucial that the user be placed at the centre of the design process, rather than at the end. As detailed in Chapter 5, Section 5.3, there are a number of publications which provide guidance on the design configuration and use of CCTV systems and control room environments; however, these guidelines are insufficient, contradict one another and are not based on evaluations with representative CCTV users. Overall, a holistic set of guidance for the development of CCTV systems is needed which considers both the social and technical aspects of a security system.

The best-practice framework presented in this chapter uses a socio-technical approach to provide CCTV security owners and practitioners with guidance on how they can achieve the following:

1. Identify the different stakeholders who belong to the CCTV security system.
2. Identify the different tasks each stakeholder will perform.
3. Define each of the stakeholders’ tasks and work environment requirements.
4. Identify and assess potential technical obstacles and stakeholder conflicts.

5. Assess the performance of the system with CCTV users and real-world tasks.

The ultimate aim of carrying out these steps is to determine whether or not the CCTV security system is fit for purpose. The TEC-VIS framework, detailed in the following sections, is available as a standalone document. To improve readability, the references in TEC-VIS are ordered numerically (e.g., [1] and [2]) and these references are provided in Section 10.4.

The following should be considered when applying the framework:

1. All phases can be implemented by all CCTV security owners and practitioners, regardless of knowledge and experience.
2. The framework can be applied to all CCTV systems, regardless of system type, cost, size, and age.
3. The technical configurations recommended in Phase 4 (see Section 10.2.4) should be considered as a *minimum requirement* and not a strict guideline. The configurations apply to the tasks described in this phase – not all CCTV observation tasks.

10.2 TEC-VIS Framework

The use of a structured framework for the deployment of a CCTV security system facilitates the identification of the various CCTV system stakeholders. A framework also enables the system owner to analyse each stakeholders' business and security-specific goals for CCTV, their tasks, and the functional operation of the system in context. The purpose of the stakeholder and system requirements and analysis processes are to identify both the social and technical factors that may reduce the performance of the CCTV security system during the early stages of deployment. Once these factors have been identified, the specifications detailing the processes, technical capabilities, and configuration can be redefined and corrected to improve the effectiveness, efficiency, and performance of the system and its users. This framework is robust, as it was developed as a result of an extensive field study and two empirical studies.

Due to the complexity and diversity of CCTV systems, the TEC-VIS framework is limited to providing guidance on the following:

1. Identification and analysis of CCTV stakeholders and their tasks.
2. Analysis of the technical capabilities of a CCTV security system.
3. Establishing system requirements to enable the production of usable video for real-time and post-event security observation tasks. Specifically, requirements are outlined for two commonly performed observation tasks (i.e., face identification and event detection) performed by CCTV users of varying skills and experience.
4. Configuration of technology used within a security work environment.
5. Configuration of CCTV cameras within a surveillance environment.

The framework does not provide guidance on specific security and surveillance technologies, nor does it address the social and privacy issues surrounding CCTV. Furthermore, guidance is not given on how to conduct a cost-benefit analysis of a CCTV deployment.

The TEC-VIS framework is intended for use by CCTV practitioners (consultants) and owners. These stakeholders are responsible for the design, specification, and deployment of security systems in which analogue, digital and networked CCTV technologies are used. The framework should be treated as a single point of reference by these CCTV stakeholders. It should be accessible to all CCTV stakeholders who belong to the system; however, the CCTV security owner(s) should manage the document and any changes made should be recorded.

The core of this framework emphasises the identification of the CCTV stakeholders within the security system, analysing their business and security goals, defining their tasks and system capabilities and identifying risks. It must be noted that no two security systems are the same; thus, user requirements and system capabilities will vary from one system to another. The TEC-VIS framework is flexible, in that it can be used for any type of CCTV security system. Where possible, examples have been provided within each phase to illustrate the specification and analyses processes.

10.2.1 Phase 1: Specify Goals of CCTV Security System

Objective: Defining the goal(s) of a system involves specifying the system's purpose. For instance, the owner should ask him/herself questions such as: "What do I want to do with the system? What do I want to achieve? Why am I implementing the system?" Often, managers, engineers, and designers involved in the design and deployment of a system tend to assume that system stakeholders will adopt the overall system goals, but this is not always the case. Users may not appreciate their assigned roles, and develop their own goals [1]. Therefore, it is very important that the goals of the system are established first. Specifying the goals will allow system owners and all other stakeholders to understand the purpose of the system.

Process: The process of defining the goals of the system should be transparent and all stakeholders should be involved. In addition, the goals should be used as a reference throughout the entire lifecycle of the system, to ensure they are being met following any organisation and/or technical changes. More than one goal can be defined, and each goal can be broken down into sub-goals. For example, a goal of the London Underground might be to use CCTV video security technology to allow operators to monitor and control overcrowding within the underground tube stations. This goal might include the following sub-goals:

- Monitor the flow of people during peak travelling hours.
- Detect accidents during all hours.
- Spot ticket touts, drunks and illegal buskers.

10.2.2 Phase 2: Identify Stakeholders and their Goals

Objective: It is very important to consider all stakeholders in the design of a system. A stakeholder is by definition *"… any group or individual who can affect or is affected by the achievement of the organisation's objectives"* [1]. In the context of a CCTV security system, it is important that the different stakeholders be accounted for, so that they can influence the development of the system along the way and help identify and avoid potential conflicts, thereby improving the overall effectiveness of the system. Figure 10.1 identifies ten possible CCTV stakeholders.

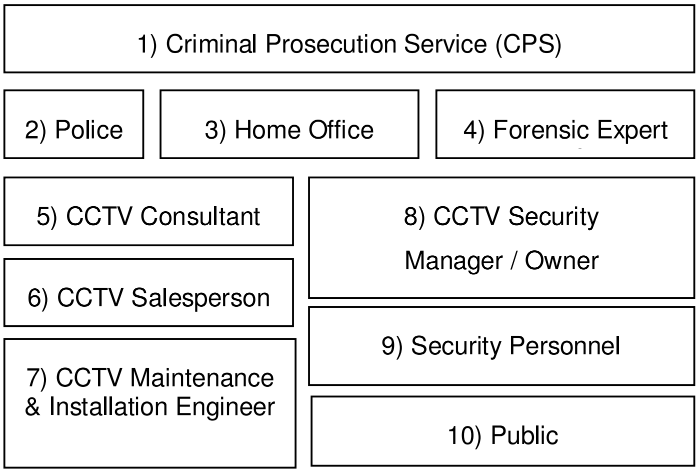

1) Criminal Prosecution Service (CPS)
2) Police
3) Home Office
4) Forensic Expert
5) CCTV Consultant
6) CCTV Salesperson
7) CCTV Maintenance & Installation Engineer
8) CCTV Security Manager / Owner
9) Security Personnel
10) Public

Figure 10.1: Key CCTV stakeholders.

The CCTV stakeholders identified in Figure 10.1 may vary from country to country. To avoid potential misinterpretations of the roles and responsibilities of these stakeholders, a basic description of each UK-based CCTV stakeholder follows:

1. **Criminal Prosecution Service (CPS)**: Those who work for the CPS in the UK are individuals within the judicial field. These individuals have the legal authority to retrieve CCTV video and images and present them as evidence in a criminal prosecution case. These users include solicitors, barristers, and court staff.

2. **Police:** The police in the UK are responsible for collecting CCTV video from CCTV owners. CCTV data is then analysed in a video lab by police investigating road traffic and crime incidents. The police also work in collaboration with community support staff, as well as CCTV control room and council staff, to deter and detect crime for the purpose of protecting the public for safety and security. The police are CCTV users as well as CCTV owners, as they can own a CCTV system and use it in their policing activities.

3. **Home Office**: The Home Office is the government department responsible for leading the national effort to protect the public from terrorism, crime and anti-social behaviour in the UK. The Home Office Scientific Development Branch (HOSDB) is a Home Office specific

department where research is undertaken to provide science and technology based solutions. The HOSDB provide a wide range of CCTV publications including: recruitment and selection of operators, layout for control rooms, CCTV performance, and how to retrieve video from CCTV systems[31]. The most widely used CCTV publication is the Operational Requirements ([2] and [4]).

4. **Forensic Expert:** In the UK, there are approximately 20 professional forensic experts trained to examine CCTV video and images. These CCTV users are highly trained in analysing 'difficult' CCTV video and imagery provided by the police. The CPS and the police typically provide CCTV video and images to forensic experts when there are doubts about the images contained. Forensic experts provide subjective assessments based on their expertise. They also use objective techniques such as photogrammetry to answer questions about unknown targets.

5. **CCTV Consultant**: A CCTV consultant is an individual who has expert knowledge and extensive experience with CCTV systems. CCTV consultants are likely to be involved in the early stages of a CCTV deployment, to assist the owner in designing and deploying the system. CCTV consultants can also be involved in the redesign of an existing CCTV system. Redesigns typically occur during system upgrades. CCTV consultants are neither CCTV users nor CCTV owners, but they do influence a CCTV system.

6. **CCTV Salesperson**: The CCTV salesperson's role is to market and sell CCTV products and systems to potential and existing CCTV system owners. They typically publicise their services and products over the Internet and through direct retailing. Larger commercial companies may showcase their CCTV products and services at security trade shows and exhibitions.[32] Some CCTV sales companies also provide additional services to CCTV owners, such as installation, repairs, and maintenance.

7. **CCTV Maintenance and Installation Engineer**: A CCTV maintenance engineer is paid to maintain an existing CCTV security system and to ensure that it is functioning correctly. A maintenance engineer may also be involved in configuring the hardware and system network.

8. **CCTV Security Managers/Owner**: There are two groups of CCTV owners: 1) public and 2) private. Public owners own private CCTV control room systems and manage them on behalf of the public local authority and in collaboration with the police. Private owners are individuals who own a CCTV security system to protect their business or home. Both public and private CCTV owners are responsible for managing their security system, including all CCTV stakeholders, the hardware/software of the system, and the camera environment. The CCTV owner is not likely to have a great deal of CCTV technical knowledge; however, they are responsible for ensuring that the system is fit for purpose.

[31] http://scienceandresearch.homeoffice.gov.uk/hosdb/cctv-imaging-technology/

[32] For example, IFSEC (InFo for SECurity): www.ifsec.co.uk

9. **Security Personnel**: These users are typically directly involved with the CCTV system. Security personnel include CCTV operators and security staff who have been recruited by their own company or by a private security company. CCTV operators are responsible for observing CCTV video events on video monitors. The security tasks they typically perform include: monitoring, detecting, recognising and identifying [2]. They are considered to be data controllers, as they use CCTV video and react to incidents.
10. **Public**: The members of the public do not interact directly with a CCTV system, but are part of a security system designed to protect the public. Members of the public are seen as data subjects, since they are captured on CCTV video. CCTV video is subject to privacy and data protection laws, and CCTV owners are required to consider relevant legislation when deploying a CCTV security system.

Process: The process of identifying the different CCTV stakeholders involved in the security system should be completed along with the identification of their business and security-specific goals for CCTV. This process is achieved by creating a list of the potential stakeholders. The contents of a stakeholder list may vary depending upon the size and context of the security system being deployed. Once these groups and individuals are identified, the name of each stakeholder should be noted along with their contact details. If possible, each user's goals should also be stated (both business and security goals). For example:

- Stakeholder group: Metropolitan Police, Camden, London
- Stakeholder name: Chief Inspector Morris
- Contact details: 241 Camden High Street, London, N1 0UX (0207 663 6645)
- Business goal: Support local businesses in creating a safe community
- Security goals:
  - Support the local authority for crime and emergency purposes
  - Investigate crime and prosecute criminals

A CCTV security system is typically configured to record CCTV video footage of people (targets) and their activities. These images are personally identifiable and are therefore classified as personal data. In privacy literature, individuals who appear in personal data are data subjects, and those who use personal data are data users and controllers. In the context of CCTV, these users are described as follows:

- Data controllers and data users are organisations and/or individuals who own and directly use the data (e.g., CCTV control room operators and CCTV owners).
- Primary data subjects are targeted individuals (e.g., criminals and suspects).
- Secondary data subjects are members of the general public.

The CCTV security system owner should identify whether stakeholders are data users, data subjects or data controllers, to ensure that there is a clear understanding of who will directly interact with the CCTV video and who will be captured on video. In addition, the handling of personal data must be considered:

- Who will own the CCTV data?
- Who is legally responsible for the access, use and destruction of the CCTV data?
- Who will be recorded on video and why?
- How will the CCTV video be protected from non-data controllers and non-data users?

CCTV security owners and practitioners should refer to the relevant legislation in addressing data protection, management and privacy issues [3]. A comprehensive list of legal guidance relating to CCTV can be found within the Home Office Operational Requirements Manual (p 20) [4].

It is important to allow all CCTV stakeholders access to stakeholder information to aid communication, encourage feedback, and raise awareness of the different CCTV stakeholders and their goals for the system. The identification and involvement of different CCTV stakeholders should be completed each time the system is altered (e.g., when new technology is deployed or additional CCTV tasks are introduced). Stakeholder involvement can be achieved by conducting a series of stakeholder meetings to facilitate communication and feedback. A stakeholder meeting should be held every four to six months and should allow the various stakeholders to share their ideas and to work on system development openly and interactively. Overall, regular stakeholder meetings will ensure that the system's performance is optimised by assessing whether security and business goals continue to match stakeholder requirements, tools, tasks, and the security environment. Attendees may differ for small, medium, and large-scale security operations (see Table 10.1).

Table 10.1: Examples of CCTV Stakeholders for CCTV Security Systems of Different Sizes

SMALL SECURITY SYSTEM: Newspaper Shop (4 staff members)
▪ Criminal Prosecution Service (CPS) – in the event of a crime ▪ Police ▪ CCTV owner (business owner) ▪ Shop assistants ▪ CCTV salesman ▪ CCTV maintenance and installation engineer ▪ Public (customers)
MEDIUM SECURITY SYSTEM: Sock Factory (15 staff members)
▪ Criminal Prosecution Service (CPS) – in the event of a crime ▪ Police ▪ CCTV owner (business owner) ▪ Factory staff ▪ CCTV consultant ▪ Security personnel ▪ CCTV salesman ▪ CCTV maintenance and installation engineer ▪ Public (customers)
LARGE SECURITY SYSTEM: Major City Airport (25 staff members)
▪ Home Office (input in policy and strategy of CCTV system) ▪ Criminal Prosecution Service (CPS) – in the event of a crime ▪ Police ▪ CCTV owner (business owner) ▪ Airport staff ▪ CCTV consultants ▪ Security personnel (team) ▪ CCTV salesman ▪ CCTV maintenance and installation engineer ▪ Public (travellers and all airport staff)

10.2.3 Phase 3: Identify Stakeholder Tasks

Objective: A task is defined as an action, or set of actions, carried out by an individual in order to meet system goals. These actions can be physical (e.g., using control facilities to track a target captured on a CCTV camera) and cognitive (e.g., monitoring real-time CCTV video and deciding whether to react to a particular event). The objectives in identifying stakeholder tasks are to enable CCTV security owners and practitioners to accomplish the following:

- Identify whether the tasks can be appropriately supported by the available technology.
- Determine what knowledge, skills, tools, and training programmes are required for stakeholders who directly interact with the CCTV, to ensure that tasks are performed effectively. For more guidance see [5].
- Allocate tasks appropriately to the users and the system.
- Record task information for use in future designs; offer new ways to perform tasks or parts of a task.

Process: This process is achieved by reviewing each stakeholder's goals (Phase 2) and decomposing them into specific tasks, which the user will be expected to perform when using the system. Examples of CCTV operator tasks can be found in [2]. Users' tasks can be defined by conducting a hierarchical task analysis (HTA). HTA is an analytical method that involves the identification of tasks, their categorisation, the subsequent identification of sub-tasks, and the verification of the overall accuracy of the task model developed. HTA is useful in the design of new systems used by human operators, since it enables a designer to envision the goals, tasks, sub-tasks, operations and plans essential to a CCTV user's activities. The completion of a HTA allows the designer to visualise data as a hierarchical network. This is simple and easy for non Human-Computer Interaction (HCI) experts to use. For more background and guidance in completing a HTA, please refer to [6].

10.2.4 Phase 4: Identify and Assess Video Quality

Before describing the objectives and processes for this phase, an overview of CCTV technology is first provided.

Overview of CCTV Technology

There are currently three types of CCTV systems available:

1. **Analogue CCTV:** These systems record video from surveillance cameras directly onto a VCR, which is physically stored on a VHS or S-VHS tapes. Analogue CCTV systems are:
   - slow when compared with digital CCTV systems. Their inefficiency can impact access and retrieval of surveillance footage needed for investigations;
   - susceptible to human error. They require constant human intervention, as VHS tapes need to be stored and replaced for continual recording. Tapes can be unintentionally over-written or misplaced;
   - limited in image resolution (240–340 TV Lines); digital CCTV video can record up to 400 TV Lines. Digital CCTV video can also be recorded at different video resolutions; the choice depends upon the type of system and the budget available for data storage and network delivery; and
   - lacking in remote video access capabilities. Digital systems can be linked to the Internet to transmit video footage for remote monitoring and surveillance purposes.

2. **Digital CCTV:** Digital video is much easier to store and transfer, because the video is compressed. The video is stored on a computer or digital video recorder (DVR) to serve as evidence following analysis. Digital CCTV is flexible, as the data can easily be transferred across various media and can be enhanced if it is not clear enough for identification. The main issue with digital CCTV is achieving high quality video when storage and network resources are scarce.
3. **Networked CCTV:** Networked CCTV systems digitise and compress analogue video signals. The encoding process occurs within the DVR device or within an Internet Protocol (IP) CCTV camera. A standard web browser can be used to monitor CCTV in real-time. The speed and quality of the video delivered over the network depends upon the bandwidth being used. The lower the bandwidth, the slower the data transfer and the lower the video quality.

Objective: Following an installation, the performance of every new CCTV system or re-design should be tested with the intended CCTV users performing their required security tasks.

Very often, the performance of a CCTV security system is assessed when it is operating in full quality mode (e.g., when CCTV video is being fed directly from a number of cameras and observed in real-time within a control room). This assessment is useful in determining the effectiveness of a camera's positioning, angle and zoom levels; however, it is equally important to assess the CCTV users' performance in carrying out their tasks with video that has been *recorded* and *streamed over the network*, and thus, takes into account of data compression during video storage and transfer.

Process: The evaluation of a CCTV system should be carried out with the intended CCTV users (e.g., operators) using CCTV video/images that have been:

- digitally recorded on hard-disk or CD/DVD and played back for the purpose of post-incident analysis;
- transferred over a network for the purpose of remote surveillance; and
- received directly from CCTV cameras (i.e., a typical CCTV control room set-up).

The effectiveness of a CCTV system and its video quality for CCTV tasks should be assessed with the intended CCTV users by following the video display, quality and network recommendations provided in Table 10.2. These recommendations are provided for specific task scenarios, face identification and detection tasks, and are based on empirical experiments with 80 untrained CCTV users [7] and **Error! Reference source not found.**. As untrained CCTV users were used for these tests, these recommendations can be applied to CCTV systems utilised by CCTV users with different skills and experience in performing CCTV security tasks. General guidance on digital CCTV and networked CCTV systems is available in the most recent UK Home Office publication [4] and [7].

Once the CCTV system has been configured following these recommendations, the system set-up should be tested in full operational mode with the intended CCTV users to ensure that the system is fit for purpose. Any task performance problems should be noted. If the CCTV system is found to be ineffective, the **video quality configurations should be reviewed**. This can be achieved by determining maximum system performance (where budget allows) by raising the video quality configurations until the desirable task performance is achieved:

- Spatially - lowering video compression (increasing bandwidth by 50 Kbps) and increasing the display resolution to the next available level.
- Temporally - increasing the frame rate to the next level (e.g., by 2 fps).

Table 10.2: Parameters for Video Display and Transmission

Low-Risk Environment **(Low-crime rate areas & places where objects are not valuable)**			
CCTV Task Scenario	**Display Resolution**	**Temporal Resolution***	**Spatial Resolution****
Face Identification (120% Rotakin): Identify an unknown person in a CCTV video/image, when a good quality photo is available for comparison (e.g., indoor access control room system)	352x288 or higher	8+ frames per second	▪ MPEG-4 ▪ 52+ Kbps
Detection (10% Rotakin): Detect suspicious events on CCTV video. Events monitored typically over a real-time network. (e.g., events such as theft, , drug dealing, and other everyday actions).	352x288 or higher	8+ frames per second	▪ MPEG-4 ▪ 2000+ Kbps

High-Risk Environment **(High-crime rate areas & places where objects are very valuable)**			
CCTV Task	**Display Resolution**	**Temporal Resolution***	**Spatial Resolution****
Face Identification (120% Rotakin): Identify an unknown person in a CCTV video/image, when a good quality photo is available for comparison (e.g., indoor access control room system)	352x288	12+ fps	▪ MPEG-4 ▪ 2000+ Kbps
Detection (10% Rotakin): Detect suspicious events on CCTV video. Events monitored typically over a real-time network. (e.g., events such as theft, drug dealing, and other everyday actions).	352x288 or higher	12+ fps	▪ MPEG-4 ▪ 2000+ Kbps

*Temporal resolution: the level of temporal compression is referred to as the frame rate and ranges from 1–25 frames per second. CCTV video is compressed through temporal compression by a digital video recording system to increase the hard disk storage space available for recording CCTV video locally, and the speed at which video is transferred over a network for remote surveillance.

The frame rate chosen depends upon the number of CCTV cameras connected to the system, the video/image quality required, and the duration of the recording. The number of frames delivered over the network of a CCTV security system is dependent upon the bandwidth available. If the system has been configured for a low bandwidth (i.e., 32 Kbps), low-quality video will be delivered to the observer's video monitor at a slower rate than could be achieved with a CCTV system configured at a higher bandwidth (i.e., 120+ Kbps). Another important factor to consider when choosing an ideal video frame rate for recording and streaming CCTV video is the speed with which targets and objects are moving. The faster the movement within a scene, the higher the frame rate needed to capture and process the video frames.

**Spatial resolution: is accomplished through the use of a video compression device (encoder), which is found within the hardware of the security system. The video is converted from analogue to digital and is then compressed using either a hardware or software compression algorithm (CODEC). The size of the video is reduced to accommodate available hard-disk storage space requirements, thus increasing the recording time and data transfer rate. The video quality after compression will differ depending upon the type of video CODEC used and the content being captured (slow vs. fast movements). The impact of spatial video compression is illustrated in the images shown in Figure 10.2. These images were compressed using an MPEG-4 video CODEC (Windows Media CODEC Version 1) at four different video bit rates: 32, 52, 72 and 2000 Kbps. Figure 10.3 shows images taken from video encoded at the same bit rates using a Wavelet CODEC (WAVC). The compression results clearly show that video encoded using MPEG-4 (above 52 Kbps) is of better quality than Wavelet encoded CCTV.

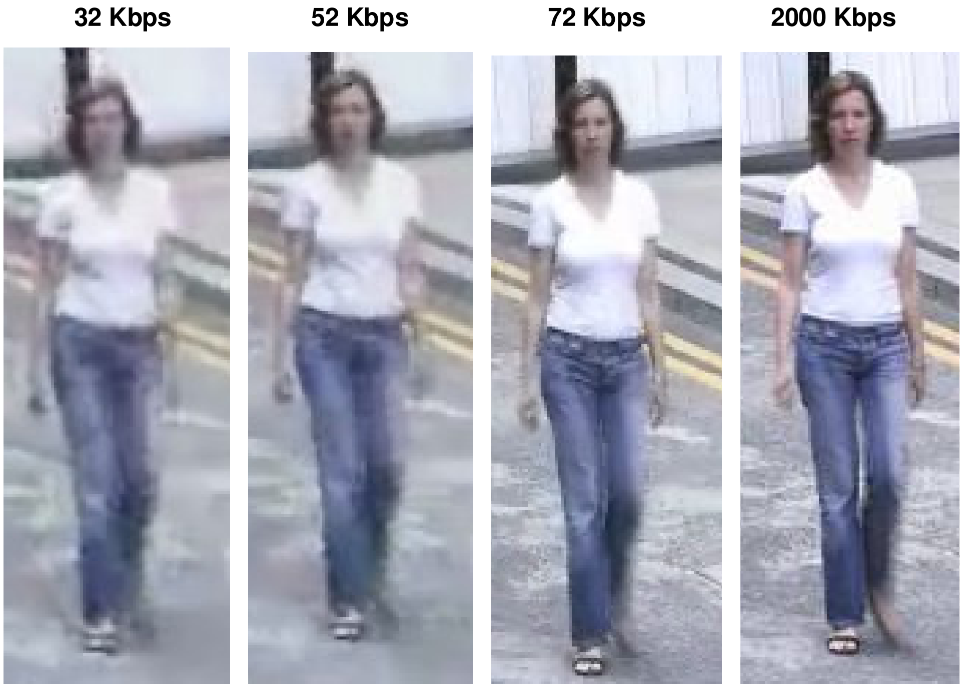


Figure 10.2: The effect of MPEG-4 video compression on image quality.

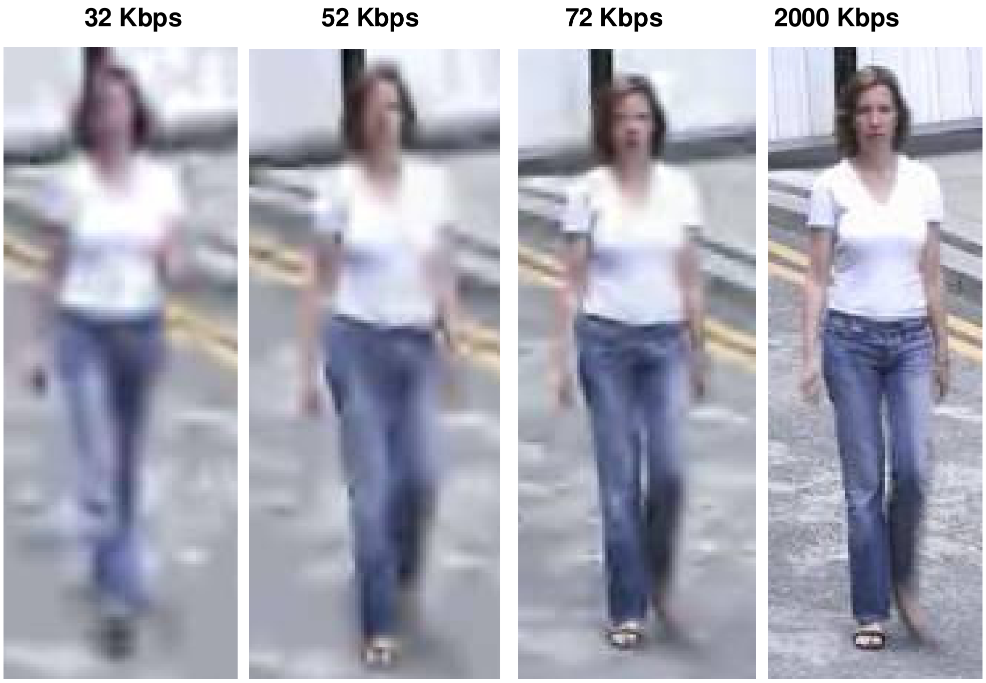


Figure 10.3: The effect of Wavelet video compression on image quality

10.2.5 Phase 5: Identify Capabilities of System

Objectives: System capabilities are identified to: 1) help CCTV owners understand the technology being used for the CCTV system, and 2) ensure that the technology used meets all stakeholder goals and task requirements.

Process: The capabilities of the system can be identified by reviewing the goals for the system, as well as the individual goals and tasks of each stakeholder (Phases 1–3). Following this, the CCTV video display, recording quality, storage, and network requirements should be reviewed (Phase 4). The individual components of the system should be assessed to ensure that they are configured correctly. This phase is achieved by following the best-practice guidance provided in Tables 10.3, 10.4, and 10.5. The recommendations provided in these tables are a result of research from the following sources:

- Detailed literature reviews (see Chapters 1-7)
- Published research and guidance
- Control room field study (see Chapter 4)

Table 10.3: Best-Practice Recommendations for the Security Environment

Internal: The CCTV Operator's Work Environment

General Security Environment Guidance

1. When CCTV video is monitored in real-time by CCTV users and security staff within a control room environment, the International Control Room Standards should be followed [14] to support the ergonomic design of the CCTV user's physical work environment.
2. All CCTV users who perform security observation tasks with CCTV video and images must be trained in their tasks and assessed every four to six months. These assessments should involve reviewing and testing the CCTV operator on the surveillance environment, security procedures, security tasks, and the use of all necessary tools and equipment.
3. All tools and equipment necessary for the CCTV user's tasks should be safely arranged, free from obstruction, easy to reach and fully functional. Any tools and equipment used for security observation tasks must be tested regularly and maintained every four to six months.
4. Faulty tools and equipment should be reported through a fault reporting system (this can be a paper or electronic system). Tools that are used regularly should be assigned a high priority for repair. During repair or maintenance, faulty tools and equipment must be replaced, so that the CCTV user is still able to perform their tasks.
5. Any tools and equipment that cannot be repaired should be properly disposed of and replaced.
6. Any non-essential tools and equipment should be placed in storage outside the work environment. Tools used infrequently but which are required for emergency use must be accessible at all times.
7. To avoid confusion and delays during incident reporting (via radio/ telephone), CCTV users and other stakeholders should use a standard communication protocol which states rules for effective audio communication. The communication protocol should include the use of the NATO phonetic alphabet and police identity codes to provide a standard vocabulary for describing names, vehicle registration number plates, street locations, target ethnicities, etc.
8. The average noise level within the control room shall not exceed 85 dB(A) during the length of the working day. Prolonged, very low or very high frequency noises should be avoided. Alarm signals should be at least 10 dB(A) over the background noise of the control room, and noise levels should not interfere with communications, warning signals or mental performance.[33]
9. If a radio system is used, CCTV users should be assigned to respond to an equal number of radio channels (based on the level of radio contact and priority assigned to calls). A detachable headset should be used if the operator is responsible for responding to more than two radios/telephones.

[33] http://www.hse.gov.uk/comah/sragtech/techmeascontrol.htm

Table 10.4: Best-Practice Recommendations for CCTV Cameras/Monitors

Internal: The CCTV Operator's Work Environment

CCTV Camera and Monitor Guidance

1. The maximum number of displays for a single control workstation should be based on a task analysis. No more than four displays (of up to 25 inch diagonal) should be used for monitoring video and non-video information and be operated [14].
2. For a detection task in which specific events are detected, such as suspicious actions and crimes, the observer must perform this task using a single video monitor (spot/incident monitor) with the monitor positioned directly in front of them (see point 7).
3. The CCTV video should be distributed equally across monitors and between CCTV users to reduce the probability of errors and information overload. CCTV cameras located in areas in which the level of crime is moderate to high should be displayed at all times to ensure that these crime hot spots are in constant view.
4. The choice of cameras on display within a row of monitors (monitor bank) should be reviewed on a weekly basis by security management and the local police to ensure that incidents are being effectively monitored within the surveillance area. The choice of cameras on display should vary depending on the priority of surveillance (see point 3).
5. Video monitors placed within a monitor bank should be positioned at a minimum distance of 59.1 inches (1.5m) from the CCTV observer.
6. Video monitors placed within a monitor bank should be between 17–28 inches in size [2].
7. Detection tasks should be performed at the observer's workstation on a single spot monitor facing the user to allow for close-up inspections. Spot monitors should be positioned 29.5 inches – 39.4 inches (75 cm-1m) from the observer [17].
8. It is recommended that a maximum of four CCTV cameras be displayed on a single video monitor for a detection task performed by two observers (adapted from [2]).
9. A switching/auto-cycling camera display function will allow a set number of CCTV cameras to be displayed over a set period of time. When using this function, monitors should be set to switch at the same time to avoid a 'flicker' effect. Furthermore, the switch display should not be set to less than five minutes as it is likely to cause visual discomfort and distract the user from the video events.
10. To avoid fatigue, CCTV users should view CCTV on video monitors for 50–60 minutes and should then take a 5–10 minute comfort break [16]. Comfort breaks should be taken away from the observation area.
11. A CCTV user should be present at all times when there are more than 5 video monitors in use for a monitoring task along a monitor bank.
12. For a CCTV system that is connected to 10+ cameras, a list should be provided to the CCTV observer containing the ID of each camera, together with its type, physical location and positioning information. The location of all CCTV cameras should be easily identifiable on a geographical map.
13. CCTV camera lists containing the information identified in point 12 should be updated following any changes to the system cameras. Each CCTV camera on the list should be numbered and placed in geographical order (and not by installation date).

Table 10.5: Best-Practice Recommendations on CCTV System Strategy and Design

CCTV System Strategy and Design

1. The type of CCTV camera chosen must be suitable for the CCTV observer's task. For example, a movable camera is required for detecting a suspicious target and a fixed camera is suitable for monitoring crowds on a busy train station platform. If a new task is added, the right type of CCTV camera must be supplied for the task and the CCTV observer must be trained and assessed on the task every four to six months.
2. All CCTV cameras connected to the system should be adequately safeguarded from vandalism, accidental damage, and weather damage.
3. The lighting to the camera capture area should be adjusted until the overall scene, specific objects (i.e., vehicle registration number plates) and targets are easily recognisable and identifiable. The possible effects of unwanted and variable lighting (e.g., flashing emergency lights, bad weather conditions, Christmas lights, etc.) should be considered during the camera installation.
4. CCTV cameras placed in unplanned, inappropriate positions, and angles may result in camera blind spots. Failure to review, and when necessary, move or decommission equipment may result in incidents and targets going unnoticed or possibly misinterpreted/misidentified.
5. CCTV cameras should not be left in locations where there is no activity. A nominated individual should assess each CCTV camera every four to six months and evaluate its purpose and operational functions.
6. Depending on the configuration of the system, camera signal loss may occur for various reasons (e.g., technical failure or bad weather conditions). In the event of camera signal loss, a fault detection and repair reporting tool should be used.
7. Organisational responsibility must be identified for environmental maintenance affecting the operational performance of a CCTV system (i.e., trees, foliage and bunting). In addition to identifying relevant individuals, their roles and responsibilities, and procedures must be put in place to ensure that public CCTV cameras are not obstructed or damaged by the environment.
8. The physical camera environment should be assessed every four to six months and also following any environmental changes to ensure CCTV cameras are not affected in their performance and display.

10.2.6 Phase 6: Identify and Eliminate Potential Obstacles to System Capabilities

Objective: By completing phases 1–5, it will be possible to see where the deployment of a CCTV system may start to break down. In phase 6, reviewing the security systems' capabilities will allow the owner to identify instances in which the technology may be ineffective, insufficient, or unsuitable. Two examples are provided below:

1. **High-end CCTV system:** A new CCTV system was required to record very high-quality CCTV video to enable a face recognition system to accurately match targets from real-time CCTV video with images from a database of criminals provided by the police. Unfortunately, the budget allocated for the system was too low to accommodate these video quality requirements. This was a financial obstacle. If this obstacle had not been identified early in the planning and design stages of the system deployment, it is likely that the CCTV system would not be fit for its intended purpose – to recognise faces.

2. **Low-end CCTV system:** A CCTV security system was installed in a small shop. There were no anti-vandalism measures taken to physically protect the CCTV cameras and wiring. The CCTV cameras had been positioned within arm's reach in order to capture images of faces, with the intention of using this video to identify criminals entering the shop. However, the positioning posed a risk of vandalism to the CCTV cameras. This obstacle was created as a result of a conflict between the CCTV owner's goals and the requirement to capture good quality close-ups of faces for identification purposes.

Process: Potential obstacles which may affect the capabilities of the system should be considered as early as possible. This can be achieved by reviewing the items identified in Phases 4 and 5. Once this review is complete, countermeasures should be devised to avoid and minimise the risks associated with each obstacle identified. To decide which obstacle to deal with first, they should be rated using a rating[34] scale. The ratings in this scale should be based on the potential frequency, impact and persistency of the obstacle. Those obstacles with a rating of four and above should be treated as serious and given high priority for resolution. Those with a rating of three or below should be noted and resolved later, where time and funding permit. In addition to determining the significance of a problem, the risk of an obstacle can be determined by making a cost-benefit assessment in context. The owner should consider the impact the obstacle will have on the user and whether or not the cost will be high in terms of loss in task performance and increase in task demand.

10.2.7 Phase 7: Identify and Resolve Potential Stakeholder Conflicts

Objective: It is likely that there will be conflicts in the stakeholder requirements, as each stakeholder will have different goals for the system, which may lead to different requirements for the system. For instance, stakeholders may disagree on the priorities of different requirements. Often these disagreements are a result of functional requirements desired by some stakeholders, which may be undesirable to others. The objective of this phase is to reduce the likelihood of system and business failures caused by conflicts between stakeholder requirements. Following are two examples of potential stakeholder conflicts:

1. **Lack of trust between store staff:** When a CCTV security system is installed in a store, the staff may feel they have no privacy, as the CCTV cameras will be watching their actions continuously whilst they are at work. This may create a lack of trust between store staff and management.
2. **Conflict between CCTV owner and installer's technical requirements:** There may be a conflict between the CCTV owner's requirements and the installer's technical advice. For instance, the installer may want to position a large number of CCTV cameras in a certain area of the store; the store owner may disagree and only want to install a small number of cameras

[34] Suggested significance scale: (1 – extremely minor; 2 – minor; 3 – moderate; 4 – serious and 5 – very serious).

where he/she thinks there is a need (based on previous experience with crime on the premises).

Process: Stakeholder conflicts can be identified by reviewing Phases 1–6. Once the conflicts are identified, each should be given a rating. Conflicts rated four and above should be resolved immediately, and conflicts rated three and below should be noted and resolved at a later date. Although this process is a hypothetical exercise, it is extremely valuable for identifying risks and vulnerabilities in the early stages of a CCTV deployment.

10.2.8 Phase 8: Restate Goals and Tasks

Objective: By completing Phases 1–7, the following should now be defined:

- CCTV security system goals
- Stakeholders belonging to the system
- Stakeholder goals (business and security)
- Stakeholder tasks, activities, skills, and knowledge
- Video quality requirements (i.e., storage, display and network)
- Technical set-up of CCTV security system (system capabilities)
- Potential obstacles that may alter/disrupt the system's capabilities
- Potential stakeholder conflicts that may disrupt the system

In the final phase of the TEC-VIS framework, the system owner should perform a holistic review of the requirements for the CCTV system as identified in Phases 1–7. The objective of this review is to restate the stakeholder and system requirements if necessary, so they are in line with one another. The objective of restating the goals and tasks is to ensure that the CCTV owner does not lose sight of the primary goals for the CCTV system.

Process: Restating the CCTV goals, stakeholders, tasks, and system capabilities is an iterative and analytical process which should be carried out throughout the lifecycle of the system. This process can be achieved by conducting a series of stakeholder meetings so that the requirements for the security system accommodate *all* of the stakeholders. This exercise should determine whether the requirements are realistic, and also whether they are financially and technically feasible.

10.3 Application of TEC-VIS Framework

The TEC-VIS framework was developed following field and empirical research into CCTV effectiveness. Principally, the framework informs practice. There was no opportunity to test the framework with a real CCTV deployment. This was a limitation of the framework. Instead, to demonstrate how this framework can be applied in practice, an *example of a CCTV deployment* scenario is next presented and the completion of the 8-phases in TEC-VIS is illustrated.

In this scenario, a business owner wishes to deploy a digital CCTV security system. This scenario is based on a real case study.

Scenario: Jewellery Store Owner to Deploy a CCTV Video Security System

A town centre jewellery store owner, Mr Elton, is suffering from a huge financial loss, which he suspects is the result of employee/staff theft, as well as customer theft. At present, the shop uses security mirrors to periodically monitor staff and customer activity, whilst the owner serves customers at the cashier till. In an attempt to recover the losses, Mr Elton wants to replace these security mirrors with a digital CCTV security system for three main reasons: 1) to deter thieves; 2) to detect theft by monitoring blind spots within the store, which the store mirrors cannot show; and 3) to use recorded CCTV video to help the police investigate incidents and to catch and prosecute criminals quickly.

The proposed CCTV system is to be installed by a private CCTV installation company. Mr Elton also wants to sign-up to a 'shop watch' radio scheme, which will enable him and his staff to report any incidents through the use of a point-to-point radio contact system. The radio system will be linked to a CCTV control room two miles from the store. CCTV operators working at the control room will provide immediate assistance to Mr Elton and his staff once an incident is reported by radio, and will help locate perpetrators by monitoring street CCTV cameras near the store.

Phase 1

Mr Elton's security goal:

- Use CCTV as a tool in the store to protect the jewellery stock and cash

Mr Elton's business goals:

- Avoid making insurance claims
- Reduce the costs of employing security staff to monitor activity in the shop

Mr Elton's sub-security goals:

- Deter thieves/intruders
- Make himself and the staff members feel safe
- Make his customer feel safe

Table 10.6: Identification of Stakeholders' Goals and Tasks

Phase 2	Phase 2	Phase 3
Stakeholders	**Stakeholders' Security & Business Goals**	**Stakeholders' Tasks**
CCTV System Owner .	▪ Reduce stock loss ▪ Deter thieves/other criminals ▪ Feel safe and secure ▪ Lower insurance premiums ▪ Ensure customers feel safe ▪ Monitor activities of shop staff ▪ Record CCTV for evidential purposes ▪ Buy a CCTV system that is affordable, usable and effective	▪ Monitor shop for staff and customer theft. ▪ Record CCTV video to serve as evidence for the police.
Staff	▪ To feel safe	▪ Monitor shop for customer theft. ▪ Communicate with CCTV operators to catch thieves.
Customers	▪ To feel safe	▪ Use shops where steps have been taken to minimise risk of robbery or violence.
Salesperson	▪ Gain a sale ▪ Gain customer loyalty through future upgrades	▪ Promote additional security products to the customer. ▪ Establish regular customer contact to build good rapport for future sales and advice.
Installer	▪ Install as soon as possible ▪ Make further sales through additional upgrade installation work	▪ Survey the store. ▪ Establish customer requirements. ▪ Install system. ▪ Test functionality against task performance and error criteria. ▪ Provide store owner with installation advice. ▪ Provide after-installation support.
CCTV Operator	▪ To support store staff using public CCTV cameras to locate and identify thieves.	▪ Respond to radio calls from store staff and locate perpetrators using external street CCTV cameras. ▪ Contact shop to warn if a known or suspicious perpetrator is nearby.
Police	▪ Ensure shop is recording CCTV video for evidential purposes.	▪ Seize CCTV video footage following an incident. ▪ Detect incidents on recorded CCTV video footage. ▪ Identify perpetrators for prosecution using post-event recorded CCTV video.

Table 10.7: Identification of System Capabilities and Obstacles

<table>
<tr><th colspan="2">Phase 4: Identify & Assess Video Quality</th></tr>
<tr><td colspan="2">The security system will be used for detecting theft and identifying thieves (event detection and face identification tasks) within a high-risk environment. The CCTV video display and network for this system should be configured with a temporal resolution set at 12 fps, resolution 352x288, and spatial video quality at 2000 Kbps (MPEG-4).</td></tr>
<tr><th>Phase 5: Identify Capabilities of System</th><th>Phase 6: Identify and Eliminate Potential Obstacles to Achieving System Goals</th></tr>
<tr><td>1. Observe internal and external store activity on video monitors.
2. Allow CCTV operators to access the external CCTV cameras to support any incidents.
3. Use Pan, Tilt, and Zoom (PTZ) cameras and controls to inspect customer and staff in the store who appear suspicious.
4. For observing live CCTV video, the following target sizes should be applied [2]:
o Monitor and control: 5% of screen height
o Detect: 10% of screen height
o Identify: 50% of screen height
o Recognise: 120% of screen height
5. Record digital CCTV video at 12 frames per second or above.
6. Use a motion sensor during store closing time, which will automatically record CCTV video in full frame rate if any movement is detected.
7. Use a two-way radio system to communicate directly with CCTV operators at a local CCTV control room, so that support is available during, before, and after an incident in the store.</td><td>1. The owner has high budgetary constraints for CCTV deployment. As a consequence, the digital recording system has limited storage space and cannot record CCTV video in full quality; therefore, the video cannot be used for investigating incidents.
Solution
Develop a site plan; decide on an initial number of essential CCTV cameras; and configure the CCTV system to full quality. Invest in more CCTV cameras later, when funding is available.
2. The owner has little knowledge of CCTV systems and the CCTV salesperson sold the owner a system that is not very usable and lacks sufficient operating instructions. Consequently, the system’s performance is poor, since it is ineffectively configured and receives limited system maintenance.
Solution
The CCTV owner should attempt to understand the system following its installation and should liaise with the technical installation team as closely as possible during the sale.
The CCTV owner should test the system and practice using the digital video recorder to become familiar with the system.
3. Decorative banners/signage, hanging baskets and overgrown trees may occlude CCTV cameras outside the store and limit surveillance coverage.
Solution
In advance, consult those responsible for environmental maintenance and request they cut back trees or remove them altogether.
4. Interior shop lighting and strong sunlight may alter the picture quality.
Solution
Good lighting conditions are important for the identification of targets and for the detection of targets and target activities.
Alter the position of internal cameras and install a retractable roof on shop exterior to protect from sunlight.</td></tr>
</table>

Table 10.8: Identification of Stakeholder Conflicts and Resolutions

Stakeholder	Phase 7	Phase 7
	Stakeholder Conflicts	**Potential Resolutions**
CCTV Owner .	1. Disagreements with other stakeholders; the owner may not understand the significance of advice given by salesman and installer. 2. Technical failures with CCTV system once bought.	▪ Take time to understand the system. ▪ Obtain third-party advice and reviews. ▪ Obtain more than one proposal; identify common themes and resolve discrepancies. ▪ Resolve technical failures with 24-hour support from installers.
Staff	1. Staff may feel they have no privacy and that they are being continuously watched. There may be a lack of trust between store staff and the CCTV owner.	▪ Assure staff the CCTV system has been set-up to monitor shoplifting activities of customers and possible staff theft within the store. ▪ Assure staff that no sound will be recorded by the surveillance system (if this is the case).
Customers	1. Customers may feel that CCTV cameras are intrusive. 2. Risk of vandalism by members of public.	▪ Place CCTV signs within and outside the store informing that a CCTV system is in operation and is set-up to record and monitor events for the purpose of crime detection. A name and telephone number should also be provided. ▪ Install cameras with anti-climbing fences; use anti-graffiti paint; and install conduit to protect CCTV camera cabling and wires.
CCTV Salesperson	1. Salesperson may disagree with owner's requirements. 2. Salesperson may sell the owner an unsuitable CCTV system.	▪ Exchange owner's requirements at the start of the sales consultation. ▪ Ensure a refund can be made within 30 days of the installation date.
CCTV Installer	1. Conflicts of interest between owner's requirements and installer's technical advice; possible misunderstanding between installer and owner.	▪ Reach an understanding with the installation team and ensure that all of the CCTV owner's requirements for the CCTV system are considered.
CCTV Operator	1. Poor communication between shop staff and operators.	▪ Ensure that a radio and telephone communication protocol is in use; see [11].
Police	1. Low level of communication with CCTV control room operators. 2. CCTV video cannot be used and does not meet police requirements in terms of video quality, content and format.	▪ Police should meet with operators weekly to exchange information on crime and to increase operators' awareness/knowledge. ▪ Make documentation available to store owner and assist police staff in retrieving video data.

10.4 TEC-VIS Framework References

[1] Freeman, R. E. (1994). Strategic management: A stakeholder approach. Boston, MA: Pitman.

[2] Aldridge, J. (1994). CCTV operational requirements manual, Version 3, Police Scientific Development Branch (PSDB) Publication.

[3] CCTV Systems and the Data Protection Act - Code of Practice. (2008).

[4] Cohen, N., Gatusso, J., and MacLennan-Brown, K. (2006). CCTV operational requirements manual - Is your CCTV system fit for purpose? Version 4. Home Office Scientific Development Branch (HOSDB) Publication.

[5] Wallace, E and Diffley, C. (1998). Making it work: CCTV control room ergonomics. Police Scientific Development Branch publication.

[6] Shepherd, A. (2001). Hierarchical task analysis. London; New York: Taylor and Francis.

[7] Keval, H., Gatusso, J., and Maclennan, K. B. (2007). Did you see what happened? A study on video frame rates for detecting events from CCTV video. A poster presented at the International Crime Science Conference, London.

[8] Keval, H.U., Sasse, M. A. (2008b). Can we ID from CCTV? Image quality in digital CCTV and face identification performance. In *Proceedings of SPIE series*. SPIE.SPIE Mobile Multimedia/Image Processing, Security, and Applications, Agaian,S.S., Jassim,S. A. (ed.).

[9] UK police requirements for digital CCTV systems. (2005). Home Office Police Scientific Development Branch and ACPO.

[10] Walters, P. E. (1995). CCTV systems thinking – systems practice. In Proceeding of European Convention on Security and Detection, Brighton, UK: IEE, pp. 64–69.

[11] Keval, H. and Sasse, A. (2006). Man or gorilla? Performance issues with CCTV technology in security control rooms. In Proceedings of International Ergonomics Association, 16th World Congress on Ergonomics Conference, Maastricht, Netherlands.

[12] Keval, H. (2006). CCTV control room collaboration and communication: Does it work? In Proceedings of Human Centred Technology Workshop, Brighton, UK.

[13] Keval,H., Sasse, M. A. (2008a). "Not the Usual Suspects": A Study of Factors Reducing the Effectiveness of CCTV. Security Journal, ISSN: 0955-1662.

[14] ISO. (2004). BS EN ISO 11064-4: Ergonomic design of control centres. Part 4, Layout and Dimensions of workstations. Geneva: International Standards Organisation.

[15] Tickner, A.H and Poulton, E.C. (1973). Monitoring up to 16 synthetic television pictures showing a great deal of movement. Ergonomics, 14 (4).

[16] HSE, Health and Safety. (1992). Display screen equipment regulations, regulation 4.

[17] Wood, J. (2001). Part III Control room design: Control room mock-up trials. In J. Noyes and M. Bransby (Eds.), People in control: Human factors in control room design (pp. 189-206). London, United Kingdom: IEE.

10.5 Chapter Summary

The framework detailed in this chapter, TEC-VIS, is the guiding methodology for the effective design, configuration, and use of digital CCTV. This framework was developed from the results of the field research conducted in 14 public-space CCTV control rooms (see Chapter 4). The video quality guidance provided in Phase 4 was drawn from the research findings from two CCTV video quality experiments (see Chapters 8 and 9). The guidance also draws upon previous research and standards in CCTV security and control room design. The framework is intended for use by CCTV security practitioners and owners in identifying potential CCTV stakeholders, their business and security goals and task requirements, as well as system capabilities and associated risks (potential technical obstacles and stakeholder conflicts). The framework provides guidance in four key areas: 1) video storage, transmission and display requirements for security observation tasks; 2) internal security environment requirements; 3) CCTV camera and monitor guidance; and 4) CCTV system and strategy design.

The guidance provided in this framework should be applied during the design stages of a new CCTV deployment or during the re-design of an existing CCTV system. Once the framework has been completed, the system should be evaluated with the intended CCTV users and tasks to establish whether the chosen configuration is effective, and therefore fit for purpose. As part of the evaluation, it is emphasised that owners should consider CCTV operator factors (e.g., task environment set-up, training, knowledge, motivation and job satisfaction) to determine the biggest gains and losses in performance and where they occur. This evaluation exercise is also useful in weighing the costs associated with having a system with high performance. The TEC-VIS framework can be used as a best-practice guide to support the planning, design, and deployment phases of *any* type of CCTV security system, regardless of its context. The phases described in TEC-VIS should be completed before deployment, after deployment and every time the system is altered. The TEC-VIS framework is presented as the research contributions of this thesis. To validate the guidance contained in TEC-VIS, three international experts have critically reviewed the framework. This was a necessary part of the research process to strengthen the framework. The review is discussed in detail in the following chapter.

Chapter 11

Expert Evaluation of the TEC-VIS Framework

The aim of the TEC-VIS framework is to help designers and owners of CCTV systems to configure systems that effectively support security tasks. Ideally TEC-VIS would be evaluated by applying the framework and guidance to a real-world CCTV system deployment. However, this was not feasible within the timeframe of the thesis. Instead, an expert evaluation was conducted with three independent reviewers with knowledge and expertise in HCI, ergonomics and security. These experts were asked to complete a review of the TEC-VIS framework to validate its content and scope. The evaluation rationale, background information of each reviewer, and a discussion of the completed reviews are presented in this chapter, together with the results of the evaluation. The chapter finishes with a summary of the limitations of TEC-VIS and a number of recommendations for the future development of the framework.

11.1 Expert Evaluation Rationale

The review of the framework was carried out by three reviewers who had knowledge and experience in HCI, ergonomics and security. All reviewers had a complete academic background and one reviewer had vast commercial experience (reviewer 3). The purpose of the expert review was to identify areas requiring improvement and to identify any substantive or methodological gaps, so that TEC-VIS could be developed into a coherent, complete framework that can be applied to all types of CCTV security systems. A number of constructive comments and recommendations on specific aspects of the framework were taken forward for necessary changes to TEC-VIS. These changes are reflected in version 2, presented in Chapter 10. All of the changes required to create version 2 are listed in Table 11.1.

The framework is split into two types of guidance:

1. Conceptual based guidance, on:
    a. how to carry out a user analysis, identifying the different stakeholders, and their goals and tasks, and what support is required for each, and
    b. how to remove technical barriers and potential stakeholder conflicts that may reduce the effectiveness of the system.
2. Empirical based guidance, which provides:
    a. a prescriptive set of recommendations on digital CCTV video storage, transmission and display requirements for two key security observation tasks;
    b. recommendations for configuring the internal (CCTV camera and monitor) and the external CCTV camera environment; and
    c. guidance on CCTV system and strategy design.

11.2 Expert Reviewer Profiles

Table 11.1 provides background information on the three evaluators who reviewed TEC-VIS. All reviewers had knowledge and expertise in the field of HCI, ergonomics and security.

Table 11.1: TEC-VIS Expert Reviewers: Background and Experience

Reviewer: Background & Experience	Current Research, Interests and Expertise
Reviewer 1: Dr Iain Darker Senior Researcher Ergonomics Safety Research Institute Loughborough University, UK Experience: - Ergonomics and HCI: 10 years - Usability: 2 years - Security: 3 years	**Current area of work:** Human perception and cognition as they relate to the detection of concealed weapons via CCTV (a 5-year EPSRC project: 'Medusa'). **Areas of interest/expertise:** - Human experimental psychology (cognitive, applied, and neuroscience). - Research methods (quantitative and qualitative analysis, statistics, programming, and experimental design). - Social policy.
Reviewer 2: Dr Martin Maguire Senior Researcher and Lecturer Ergonomics Safety Research Institute Loughborough University, UK. Experience: - Ergonomics and HCI: 25 years - Usability: 20 years - Security: 3 years	**Current area of work:** - Graphical user interface (GUI) style guide for process plant systems for a public sector organisation. - Ergonomic assessments for Home Office offender management systems. - Teaching: Ergonomics and HCI for undergraduate and postgraduate students. **Areas of interest/expertise:** - Applying the processes of user-centred design (UCD) and user interface design in a practical way and with limited resources. - Usability of software user interfaces. - Designing systems for inexperienced users e.g., kiosks, websites, etc. - Accessibility of websites.
Reviewer 3: Dr Ian Nimmo President of User Centered Design Services Inc. (USA) Experience: - Ergonomics and HCI: 16 years - Usability: 38 years - Security: 10 years	**Current area of work:** - Control room and command and control centres design. - Past program director of Abnormal Situation Management Consortium. - HCI design and alarm management - Development of standards: ISA SP18, SP101, SP8, SP99. **Areas of interest/expertise:** - Best-practice requirements for alarm management, HCI, and control room design. - Directed a program with 50+ PhD students researching abnormal situations in multiple industries.

- Reviewer 1 has:
    - Academic background: BSc. Neuroscience and PhD in Cognitive psychology.
    - Membership: Institute of Engineering and Technology and member of the Applied Vision Association.
    - HCI research: 1 year post doctoral experience in social policy as a research associate and 2 years post doctoral experience in applied vision science as a research associate.
    - Specific methods and skills: human experimental psychology, quantitative and qualitative analysis, statistics, and experimental design and analysis.
    - Specific experience in CCTV security: working as a senior researcher on a 5-year CCTV research project investigating the human perception and cognitive factors when detecting concealed weapons from CCTV video. So far, gained 2 years post doctoral experience in applied vision science and gained through employment as a research associate on the MEDUSA project. MEDUSA is concerned with the identification of situations associated with gun related threats, based on behavioural interpretation of CCTV data and through combining psychological and image processing approaches.

- Reviewer 2 has:
    - Academic background: BSc. Computer Science and PhD in Ergonomics (HCI).
    - HCI research: 25 years of research experience working within a Human Sciences research group (HUSAT and ESRI at Loughborough University) as a research associate and research fellow.
    - Specific methods and skills: field research with users and participants, statistics, user interface design, observation, questionnaire design and administration, focus group organisation and moderation, conducting usability evaluations of several products, websites and software systems.
    - Specific experience in CCTV security: No specific CCTV experience but related experience in multimedia/video based systems - developing guidance for designers (1998) and a project on online security to develop an understanding of different kinds of fraud and phishing (2007).

- Reviewer 3 has:
    - Academic background: BSc. Electrical engineering and PhD in Ergonomics and Cognitive Sciences.
    - HCI research experience: International lecturer in ergonomics and HCI, Program Director for 10 years: NIST Advanced Technology Program - Abnormal Situation

Management Consortium, academic supervision of PhD students at Honeywell technology centre – Purdue University, Ohio State University, UCLA, and Toronto University, and 10 years research work in control room environment research for the US Pilots Association.

- Specific methods and skills: field research, statistics analysis, conducting ergonomics and usability evaluation within high risk and security environments.
- Specific experience in CCTV security: 8 years research consultancy in ergonomic design of control rooms in multiple industries such as: including refining, petrochemical, mining and minerals, pharmaceutical, offshore, silicon, and specialty chemicals. Specialist areas: human machine interface design, alarm management, staffing task analysis, staffing workload assessment and security.

11.3 Review Structure for Evaluation

Each reviewer was emailed a briefing form (Appendix I), which included a: 1) summary of the framework; 2) the rationale for the framework; and 3) the evaluation criteria. The feedback provided by each of the reviewers is included in Appendix J. The experts were asked to evaluate the framework by considering the criteria detailed in Table 11.2.

Table 11.2: TEC-VIS Evaluation Criteria Provided to Expert Reviewers

TEC-VIS Framework Evaluation Criteria (Questions)
1. Comprehensiveness
1a. Does the framework consider all CCTV stakeholders? If not, who is missing? 1b. Does the framework consider all of the social factors related to a CCTV deployment? If not, what is missing? 1c. Does the framework consider all of the technical factors related to a CCTV deployment? (Are all of the digital aspects of CCTV security considered? If not, what is missing?)
2. Style and Language
How well can the framework be understood by different CCTV stakeholders? (Is it suitable for all CCTV users with different levels of experience and knowledge of CCTV?)
3. Applicability
3a. Do you think the framework can be applied to a real-world CCTV deployment? 3b. Do you think the framework can be applied to all types of CCTV deployments?

11.4 A Critical Review of the Framework

The intentions of the review were to use the results to: 1) determine whether the framework could be used effectively by CCTV owners and consultants, to support in the deployment of effective CCTV systems. Specifically, the review was completed to assess how well the processes described in each of the eight phases could be applied in the real-world. The results were also used to 2) identify areas of the framework that could be improved through further research and development.

In the following section, specific feedback provided by the three reviewers, together with a discussion of the feedback are presented for each of the three criterion established for evaluation.

11.4.1 Comprehensiveness – Stakeholders

Reviewer 1 found the framework to be comprehensive (in terms of key CCTV stakeholders), but suggested the inclusion of two additional stakeholder groups:

1) Media staff: **Reviewer 1** assumed 'media users' fell under the CCTV stakeholder classification, and believed that these users were responsible for approaching data controllers to collect CCTV data; this assumption was incorrect. Media staff are *not* be responsible for collecting CCTV data directly from data controllers, such as CCTV control room operators or a private CCTV owner. Instead, the police have the responsibility and authority to request CCTV video footage from data controllers - not media staff. The police typically liaise with the media when a crime is publicised locally and nationally (e.g., on the news, through campaigns and in newspapers).

2) Support staff (e.g., cleaning staff): also thought that support staff were missing in the list of CCTV stakeholders in the TEC-VIS framework. These individuals were not included in the list of CCTV stakeholders, as they do not have an impact on a CCTV system, nor do they interact with a CCTV security system in any way. Support staff members may work within a CCTV environment, but do not have set goals for the system or perform tasks with CCTV. The only possible consideration is that cleaning rotas may change if the physical set-up of the security environment is modified. This change however is unlikely to have any impact on the performance of the CCTV security system.

Reviewer 2 believed that Phase 2 of the framework was *"…comprehensive and presents a full range of stakeholders ..."* The identification of CCTV stakeholders was considered to be a useful starting point for the framework user. It can then be used to detail the names of specific users, such as suppliers and maintainers, and be used later for reference. Recording each of the CCTV stakeholders' names was a presumed step in Phase 2; however, this procedure was not included in the process description in the framework. Recording each CCTV stakeholder's name and contact details, as well as their expertise is useful and will facilitate communication between the various CCTV stakeholders. Also, the names will allow new CCTV users to review and understand the stakeholder hierarchy and the range of expertise of users across the system. Detailing CCTV user contact information during the analysis will no doubt improve communication and support; this is most effective when the information is accurate and up-to-date.

Reviewer 3 indicated that there was one significant stakeholder group missing: IT managers/professionals (particularly for digital and networked CCTV systems). These users were assumed the same as a CCTV consultant, installation, and maintenance users. The framework did not describe the exact role of each stakeholder listed; therefore, **Reviewer 3** assumed that IT managers/professionals were not considered, thus included in the list of CCTV stakeholders. In the UK, the management of CCTV data and a network is a role and responsibility which can be shared amongst several CCTV stakeholders (e.g., CCTV owner, consultants and security personnel). The roles and responsibilities for CCTV users in the US may differ from UK CCTV users. It is likely that **Reviewer 3** made this assumption as he is a US citizen and may not be aware that several CCTV stakeholders could have potentially more than one role. To ensure that there are no misunderstandings about the roles of the different stakeholders for a CCTV security system, version 2 of TEC-VIS describes the basic role of 10 key CCTV stakeholders.

11.4.2 Comprehensiveness – Social Factors

Reviewer 1 provided positive feedback indicating that the framework considered all of the social factors within the stakeholder analysis (Phase 2). However, both **Reviewers 1 and 2** suggested that the framework could have discussed some of the specific social issues related to CCTV and task performance. For instance, **Reviewer 2** commented that the framework could have included guidance on CCTV operator recruitment and training, and selection of CCTV users. These aspects are important, as they affect a CCTV user's task performance.

Specific social factors that alter a CCTV user's security observation task performance, such as staffing relationships, job motivation and job satisfaction, were not areas intended for discussion. These factors were excluded from the framework, since they were not part of the scope of the research for this thesis. Instead, the Human-Computer Interaction (HCI) factors concerning the design, configuration, and use of CCTV systems were examined. Although the research in this thesis focuses on the technical aspects of CCTV and their effect on different CCTV stakeholders, the TEC-VIS framework does acknowledge these social issues (i.e., recruitment, training and selection) within Phase 5, and references previous research and guidance (Wallace, Diffley, Baines and Aldridge, 1997).

Reviewer 1 offered two examples of other social factors which could have been considered:

> *"High-profile examples could be used in order to highlight the noteworthy, socio-technical approach adopted by TEC-VIS. For instance, consideration might be given to the advantages inherent in facilitating co-operation between CCTV operators in tasks such as tracking targets across multiple CCTV cameras, or to the importance of providing feedback to CCTV operators on their performance."*

In response to this comment:

1. The framework was not developed specifically for CCTV control room owners, but for all CCTV owners. It is possible that the framework was written with a stronger emphasis on CCTV control room systems, due to the examples and references presented. (Phase 5 provides specific recommendations that apply primarily to CCTV control room users). As a result, **Reviewers 1 and 2** perceived TEC-VIS to be a framework on CCTV control rooms. The introduction at the beginning of Phase 1 in Version 2 now clearly states that the framework is designed for all CCTV owners and security practitioners, and is not provided exclusively for CCTV control room owners.
2. The two examples given by **Reviewer 1** are noteworthy; however, these examples were not included in the framework as these aspects were not covered for this research thesis. The examples (highlighting the advantages of facilitating operators when tracking targets across multiple CCTV cameras and the importance of providing feedback to operators on their performance) are both regarded as highly important in supporting CCTV user performance, as well as other CCTV users who perform security tasks using CCTV video and tools. To cover these issues and provide relevant guidance requires further research.

Reviewer 3 made no comment under question 1b.

11.4.3 Comprehensiveness – Technical Factors

All reviewers felt that the framework addressed the main technical issues. The only exception was the lack of guidance on video storage and physical space when recording CCTV video (Phase 4).

Reviewer 1:

> *"The framework considers some factors including recorded image quality, but does not consider other technical issues which might be related to such factors. For example, the calculation of the digital storage capacity necessary to accommodate these images is not addressed."*

Reviewer 2:

> *"Some other technical factors you could cover are…what amount of storage is needed for data recording."*

The specific amount of storage required for recording CCTV video was intentionally not covered in the guidance provided in Phase 4, as the amount of storage required depends on many factors, such as the type of recording system, type of video CODEC used, processing power, and so on. Instead, guidance on the minimum video compression (bit rate) and frame rates were specifically provided for specific CCTV security observation tasks.

Reviewer 2 acknowledged that guidance on video storage for digital CCTV systems exists and recommends that *"…there may be value in directing users of TEC-VIS to the work of Cohen et al. (2007) whilst highlighting where TEC-VIS extends this work."*

This point was well made. Within Phase 4 (version 2 of TEC-VIS), a reference to existing guidance on digital CCTV video storage, replay, and disposal is added. In addition, a note has been made that TEC-VIS offers an extension to the existing guidance by providing concrete recommendations for two common security observation tasks (face identification and detection).

Reviewer 3 made the comment that *"Phase 4 video storage should also discuss technologies from tape to video digital recorders to video streaming to a server and the advantages and disadvantages of the technologies. Perhaps it would help to identify who should provide information for Phases 5 and 6."*

As mentioned previously, the framework was unintentionally geared towards CCTV control room managers. The framework content may have included technical jargon, and in some places CCTV technology knowledge was assumed. In version 2, Phase 5 now provides a short overview of analogue (tape-based), digital and networked CCTV systems, as well as the advantages and disadvantages of each.

All reviewers found that the framework did not consider the issues surrounding the configuration of video monitors for security observations tasks. For instance, there was no guidance given on selecting monitors (e.g., LCD, CRT, or large display screens), as the research for this thesis did not assess the effectiveness of different video monitors for security observation tasks. The ISO (2004) guidance only mentions the following on displays:

"Control workstations equipped with multiple displays, i.e. typical table top or console mounted VDU´s (CRTs, Flat Panel LCDs) etc., require special attention regarding placement and layout."

If guidance had been provided, it would have been based on anecdotal evidence, and would have been unreliable. To date, there have been no evaluations carried out to assess task performance using different video monitors. In addition to monitor and screen guidance, **Reviewer 3** felt that the framework should have considered automated tools, such as automatic number plate recognition (ANPR) systems, and other tools used in conjunction with CCTV. Again, as the research presented in this thesis did not address the effectiveness of automated CCTV tools and systems, no guidance was provided in this area.

11.4.4 Style and Language

In general, **Reviewers 1 and 2** found the framework style and language clear, concise, and practical. However **Reviewers 2 and 3** commented on the occasional use of technical jargon and language, which they felt could discourage users who are not CCTV expert users. **Reviewer 3,** in particular, perceived the framework to be used by every single CCTV stakeholder; this is not the case. Version 1 does in fact state that the framework is intended for security practitioners, designers and CCTV system owners who are responsible for the design, development, and deployment of a security control room – where CCTV technology is used (both analogue and digital) (see Appendix H, Section 2).

Reviewer 2 noted that the document was mainly free of jargon; however, there was some use of technical jargon that related to the research discipline on which the framework was based, Human-Computer Interaction. **Reviewer 2** also commented that the academic style of referencing may not be suitable for practitioner audiences. In retrospect, the framework style and language was considered too technical and research oriented in places. In version 2, the style and language is altered to accommodate non-CCTV experts and non-academic readers. For instance, technical concepts and jargon have been described and the academic Harvard reference style has been altered (author, year) to a simple number system (e.g., Keval et al*.,* 2007 to [1]). This has made TEC-VIS a user friendly document for all CCTV stakeholders to comprehend and use.

Reviewer 3 said that:

> *"It is suitable for a large company security manager, but a store owner would find many of the concepts foreign. CCTV professionals will understand this and may be able to use it as a guideline for new customer projects but a new owner would find it difficult in its current format. I think you should decide who the user of the document is and write specifically for them. It would be difficult for other stakeholders to comprehend."*

The main difficulty with presenting technical guidance for the deployment of a CCTV system is that CCTV users have different knowledge and skills. Therefore, striking the proper balance in providing guidance on such a complex and technical field can be very challenging. In TEC-VIS, this challenge was overcome by simplifying each of the eight phases and describing them under 'objectives' and 'processes' for each phase. Under the objectives, the point of the phase is given (what is it) and under the process heading, the process involved to complete the phase is detailed (how you achieve it). Within the framework, Phase 2 emphasises the importance of considering all of the CCTV stakeholders during the planning of a CCTV security system. Collecting input from different CCTV stakeholders through regular stakeholders meetings is essential for effective communication and collaboration for any socio-technical system.

The TEC-VIS framework is intended for completion by CCTV owners and security practitioners (if applicable), as they are responsible for the operation and management of the system. It is the CCTV owner's role and responsibility to complete the TEC-VIS framework documentation and share it with all CCTV stakeholders involved within the system, as identified in Phase 2, for the purpose of improving communication and awareness. In version 2, a note has been added to the introduction and conclusion sections which clearly state that the TEC-VIS document should be accessible to all CCTV stakeholders. In addition, the following note has been added: 'The CCTV security owner(s) should have the right to manage the framework document and any changes made should be recorded.

11.4.5 Applicability

All reviewers found the TEC-VIS framework applicable to real-world CCTV deployments.

Reviewer 1:

> *"The framework appears sufficiently generic to facilitate its application in many contexts."*
> *"TEC-VIS appears applicable to real CCTV deployments."*
> *"The illustrative example provided is very useful in demonstrating the potential for applying TEC-VIS in practice."*

Reviewer 2:

> *"I am sure that the framework would be of interest to anyone deploying CCTV. It is perhaps most appropriate for larger scale operations where human factors expertise is available."*

Reviewer 3:

> *"I think the framework is very useful and applicable to larger security projects."*

It is worth noting that **Reviewers 2 and 3** found TEC-VIS *more* applicable to large-scale deployments than to smaller set-ups.

TEC-VIS was developed with the aim of providing both conceptual and best-practice recommendations for CCTV owners and security practitioners. The guidance provided within the framework was intended for all CCTV users working on all types of security systems. No similar guidance exists, and this is likely since it is very difficult to provide generic guidance on the design, configuration, and use of every type of CCTV system – small, medium and large. To overcome any misconceptions, the following notes have been added to the framework to emphasise that:

1. not all CCTV security systems operate in the same way,
2. all processes for each phase are intended for implementation by CCTV owners and security practitioners, regardless of system size or context, and
3. in Phase 4, the video quality recommendation should be considered a minimum recommendation and not an absolute requirement.

Furthermore, throughout the document, examples have been added where possible for small, medium and large CCTV security systems to illustrate the concepts and processes for all CCTV owners. For instance, in Phase 2, examples of the likely attendees for a stakeholder feedback session are now stated for small, medium and large scale operations. This illustration was noted as it was a point raised by all three reviewers:

Reviewer 1:

> *"When it comes to performing a stakeholder analysis and holding regular stakeholder meetings, you might indicate what is practical for say the smaller CCTV owner (say for a single shop) as opposed to the larger scale operation. Maybe you could indicate an example group of individuals who could be invited to the stakeholder meeting."*

Reviewer 2:

> *"TEC-VIS users might benefit from multiple examples covering the most common applications. For instance, an example of the use of TEC-VIS in a local authority CCTV control room would make the framework more accessible to a large sector of CCTV users."*

Reviewer 3:

> *"I think the framework is very useful but needs to be put into context for each of the stakeholders and a generic document would have to be more descriptive and have clearer definitions as to who does what."*

11.5 Positive Elements of TEC-VIS Framework

All of the expert reviewers indicated that the TEC-VIS framework met all of the criteria identified in Table 11.2. All of the reviewers also found the framework to be a valuable and highly important contribution.

Reviewer 1 acknowledged the merit of the TEC-VIS framework contribution, particularly Phase 2 (detailed stakeholder analysis) and the scenario that demonstrated its applicability:

> *"The detailed stakeholder analysis is a particular strength of TEC-VIS. Indeed, it is perhaps the most important and novel contribution of TEC-VIS to provide knowledge and guidance relating to the deployment of CCTV systems. In this respect, TEC-VIS makes a contribution over and above that offered by the latest UK government CCTV Operational Requirements Manual (Cohen, Gattuso, and Maclennan, 2007)."*

Reviewer 1 also commented on the use of a scenario to illustrate the application of the framework:

> *"The illustrative example provided is very useful in demonstrating the potential for applying TEC-VIS in practice."*

Reviewer 2 made a positive remark on the framework, stating that it is a practical guide for applying a user-centred design method to CCTV systems:

> *"It was interesting and enjoyable to read this document as it offers a practical guide to applying a user-centred design method to CCTV systems. Many other works just provide general advice or principles to follow."*

Reviewer 3 noted that the research work detailed in the framework is valuable for 'end users,' and is also a starting point for educating CCTV suppliers:

> *"The document you provided is work which is valuable. I see plenty of growth in the field you are addressing and the need for advice to end users will be extremely important, unfortunately most of the suppliers also need educating and what you have provided will be a great start for that. Well done and keep up the good work."*

11.6 Summary and Discussion of Evaluation Results

Reviewers were not provided with the field and empirical research chapters of this thesis, or the published papers from the research (see Chapter 12, Section 12.3.3) which describe the research and results used to develop TEC-VIS. This led to some reviewers expecting additional guidance in the framework, such as work psychology and monitor guidance. Many factors can affect the performance and effectiveness of a CCTV system (see Chapter 5, Section 5.1.2, Table 5.1).

The evaluation results were examined and a number of changes relevant to the scope of the framework were made to improve TEC-VIS further. These changes included the inclusiveness of all CCTV stakeholders and a consideration of the main social and technical factors necessary for the design, configuration, and use of a CCTV system. The framework was also altered so that the writing style and language was more attuned to CCTV professionals, rather than HCI and ergonomics professionals.

All reviewers found that the TEC-VIS framework applicable for real-world CCTV practice. Reviewers 2 and 3 however commented that it was most suitable for large scale CCTV operations and projects. This result was possibly due to the fact that much of the guidance within framework, particularly Phase 5 was geared towards CCTV control rooms systems (large CCTV systems). The evaluation of TEC-VIS was an extremely invaluable exercise and this validation made it possible to assess whether it was comprehensive, understood, and applicable in the real-world. Reviewer 1 was particularly appreciative that a scenario was used to exemplify the applicability of the framework.

The evaluation exercise was specifically undertaken to determine whether the framework could be used effectively by CCTV owners and consultants, to support them in deploying and configuring an effective CCTV system. A number of recommendations were taken forward to improve TEC-VIS further. Table 11.3 details the specific changes considered for the revised version of the framework (version 2), which is presented in Chapter 10.

11.7 Further Research

TEC-VIS was developed based on the research conducted for this thesis and presented in Chapters 4, 8, and 9. These studies were conducted to fulfil the research goals outlined in the introduction of this thesis (see Chapter, Section 1.3). The framework can be further developed through future research, both field and empirical, to address all of the social and technical factors that may reduce the effectiveness of a CCTV system. Over the course of the research presented, it has been realised that the scope of the framework could be broadened to provide the following additional guidance in support of CCTV security owners and consultants deploying or redesigning a CCTV system:

1. CCTV video quality requirements for a wider range of security observation tasks performed by a wider range of CCTV users in different security environments – a set of video quality requirements is needed for intelligent surveillance systems (automatic number plate recognition, face recognition, anomaly detection, etc.) to ensure that computer-based security systems also perform at a high level of accuracy.
2. Detailed guidance on the optimum configurations for specific CCTV technology, such as CCTV cameras (e.g., dome, shoe box and 3D), video recorders, camera controls, electronic geographical maps, camera databases, etc.
3. Detailed guidance on communication systems and tools (e.g., control room radio systems, telephones and handsets).
4. Detailed guidance selecting different video monitor displays for use in real-time and post-event security observation tasks.
5. Best-practice guidance on the social factors affecting CCTV users. Additional field and empirical research is needed to assess the impact of the following social factors on task performance: job skill, mental workload, shift work, levels of training, job motivation and stress.
6. Guidance on how a CCTV owner can carry out a cost-benefit analysis when deploying a CCTV system.

Table 11.3: Changes Made to TEC-VIS (version 1)

Reviewer No.	Criterion	Phase/Page No (Vers. 1)	Specific Change (Vers. 2)
2	Comprehensiveness 1a) Stakeholders	Phase 2/Page 5	In the process, the list of CCTV stakeholders identified now notes them by name, contact details, role, and expertise.
3	Comprehensiveness: 1a) Stakeholders	Phase 2/Page 4	In the objective section where Figure 1 is referenced, the roles of each CCTV user shown in the hierarchy are now described.
2	Comprehensiveness: 1a) Stakeholders	Throughout	The document now emphasises that TEC-VIS can be used by all types of CCTV owners and applied to all types of CCTV systems - not just CCTV control room systems.
1 and 2	Comprehensiveness: 1b) Social Factors	Phase 4/Page 8	In Table 2, a reference is now made to published research and guidance on the social factors that can improve CCTV user performance. Specifically, reference is now made to the CCTV operator recruitment and selection Home Office publication (Wallace et al., 1998).
1	Comprehensiveness: 1c) Technical Factors	Phase 4/Page 7	In Table 1, a reference is now made to published guidance on video storage for digital CCTV video (Cohen et al., 2007).
3	Comprehensiveness: 1c) Technical Factors	Phase 4/Page 7	A short overview is now included that describes the distinction between the three main types of CCTV security systems: (1) analogue; (2) digital; and (3) IP CCTV systems.
3	2) Style and Language	Introduction/Page 1 Summary/Page 20	The framework is now described as an open access document for all CCTV stakeholders; however, only system owners and authorised users can edit the document.
2 and 3	2) Style and Language	Throughout	Any technical CCTV and academic terms and acronyms are now explained. All published research and standards are now referenced in a numerical format (i.e., [X]).
2 and 3	3a) Applicability	Introduction/Page 1	To acknowledge that security systems are not all the same, the framework recommends that all processes be followed by all CCTV owners, regardless of the size and context of the system, and in Phase 4, the video quality requirements should be used as a minimum guide.
2	3a) Applicability	Throughout	Under each Phase, where applicable, examples have been provided to reflect the true set-up of small and large security systems.

To further illustrate the value and applicability of TEC-VIS, the framework's principles and guidance should be tested with a real-world CCTV security system deployment. The results of this evaluation can be used to further refine the framework and demonstrate the usefulness of TEC-VIS highlighting the key benefits of conducting a comprehensive and holistic analysis of a CCTV system deployment before an investment is made. If the TEC-VIS framework is used with a real-world CCTV system deployment (new or existing), each of the eight phases presented in the framework should be completed and documented by the CCTV owner or consultant, so that it can be used as a case study for other CCTV owners. The case study can then be used by those new to CCTV to gain an understanding on how to establish the requirements for a CCTV system. The use of case studies to further evaluate the framework was suggested by **Reviewer 1**: *"Future work might consider a case-study of a complete application of TEC-VIS."*

A case study is indeed an effective means of validating a conceptual framework, and is often used in the field of business and computing to test the usefulness of methods, processes, and guidance. Due to time and practical constraints, it was not possible to complete a real-world case study to further illustrate the usefulness and real-world applicability of TEC-VIS. The framework, however, has been used by a US-based control room design company which specialises in ergonomic design (Winsted Corp, Ltd). This company has chosen to use the TEC-VIS framework to inform the design of control rooms and centres (of all types, not just CCTV security control rooms). They plan to use TEC-VIS as a real-life (case-study) for security defence control centres in the UK and Dubai.

11.8 Chapter Summary

In this chapter, a critical review of the TEV-VIS evaluation has been presented. Three experts were selected to review the TEC-VIS framework (version 1, which is provided in Appendix H). The review was conducted to validate the research contributions made by this thesis. Overall, there was a consensus amongst reviewers that the TEC-VIS framework is an extremely valuable and important contribution to the CCTV and HCI research fields. An analysis of the reviewers' responses resulted in a number of changes to TEC-VIS. The key changes considered for version 2 of TEC-VIS included: altering the document style and language to cater for non-CCTV and non-academic readers, the provision of examples throughout the framework to illustrate the processes involved in each phase, and the inclusion of additional references on other research and guidance on digital CCTV storage, replay, disposal and data protection issues. A majority of the critical feedback of TEC-VIS related to a lack of clarification (mainly in Phases 2 and 4). In places, parts of the framework were misunderstood, due mainly to reviewers being given a shorter version of the framework for time and practical reasons. The changes made to TEC-VIS involved elaborating and clarifying content and making additional references to create a more coherent framework.

A number of areas have been identified for the future development of TEC-VIS (see Section 11.7). Much of the research needed to identify additional requirements will require additional empirical evaluations with users, other security observation tasks, and with a wider range of CCTV systems.

Chapter 12

Conclusions

This final chapter recalls the goals and motivations for the research carried out in this thesis. A summary of the substantive and methodological contributions is also given, followed by a critical assessment of these contributions. This thesis concludes by presenting an agenda for future research to improve the effectiveness of CCTV security systems.

12.1 Introduction

The research carried out in this thesis was motivated by the need to understand how CCTV is used in the management of security tasks, and the factors that reduce its effectiveness. The overarching goals for this research were: 1) theoretical and 2) practical. Firstly, an understanding of security observation tasks was needed as this research does not currently exist. Secondly, by building an understanding of the context and use of CCTV and other technologies used for security observation tasks, a practical framework can be developed to support CCTV practitioners and owners. The purpose of such a framework is to improve the design, configuration, and set-up of CCTV systems to ensure they are fit for purpose. The field study in this thesis (presented in Chapter 4) demonstrated the importance of taking into account and correctly configuring the: CCTV system environment, CCTV user's workstation set-up, task requirements (number of video monitor displays, display type, video quality etc.), and stakeholder communication (see Chapter 4, Section 4.4.2.5). The empirical studies (presented in Chapters 8 and 9) investigated the effect of using low-quality CCTV video for a face identification and event detection task performed by human observers. The results from these experiments provide minimum recommendations on video quality required for achieving effective task performance when using digital CCTV systems (in real-time via a network or when in play-back mode).

This research was extremely timely as there is a lot of investment in CCTV deployments and a large interest in using it for a growing range of security purposes. One of the key reasons for the high interest in CCTV is because of events such as the terrorist attacks in London (July 2005). In addition to terrorism, a number of changes led to people investing more and more in CCTV deployments. Firstly, there was a change in the way in which society perceived and utilised security systems (particularly CCTV). There were also changes in people's attitudes towards security in response to crime of all types. The number of CCTV deployments and the overall interest in using CCTV for security purposes was further encouraged by the rapid developments in the CCTV technology market. Furthermore, there have been technical developments in video and networking that transformed the way in which CCTV is being used. As a result, CCTV is being applied to new application areas and there are a large number of CCTV owners and a more heterogeneous group of users interacting with CCTV systems.

It has been a topic of debate for many years whether CCTV is effective or not. Several studies in the field of criminology have examined the effectiveness of CCTV from a sociological and political perspective - assessing whether CCTV has reduced, deterred and/or displaced crime. Research has also been conducted to establish whether CCTV has made society feel safer and whether the technology works well enough to support police activities (see the control room study by Gill et al., 2005: Chapter 2, Section 2.2). These studies are important, as they provide policy-makers (such as Home Office and local authorities) with an understanding of CCTV and the key issues which reduce the effectiveness of public surveillance systems. These studies do not provide CCTV owners and consultants with meaningful guidance on how CCTV systems can be improved in terms of effectiveness. Therefore, further research was needed to examine both the social and technical factors surrounding CCTV design and usage to develop objective and therefore quantifiable guidance.

A review of the research in HCI and security revealed that very few studies examine the effectiveness of CCTV security from the CCTV *user's* perspective and how their task performance is affected by people and technology. The only study which looked at performance to some extent was the CCTV effectiveness study by Gill et al., 2005. This study identified a number of operator factors which reduced the effectiveness of CCTV: low-quality equipment, too few operators, low operator to monitor/video display ratio, insufficient operator training (these findings are discussed in more detail in Chapter 2, Section 2.2).

Field studies at non-security control rooms which examine operator tasks, found that operator performance is reduced as a result of poor configuration and maintenance of equipment within the control room and the CCTV camera environment (see Chapter 2, Section 2.3). In addition, operator performance was reduced as a result of ineffective peer-to-peer communication (McCarthy et al., 1997; Luff et al., 2000). The implications of these findings were not discussed in much detail and the research was not taken further to improve the way in which CCTV and other technologies are used by operators within control rooms and other security work environments. Consequently, subsequent CCTV system deployments are being ignored, the various problems within CCTV control rooms and their designs still exist, and stakeholders are not paying attention to these problems – simply because they are not aware of the problems or solutions. The field study carried out in this thesis identified a number of issues which were discovered in the previous control room studies (see Chapter 4, Section 4.5.2), such as: 1) operators being overloaded with too many CCTV cameras (Gill et al., 2005); 2) CCTV video being recorded at low-quality (Gill et al., 2005); 3) an ineffective CCTV camera environment (Luff and Heath, 2001); and 4) poor radio communication (McCarthy et al., 1997). In addition to these findings, the field study presented in this thesis identified a wider number of operator performance issues: operators being overloaded with audio communication and other audio alerts; ineffective equipment set-up and layout; ineffective camera and mapping systems; and various technical issues were identified with CCTV cameras and operator radios. These findings are detailed in Chapter 4, Section 4.4.2.5.

Prior to this field research (Chapter 4), there was a limited awareness and understanding of what CCTV technology was being used for, and what security observation tasks were being performed by operators and other CCTV stakeholders. The BS EN ISO 11064 standard (ISO, 2004) is the only guidance that considers the human element of control room design. The standard is based on ergonomic principles and provides guidance on the physical aspects of control rooms such as: workstation arrangements, control room layout, use of displays and controls, and maintenance. Additional guidance is needed which details human-centred guidance on the set-up of digital CCTV and other technology used by human operators. Furthermore, greater emphasis is required on the configuration of CCTV systems as a whole.

There has been little empirical research on the effectiveness of digital CCTV video for security observation tasks performed by human observers. The guidance which does exist (UK Home Office Operational Requirements: Aldridge, 1994; Cohen et al., 2007) offers very limited guidance on digital and networked CCTV systems. A number of empirical research studies have been conducted by human centred multimedia researchers to assess the impact of video quality on human task performance with a number of video applications. These studies were carried out to identify the minimum video quality requirements for networked multimedia applications. In both Internet and mobile applications and services, it is desirable to reduce the amount of bandwidth needed without affecting the user's perception and performance with the interface. This type of evaluation has not been conducted for CCTV applications.

This thesis provides field and empirical research on the problems associated with CCTV system design and specifically the performance limitations of digital CCTV video when used by human observers for security observation tasks. This research involved an investigation which required reviewing a number of interdisciplinary areas related to CCTV (HCI, psychology, computer science, sociology, human factors, as well as legislation and guidelines). From the research findings, a number of substantive and methodological contributions were made by applying HCI knowledge and methods to improve CCTV practice, in particular the design and deployment of CCTV security systems and their context of use. These contributions have led to the development of a best-practice framework for CCTV deployment (TEC-VIS), which provides guidance for a new CCTV deployment or a redesign. Specific guidance is also provided to CCTV owners on the configuration of CCTV security systems (e.g., the user's tasks, work environment and the technology used to support their tasks).

12.2 Research Goals Revisited

The three research goals defined at the beginning of this thesis were:

12.2.1 Research Goal 1

To investigate the social and technical problems associated with digital CCTV and other technologies used by human operators *within a real-world context*.

This was a broad research goal to identify the underlying factors that reduce the effectiveness of CCTV, in particular the technical failures with CCTV technology. The research on CCTV is spread across a number of different disciplines, such as law, sociology, psychology, and computer science. Whilst a number of research studies have examined the different problems that affect operator performance within control rooms where CCTV technology is used, they do not provide any solutions (i.e., guidance or recommendations) to improve the design of control rooms, systems or user tasks. This research gap was identified by a review of the relevant literature on CCTV effectiveness. Research goal 1 was met in Study 1 through exploratory contextual field research at 14 CCTV control rooms in London and surrounding areas (see Chapter 4).

12.2.2 Research Goal 2

To identify the minimum video compression level required for an observer to identify targets (unknown to them) from CCTV images.

Another area of CCTV research which has been largely neglected is the effectiveness of CCTV video when used for security observation tasks carried out by human observers. There have been a number of CCTV stakeholders who anecdotally claim that, *"CCTV systems are not fit for purpose as a result of low quality video and images."* Despite these claims, there has not been any research to investigate in detail to examine: 1) why CCTV is not effective; 2) what factors are reducing the effectiveness of CCTV; and 3) what can be done to improve the effectiveness of CCTV. Thus, a framework was needed to support CCTV practitioners and owners in achieving effective CCTV in terms of design, configuration, and use.

A number of experiments have examined the impact of reducing temporal and spatial video quality for a number of tasks with different video applications, but only one study (van Voorthuijsen et al., 2005) has investigated the impact of low-quality CCTV video quality (temporal video quality) on a security observation task performed by humans (this study is reviewed in Chapter 6, Section 6.3). Previous studies in psychology have assessed the performance of face recognition with CCTV images taken from an analogue CCTV system, but these experiments focus on the observer's ability to recognise faces and the effect of person familiarity on task performance, rather than the impact of low-quality digital CCTV video on task performance. This research gap was identified following a review of the prior research on CCTV video quality and other video quality experiments conducted by human centred multimedia researchers (see Chapter 6). Research goal 2 was addressed through an empirical experiment with 80 participants conducted to assess the impact of excessive video compression on a face identification task (see Study 2 in Chapter 8).

12.2.3 Research Goal 3

To identify the minimum video compression level required for an observer to detect events from CCTV video.

The research carried out in both studies 2 and 3 were motivated by the same factors. For the same reasons given in the previous section, research goal 3 was investigated through an empirical experiment to address the gap in CCTV video quality research for detection tasks. This study specifically examined the impact of lowering the frame rate of digital CCTV video when used by untrained CCTV users for a detection task (see Study 3 in Chapter 9).

12.3 Contributions

The research carried out for this thesis has resulted in a number of contributions to the HCI field. This research was undertaken in response to a number of changes taking place within the field of CCTV: 1) changing social perceptions and attitudes towards security; 2) an increase in CCTV technological developments; 3) an increase in the number and variety of CCTV users; and 4) an increase in the utility of CCTV.

12.3.1 Substantive Contributions

As a result of this research, a best-practice framework on CCTV has been developed. The framework provides CCTV security owners and practitioners with a conceptual and methodological guide on the process of defining the security goals, stakeholders, tasks, system configurations, and video quality requirements for new and existing CCTV systems. The framework, Task-Effective CCTV for Video in Security (TEC-VIS), was developed using a socio-technical approach which is designed for CCTV practitioners and owners to consider step-by-step the importance of: 1) understanding the different users who belong to a CCTV security system; 2) understanding the different tasks a CCTV user should perform and the stakeholders; 3) defining the user and system requirements; 4) assessing the risks within the system; and 5) testing the performance of the system with real users and tasks to determine whether the system is fit for purpose. More specifically, the framework details guidance on: 6) conducting regular stakeholder meetings; 7) effectively managing administrative work to avoid delays in equipment repair and purchases; and 8) regular equipment maintenance (inside and outside the control room). It can be acknowledged that the TEC-VIS framework is still at a conceptual level, and to further validate its use, there is a need to test it with a real-world CCTV security system deployment. Testing the framework will ensure that the framework (once revised following the test/s), can be used with confidence by CCTV security consultants and owners in practice.

This research has also helped build a better understanding of the use of digital CCTV images for an identification task (with unknown targets) and a detection task. This research was particularly motivated by two main reasons. Firstly, there have been many anecdotal claims that CCTV video is too poor to be used for police investigations. Secondly, there are no specific guidelines or recommendations on CCTV video quality. As there is currently no empirical basis of the effects of digital CCTV compression on the performance of security observations tasks, CCTV owners have no guidance on where to set the thresholds for real-time and recorded CCTV video, and given that high video quality costs money, the temptation is to set the quality lower. Minimum video quality compression rates (bit rate and frame rate) have been identified for a face identification and event detection task and this contribution has been incorporated into the TEC-VIS framework (this can be found in Chapter 10, Section 10.2.4 - Table 10.2). It must be noted that these recommendations on video quality apply to specific task scenarios, thus they should be adopted following an evaluation within the real-world environment with real users. The desired performance of the system must be decided upon by the system owner and the video quality levels should be raised accordingly.

12.3.2 Methodological Contributions

In prior video quality experiments, researchers have largely relied on subjective assessment (see Chapter 6, Section 6.2 and 6.3). By relying on subjective measures alone (i.e., Likert scales, questionnaires, and interviews), these methods exclude the contextual factors and variables related to the user's task, the system and the user, thus reducing the ecological validity of the research. The task-oriented approach applied in the two empirical studies conducted for this thesis considers the relationships and variables associated with the task, application (system) and the users in order to derive realistic, meaningful and replicable results.

The research presented in this thesis contributes a theoretical framework to HCI knowledge. The framework provides two components:

1. **A description of contextual factors** (see page 100): based on a detailed literature review (see the literature reviews in Part 2) and the field study (see Chapter 4), a number of contextual factors which affect the performance of a CCTV system are identified. These factors relate to the target captured on CCTV video, the camera environment, CCTV system owner and user. These factors provide HCI practitioners and researchers with an awareness and understanding when researching, evaluating, and designing video systems – particularly security surveillance systems utilised by humans.
2. **A task-oriented approach** (see exemplification in Chapters 8 and 9): for investigating the effectiveness of CCTV video and images for security observation tasks. This method was put forward so that the requirements for security observation tasks under evaluation can be assessed objectively and the results can be applied to real-world tasks. This task-oriented methodology involves measuring the user's performance using classic HCI measures for CCTV tasks (task performance: correct number of detections and incorrect detections) so that performance results can be easily compared and these experiments can be easily replicated. Task performance measures can be contextualised to increase ecological validity. This can be achieved by matching laboratory tasks as closely as possible to the tasks performed in the real-world, by taking into account the context of the:
    - **task**, for example, in the context of a remote detection task performed over the Internet in real-time mode;
    - **users**, for example, untrained CCTV users were recruited for Study 2 and 3 as these users are increasingly used for identification and detection tasks with CCTV video/images to support security personnel; and
    - **application**, for example, the video stimuli used in Study 2 and 3 were closely matched to the content typically recorded on everyday CCTV systems (e.g., mixed ethnic faces, crime and non-crime activities). Also, the quality of the video was lowered to match the typical quality of video produced by low-budget CCTV systems.

The consideration of the contextual factors provided in this thesis (see page 100) and the adoption of a task-oriented approach will enable HCI researchers to uncover task requirements in context to improve current CCTV system design and future developments. In addition to these contributions, the research in this thesis demonstrates the applicability of HCI techniques and methods in the domain of CCTV security for shaping the next generation.

12.3.3 Research Publications

Chapter 4, Study 1: Control Room Field Study - Factors Reducing the Effectiveness of CCTV:

- Keval, H., and Sasse, M.A. (2006). Man or gorilla? Performance issues with CCTV technology in security control rooms. 16th World Congress on Ergonomics Conference. In *Proceedings of International Ergonomics Association*, Maastricht, Netherlands.
- Keval,H., Sasse, M. A. (2008a). "Not the Usual Suspects": A Study of Factors Reducing the Effectiveness of CCTV. Security Journal, ISSN: 0955-1662.

Chapter 8, Study 2: Face Identification with Compressed CCTV Images:

- Keval, H.U., Sasse, M. A. (2008b). Can we ID from CCTV? Image quality in digital CCTV and face identification performance. In *Proceedings of SPIE series*. SPIE. SPIE Mobile Multimedia/Image Processing, Security, and Applications, Agaian,S.S., Jassim,S. A. (ed.).

Chapter 9, B – Study 3: Detecting Events from CCTV Video with Low Frame-Rate CCTV Video:

- Keval, H., Gatusso, J., and Maclennan, K. B. (2007). Did you see what happened? A study on video frame rates for detecting events from CCTV video. A poster presented at the International Crime Science Conference, London.
- Keval, H. and Sasse, A. (2008c). To Catch a Thief - you need at least 8 frames per second: The impact of frame rates on user performance in a CCTV detection task. In *Proceedings of the 16th ACM International Conference on Multimedia.* Vancouver, British Columbia, Canada: ACM, pp. 941-944.

12.4 Limitations

Whilst there have been a number of successes in the research carried out in this thesis, it is also important to identify areas of limitations.

There was some difficulty in the initial planning stages when trying to approach security control room managers in central London about participating in the field study. This resistance was expected, as a year into the research, London was hit by two terrorist attacks (July bombings in 2005). In fact, on a number of occasions, security managers thought that the research queries were from a journalist rather than a PhD researcher. The resistance from security managers caused some delays at the start of this research and at a number of control rooms during the field research. Consequently very little data could be gathered from five out of the 14 control rooms visited since CCTV managers and operators were not very open or willing to report their difficulties with their work. Data collection was also limited at three control rooms as management limited the duration of the observation period to ~3-4 hours. Although this limitation was unavoidable, obtaining access to a wide range of control rooms allowed for the discovery of a wide range of task performance issues when using CCTV technology.

One of the biggest challenges of conducting field research in the security domain is gaining access to security workplaces and security stakeholders, as these environments hold sensitive data (personally identifiable data which is subject to data protection and privacy legislation). In Study 1 (Chapter 4), a relatively large sample of control rooms were studied (14) in comparison to previous research studies in control rooms (see Chapter 2). There are indeed many more control rooms that could have been included in the research study; however due to time and access constraints this was not practically feasible. Although a wide range of CCTV control rooms were visited (large/small, busy, short-staffed, urban/rural, old/new etc), the results from the study can be generalised to the 14 control rooms visited and other control rooms in the UK of similar set-up and organisation.

The main limitation identified in the first empirical study (see Chapter 8) was the use of unrealistic CCTV images. In Study 1, the face images were created from video recorded using a mini-DV camera recorder using a walking mask method (see Chapter 8, Section 8.6). This method was used to standardise each of the 64 face images, nevertheless the creation of the face images inadvertently created unnatural CCTV images which potentially increased task difficulty as well as the user's perceptions in task difficulty and video quality. For instance, the subjective and task performance results showed that the Afro-Caribbean targets were the hardest to identify (see Chapter 8, Section 8.4.1.3). This result may have been exacerbated, since 2d face images were used which captured less detail with darker skinned targets in comparison to lighter skinned targets. Despite this limitation, the results reported in Study 2 are still valid and the video quality recommendation can be applied to real-world CCTV systems because the images were created in the similar way to a typical low-cost digital CCTV system (record video signals digitally and encode using common video CODECs). Furthermore, in addition to recommending a minimum video quality level and CODEC type, additional lighting to the CCTV camera capture area was also recommended to ensure that dark skinned targets are recorded at high spatial detail.

Another criticism related to the empirical studies (see Chapters 8 and 9) was the lack of comparison in task performance between trained and untrained CCTV users. For the benefit of obtaining results which held high statistical power, a large sample size was needed for both studies; this could not be achieved easily if trained and untrained CCTV users were to be both recruited. Despite this practical limitation, the use of untrained observers does not affect the validity of the results as they two important contributions: 1) a detailed example of how a task-oriented approach can be used to assess CCTV video quality and 2) the development of specific video quality requirements for a face identification and detection task performed by untrained CCTV users in real-time. Prior to these experiments, video quality requirements for these tasks did not exist and was required given the increase in the number of untrained CCTV users being used to support security personnel in real-time monitoring of public places.

Following a review of the key research contributions (the review of TEC-VIS framework in Chapter 11), three reviewers highlighted a number of limitations, which related to the lack of guidance provided on the cost-benefit analysis of CCTV, the performance of security tasks with different types of video monitors, and the social factors affecting CCTV users' performance with CCTV video. These issues were not investigated as they were not included within the scope of the research presented in this thesis. These issues were *not* explored in this thesis, although they are considered to be valuable areas for further research in CCTV.

12.5 Directions for Future Research

CCTV technology will continue to develop, the number of deployments will continue to increase and there is likely to be more third generation CCTV systems designed with an increasing number of intelligent capabilities to support and improve observer vigilance (e.g., the addition of microphones and speakers to CCTV cameras in producing 'listening' and 'talking' CCTV cameras). With these changes, it is vital that various human factors, such as user knowledge, training, and skills be considered to meet the challenges with current and future CCTV systems. It is also essential to understand the entire socio-technical system and determine whether it is technically capable of performing as it should. Given that this is the first research study to have examined the social and technical factors which reduce the effectiveness of CCTV using HCI methods and techniques, further work into this area would be invaluable to increase this knowledge base. In the final sections of this thesis, a programme for future research that has been formulated to improve the effectiveness of CCTV is presented.

12.5.1 Sociological Research – Identification of Crime Patterns

In Study 1, operators were interviewed whilst they performed proactive surveillance tasks to understand *how* they performed their observation tasks and in *what* elements of the scene they were particularly interested. Specifically, they were asked what strategies they adopted when detecting suspicious events and targets from CCTV on their spot monitors. From the interviews with operators, a number of crime patterns emerged: e.g.: *"Criminals don't get out of bed until about lunchtime"* and *"Shoplifting in the town is likely to happen after 3:30 when school kids finish from school or during the Easter and summer holidays"* (see Chapter 4, Section 4.4.2). It would have been valuable to research crime patterns further through a detailed analysis. This could be achieved by conducting a sociological study to identify a whole range of crime patterns used by trained CCTV observers when detecting events and identifying suspicious targets from real-time CCTV video. The identification of crime patterns could then be examined in more detail by CCTV user type (e.g., security and surveillance knowledge and experience) and by geographical regions. These patterns can be identified through field research (semi-structured interviews) with experienced CCTV operators and other security personnel. Insights into criminal behaviour over time and space can also be elicited by interviewing criminals and repeat offenders who are open, honest, and willing to describe their thoughts and motivations about planning a crime where CCTV cameras are present. Interviewing both CCTV observers and criminals can strengthen the data and increase the reliability of the formulated crime patterns.

These crime patterns can be applied in a number of ways. For instance, they can be used to inform the research and design of intelligent CCTV systems to improve the accuracy and speed of automatic detection and recognition algorithms. In addition to improving the design of intelligent CCTV systems, crime patterns can also be used to provide inexperienced CCTV users with the necessary training, knowledge, and skills to perform security observation tasks with real-time CCTV video. CCTV owners and managers can also benefit from these findings as they can use common crime patterns specific to geographical locations and specific time periods to assist in the deployment of CCTV cameras - or for an entire CCTV security system. The police could also use crime patterns to support their intelligence.

12.5.2 Psychological Research – Organisational Factors

Although prior research has examined the organisational factors surrounding CCTV operator performance (Wallace and Diffley, 1998), further research is needed to consider existing CCTV control room environments. This need was also identified in the expert review of TEC-VIS (see Chapter 11), where guidance on control room organisational ergonomics was found lacking. Specific guidance within this area of design is needed to support CCTV users working within security environments to improve their work performance. For example, through field and empirical research, it will be possible to determine the optimum number of hours an operator should perform proactive and reactive surveillance tasks, how long a rest break should be, how frequently tasks should be rotated, and how much training and feedback is required to achieve effective performance. The research will benefit CCTV owners by providing them with the necessary knowledge about the people factors associated with deploying a CCTV security system. If these factors are considered during the design of the CCTV security system, operators (and other CCTV stakeholders) will be better supported and managed.

12.5.3 Computer Science and HCI Research – Video Quality Studies

The experiments carried out in this thesis were geared towards evaluating the effect of video compression on two commonly performed security observation tasks. However, these experiments are by no means limited in potential, as many variables can be tested under the same experimental paradigm used for the empirical experiments. There are a multitude of factors which can affect CCTV video quality and observer performance (see Chapter 5, Section 5.1.2, Table 5.2) when carrying out security observation tasks and these factors can be assessed in a number of ways. Further experiments can be carried out following the task-oriented approach proposed in the methodological contributions in this chapter (see Section 12.3.2) to identify the requirements for other security observation tasks (monitoring and recognition tasks). Additional task performance experiments can be carried out to determine the optimum CCTV video quality for all four security observation tasks (monitor, detect, recognise, and identify):

- Assess task performance using different video monitors (TFT vs. LCD, plasma vs. projected screens, etc.) to determine the optimum display for CCTV and non-CCTV video based tasks.
- Examine the impact of excessive video compression using video CODECs other than MPEG-4 and Wavelet (e.g., H264 and M-JPEG).

- Compare user task performance when CCTV video is observed under different video resolutions.
- Compare task performance with trained and untrained CCTV users, and under different security contexts (e.g., casino, airport, shopping centre, business premises, and extremely crowded places).

As a result of time and practical limitations during the course of this research, it was not possible to install a test CCTV system to gather real-world CCTV video stimuli for the face identification and detection studies. For the purpose of achieving high ecological validity in future CCTV experiments, a 'test' CCTV system can be installed to capture natural video footage of everyday activities. This set-up will allow full control in creating specific video content for the task scenarios (e.g., theft detection within a shop, monitoring overcrowded areas, targets carrying a concealed weapon or drugs). Using the recorded footage, the quality of the video can be altered by the experimenter. This can be achieved by employing open-source video CODECs typically used in commercial digital and networked CCTV systems to obtain a range of video quality conditions for task assessment. The real-world CCTV video footage can be used for future experimentation for security observation tasks performed by both humans and computers.

Bibliography

Agrafiotis, D., Canagarajah, N., Bull, D., Dye, M., Twyford, H. E., Kyle, J. G. *et al.* (2003). Optimized sign language video coding based on eye-tracking analysis. In *Proceedings of Visual Communications and Image Processing* (pp. 1244-1252). University of Italian Switzerland, Lugano, Switzerland.

Aldridge, J. (1989). *The Rotakin - A test target for CCTV security systems* (Rep. No. 16/89). Home Office, Police Scientific Development Branch (PSDB).

Aldridge, J. (1994). *CCTV operational requirements manual, Version 3* (Rep. No. 17/94). Home Office, Police Scientific Development Branch (PSDB).

Aldridge, J. and Gilbert, C. (1995). *Performance testing of CCTV perimeter surveillance systems* (Rep. No. 14/95). Home Office, Police Scientific Development Branch (PSDB).

Aldridge, J. (2007). Personal communication.

Armitage, R. (2002). *To CCTV or not to CCTV? A review of current research into the effectiveness of CCTV systems in reducing crime.* Community safety practice briefing. London: NACRO.

Bachmann, T. (1991). Identification of spatially quantised tachistoscopic images of faces: How many pixels does it take to carry identity? *European Journal of Cognitive Psychology, 3,* 107.

Barber, P. J. and Laws, J. V. (1994). Image quality and video communication. In R. Draper, W. Hall, and J. Richards (Eds.), *Proceedings of IEE International Symposium on Multimedia Technologies and Their Future Applications* (pp. 165-175). London: Pentech Press.

Bentley, R., Hughes, J., Randall, D., Rodden, T., Sawyer, P., Shapiro, D. et al. (1992). Ethnographically-informed systems design for air traffic control. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work* (pp. 123-129). Toronto, Ontario: ACM Press.

Bertelsen, O. W. (2004). The activity walkthrough: An expert review method based on activity theory. In *Proceedings of the Third Nordic Conference on Human-Computer Interaction: NordiCHI 2004* (pp. 251-254). New York: ACM.

Beyer, H. and Holtzblatt, K. (1998). *Contextual design: defining customer-centered systems.* San Francisco, CA: Morgan Kaufmann Publishers.

Blandford, A. and Furniss, D. (2005). DiCoT: A methodology for applying distributed cognition to the design of team working systems. In *Proceedings of DCVIS* (pp. 26-38). Springer.

Bouch, A. and Sasse, A. (1999). Network Quality of Service: What do users need? In *Proceedings of the 4th International Distributed Conference (IDC'99)*, Madrid, Spain (pp. 21-23).

Bouch, A., Sasse, A., and Wilson, G. (2001). A 3-dimensional approach to assessing end-user quality of service. In *Proceedings of the London Communications Symposium*, London (pp. 47-50).

Bridger, R. S. (1995). *Introduction to ergonomics* (2nd ed.). London: Taylor and Francis.

Bromby, M. (2002). To be taken at face value? Computerised identification. *Information Communication Technology Law Journal, 11,* 63-73.

Bruce and Young. (1986). Understanding face recognition. *British Journal of Psychology, 77, 305-327.*

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J. B., Burton, A., and Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied, 5,* 331-338.

Burton, A. M., Wilson, S., Cowan, M., and Bruce, V. (1999). Face recognition in poor quality video: evidence from security surveillance. *Psychological Science, 10,* 243-248.

Card, S. K., Moran, T. P., and Newell, A. (1986). *The model human processor: an engineering model for human performance: Handbook of perception and human performance*. New York, NY: Wiley.

Carroll, J. M. (1990). Infinite detail and emulation in an ontologically minimized HCI. In J. C. Chew and J. Whiteside (Eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 321-328). New York, NY, USA: ACM Press.

CCTV Systems and the Data Protection Act - Code of Practice (2008).

Chance, J. E. and Goldstein, A. G. (1996). The other-race effect and eyewitness identification. In S.L.Sporer, R. S. Malpass, and G. Koehnken (Eds.), *Psychological issues in eyewitness identification* (pp. 157-176). Mahwah: NJ: Lawrence Erlbaum Associations.

Chen, F., Choi, E. H., Ruiz, N., Shi, Y., and Taib, R. (2005). User interface design and evaluation for control room. In *Proceedings of the 19th Conference of the Computer-Human Interaction, Special Interest Group (Chisig) of Australia on Computer-Human Interaction: Citizens Online: Considerations for Today and the Future* (pp. 1-4). Canberra: Australia

Chiroro, P. and Valentine, T. (1995). An investigation of the contact hypothesis of the own-race bias in face recognition. *Quarterly Journal of Experimental Psychology: A Human Experimental Psychology, 48A,* 879-894.

Claypool, M., Claypool, K., and Damaa, F. (2006). The effects of frame rate and resolution on users playing first person shooter games. In *Proceedings of ACM/SPIE Multimedia Computing and Networking (MMCN)*, San Jose, California, USA.

Cohen, N (2004). Personal communication.

Cohen, N., Gatusso, J., and MacLennan-Brown, K. (2006). *CCTV operational requirements manual: Is your CCTV system fit for purpose? Version 4* (Rep. No. 55/06). Home Office, Scientific Development Branch (HOSDB) Publication.

Cohen, N (2006). Personal communication.

Cooligan, H. (1999). *Research methods and statistics in psychology* (3 ed.) London: Hodder and Stoughton.

Costen, N. P., Parker, D. M., and Craw, I. G. (1994). Spatial content and spatial quantisation effects in face recognition. *Perception, 23,* 129-146.

Cowen, L. (2001). *An eye movement analysis of web page usability.* Unpublished master's thesis: Lancaster University, UK.

Cucchiara, R. (2005). Multimedia surveillance systems. In *Proceedings of the Third ACM International Workshop on Video Surveillance and Amp; Sensor Networks* (pp. 3-10). New York, NY: ACM.

Dix, A., Finlay, J., Abowd, G., and Beale, R. (2003). *Human-Computer Interaction* (3rd ed.), Prentice-Hall, Inc, Upper Saddle River, NJ, USA.

Donald, C. (2001). Part I Human Performance: Vigilance. In J. Noyes and M. Bransby (Eds.), *People in control: Human factors in control room design* (pp. 35-49). London, United Kingdom: IEE.

Elliot, E. S., Wills, E. J., and Goldstein, A. G. (1973). The effects of discrimination training on the recognition of white and oriental faces. *Bulletin of the Psychonomic Society, 2,* 71-73.

Ellis, S., Candrea, R., Misner, J., Craig, C. S., Lankford, C. P., and Hutchinson, T. E. (1998). Windows to the soul? What eye movements tell us about software usability. In *Proceedings of the Usability Professionals' Association Conference,* Washington, DC, USA (pp. 151-178).

Ellis, H.D., Shepherd, J.W., and Davies, G.M. (1979). Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition, *Perception, 8,* 431-439.

Fléchias, I. (2005). *Designing secure and usable systems.* Unpublished doctoral thesis: University College London.

Foley, M. A. and Foley, H. J. (1998). A study of face identification: Are people looking "beyond" disguises? In M. Intons-Peterson and D. Best (Eds.), *Challenges and controversies: Memory distortions and their prevention* (pp. 29-47). Mahwah, NJ: Lawrence Erlbaum.

Freeman, R. E. (1984). *Strategic management: A stakeholder approach*. Boston, MA: Pittman.

Fry, P (2005). Personal communication.

Gerrard, G., Parkins, G., Cunningham, I., Jones, W., Hill, S., and Douglas, S. (2007). *National CCTV strategy*. Home Office and ACPO publication, October 2007.

Ghinea, G. and Thomas, J. P. (1998). QoS impact on user perception and understanding of multimedia video clips. In *Proceedings of ACM Multimedia* (pp. 49-54). Bristol, England, UK.

Ghinea, G. and Thomas, J. P. (1999). An approach towards mapping quality of perception to quality of service in multimedia communications. In *Proceedings of the IEEE Workshop on Multimedia Signal Processing* (pp. 497-502). Copenhagen, Denmark.

Ghinea, G. and Magoulas, G. D. (2001). Perceptual considerations for quality of service management: An integrated architecture. In *8th International Conference on User Modelling, Lecture Notes in Computer Science* (pp. 234-236). Sonthofen, Germany: Springer.

Gill, M. and Spriggs, A. (2005). *Assessing the impact of CCTV, Home Office Research Study 292* (Rep. No. 15/05). Home Office Research, Development and Statistics Directorate.

Gill, M., Spriggs, A., Allen, J., Jessiman, P., Swain, D., Hemming, M. et al. (2005). *Control room operation: Findings from control room observations* (Rep. No. 17/05). Home Office Scientific Development Branch Publication.

Gill, M. (2006). *The Handbook of Security*. New York, NY: Palgrave Macmillan.

Girod, B. (1993). What's wrong with mean-squared error? In Watson, A. B. (Ed.), *Digital images and human vision* (pp. 207-220). Cambridge, MA, USA: MIT Press.

Green, M. (2004). Errors in eyewitness identification procedures. Visual Expert - Human Factors Available: http://www.visualexpert.com/Resources/mistakenid.html

Green, D. and Swets, J. (1966). *Signal detection: theory and psychophysics*. New York: Wiley.

Gulliver, S. R. and Ghinea, G. (2004). Stars in their eyes: What eye-tracking reveals about multimedia perceptual quality. *EEE Transactions on Systems, Man and Cybernetics, Part A, 34,* 472-482.

Heath, C., Jirotka, M., Luff, P., and Hindmarsh, J. (1993). Unpacking collaboration: The international organisation of trading in a city dealing room. In *Proceedings of the Third Conference on*

European Conference on Computer-Supported Cooperative Work, Milan, Italy, G. de Michelis, C. Simone, and K. Schmidt, Eds. ECSCW. Kluwer Academic Publishers, Norwell, MA, (pp 155-170).

Henderson, Z., Bruce, V., and Burton, A. M. (2001). Matching the faces of robbers captured on video. *Applied Cognitive Psychology, 15,* 464.

Hill, N., Brierley, J., and MacDougall, R. (1999). *How to measure customer satisfaction*. Hampshire, UK: Gower Publishing.

Hollnagel, E. and Woods, D. D. (1983). Cognitive systems engineering: New wine in new bottles. *International Journal of Man-Machine Studies, 18,* 583-600.

Holtzblatt, K. and Jones, S. (1993). Contextual inquiry: A participatory technique for system design. In D. Schuler, A. Namioka (Eds.), *Participatory design: Principles and practice* (pp. 180-193). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Horn, D. B., Karasik, L., and Olsen, J. S. (2002). The effects of spatial and temporal video distortion on lie detection performance. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems* (pp. 714-715). New York, NY: ACM.

HOSDB. (2005). *UK police requirements for digital CCTV systems* (Rep. No. 09/05). Home Office Police Scientific Development Branch and ACPO.

Hughes, J. A., King, V., Rodden, T., and Anderson, A. H. (1994). Moving out from the control room: Ethnography in system design. In *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work* (pp. 429-439). New York NY: ACM Press.

Ikehara, C. and Crosby, M. (2005). Assessing cognitive load with physiological sensors. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences,* Washington, DC: IEEE, Computer Society (pp 1-9).

ISO. (1997). *ISO 9241-11: Ergonomic requirements for office work with visual display terminals (VDTs), Part 11 - guidelines for specifying and measuring usability*. Geneva: International Standards Organisation.

ISO. (1999). *ISO 13407: Human-centred design processes for interactive systems*. Geneva: International Standards Organisation.

ISO. (2004). *BS EN ISO 11064-4: Ergonomic design of control centres*. Part 4, Layout and Dimensions of workstations. Geneva: International Standards Organisation.

ITU. (2002). *ITU-R Recommendation BT.500-1: Methodology for the subjective assessment of the quality of television pictures*. Geneva, Switzerland: International Telecommunication Union.

John, B. and Marks, S. (1997). Tracking the effectiveness of usability evaluation methods. *Behaviour and Information Technology, 16,* 188-202.

Johnson, B. F. and Caird, J. K. (1996). The effect of frame rate and video information redundancy on the perceptual learning of American Sign Language gestures. In M. J. Tauber (Ed.), *Companion on human factors in computing systems: Common ground.* NY, New York: ACM Press.

Just, M. A. and Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87,* 329-354.

Kassin, S., Tubb, V. A., Hosch, H. M., and Memon, A. (2001). General acceptance of eyewitness testimony research: A new survey of the experts. *American Psychologist, 56,* 405-416.

Keval, H. and Sasse, A. (2006). Man or gorilla? Performance issues with CCTV technology in security control rooms. In *Proceedings of International Ergonomics Association.* Maastricht, Netherlands: IEA.

Keval, H. (2006). CCTV control room collaboration and communication: Does it work? In *Proceedings of Human Centred Technology Workshop*, Brighton.

Keval, H., Gatusso, J., and Maclennan, K. B. (2007). Did you see what happened? A study on video frame rates for detecting events from CCTV video. A poster presented at the *International Crime Science Conference,* London.

Keval,H., Sasse, M. A. (2008a). "Not the Usual Suspects": A Study of Factors Reducing the Effectiveness of CCTV. Security Journal, ISSN: 0955-1662.

Keval, H.U., Sasse, M. A. (2008b). Can we ID from CCTV? Image quality in digital CCTV and face identification performance. In *Proceedings of SPIE series.* SPIE.SPIE Mobile Multimedia/Image Processing, Security, and Applications, Agaian,S.S., Jassim,S. A. (ed.).

Keval, H. and Sasse, A. (2008c). To Catch a Thief - you need at least 8 frames per second: The impact of frame rates on user performance in a CCTV detection task. In *Proceedings of the 16th ACM International Conference on Multimedia.* Vancouver, British Columbia, Canada: ACM, pp. 941-944.

Klatzty, Forest. (1984). Recognising familiar and unfamiliar faces. *Memory and Cognition, 12,* 60-70.

Knoche, H. and de Meer, H. (1999). Utility curves: Mean opinion scores considered biased. In *Proceedings of IWQoS.* London, UK.

Knoche, H., McCarthy, J. C., and Sasse, M. A. (2008). How low can you go? The effect of low resolutions on shot types. *Multimedia Tools and Applications Series, 36,* 145-166.

Korshunov, P. and Ooi, W. T. (2006). Critical video quality for distributed automated video surveillance. In *Proceedings of the 13th Annual ACM international Conference on Multimedia* (pp. 151-160). ACM Press, New York, NY.

Laughery, K. R., Alexander, J. F., and Lane, A. B. (1971). Recognition of human faces: Effects of target exposure time, target position, pose position, and type of photograph. *Journal of Applied Psychology, 55,* 483.

Lewin, C. and Herlitz, A. (2002). Sex differences in face recognition- women's faces make the difference. *Brain and Cognition, 50,* 121-128.

Leydon, J. (2006). Holidaymaker uses laptop to nab burglars. The Register Available: http://www.theregister.co.uk/2006/09/14/holidaymaker_foils_burglary/

Light, L., Kayra-Stuart, F., and Hollander, F. (1979). Recognition for typical and unusual faces. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 212-218.

Luff, P and Heath, C. (2001). Part III Methods: Naturalistic analysis of control room activities. In J. Noyes and M. Bransby (Eds.), *People in control: Human factors in control room design* (pp. 151-165). London, United Kingdom: IEE.

Maguire, M. (2001a). Context of use within usability activities. *International Journal of Human-Computer Studies*, *55*, 453-483.

Maguire, M. (2001b). Methods to support human-centered design. *International Journal of Human-Computer Studies*, 55, 587-634.

Malpass, R. S. and Kravitz, J. (1969). Recognition for faces of own and other race faces. *Journal of Personality and Social Psychology, 13,* 330-334.

McCahill, M. and Norris, C. (2003). CCTV and Public Attitudes. Report to the European Commission Fifth Framework RTD as part of the UrbanEye: On the threshold of the urban panopticon project.

McCarthy, J. C., Wright, P. C., Healey, P., Dearden, A., and Harrison, M. D. (1997). Locating the scene: The particular and the general in contexts for ambulance control. In *Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work: The Integration Challenge* (pp. 101-110). Phoenix, Arizona, United States: ACM Press.

McCarthy, M., Sasse, A., and Miras, D. (2004). Sharp or smooth? Comparing the effects of quantization vs. frame rate for streamed video. In *Proceedings of CHI, ACM, New York, USA* (pp. 535-541).

Mead, L. (1998). *Usage of video recordings in surveillance, the value of such as evidence and potential problems which can arise.* In *Proceedings of 13th Annual BILETA Conference: 'The Changing Jurisdiction', British and Irish Legal Education Technology Association,* Trinity College, Dublin (pp 1-6).

Millen, D. R. (2000). Rapid ethnography: time deepening strategies for HCI field research. In *Proceedings of the 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques* (pp. 280-286). New York, N.Y: ACM.

Miras, D. (2004). *On quality aware adaptation of Internet video.* Unpublished doctoral thesis. University College London, London.

Moore, A. D. (2000). Employee monitoring and computer technology: Evaluative surveillance v privacy. *Business Ethics Quarterly, 10,* 697-709.

Morse, J. M. (1995). *Critical issues in qualitative research methods*. Thousand Oaks, CA: Sage Publications.

Morton, S. (2004). *Recording systems, not cameras, limit image quality in video surveillance systems* (Rep. No. 28), CCTV Focus, May 2004.

Muir, L., Richardson, I., and Hamilton, K. (2005). Visual perception of content-prioritised sign language video quality. In *Proceedings IEE Visual Information Engineering*. University of Glasgow, Glasgow pp.17-22.

Neilson, J. (1993). *Usability engineering*. San Francisco: Morgan Kaufmann.

Ney, S. and Pichler, K. (2002). *Video surveillance in Austria,* Working Paper No.7. Urban Eye Project.

Norman, D. (1983). Some observations on mental models. In R.M. Baecker (Ed.), *Human-computer interaction: A multidisciplinary approach* (pp. 241-244). San Francisco, CA: Morgan Kaufmann Publishers.

Norros, L. and Nuttinen, M. (2005). Performance based usability evaluation of a safety information and alarm system. *International Journal of Human-Computer Studies, 63,*328-361.

O'Reilly, K. (2005). *Ethnographic methods*. Oxon: Routledge.

O'Toole, A., Jiang, F., Roark, D., and Abdi, H. (2006). Predicting human performance for face recognition. In W. Zhao and R. Chellappa (Eds.), *Face processing: Advanced modelling and methods* (pp. 293-319). New York: Academic Press.

Oxlee, G (2004). Personal communication.

Paay, J. (2007). From ethnography to interface design. In J Lumsden, *Handbook of research on user interface design and evaluation for mobile technology,* National Research Council of Canada, Canada (pp 1-15).

Patton, M. Q. (1990). *Qualitative evaluation and research methods*. Newbury Park, London: Sage Publications.

Phillips, C. (1999). A review of CCTV evaluations: Crime reduction effects and attitudes towards its Use. In K. Painter and N. Tilley (Eds.), *Surveillance of public space: CCTV, street lighting and crime prevention, Crime prevention studies, Vol. 10* (pp. 123-155). New York: Criminal Justice Press.

Police guidelines for CCTV security systems ignored (2007). HomeSecurityWatch.Org.

Prior, M . (2002): Big Brother - So What? A Study of the Attitudes of Young People to Workplace Surveillance. In *Proceedings of the sixth ETHICOMP Conference* Lisbon, Portugal. Lisbon: Universidade Lusiada (pp 259-269).

Rail safety man barred from phoning transport police (2008). Kent Online.

Ramachandran, D., Kam, M., Chui, J., Canny, J., and Frankel, J. F. (2007). Social dynamics of early stage co-design in developing regions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1087-1096). New York, USA: ACM.

Rehnman, J. and Herlitz, A. (2006). Higher face recognition ability in girls - Magnified by own-sex and own ethnicity bias. *Memory, 14,* 296.

Rights group criticises 'Asbo TV' (2006, January 10). BBC News. Retrieved date 15/01/06, from http://news.bbc.co.uk/1/hi/england/london/4597990.stm

Rimmel, A. N., Keval, H., Mansfield, N. J., and Hands, D. (2005). Quantification of error perception for full-face digital video. *IEEE Letters, 41,* 902-903.

Robertson, I. L. and Monro, D. M. (1997). Video Surveillance using low bandwidth, high compression systems. In *Proceedings of European Conference on Security and Detection* (pp. 31-35).

Schwaninger, A., Lobmaier, J., and Collishaw, S. M. (2002). Role of featural and configural information in familiar and unfamiliar face recognition. In H. H. Bülthoff, S. Lee, T. Poggio, and C. Wallraven (Eds.), *Proceedings of the Second International Workshop on Biologically Motivated Computer Vision, Lecture Notes In Computer Science* (pp. 643-650). London: Springer-Verlag.

Serif, T., Gulliver, S. R., and Ghinea, G. (1999). Perceptual impact of multimedia QoS: A comprehensive study of pervasive and traditional computing devices. In *ACM Symposium on Applied Computing*, Nicosia, Cyprus, pp 43-51.

Shackel B, 1984. The Concept of Usability. In J Bennett, D Case, J Sandelin, and M Smith (eds), *Visual Display Terminals: Usability Issues and Health Concerns*, 45-87. Englewood Cliffs, NJ: Prentice Hall.

Sharp, H., Rogers, Y., and Preece, J. (2007). *Interaction design: Beyond human-computer interaction.* (2nd ed.) USA: John Wiley and Sons, Inc.

Sheperd, J. W., Deregowski, J. B., and Ellis, H. D. (1974). A cross-cultural study of recognition memory for faces. *International Journal of Psychology, 9,* 212.

Shepherd, A. (2001). *Hierarchical task analysis*. London; New York: Taylor and Francis.

Sirohey, S., Wilson, C., and Chellapa, R. (1995). Human and machine recognition of faces: A survey. In *Proceedings of IEEE, 85* (pp. 705-740).

Stacey, P.C., Walker, S., and Underwood, J.D.M. (2005). Face processing and familiarity: Evidence from eye-movement data. *British journal of Psychology, 96,* 407-422.

Super cops seek fixed link to town centre CCTV (2006). The Register. Retrieved 25/09/06 from http://www.theregister.co.uk/2006/09/21/soca_CCTV/

Swartz, M. and Wallace, D. (1993). Effects of frame rate and resolution reduction on human performan*ce.* In Proceedings of lS&T's 46th Annual Conference, Munich, Germany (pp

SWGIT. (2004). *Recommendations and guidelines for the use of CCTV security systems for commercial institutions* (Rep. No. 2.1). Scientific Working Group on Imaging Technology.

Taylor, N. (2002). State surveillance and the right to privacy. *Surveillance and Society 1*(1)*,* 66-85.

Tickner, A. H. and Poulton, E. C. (1973). Monitoring up to 16 synthetic television pictures showing a great deal of movement. *Ergonomics, 14,* 171-175.

Troscianko, T., Holmes, A., Stillman, J., Mirmehdi, M., Wright, D., and Wilson, A. (2004). What happens next? The predictability of natural behaviour viewed through CCTV cameras. *Perception, 33,* 87-101.

Twidale, M., Randall, D., and Bentley, R. (1994). Situated evaluation for cooperative systems. In *Proceedings of the Conference on Computer-Supported Cooperative Work* (pp. 441-452). Chapel Hill, North Carolina, USA: ACM press.

van Voorthuijsen, G., van Hoof, H., Klima, M., Roubik, K., Bernas, M., and Pata, P. (2005). CCTV effectiveness study (pp. 105-108). In *Proceedings of* Security Technology, CCST '05, 39th Annual 2005 International Carnahan: IEEE.

Veltman, J. and Galliard, A. (1998). Physiological workload reactions to increasing levels of task difficulty. *Ergonomics, 41,* 656-669.

Walker, S and Cohen, N (2006). Personal communication.

Wallace, E., Diffley, C., Baines, E., and Aldridge, J. (1997). Ergonomic design considerations for public area CCTV and security applications. In *Proceedings of International Ergonomics Association Congress.*

Wallace, E. and Diffley, C. (1998). CCTV: Making it work. *CCTV Control Room Ergonomics*, Police Scientific Development Branch, United Kingdom, Home Office, Publication Number 14/98.

Walker-Smith, G.J., Gale, A.G., and Findlay, J.M. (1977). Eye movement strategies involved in face perception, *Perception, 6,* 313-326.

Walters, P. E. (1995). CCTV systems thinking - systems practice (pp. 64-69). In *European Convention on Security and Detection*, Brighton, UK: IEE, Wiley

Wang, D., Speranza, F., Vincent, A., Martin, T., and Blanchfield, P. (2004). Towards optimal rate control: A study of the impact of spatial resolution, frame rate and quantization on subjective quality and bit rate. In *Proceedings of Visual Communications and Image Processing (VCIP)* (pp 198-209). Lugano, Italy: SPIE.

Watson, A. and Sasse, A. (1998). Measuring perceived quality of speech and video in multimedia conferencing applications. In *Proceedings of ACM Multimedia* (pp. 55-60). Bristol: ACM.

Web users to 'patrol' US border (2006, June 2). BBC News. Retrieved date 2/06/06, from http://news.bbc.co.uk/1/hi/world/americas/5040372.stm

Whitefield, A., Wilson, F., and Dowell, J. (1991). A framework for human factors evaluation. *Behaviour and Information Technology, 10,* 65-79.

Wilson, F. and Descamps, P. T. (1996). Should we accept anything less than TV quality? In *Proceedings of Visual Communication, International Broadcasting Convention*, Netherlands, Amsterdam (pp. 606-611).

Wilson, G. M. and Sasse, M. A. (2000). Do users always know what's good for them? Utilising physiological responses to assess media quality, In McDonald, S., Waern, Y., Cockton, G. (Eds.), People and Computers XIV-Usability or Else! (pp. 327-339). *Proceedings of HCI 2000*, Berlin: Springer.

Wilson, G. M. and Sasse, M. A. (2004). From doing to being: getting closer to the user experience. *Interacting with Computers, 16,* 697-705.

Wood, J. (2001). Part III Control room design: Control room mock-up trials. In J. Noyes and M. Bransby (Eds.), *People in control: Human factors in control room design* (pp. 189-206). London, United Kingdom: IEE.

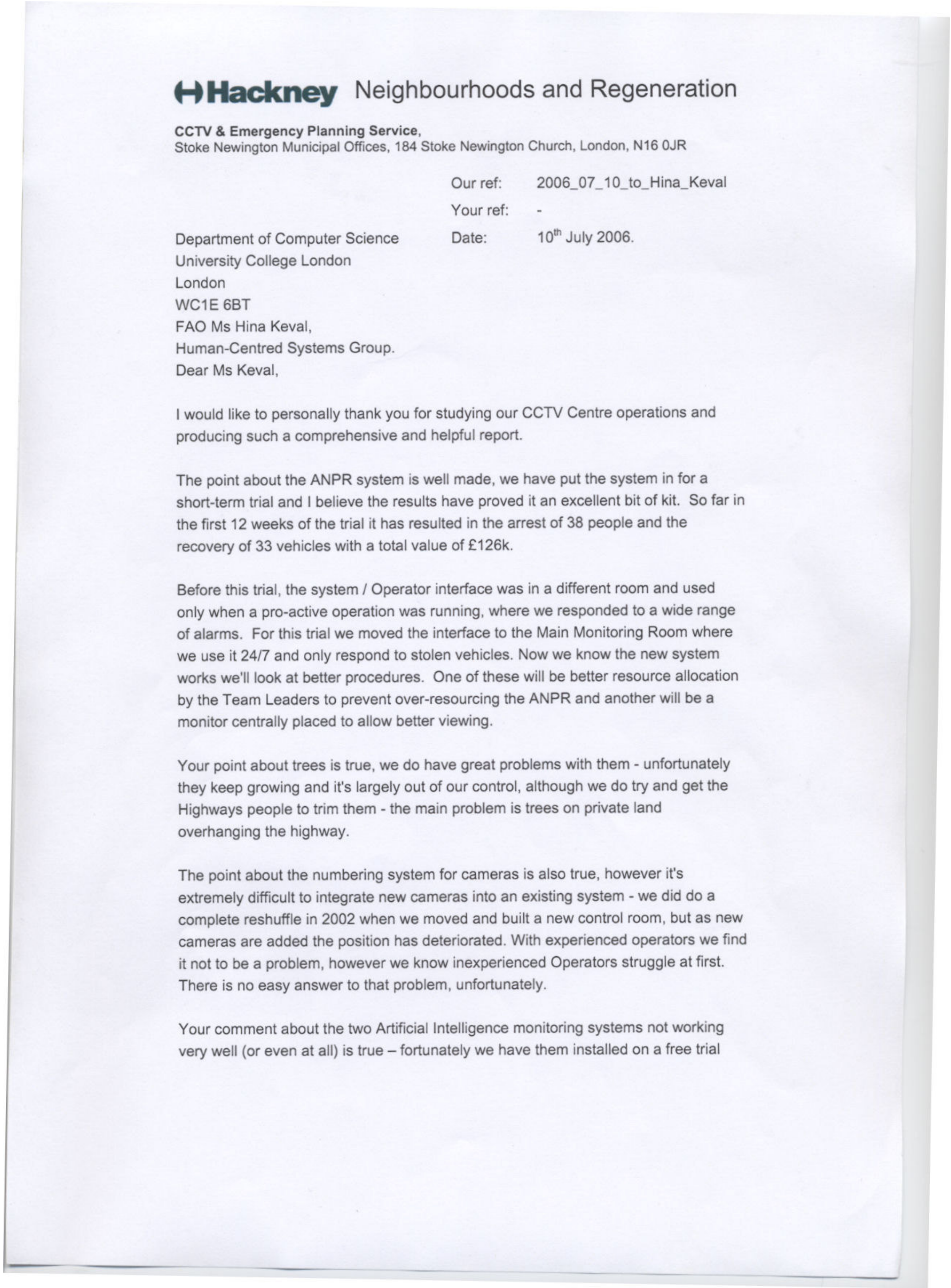
Young, A.W., Hay, D.C., McWeeney, K.W., Flude, B.M. (1985). Matching familiar and unfamiliar faces on internal and external features, *Perception, 14*, 737-746.

APPENDICES

Appendix A

Study 1: Letter from Hackney CCTV Emergency & Planning Service

This letter was received in receipt of a short CCTV control room evaluation report. The manager of this control room was interested in the findings of the 12-hour observation visits and requested that a short report be prepared after the visits. Visits were made during two different operator shifts to maximise the evaluation work; one visit was made during an afternoon shift, 13:00–19:00, and the second visit was made during an evening shift, 18:00–24:00.



Hackney Neighbourhoods and Regeneration

CCTV & Emergency Planning Service,
Stoke Newington Municipal Offices, 184 Stoke Newington Church, London, N16 0JR

Our ref:	2006_07_10_to_Hina_Keval
Your ref:	-
Date:	10th July 2006.

Department of Computer Science
University College London
London
WC1E 6BT
FAO Ms Hina Keval,
Human-Centred Systems Group.
Dear Ms Keval,

I would like to personally thank you for studying our CCTV Centre operations and producing such a comprehensive and helpful report.

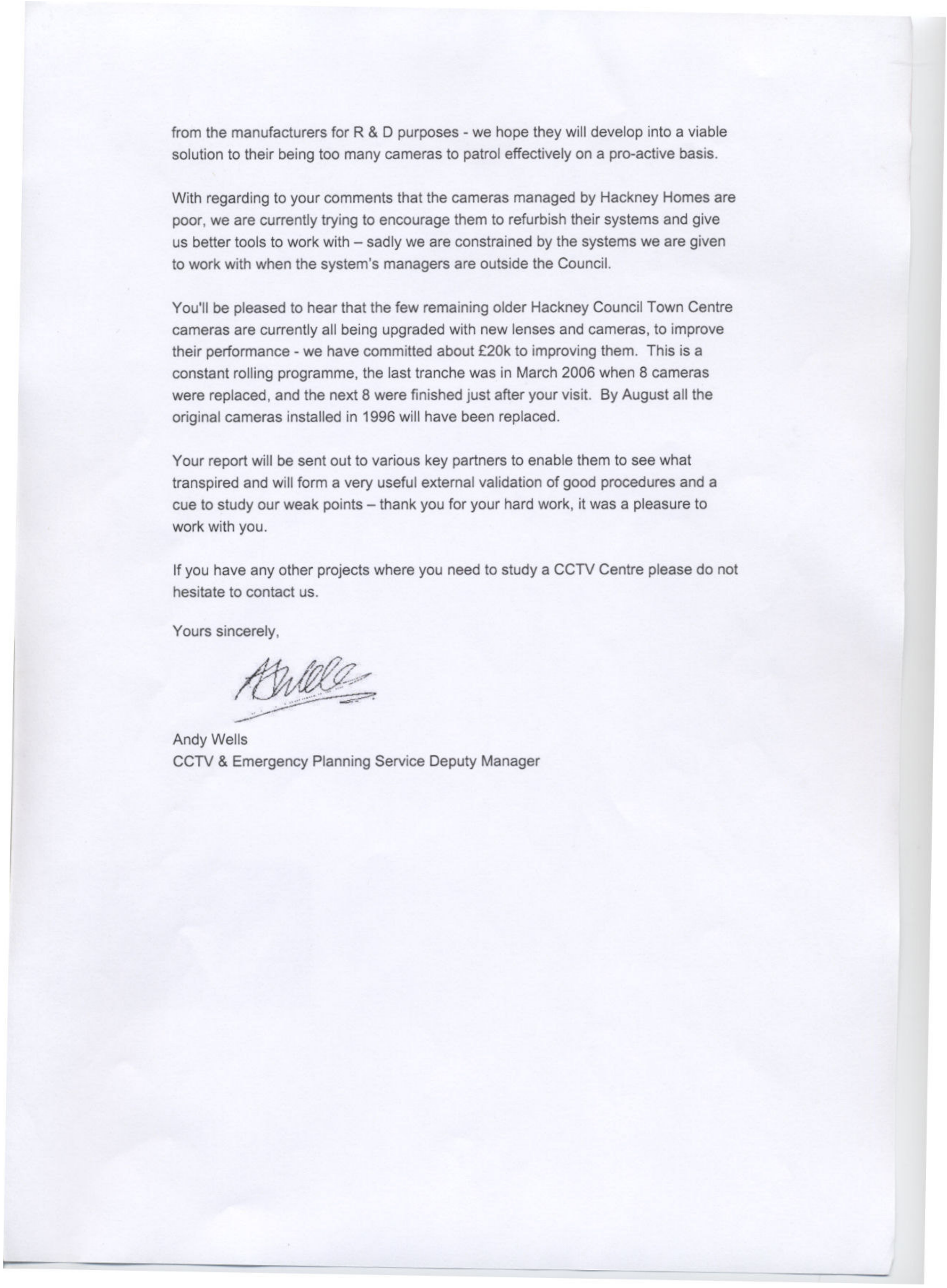
The point about the ANPR system is well made, we have put the system in for a short-term trial and I believe the results have proved it an excellent bit of kit. So far in the first 12 weeks of the trial it has resulted in the arrest of 38 people and the recovery of 33 vehicles with a total value of £126k.

Before this trial, the system / Operator interface was in a different room and used only when a pro-active operation was running, where we responded to a wide range of alarms. For this trial we moved the interface to the Main Monitoring Room where we use it 24/7 and only respond to stolen vehicles. Now we know the new system works we'll look at better procedures. One of these will be better resource allocation by the Team Leaders to prevent over-resourcing the ANPR and another will be a monitor centrally placed to allow better viewing.

Your point about trees is true, we do have great problems with them - unfortunately they keep growing and it's largely out of our control, although we do try and get the Highways people to trim them - the main problem is trees on private land overhanging the highway.

The point about the numbering system for cameras is also true, however it's extremely difficult to integrate new cameras into an existing system - we did do a complete reshuffle in 2002 when we moved and built a new control room, but as new cameras are added the position has deteriorated. With experienced operators we find it not to be a problem, however we know inexperienced Operators struggle at first. There is no easy answer to that problem, unfortunately.

Your comment about the two Artificial Intelligence monitoring systems not working very well (or even at all) is true – fortunately we have them installed on a free trial



from the manufacturers for R & D purposes - we hope they will develop into a viable solution to their being too many cameras to patrol effectively on a pro-active basis.

With regarding to your comments that the cameras managed by Hackney Homes are poor, we are currently trying to encourage them to refurbish their systems and give us better tools to work with – sadly we are constrained by the systems we are given to work with when the system's managers are outside the Council.

You'll be pleased to hear that the few remaining older Hackney Council Town Centre cameras are currently all being upgraded with new lenses and cameras, to improve their performance - we have committed about £20k to improving them. This is a constant rolling programme, the last tranche was in March 2006 when 8 cameras were replaced, and the next 8 were finished just after your visit. By August all the original cameras installed in 1996 will have been replaced.

Your report will be sent out to various key partners to enable them to see what transpired and will form a very useful external validation of good procedures and a cue to study our weak points – thank you for your hard work, it was a pleasure to work with you.

If you have any other projects where you need to study a CCTV Centre please do not hesitate to contact us.

Yours sincerely,

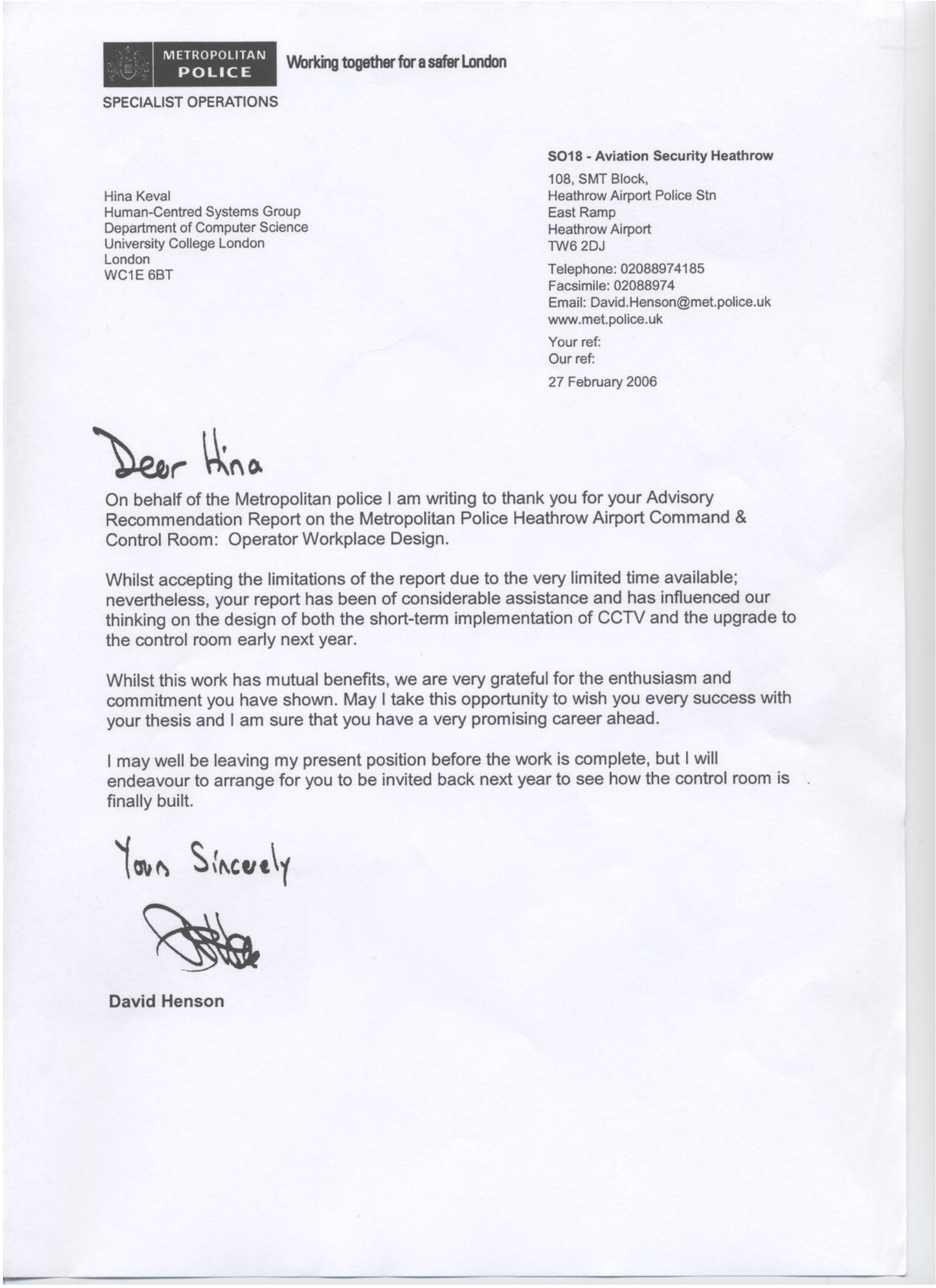
Andy Wells
CCTV & Emergency Planning Service Deputy Manager

Appendix B

Study 1: Letter from Metropolitan Police (Heathrow Airport)

This letter was received in receipt of an advisory recommendation report sent to the SO18 Aviation Security department at Heathrow airport. This report was prepared at the request of the command and control management Metropolitan Police team at Heathrow airport. The team requested a 3-month evaluation exercise of the command and control room to evaluate the control room set-up (work environment) and the technology systems in use; a set of recommendations were requested based on 2-day observations to inform the team on the redesign of their control room during the time of its upgrade (digital and network part of the CCTV system).

Although the evaluation was a 3-month study, observations were only conducted at the command and control centre for 10 hours and 2 hours at one of the three brand new C3i police control rooms (Hendon, London). The observation time was restricted due to information privacy issues as well a limited project time being made available for the research. Three operators and two supervisors were informally interviewed using unstructured interviews in order to identify: (1) their existing methods of work; (2) the tools and systems they used; and (3) the tasks they performed using these tools and systems. The proposed recommendations were presented at a meeting with the senior control room manager, two project managers and the strategic CCTV manager (Chief Inspector, Metropolitan Police).



METROPOLITAN POLICE Working together for a safer London

SPECIALIST OPERATIONS

SO18 - Aviation Security Heathrow

108, SMT Block,
Heathrow Airport Police Stn
East Ramp
Heathrow Airport
TW6 2DJ

Telephone: 02088974185
Facsimile: 02088974
Email: David.Henson@met.police.uk
www.met.police.uk

Your ref:
Our ref:

27 February 2006

Hina Keval
Human-Centred Systems Group
Department of Computer Science
University College London
London
WC1E 6BT

Dear Hina

On behalf of the Metropolitan police I am writing to thank you for your Advisory Recommendation Report on the Metropolitan Police Heathrow Airport Command & Control Room: Operator Workplace Design.

Whilst accepting the limitations of the report due to the very limited time available; nevertheless, your report has been of considerable assistance and has influenced our thinking on the design of both the short-term implementation of CCTV and the upgrade to the control room early next year.

Whilst this work has mutual benefits, we are very grateful for the enthusiasm and commitment you have shown. May I take this opportunity to wish you every success with your thesis and I am sure that you have a very promising career ahead.

I may well be leaving my present position before the work is complete, but I will endeavour to arrange for you to be invited back next year to see how the control room is finally built.

Yours Sincerely

David Henson

Appendix C

Study 1: Observation Checklist

Observation Checklist used at CCTV Control Rooms

1.0 Operator Tasks in Context

1.1 What hours do operators work?

1.2 What tasks do operators perform?

1.3 How frequently are these tasks carried out (day/night/weekend)?

1.4 Is the operator’s workstation suitable for their tasks?

2.0 Artefacts

2.1 Is the operator able to communicate effectively with other staff?

2.2 What systems are used to perform the tasks?

2.3 Are the video signal from street cameras reliable for display?

2.4 Is the operator able to communicate effectively using their tools?

3.0 Situation Awareness - System Issues

3.1 Is the number of cameras manageable for their tasks?

3.2 What tools support the operator's situation awareness?

3.3 Were the information displays well located for the operator to react?

3.4 Were there too many displays per operator to search or view?

3.5 Were operators able to view the entire camera scene?

3.6 Were the camera controls easy to use and usable?

3.7 Could the operator access equipment and information when needed?

4.0 Processing of CCTV video footage

4.1 Could the operator retrieve and make copies of CCTV video?

4.2 Was the CCTV video usable for investigating crime?

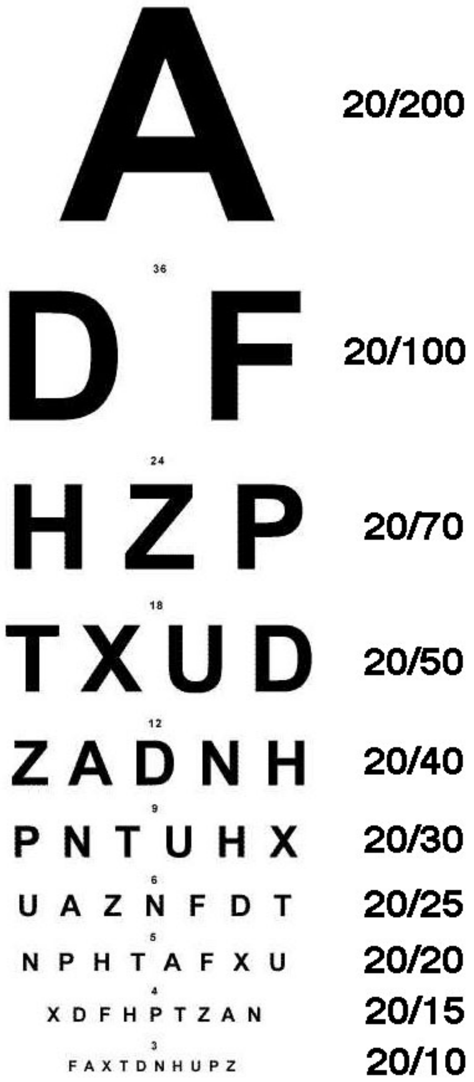
4.3 Was the operator able to change tapes using the video recorder?
(This is applicable where analogue CCTV systems are in use.)

Appendix D

Study 2 and 3: Standardised Snellen Eye Chart Test

Instructions:

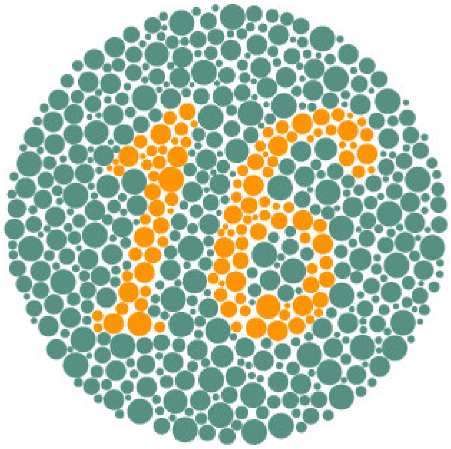
The following eye chart was printed to the correct scale and affixed to a plain white wall on the computer science laboratory at eye level. When checking visual acuity, one eye at a time is covered and the vision of each eye is recorded separately, as well as both eyes together. In the Snellen fraction 20/20, the first number represents the test distance, 20 feet. The second number represents the distance that the average eye can see the letters on a certain line of the eye chart. Thus, 20/20 means that the eye being tested can read a certain size letter when it is 20 feet away. The criteria was used to accept participants for the video experiments was 20/20 vision = 100% visual acuity.

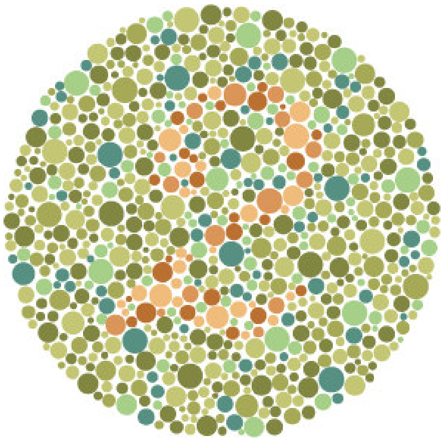


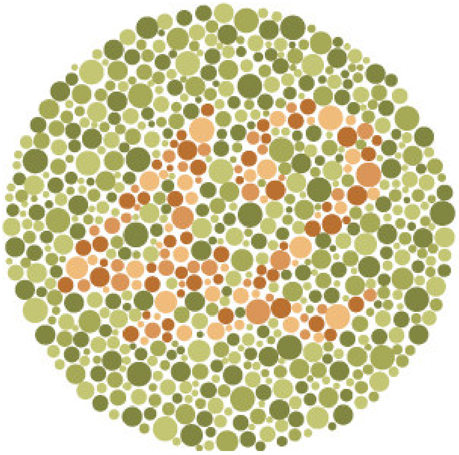
Appendix E

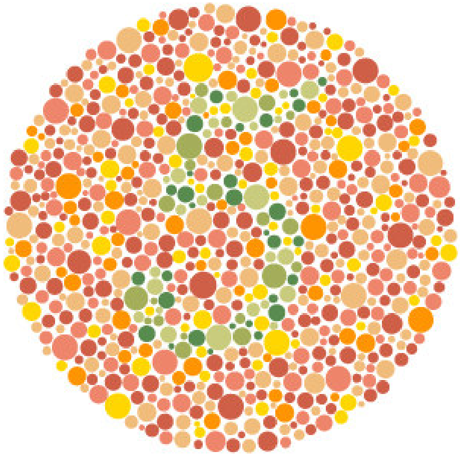
Study 2: Ishihara Colour Perception Test

Please say what numbers you see in each of the mosaics (each mosaic was presented one-by-one):









Appendix F

Study 2: Post-Experiment Questionnaire

1a) Please rank the following ethnic groups in the order in which you found them difficult to identify in the test. A score of 1 = very easy to identify, and 4 = very difficult to identify:

- Afro-Caribbean
- Oriental
- Indian-Asian
- White-Caucasian

1b) Please state your reasons for your rank choices in question 1a.

____

____

2a) Did you find male or females faces easier to recognise?

male / female / neither

2b) If you answered male or female in question 2a, please explain your choice.

____

____

3) What features did you use to help make your decisions in identification in the task?

____

4) Overall, please state whether you found his test easy or difficult to complete:

- Easy
- Difficult

Any comments:

____

____

____

Appendix G

Study 3: Detection Scenario Descriptions

Clip	CCTV Scene Content	Option 1	Option 2	Option 3
1	Appears to be a wallet grab	Two **males walk into the canteen separately. Male on left taps his friend, surprises him and says hello. His friend is surprised and smiles.**	Two males enter canteen separately. Male on left covertly takes something out of his friend's pocket, and then pretends to tap him to get his attention to say hello.	Two males enter canteen separately. Male on the left taps his friend to say hello and then leaves the canteen.
2	Drug exchange – then theft	Two males talk about a deal on the street. They discuss who should get what share of the money. The male on the left is in a rush and leaves quickly without saying bye to his associate.	**Two males talk about a deal of some sort on the street. The male on the right tries to sell the other male something, being as convincing as possible. The male on the left is not convinced and is reluctant to pay up. The male on the right sees his reluctance, grabs his money and runs off.**	Two friends talk about work. After catching up, the male on the left takes his friend's money and runs off as a joke.
3	Wallet grab from rucksack	Two males walk down the street. The male on the right bumps into the male on the left by accident. They continue to walk down the same road in the same direction.	Two males walk down the street. The male on the right tries to put his hand into the male on the left's pocket. He then stops himself as he hears another person coming up behind him.	**Two men walk down the street. The man on the left talks on his phone and the man on the right attempts to take something out of the other man's pocket. He succeeds, putting the item in his jacket pocket.**
4	Handbag theft	A female hugs her male friend. A cleaner walks past and then a male walks past.	**A female hugs her male friend, leaving her bag on the floor. A male walking towards the girl sees the bag and grabs it from the floor as he walks past. He walks off casually in the same direction.**	A female and male hug and start to chat. A male walks past, dropping his bag, then picks it up and continues walking past the couple chatting.
5	Rucksack theft	**A female walks down the street and a man brushes past her to steal something from her rucksack pocket. He then passes the item to another man who is walking in the opposite direction.**	A female walks down the street and a male who is in a rush walks past the girl as quickly as he can.	A male talks on the phone and quickly walks to find his friend.

6	Distressful argument, feud, money snatch	Two females are talking to each other catching up on old times.	**Two females are discussing an issue, the female wearing the white coat looks very distressed and decides to insult the other female and then runs off.**	Two friends are catching up on gossip and then the one wearing the white coat rushes off not realising the time.
7	Appears to be a wallet grab	Two males walk into the canteen separately. The male on the left taps his friend, surprises him and says hello. His friend is surprised and smiles.	Two males enter canteen separately. Male on left covertly takes something out of his friend's pocket, and then pretends to tap him to get his attention to say hello.	**Two males enter canteen separately. The male on the left taps his friend to say hello and then leaves the canteen.**
8	Drug exchange – then theft	Two males talk about a deal on the street. They discuss who should get what share of the money. The male on the left is in a rush and leaves quickly without saying bye to his associate.	**Two males talk about a deal of some sort on the street. The male on the right tries to sell the other male something, being as convincing as possible. The male on the left is not convinced and is reluctant to pay up. The male on the right sees his reluctance, grabs his money and runs off.**	Two friends talk about work. After catching up, the male on the left takes his friend's money and runs off as a joke.
9	Wallet theft from friend's pocket	**Two friends talk and then hug. The male on the right takes an opportunity to steal something out of his friend's jacket pocket when hugging him.**	Two colleagues are discussing a meeting place, they agree and then hug and say goodbye to each other.	Two colleagues are discussing a meeting place. They then decide to hug and at the same time they exchange a suspicious item.
10	Wallet grab from rucksack	Male wearing a rucksack looks around suspiciously and then asks a girl for the time.	**Male wearing a rucksack looks around suspiciously and then asks a girl for the time to deliberately distract her. At this point, a man steals something from the girl's rucksack pocket and walks away.**	A male wearing a rucksack is waiting for his friend and looks around to see if he can see him. He then asks a girl who passes him on the street for the time.
11	Drug exchange	Two male friends walk past each other on the street and briefly shake hands to acknowledge each other.	Two males walk past each other.	**Two males covertly shake hands and exchange illegal drugs during their brief handshake.**

12	Mobile theft from handbag on bench	A female takes a seat on a bench where two other females are seated and talking with one another. The female touches the bag of the girl beside her and finds her mobile phone. She steals it and walks off.	**A female takes a seat on a bench where two other females are seated and talking with one another. She then takes her mobile and decides to get up from the bench and walk off.**	A girl sits on a bench where there are other girls. She waits to see what she wants to do and then decides she wants to go somewhere else. She then leaves the bench shortly afterwards.
13	Mobile theft from handbag on bench	**Two females are talking together in a deep conversation on the bench. Another girl (oriental) is also seated on the bench, relaxing waiting for her friend to meet her. The girl then decides to dip her hand into the girl's handbag which is beside her and steals her mobile.**	Two females are talking together in a deep conversation on the bench. Another girl also seated on the bench is alone relaxing waiting for her friend to meet her.	Two females are talking together in a deep conversation on the bench. Another girl also seated on the bench is alone, relaxing waiting for her friend to meet her. She checks her phone and then leaves without her handbag.
14	Wallet grab from pocket	A female walks into canteen area alone and then a man walks to the canteen area too.	**Female walks into canteen area on her own and a male follows her and covertly removes her wallet from her back trouser pocket.**	A female walks into canteen area on her own and then a male follows her into the canteen area acting suspicious.
15	Wallet theft from pocket	A group of people are walking together; nothing interesting happens in the recording.	A group of people are walking together. A girl behind the group rushes to catch up with the group and tries to get one person in the group's attention. She then walks with the group in the same direction.	**A group of people are walking together. A female comes very close to another girl in the group and grabs her wallet from her pocket discreetly.**
16	Wallet theft from pocket	A group of people are walking together; nothing interesting happens in the recording.	**A group of people are walking together. A female comes very close to another girl in the group and grabs her wallet from her pocket discreetly.**	A group of people are walking together. A girl behind the group rushes to catch up with the group and tries to get one person in the group's attention. She then walks with the group in the same direction.
17	A girl unzips her bag and removes her mobile phone.	**A girl unzips her bag and removes her mobile phone from her bag.**	Girl fidgets with her bag whilst talking with her friends.	Girl removes her wallet from her bag to give some money to her friends.

18	Friends handshaking	A group of three friends are standing chatting to each other on a busy street.	**A group of three friends are standing chatting to each other on a busy street, and then they all shake hands with each other.**	The man with a striped woolly hat attempts to pick pocket from a man standing on the street who is chatting to his friends.
19	Checks wallet contents and phone	**Someone tried to steal something from the pocket of someone in the group of three standing and chatting.**	A group of three stand talking. The girl on the far right in the black jacket checks her watch for the time and waits for another person to join the group.	The girl in yellow jacket checks her mobile phone and answers a telephone call.
20	Friend touches friend’s bag	A group of friends are laughing and talking in the street.	A group of friends are laughing and talking in the street. The man on the left comments on the girl’s bag and says he likes it.	A group of friends are laughing and talking in the street. The man on the left comments on the girl’s bag and says he likes it, and then discreetly drops something in her bag.
21	Walking	Two men shake hands at a street entrance.	Two men exchange something when shaking hands with each other.	Two men walk past suspiciously and exchange something. An elderly man checks his bag as he thinks something is missing.
22	Girl asks man for directions	A couple are laughing and chatting together in the street and pointing at people.	Girl approaches a man, who is standing waiting for someone, to ask for directions.	A girl sees her friend and approaches him, then is very happy and shows him her friends who are standing on the end of the street.
23	Girl gives a leaflet to a passerby	A girl tries to sell a passer-by a promotion and she doesn't take any notice of the passerby ignoring her.	A girl offers a leaflet to a passerby on the street, but the passerby doesn't pay attention to her.	A passerby on the street takes a leaflet from a girl.
24	Same as above, girl gives wallet	Two girls are chatting on a park bench. The girl on the right passes her wallet to her friend sitting next to her.	Two girls are chatting on a park bench. The girl on the right passes her phone to her friend sitting next to her.	Two girls chatting on park bench.
25	People chatting in street, man checks his mobile phone	A man walks down the street, removes his mobile phone from his pocket.	A man walks down the street, removes his mobile phone from his pocket and sends a text.	A man walks down the street, thought his phone rang, removed from his pocket to check.

26	Girl on far right of bench untangles her music player headphones	Girl on far right of bench tries to use her mobile phone.	Girl on far right of bench untangles her music player headphones.	Girl on far right of bench shows her necklace to her friend.
27	Boy picks keys up and gives to girl	Four people are chatting on the street. A big cement truck passes by.	Four people are chatting on the street. The boy thinks something dropped and checks by trying to pick it up. He then shakes the girl's hand.	Four people are chatting on the street. A big cement truck passes by. The boy picks up something which dropped and gives it to the girl.
28	Boy removes paper from his bag	A group of people are chatting. The girl on the far right removes a book out of her bag and opens it.	A group of people are chatting. A boy then removes a newspaper out of his bag and reads it.	A group of people are chatting. The girl on the far right removes a box of cigarettes from her bag.
29	Boy drops keys	A group of three are chatting on the street. The boy accidentally drops his keys on the floor.	A group of three are chatting on the street.	A group of three are chatting on the street. The girl in the centre (black jacket) shakes the boy's hand.
30	Girl checks her mobile	A group of girls are standing chatting to one another. Nothing interesting happens.	A group of girls are standing chatting to one another. The girl in the centre (white jacket) reads her text from her mobile and appears shocked.	A group of girls are standing chatting to one another. The girl in the centre (white jacket) puts her mobile phone in her bag and then yawns.
31	A man reads a bus timetable, and checks change in his wallet.	A man reads a bus timetable, and checks change in his wallet.	A man reads checks change in his wallet.	A man reads a bus timetable, and throws a receipt from his pocket on the floor.
32	Girl checks her phone for a text	Girl in a black logo top places her mobile phone in her bag.	Girl in a black logo top checks her phone for a text.	Girl in the black logo top checks her diary and then places it back in her bag.

Appendix H

TEC-VIS Framework (Version 1)

1.0 Introduction

This document is aimed at security practitioners (CCTV control room designers, consultants, and owners), and serves to communicate a background in the planning, design and configuration of a security control room. The document provides a best-practice framework on the design of security control rooms in specific. The framework - the Task-Effective CCTV for Video in Security (TEC-VIS) was developed as a result of 3-years extensive PhD and consultancy research in London by the document author. The research was undertaken using methods and techniques taken from cognitive psychology, human-computer interaction (HCI), computing, and sociology disciplines. The framework details 8 phases which emphasises the importance of understanding the users within the control room and outside the control room system, operator tasks, stakeholder and system requirements, and then assessing the risks within the system.

Security control rooms are human operated work environments in which several tools and systems are used by a handful of operators to monitor and control events on the outside. Over the last few decades, the vast developments in the security and surveillance technology market has revolutionised security control rooms today creating complex systems and operator tasks. For the design of a control room – like with any workplace design, it is crucial that the user (operator) is placed at the centre of the design process rather than at the end when the environment has been created and the system has been implemented. The ISO 11064 standard is currently the only documentation which provides guidance on the design of control rooms. The standard is based on ergonomic principles and covers guidance on the physical aspects of control rooms such as: workstations arrangements, control room layout, use of displays and controls, and maintenance. The main drawback of the ISO standard is that it does not provide guidance on the set-up of systems within control rooms, and disregards the cognitive factors which can affect operator task performance. Apart from this standard, there is no other publicly available guidance for security practitioners to support the design of control rooms.

In this document, the Task Effective CCTV for Video in Security (TEC-VIS) framework is presented. The framework was developed by taking a user-centred design (UCD)[35] approach which is designed to support security practitioners (designers and owners) during the planning and design stages of a security control room deployment. The framework was developed in 2007 following 3 years PhD research as well as external consultancy at 14

[35] UCD is a design approach in which the emphasis is on the user and through which a high level of usability is achieved.

public-space CCTV control rooms in the UK (see the following publications: Keval, 2006; Keval and Sasse, 2006; Keval, 2005; and Keval, 2007). Also, the technical guidance was formed as result of two lab experiments where the performance of security tasks was evaluated with 160 participants at University College London (Computer Science, Human Centred Systems).

A brief overview of the research disciplines carried out in this study is described in Section 2; following this, the framework is described (see Section 3). Section 4 summarises the framework in the form of an 8-phase checklist. Section 5 and 6 includes the acknowledgements and reference list.

1.1 Overview of Ergonomics and Human-Computer Interaction

Ergonomics is a scientific discipline which is concerned with the design of everyday products and systems with the aim of optimising human use, taking into account health, safety, comfort, and performance. Ergonomics is also known as human factors engineering, this term is more common in North America, whilst Ergonomics is a European favoured term. The discipline is broadly split into three domains by the International Ergonomics Association[36]: (1) organisational; (2) physical; and (3) cognitive.

(1) Organisational ergonomics (also known as macroergonomics) is about optimising the socio-technical system, these include: work policies, organisation structure, and processes. In the context of a security control room the organisational ergonomic factors a designer should consider when optimising the operator's system include: job satisfaction; shift work; supervision; safety culture; teamwork; and ethics.

(2) Physical ergonomics is concerned with the anatomical, anthropometric, biomechanical, and physiological characteristics of human users, thus relates to the user's physical activity at work. In the context of a security control room, the following aspects of the operator's physical workplace should be considered during the design process: the operator's working postures; their repetitive actions, and the layout of their personal workstation. These physical factors should be considered in the control room design process, with the aim to maintain the operator's health and safety to a high standard. As a control room is a shared work environment, where tasks and activities are carried out collaboratively, the design of the operator's physical environment should also be considered. For instance, the designer should consider the required levels of lighting and space, as well as the minimum levels of noise.

(3) Cognitive ergonomics is concerned with the human mental processes (human cognition) during the performance of tasks in the workplace. Human cognition consists of: attention,

[36] http://www.iea.cc/

memory, audio/visual perception, decision-making, and motor response. These all affect the interactions amongst human users and the various elements of the system.

When assessing the CCTV operator's performance in their tasks, the cognitive factors that can be used to make the assessment include measuring their mental workload (*i.e.* how much information they absorb when dealing with radio communication, vigilance: how they decided whether to react to incidents observed on video monitors). The purpose of assessing the user's cognitive capabilities is to make it possible to determine whether the user has understood and digested the information available to them, and to ensure that are able to make the correct decisions to communicate and execute their tasks. The outcome of such an evaluation can help in the design of security tasks, the control room environment, and the systems to allow for the proper allocation of tasks for operators and across the system.

Human-Computer Interaction (HCI) is a branch of computer science which involves the study of how users interact with computer systems. HCI goes beyond designing screens and menus that are easier to use, an is largely concerned with the long-term effects that systems will have on humans, and explores the reasoning behind building specific functionality into computers. HCI methods and techniques were applied for the field and empirical research which contributed to the framework which is presented in the next Section.

2 TEC-VIS Framework

Having a framework for the deployment of a CCTV security control room allows security practitioners, designers, and CCTV owners to understand each of the different stakeholders involved with the security system. Also, it enables an analysis of the stakeholders interacting with the system and an analysis of the technical capabilities of the system in context to identify factors which may reduce the effectiveness of the system. Due to the complex nature of video security systems, and the multitude of factors affecting the use of security (and other technologies). This framework is limited in providing guidance on:

- How to analyse the operators of the working security system
- How to analyse the technical capabilities of the system
- The requirements for recording, distributing, and displaying digital CCTV video.
- The physical set-up of the security control room
- The configuration of the CCTV camera environment

The TEC-VIS framework is intended for ***security practitioners (designers*** **and CCTV security owners)** who are responsible for the design, development, and deployment of a security control room - where CCTV technology is used (both analogue and digital). The framework is designed to support the deployment of any type of security control system regardless of its function, size, and location. In the first two phases, the guidance provides steps to allow the reader to identify the essential building blocks of the CCTV security system before deciding on the task and system requirements. These phases incorporate a user-analysis method: a requirements analysis method used during the design of a product or system. Phases 3-8 describe the steps required to identify and analyse both the stakeholder and technical requirements for the system taking a socio-technical approach. The remainder of this Section describes each of these phases individually.

Phase 1: Specify Goals of Security System

Objective:

In the discipline of HCI and ergonomics, a goal for a system is defined for the purpose of explicitly identifying what the user wants to do with the system, thus asking the question: what do they want to achieve? Managers, engineers and designers who are involved in the design and deployment of a system tend to assume that the stakeholders within the team will adopt the goals of the wider system - but this is *not* true. This is a risky assumption, as people may be perverse or might not appreciate what they are supposed to do (Shepherd, 2001). In this way, for a CCTV system deployment - it is important to *first* define the goals of the control room to determine what the system will actually do.

Process:

The goals of a system define what the system will do to fulfil the purpose/aims for the system. An example of a CCTV goal could be to: to set-up a control room for operators to monitor live video to manage and control crowds at an underground tube station. More than one goal can be defined, and these can be decomposed into sub-goals which are goals which are related to the primary goal. For example, for crowd control, the sub-goals may include:

- Monitor the flow of people during peak hours
- Monitor for incidents during peak hours
- Monitor for ticket touts etc.

Phase 2: Identify Stakeholders in Security System

Objective:

Stakeholders play very important roles in the design of systems - both in the software and hardware of systems. In strategic management, Freeman (1984) views a stakeholder in an organisation (by definition) as *"…any group or individual who can affect or is affected by the achievement of the organisation's objectives."* In the context of a security control room system, a number of stakeholder groups exist. Figure 1 illustrates the hierarchy of the different stakeholder which may be involved and interact with operators working in the control room.

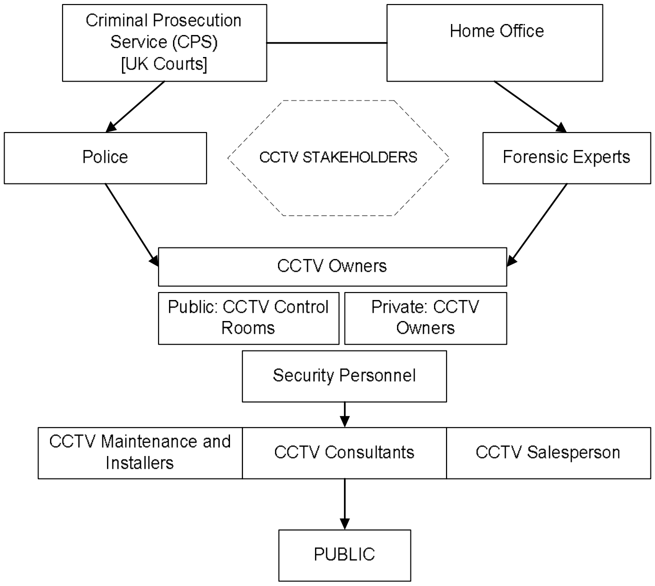


Figure 1: A hierarchy of CCTV stakeholder groups

Process:

The different stakeholders who interact with the CCTV system must be considered throughout the planning and design stages of the system, as *"…it will allow them to influence the development of the system along the way results in the correcting of misconceptions and evaluating intermediate solutions,"* (Beale et al, 2005).

The identification and involvement of the different CCTV stakeholders is an iterative process which should be facilitated through a series of stakeholder communication and feedback sessions. This can be achieved by running regular stakeholder meetings, which are ideally held throughout the lifecycle of the system (every 6-months). Regular meetings will ensure that system performance is optimised, and that the system goals are matched with the user's requirements, the tasks, the security environment, and the tools used for security.

A typical CCTV system is set-up to record surveillance video and images containing individuals and car number plates, therefore the CCTV data is classified as personal data. In privacy literature, where personal data is used, the stakeholders are described as 'data users' and 'data subjects.' In the context of CCTV, these are:

- Data controllers and users: organisations and/or individuals who own and directly use the data (*i.e.* CCTV control rooms).
- Primary data subjects: specific targeted individuals (*i.e.* criminals and suspects).
- Secondary data subjects: non specific targeted individuals (*i.e.* members of the public).

The data users and subjects should be defined for each stakeholder group across the system, to ensure that there is a clear understanding of who is using the CCTV video and who will be captured on video. The classification of these users is also useful for understanding the data handling issues such as:

- Who will own the CCTV data?
- Who is legally responsible for the access, use, and destruction of CCTV data?
- Who will be recorded on CCTV video and images, and why?
- How will the CCTV video be protected from non-data controllers and users?

Phase 3: Identify Stakeholder Tasks Required by the System

Objective:

Tasks as goal oriented actions are one of the most important aspects of HCI design and research. In HCI and ergonomics literature, a task is described as an action or set of actions carried out by an individual in order to meet the system goals. This action can be 1) physical, for instance an operator using control facilities to track a target on a camera and 2) cognitive, for instance an operator monitoring real-time CCTV video, and deciding whether to react or not to particular events. The goals, and sub-goals identified in Phase 1 should be decomposed into stakeholder tasks. The objective of defining these tasks is to enable the system owner to:

- Determine whether the tasks can be appropriately supported by the available technology
- Determine what knowledge, skills, tools, and training is required for users to perform the task effectively.
- Allocate these tasks appropriately to humans and the systems available.
- Input into the (future) design of supporting systems to offer new ways of performing parts of a given task.

Process:

A key issue in usability is that of understanding users, and a key part of user-centred design is that of describing the tasks that the users expect to be able to accomplish using the designed system. This can be achieved by conducting a Hierarchical Task Analysis (HTA). HTA involves identifying tasks, categorising them, identifying the subtasks, and checking the overall accuracy of the model. HTA is useful for the design of systems as it provides a model for task execution, enabling the designer to envision the goals, tasks, subtasks, operations, and plans essential to operator's activities. HTA is useful for decomposing complex tasks, but has a narrow view of the task, and normally is used in conjunction with other methods of task analysis to increase its effectiveness. HTA serves as both an analytical framework and a practical tool for designers. For a more detailed overview of HTA (see Shepherd, 2001).

Phase 4: Identify Storage, Network and Display Requirements

Objective:

Very often, the performance of a CCTV system is assessed when it operates in real-time mode (*i.e.* when the operator is observing real-time CCTV video directly from street cameras in a CCTV control room on a video wall. Although, performance is necessary under real-time operation, the performance of the system should also be assessed with real users and tasks with CCTV video which has been recorded and streamed (*i.e.* used for remote surveillance). By assessing the effectiveness of recorded and streamed CCTV video, the owner will be able to determine whether the system can produce usable recordings for police investigations, and perform real-time reactive and proactive surveillance over a network without error.

Process:

The performance of the security observation tasks should be assessed with the intended CCTV users based on tested guidance under two conditions (depending on the set-up chosen): 1) with recorded CCTV video and imagery and 2) with networked CCTV video accessed remotely. This evaluation process can be facilitated by following the recommendations on the image resolution, temporal and spatial resolution for CCTV (see Table 1). The recommendations given are for two of the most commonly performed security observation tasks: (1) face identification and (2) detection. These recommendations are based on findings from two empirical experiments in CCTV video quality (Keval, Gatusso., and Maclennan (2007); Keval and Sasse (2008b: 2008c). These tasks can be performed by CCTV users of any skill and level of experience. (Note: the two experiments were carried out with untrained CCTV users) to determine a conservative set of video quality requirements for these tasks.

CCTV Task Type	Image Resolution	Temporal Resolution*	Spatial Resolution**
Face Identification Identify an unknown person from a CCTV image with the use of a good quality image	Low Risk[37] Environments Common Image Format (CIF) resolution (352 x 288)	8+ fps	MPEG-4 Video CODEC No less than 52 Kbps
	High Risk[38] Environments Above CIF resolution (352 x 288)	12+ fps	MPEG-4 Video CODEC No less than 52 Kbps
Detection of Events Detect an crime event (suspicious/crime event from CCTV video)	Low Risk Environments CIF resolution (352 x 288)	8+ fps	MPEG-4 Video CODEC 2000 Kbps
	High Risk Environments Above CIF resolution (352 x 288)	12+ fps	MPEG-4 Video CODEC 2000 Kbps

Table 1: Recommended parameters for video transmission, storage and display

Phase 5: Identify Functional Capabilities of the CCTV Security System

Objective:

By treating the CCTV system as a part of a wider security system, the functionality required for the system can be defined (Walters, 1995). In this phase, the system owner should determine what exactly the system control room should be capable of in terms of function. The objective of this is to make sure the technology used is set-up effectively for each of the stakeholders, and their tasks.

Process:

The functional capabilities of the system can be identified by reviewing the system goals, stakeholders, and their tasks. Following this, the owner should consider the intended set-up of the system to ensure that the control room environment is effectively designed as well as the external control room environment. This configuration process can be facilitated by following the best-practice recommendations detailed in Tables 2, 3 and 4. These recommendations

[37] Low risk: where crime is very unlikely (i.e. low crime rate areas and places where objects are not valuable).

[38] High risk: where crime is very likely (i.e. town and city centres, secluded/hidden spots, places where objects are highly valuable).

were devised based on previous standards and research, but largely from the CCTV control room field research (Keval and Sasse, Forthcoming) and following two CCTV video quality experiments Keval et al. (2007) and Keval and Sasse (2008a:2008b).

Internal: The CCTV Operator’s Work Environment

General Security Environment Guidance:

1. Where CCTV video is monitored in real-time by CCTV operators and security staff within a control room environment - the International control room standards should be followed (BS EN ISO 11064: Ergonomics Design of Control Centres, 2001).

2. All CCTV users who perform security observation tasks with CCTV video and images *must* be trained and regularly assessed (every 6-8 months) on their tasks, the surveillance environment, security procedures, and the use of tools and equipment.

3. Any tools and equipment used for security observation tasks must be regularly tested and maintained periodically every 6-8 months. Tools should also be positioned around the CCTV user at their workstations. Tools should be safely arranged, free from obstruction and must be easy-to-reach, and adequate for the user’s tasks.

4. Any essential tools and equipment that are faulty should be repaired quicly, or if necessary disposed of safely and replaced.

5. Any non-essential tools and equipment which are not used frequently should be placed in storage outside the work environment. Tools used less frequently but required for emergency use must be accessible at all times.

6. To avoid confusion and delays during incident reporting over the radio or telephone, a standard communication protocol should be used which sets out a standard set of rules for communicating information over radio and phone. The protocol should include the use of the NATO phonetic alphabet and police identity codes.

7. Noise within a security observation environment should be kept to a minimum to avoid task interference. If a radio system is used, operators should be responsible for responding to radio channels equally (based on the level of radio contact and priority required for calls). A detachable headset should be used if the operator is responsible for several radio and telephones.

Table 2: Best-practice recommendations for general security

Internal: The CCTV Operator’s Work Environment

CCTV Camera and Monitor Guidance:

1. Where several CCTV cameras are observed by operators 5+ hours, no more than 5 video monitors should be used to observe general CCTV activity at a time when the observer is required to monitor scenes (figure adapted from Tickner and Poulton, 1973).

2. For a detection task – where specific events are detected such as suspicious actions and crimes, the observer must perform this task on a single video monitor.

3. The level and complexity of the CCTV video should be arranged in the monitor bank equally between operators, to reduce the probability of errors and information overload (reference). CCTV cameras located in areas where the levels of crime are moderate to high should be displayed at all times to ensure that the crime hotspots are in constant view.

4. The choice of camera displays on the monitor banks should be reviewed on a weekly basis by security management and the local police. The choice of camera displays will vary depending on the priority of surveillance.

5. Video monitors placed in the bank positions should be positioned at a minimum distance of 1.5 m or more from the observer.

6. Video monitors placed in the bank positions should be in the size range between 17-28 inches for a monitoring task (Wallace and Diffley, 1998).

7. Detection tasks should be ideally performed at the observer’s workstation on a spot / incident monitor which is a monitor used by operators for a close up inspection of events. Spot/incident monitors should be positioned 0.5 -1.5m from the operator (Wallace and Diffley et al, 1998).

8. A maximum of four pictures per monitor should be used for picking details from a scene which has been detected as suspicious by two observers (adapted from Wallace and Diffley, 1998).

9. By using a switching/auto-cycling camera display function, the displays of a set number of CCTV cameras can be displayed over a set time in groups. The timing of the display from one set of cameras to another set should occur at the same time and should switch display no less than 5 minutes at a time.

10. To avoid fatigue, CCTV operators and security personnel should view CCTV video activity on monitors between 50-60 minutes and should take a 5-10 minute comfort break every hour (HSE, 1994). Comfort breaks should be taken away from control room, and an operator should never be left alone unsupported during break times.

11. For a CCTV system which is linked to 8-10+ cameras, a list of the camera ID, physical street location, type, and position should be accessible on a computer database and identifiable on a geographical map. Lists should be updated following any changes. Each CCTV camera in the lists should be numbered logically in its geographical order, rather than the order of its installation date.

Table 3: Best-practice recommendations for configuring CCTV camera and monitors

CCTV System Strategy and Design:

1. The type of CCTV camera chosen must be suitable for the observer's task. For example a movable camera is required for tracking a suspicious target and a fixed camera is suitable for monitoring crowds outside a train station. If a task is added, the right CCTV camera must be supplied.

2. All of the CCTV cameras connected to the system should be adequately safeguarded from vandalism, accidental and weather damage.

3. The lighting to the camera capture area should be adjusted until the scene and individuals are clearly recognisable and identifiable. The possible effects of unwanted and variable lighting (*i.e.* flashing emergency lighting/weather conditions, Christmas decorations) should be considered during the camera installation.

4. Leaving unused CCTV cameras connected to the system will lengthen the camera list unnecessarily, making it difficult for the observer to locate the right camera at the right time. CCTV cameras placed in inappropriate positions may result in camera blind spots. Failure to review and – where necessary – move or decommission equipment may result in incidents and individuals may go unnoticed or mis-identified/mis-interpreted.

5. CCTV cameras should not be left in operation in places where no activity or requirement for CCTV is needed. A nominated individual should review and assess each CCTV camera's purpose and operational functions, periodically every 4-6 months.

6. Where 10+ CCTV cameras are used for security and surveillance tasks depending on the configuration of the technology the camera signal loss may occur for various reasons (*i.e.* technical failure or bad weather). In the event of camera signal loss a fault detection and repair reporting tool should be used, which automatically provides a report of camera maintenance required, as well as repairs for the system owner to respond and resolve.

7. Individuals who are responsible for maintaining trees, foliage, and bunting should ensure that they do not obstruct, or damage any CCTV cameras. This should be assessed following any developments carried out near to any public CCTV cameras.

Table 4: Best-practice recommendations for CCTV system strategy and design

Phase 6: Identify and Eliminate Potential Obstacles to Functional Capabilities

Objective:

By completing phases 1-5, it will be possible to oversee where the deployment of the CCTV system may start to fall down. In this phase, a review of the functional capabilities of the CCTV system may reveal instances where the technology is not available/or suitable for the user's tasks.

Process:

Obstacles can be identified by reviewing the Phases 1-5; each obstacle should be rated using a severity rating[39] scale. The ratings relate to the potential frequency, impact and the persistence of the problem. Obstacles with ratings 4 and above should be given priority to resolve, and the remaining should be noted and resolved later (time and cost permitting).

[39] Suggested severity scale: (1 – extremely minor; 2 – minor; 3 – moderate; 4 – serious and 5 – very serious)

Phase 7: Identify and Resolve Potential Stakeholder Conflicts

Objective:

It is likely that there will be conflicts in the stakeholder requirements as each stakeholder will have different goals in the system, which may lead to different requirements for the system. Stakeholders may disagree on the priorities of different requirements. Often these disagreements are as a result of functional requirements desired by some stakeholders, which may be undesirable to others. The objective of this phase is to reduce the likelihood of a system failure, caused by the conflicting stakeholder requirements.

Process:

Potential stakeholder conflicts should be addressed and resolved during the planning stages of the system design following Phases 1-6. Once these conflicts are identified, a level of severity should be using the ratings (same as Phase 6). The conflicts with a rating of 4 and above, should be resolved immediately, and the remaining should be noted and resolved later (time and cost permitting). Although this process (like in Phase 6) is a hypothetical exercise, the process is extremely valuable. For instance, if conflicts can be foreseen at the early stages of the security system design, forward planning can be carried out so that future conflicts between stakeholders can be avoided.

Phase 8: Restate Security Goals and Stakeholder Tasks

Objective:

By completing Phases 1-7, the following should now have been defined:

- Goals for the security control room
- Stakeholders belonging to the control room system
- Tasks and sub-tasks the different stakeholders will perform
- Data quality requirements (*i.e.* storage, network and display)
- Functional capabilities of the CCTV security system
- Potential obstacles to functional capabilities
- Potential stakeholder conflicts

In the final phase, the system owner should perform a holistic review of the requirements identified in Phases 1-7 to re-define them in line with the findings identified across the phases. The objective of re-defining the goals, task, and the user/system requirements is to ensure that the owner does not lose sight of their primary goals for the system.

Process:

This phase is iterative and the analysis should be carried out through a series of stakeholder meetings - so that the requirements for the control room accommodate all of the stakeholder groups – and that the requirements are feasible, both financially and technically.

3 Summary of TEC-VIS

The TEC-VIS framework which was described in the previous Section is summarised in the 8-phase checklist below:

TEC-VIS: A Best-practice Framework for CCTV Effectiveness

Phase 1 **Specify Goals of CCTV Security System**

1.1 Identify the primary goals of the security control room.

1.2 Identify the sub-goals (related to primary goals) of the security control room.

Phase 2 **Identify Stakeholders**

2.1 Identify every stakeholder who has an involvement in the CCTV system.

2.2 Place each stakeholder in a hierarchical order as shown in Figure 1.1.

2.3 Identify the data users and data subjects for each of the stakeholders.

2.4 Consider the data handling issues concerning privacy and usage of CCTV.

Phase 3 **Identify Stakeholder Tasks**

3.1 For each stakeholder identify the tasks they are required to perform.

3.2 Complete a Hierarchical Task Analysis and decompose the security goals into user tasks.

3.3 Determine whether the tasks are suitable for each of the users and that they are properly understood.

Phase 4 **Determine Video Storage, Network, and Display Requirements**

4.1 In low risk environments, for a face identification task – the video recording system should be set to record video at a minimum recording video resolution of 352x288 (CIF), compression quality not less than 52 Kbps (MPEG-4), and at 8 fps or higher.

4.2 In high risk environments, for a face identification task (performed with recorded CCTV video) - the requirements should follow the same as above, but with video recorded at 12 fps or higher.

4.3 In low risk environments, for a detection task - the video recording system should be set to record video same as above, but with a compression quality of no less than 2000 Kbps (MPEG-4).

4.5 The system should be set-up on a trial basis and should be evaluated with observers against the intended security observation tasks to assess the likelihood of errors.

Phase 5 **Identify Functional Capabilities**

5.1 Identify what the CCTV security should do in terms of functional capabilities, based on the analysis of the security goals, 1 stakeholder tasks, and the technology available.

5.2 Follow the best-practice recommendations for configuring the observer's internal and external CCTV environment.

Phase 6 **Identify and Eliminate Obstacles to Functional Capabilities**

6.1 Identify obstacles that may reduce or remove the effectiveness of the capabilities.

6.2 Prioritise the obstacles by assigning a severity for each one, and then attempt to eliminate these obstacles in priority order.

Phase 7 **Identify and Resolve Stakeholder Conflicts**

7.1 Identify conflicts between stakeholders, *i.e.* their requirements which may reduce or impair the operational performance of the CCTV System.

7.2 Prioritise the stakeholder conflicts by assigning a severity rating for each one, and then attempt to eliminate these obstacles in priority order.

Phase 8 **Re-define Goals, Tasks, and Requirements**

8.1 Based on the analysis carried out in phases 1-7, review the security goals, stakeholders, tasks, and functional capabilities of the control room system.

8.2 This phase should be completed each time the security control room has been altered.

4 Framework Document Summary

In this document, a framework is presented detailing an eight-phase best-practice guide for the deployment of a security control room system, where CCTV technology is used. The guidance is designed to support security practitioners and designers and provides high level guidance on carrying out a user-analysis and system requirements analysis for a CCTV deployment.

Also, low level guidance is provided in the form of recommendations, to ensure that the system is configured to record, transmit, and display CCTV video and images that are usable for human operator security tasks. In addition to this, best-practice guidance is given on the set-up of both the operator's control room environment as well as the camera environment.

In summary, the owner of the CCTV control room should explicitly define the security goals and user tasks, and – until guidelines based on scientific data become available – they should test whether their chosen system configuration is fit for purpose. The evaluation must consider the operator factors (*i.e.* tasks, environment, training, knowledge, motivation and job satisfaction etc.), to determine what and where the biggest gains and losses in system performance are, and to weigh the cost of higher image/video quality against the cost of a drop in performance.

5 Acknowledgements

This research was not possible without the support and time generously given by a number of academic and commercial contributors and collaborators. Their efforts and contributions are very much appreciated:

- Prof. Angela Sasse (University College London, UK)
- Dr. Martin Gill (Ergonomics Safety Research Institute, Loughborough University, UK)
- Dr. Iain Darker (Ergonomics Safety Research Institute, Loughborough University, UK)
- Ian Nimmo (User Centered Design Services Inc, USA)
- Peter Fry (CCTV User Group Chairman, UK)
- 14 public-space CCTV control room managers and staff (UK).
- Metropolitan Police, UK
- Home Office Scientific Development Branch, UK
- Forensic experts, UK

6 Application of TEC-VIS Framework

The TEC-VIS framework was developed as a result of the findings from the CCTV control room field study and the CCTV observation task experiments, as well from previous research literature. The framework can be used a guide to support the planning, design and deployment phases of *any* CCTV security system. The framework should enable the security practitioners, security designers, and system owners to:

- Identify and acknowledge the intended security goals, stakeholders, tasks and technical capabilities the system will deliver.
- Identify and reduce the factors which may reduce the system's performance, effectiveness and efficiency (*i.e.* stakeholder conflicts and technical obstacles).

Phases 1-7 should be completed in numerical order. The TEC-VIS process is iterative, therefore the phases should be completed before the system is deployed, after it is deployed, and every time the system is altered (*i.e.* if a new piece of technology, stakeholder group, or a task introduced into the system). The TEC-VIS exercise must be documented and accessible to all CCTV stakeholders.

To understand how this framework can be applied in practice, an example of CCTV deployment scenario is described to illustrate how the framework should be completed. In this scenario, a hypothetical situation is described in which a business owner has found an opportunity to deploy a CCTV security system. The scenario given is based on a true situation and was adapted purposely so that the application of the TEC-VIS framework could be best illustrated.

Phase 6 (analysis of stakeholder conflicts), Phase 7 (analysis of obstacles to functional capabilities), and Phase 8 were not completed for the scenario. Phases 6 and 7 however, must be followed for a real CCTV deployment - as performing a risk analysis for the CCTV security system is an essential part of the evaluation process. Phase 8 is the final phase detailed in the TEC-VIS framework, which is a compulsory and iterative phase, which can only be carried out once Phases 1-7 are completed and reviewed.

Scenario: Jewellery store owner to deploy a CCTV video security system:

A town centre jewellery store owner (Mr. Elton) was suffering from a huge financial loss and this was potentially as a result of employee theft, customer shoplifting, and vendor fraud. The shop at present uses security mirrors periodically to monitor staff and customer activity whilst the owner is serving customers at the cashier till.

In attempt to recover the losses, Mr. Elton wanted to replace these security mirrors with a CCTV security system for three main reasons:

- To deter thieves
- Detect theft through monitoring the shop blind spots the mirrors cannot show
- To use recorded CCTV video footage to help the police investigate incidents, and prosecute criminals.

The proposed CCTV system was to be installed by a private CCTV installation company. Mr. Elton was also considering in signing up to a ‘shop watch’ scheme which would enable him and his staff to report incidents such as theft and violence in the shop using a point-to-point radio contact system. The radio system would be linked to a CCTV control room 2-miles away from the shop, where CCTV operators can provide immediate assistance to Mr Elton and his staff by locating perpetrators on street CCTV cameras near to the jewellery store.

Each of the eight phases from TEC-VIS was followed step-by-step, and these are detailed in the remainder of this Section (see Chapter 10: Tables 10.5, 10.6, and 10.7.

Appendix I

TEC-VIS Framework Evaluator Briefing Form

1 Rationale

The TEC-VIS is a framework which was developed by Ms Hina Keval for her PhD research, as part of her main thesis contributions (at the Department of Computer Science, University College London). The framework is to be reviewed by 3 experts who have practical experience and/or have research experience in the following areas of work: Human-Computer Interaction (HCI), usability, security (CCTV, control rooms etc.) The purpose of the review is to provide the author with constructive unbiased feedback as means to validate the research contributions to -define the framework following these evaluations, and to put forward a future program of research needed in CCTV effectiveness in HCI.

2 Goals of TEC-VIS

What is TEC-VIS? It is a framework which should be followed when deployment a new CCTV system, or when planning a re-design of an existing CCTV system. The framework is designed to support CCTV practitioners these include CCTV owners, managers, as well as security consultants and designers. The goal of the framework is to improve the set-up and use of CCTV, to improve the effectiveness and performance of the system. Two sets of guidance are detailed in the framework:

1. High-level guidance:

- Provides a stakeholder framework detailing the design considerations for a CCTV system.
- The framework encourages the system owner to carry out a user-analysis, which requires them to consider the goals, tasks and the individual users (stakeholders) of the system.
- In this framework, the owner is also required to carry out an analysis of the stakeholder requirements, and the technical capabilities, to remove user conflicts and technical barriers which may reduce the effectiveness of the system.

2. Low-level guidance:

- Provides a prescriptive set of recommendations on the best-practice for digital CCTV video storage, transmission, and display requirement for observation tasks.
- Also provides a number of recommendations for configuring the internal and external CCTV camera environment.

3 Evaluator Background Information

Please provide the following information about you and your background.

Name:

Organisation/Institution:

Please state if you have experience in the following fields, and state the month/years:

1. Ergonomics/HCI Y/N Length of Experience:
2. Usability Y/N Length of Experience:
3. Security Y/N Length of Experience:

Please list areas of work you are currently involved in:

Please list your areas of interest and expertise:

4 Evaluation Requirements

Please review the TEC-VIS framework (see attached), by providing your comments and opinions (as a guide 1-2 pages), using the following criteria:

7. **Comprehensive:**
    - 1a: Does the framework consider all CCTV stakeholders? If not, what is missing?
    - 1b: Does the framework consider all the social factors related to a CCTV deployment? If not, what is missing?
    - 1c: Does the framework consider all the technical factors related to a CCTV deployment? (i.e. are all the digital aspects of CCTV considered). If not, what is missing?
8. **Style and language:**
    - 2: How well is the framework understood for all CCTV stakeholders (i.e. is it suitable for all CCTV users at different all levels of experience and knowledge of CCTV)?
9. **Application:**
    - 3a: Do you think the framework is applicable for real CCTV deployments?
    - 3b: Do you think the framework can be applied to all types of CCTV deployments?

Appendix J

TEC-VIS Framework Reviewer Responses

Reviewer 1

Comprehensive

1a. Does the framework consider all CCTV stakeholders? If not, what is missing?

The framework appears to consider all principle stakeholders. However, if the list of stakeholders were to be exhaustive it might also include the media and support staff who work within the CCTV control room. The media might approach a data controller for CCTV images. Support staff who work within the CCTV control room, but who do not work directly with the CCTV equipment (for instance, cleaning staff), might be affected by the set-up of the control room.

1b. Does the framework consider all the social factors related to a CCTV deployment? If not, what is missing?

The emphasis on a stakeholder centred analysis could, potentially, cover all social factors related to a CCTV deployment. The detailed stakeholder analysis is a particular strength of TEC-VIS. Indeed, it is perhaps the most important and novel contribution of TEC-VIS to knowledge and guidance relating to the deployment of CCTV systems. In this respect, TEC-VIS makes a contribution over and above that offered by the latest UK government CCTV Operational Requirements Manual (Cohen, Gattuso, and Maclennan, 2007).

However, it might be of benefit to illustrate the implementation of TEC-VIS with specific social aspects of a CCTV deployment, alongside a consideration of how the technology interacts with these factors. This is achieved to some degree in the example of the application of TEC-VIS, but its importance is somewhat understated. High profile examples could be used in order to highlight the noteworthy, socio-technical approach adopted by TEC-VIS. For instance, consideration might be given to the advantages inherent in facilitating co-operation between CCTV operators in tasks such as tracking targets across multiple CCTV cameras, or to the importance of providing feedback to CCTV operators on their performance (see Gill and Spriggs, 2005; Gill et al., 2005).

1c. Does the framework consider all the technical factors related to a CCTV deployment (i.e., are all the digital aspemcts of CCTV considered)? If not, what is missing?

The framework considers some factors including recorded image quality, but does not consider other technical issues which might be related to such factors. For example, the calculation of the digital storage capacity necessary to accommodate these images is not addressed (cf. Cohen, et al., 2007). Cohen et al. (2007) consider a comprehensive range of technical issues. It may be not be useful to reiterate these issues in detail within the TEC-VIS framework, but there may be value in directing users of TEC-VIS to the work of Cohen et al. (2007) whilst highlighting where TEC-VIS extends this work.

Style and language

2. How well is the framework understood by all CCTV stakeholders (i.e., is it suitable for all CCTV users at all levels of experience and knowledge of CCTV)?

The TEC-VIS framework, whilst mainly clear and free of jargon, does contain some technical language. This may be unavoidable and at points useful explanations are provided for the layman. However, occasional technical acronyms such as “HCI” are included. These may put off the lay reader. Additionally, the use of academic-style citations and references may not be appropriate for practitioner audiences. TEC-VIS is described as offering high-level and low-level guidance. Reconciling these two aims in a single, user-friendly document may not be possible. TEC-VIS could benefit from multiple formulations which provide selected information in a manner tailored to a specific type of reader.

Application

3a. Do you think the framework is applicable for real CCTV deployments?

TEC-VIS appears applicable to real CCTV deployments. The illustrative example provided is very useful in demonstrating the potential for applying TEC-VIS in practice. Future work might consider a case study of a complete application of TEC-VIS.

3b. Do you think the framework can be applied to all types of CCTV deployments?

The framework appears sufficiently generic to facilitate its application in many contexts. However, its generic nature is both a strength and a weakness. TEC-VIS users might benefit from multiple examples covering the most common applications. For instance, an example of the use of TEC-VIS in a local authority CCTV control room would make the framework more accessible to a large sector of CCTV users.

References

Cohen, N., Gattuso, J., and MacLennan-Brown, K. (2007). *CCTV operational requirements manual: Is your CCTV system fit for purpose.* (Publication No. 55/06). Sandridge, UK: Home Office Scientific Development Branch.

Gill, M., and Spriggs, A. (2005). *Assessing the impact of CCTV.* (Home Office Research Study No. 292). London, UK: Home Office Research, Development and Statistics Directorate.

Gill, M., Spriggs, A., Allen, J., Hemming, M., Jessiman, P., Kara, D., *et al*. (2005). *Control rooms: Findings from control room observations.* (Home Office Online Report No.14/05). London, UK: Home Office.

Reviewer 2

General Comments

It was interesting and enjoyable to read this document as it offers a practical guide to applying a user-centred design method to CCTV systems. Many other works just provide general advice or principles to follow.

The document doesn't say who the framework is aimed at. While it is a key document for the CCTV owner, would they be the best person to apply the framework? Maybe it should be the CCTV consultant, provided they have some human factors knowledge. Maybe the framework could give some advice on this.

Comprehensive

1a. Does the framework consider all CCTV stakeholders? If not, what is missing?

The framework seems to be comprehensive and presents a full range of stakeholders identified in Figure 1.1. The user of the document could certainly take this list as a starting point and then add specific information for their own situation. They could for example put the names of the CCTV supplier and maintainer into the framework.

1b. Does the framework consider all the social factors related to a CCTV deployment? If not, what is missing?

Regarding social factors, you might want to consider operator abilities, skills and training. Aspects to include might be: selection of operator staff, eyesight testing, course attendance and on-site testing for owners on managing security staff. Is it also necessary to perform some kind of check into the background of recruits who want to take up jobs as security staff?

1c. Does the framework consider all the technical factors related to a CCTV deployment (i.e., are all the digital aspects of CCTV considered)? If not, what is missing?

Do you need to say something about the Data Protection Act and its implications for CCTV? This would include any limits on data recording or length of time recordings can be stored.

The framework is helpful regarding technical factors and gives clear advice about the requirements for equipment selection. I wasn't sure whether there needs to be a distinction between monitoring only applications, and monitoring and storage applications. Maybe only the latter type of system is generally used.

Some other technical factors you could cover are:

- Whether a colour screen is needed.
- Are LCD screens as effective as traditional CRTs
- What amount of storage is needed for data recording
- The use of dedicated security hardware versus a general PC with security cameras and equipment added

Style and language

2. How well is the framework understood for all CCTV stakeholders (i.e. is it suitable for all CCTV users at different all levels of experience and knowledge of CCTV)?

The style and language seemed fine. The guide is also concise which is helpful if it is to be used in practice.

Application

3a. Do you think the framework is applicable for real CCTV deployments?

3b. Do you think the framework can be applied to all types of CCTV deployments?

When it comes to performing a stakeholder analysis and holding a regular stakeholder meeting, you might indicate what is practical for say the smaller CCTV owner (say for a single shop) as opposed to the larger scale operation. Maybe you could indicate an example group of individuals who could be invited to the stakeholder meeting.

In order to cover some of the higher level stakeholders who are unlikely to be involved directly such as the Home Office or forensic experts, perhaps the needs of these groups could be included (briefly) in the framework as a reference for the CCTV owner when specifying their system. (However you do this within the example scenario.)

I am sure that the framework would be of interest to anyone deploying CCTV. It is perhaps most appropriate for larger scale operations where human factors expertise is available. For smaller scale operators they would perhaps need advice on who should apply the framework on their behalf.

<u>Reviewer 3</u>

General Comments

The document you provided is work which is valuable. I see plenty of growth in the field you are addressing and the need for advice to end users will be extremely important, unfortunately most of the suppliers also need educating and what you have provided will be a great start for that. Well done and keep up the good work.

Comprehensive

1a. Does the framework consider all CCTV stakeholders? If not, what is missing?

The framework is missing a significant stakeholder the IT Manager/Professional. As cameras and CCTV is becoming more network orientated and data is often web enabled the framework should consider many elements associated with the IT world from security access to data integrity.

1b. Does the framework consider all the social factors related to a CCTV deployment? If not, what is missing?

1c. Does the framework consider all the technical factors related to a CCTV deployment (*i.e.,* are all the digital aspects of CCTV considered)? If not, what is missing?

The framework does not consider the CCTV environment in enough detail. ISO 11064 is a cop out; while it does consider console and screen ergonomic it does not provide enough guidance for dealing with common problems associated with multiple and large screen displays used in video surveillance. Consider a change in the tradition use of multiple monitors in the Monitor Guidance Section. We have found a hierarchical HCI using a large screen display as an overview of all cameras patched together then having four desktop monitors providing area view (a sub Section of the total overview) a detail view for spot/incident and then workstations for tracking maps and one for facial recognition software and zoom capability. This would require the framework provide some guidance on the use of LSD or video walls. The example is a good example but it falls short of expectations; in phase 3 it does not represent an HTA.

Phase 4 - video storage should also discuss technologies from tape to video digital recorders to video streaming to a server and the advantages and disadvantages of the technologies. Perhaps it would help to identify who should provide information for Phases 5 and 6.

The framework should also consider automated tools such as license plate recognition; the Section that references tools is too vague.

Style and language

2. How well is the framework understood by all CCTV stakeholders (*i.e.,* is it suitable for all CCTV users at all levels of experience and knowledge of CCTV)?

I think the framework will be outside the understanding level of the stakeholders identified, it is suitable for a large company security manager, but a store owner would find many of the concepts foreign. The CCTV professionals will understand this and may be able to use it as a guideline for new customer projects, but a new owner would find it difficult in its current format. I think you should decide who the user of the document is and write specifically for them. It would be difficult for other stakeholders to comprehend.

Application

3a. Do you think the framework is applicable for real CCTV deployments?

I think the framework is very useful and applicable to larger security projects, and a more prescriptive framework would be required for lower-end stakeholders.

3b. Do you think the framework can be applied to all types of CCTV deployments?

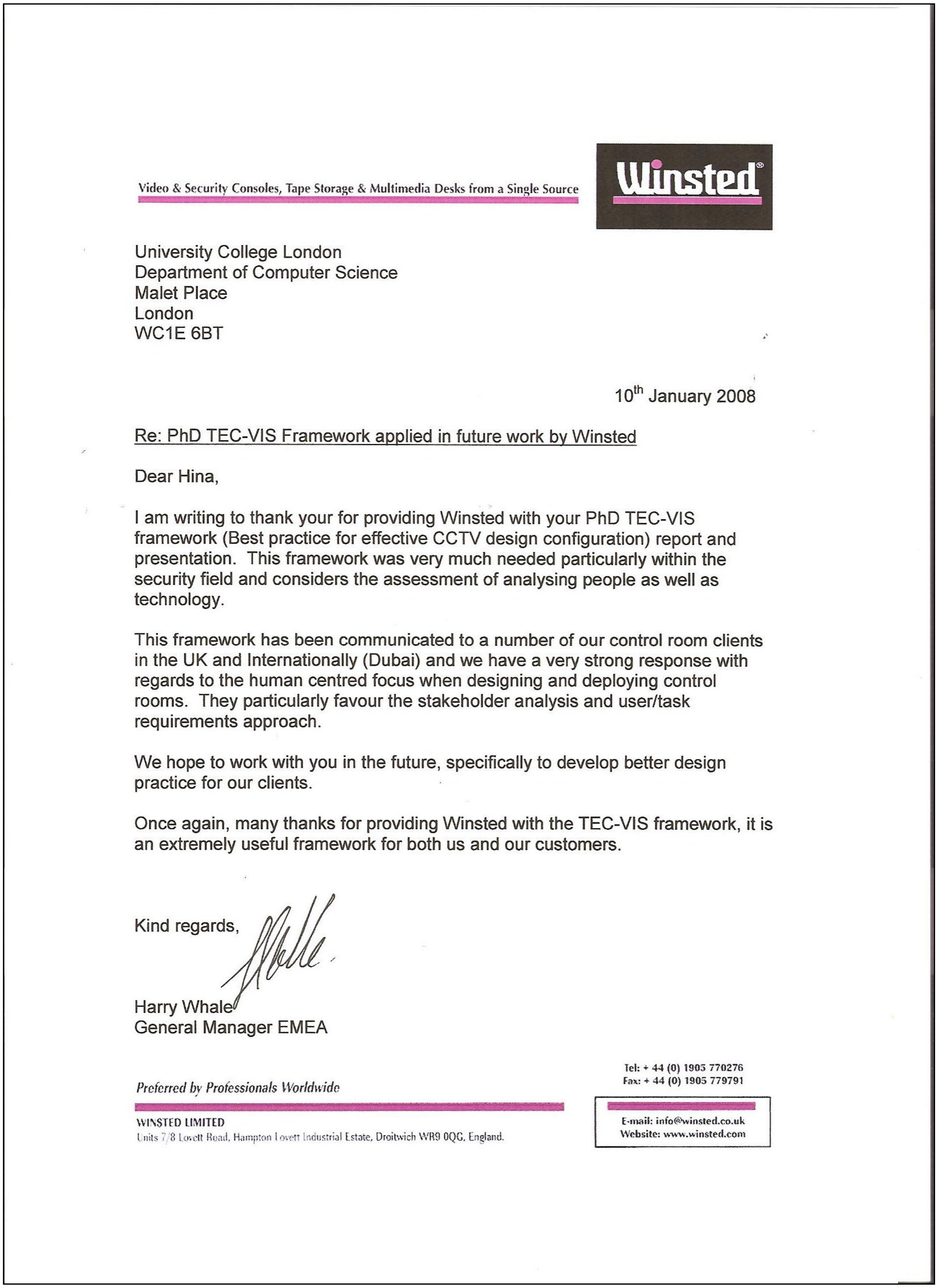
If the framework is intended to provide a design that considers what tasks are performed by operators, the Phase 3 Section and examples are too weak to demonstrate improvement in this area. A multiple system monitoring company that gets an alert of a real-time problem has several challenges from orientation, evaluation to action that are not identified in the framework.

While the framework is trying to stay away from being just a project guide it has some elements of a project plan but does not have project cost/benefit which needs to input into Phase 7.

I think the framework is very useful but needs to be put into context for each of the stakeholders and a generic document would have to be more descriptive and have clearer definitions as to who does what.

Appendix K

Letter from Winsted Ltd In Support of TEC-VIS Framework



Video & Security Consoles, Tape Storage & Multimedia Desks from a Single Source

Winsted®

University College London
Department of Computer Science
Malet Place
London
WC1E 6BT

10th January 2008

Re: PhD TEC-VIS Framework applied in future work by Winsted

Dear Hina,

I am writing to thank your for providing Winsted with your PhD TEC-VIS framework (Best practice for effective CCTV design configuration) report and presentation. This framework was very much needed particularly within the security field and considers the assessment of analysing people as well as technology.

This framework has been communicated to a number of our control room clients in the UK and Internationally (Dubai) and we have a very strong response with regards to the human centred focus when designing and deploying control rooms. They particularly favour the stakeholder analysis and user/task requirements approach.

We hope to work with you in the future, specifically to develop better design practice for our clients.

Once again, many thanks for providing Winsted with the TEC-VIS framework, it is an extremely useful framework for both us and our customers.

Kind regards,

Harry Whale
General Manager EMEA

Preferred by Professionals Worldwide

WINSTED LIMITED
Units 7/8 Lovett Road, Hampton Lovett Industrial Estate, Droitwich WR9 0QG, England.

Tel: + 44 (0) 1905 770276
Fax: + 44 (0) 1905 779791

E-mail: info@winsted.co.uk
Website: www.winsted.com