Evaluating Human Interaction with Advanced Air Traffic Management Automation

Brian G. Hilburn, Ph.D.
National Aerospace Laboratory NLR
Anthony Fokkerweg 2
1059 CM Amsterdam
THE NETHERLANDS
+31 20 511 3642 (tel); 3210 (fax)

hilburn@nlr.nl

SUMMARY

The global Air Traffic Management (ATM) system represents one of the most complex human-machine systems known. As the system has evolved in step with traffic demands, automation has been increasingly called upon to augment the capacities of the human operator, namely the air traffic controller. Future generations of ATM automation seek to augment, or even replace, such human faculties as decision making and strategic control. As such, they represent some of the most ambitious attempts at automation conceivable.

Against this backdrop, the NLR is actively involved in the development and evaluation of new forms of ATM automation, both in the cockpit and on the ground. This effort centres on:

- *Exploring new concepts for advanced automation, such as adaptive automation, in which a system is capable of dynamically adjusting its level of assistance on the basis of some measured or otherwise inferred need for task assistance; and*

- *Refining methods for evaluating human interaction with advanced automation. One of the chief techniques used in this regard has been realtime human-in-the-loop simulations, used in conjunction with objective psychophysiological measures of human performance.*

This paper summarises results of two studies conducted by the NLR. The first study examined air traffic controllers' attitudes toward possible new forms of automation. In the second study, a series of human-in-the-loop simulations was conducted to evaluate the potential benefits of advanced ATM automation on human-machine system performance. Evaluations relied on a set of objective (including psychophysiological) measures of human performance.

The common thread linking the two studies is the underlying notion of real time strategic decision aiding automation. That is, the studies were based on two different developmental systems, each of which aims to both detect and resolve air traffic conflicts in real time. The paper concludes with a discussion on lessons learnt, both as they relate to automation design per se, but also to the techniques used to evaluate human performance with complex automation.

AUTOMATION IN ATM

Widespread use of ATM automation first appeared in the 1960s, to assist with flight data processing. Around that time, computers began to assume responsibility for processing flight data, producing flight data strips (paper forms used by controllers to track and notate the progress of each flight), and checking filed flight plans for obvious errors (e.g., aircraft type, route information). Until very recently, automation was largely confined to this role, as a semi-skilled adjunct to the human controller, assigned only the relatively routine and mundane aspects of the controller's job. The days of such a relationship, however, now appear to be numbered (Hopkin, 1994).

Given the enormous projected growth in air traffic, both industry and civil aviation authorities have been forced to acknowledge the inevitable trend that ATM of the future will follow: a system comprised of more aircraft, shorter permissible delays, closer spacing, and less controller time spent per aircraft (ICAO, 1993). In addition to there being more aircraft in the sky, such advances as digital data link will also make more information available about each aircraft. This abundance of information will certainly exceed the capacity of current systems. As a result, a number of specific ATM automation efforts are under development or under consideration world-wide. Writing nearly 20 years ago, Gosling and Hockaday (1984) identified seven alternative strategies for ATM control, to help them evaluate the potential benefits of automation in ATM. The framework shown in the figure below, loosely based on that of Gosling and Hockaday, is a useful way to describe the current evolution of ATM automation. At its lowest level, ATM relies on the *Visual Flight Rules* (VFR), or "see-and-avoid" concept that underlies most general aviation. The final step along the current evolutionary path of ATM represents a fully deterministic, or fully automated, system. At the moment, ATM has evolved to the point that it falls roughly between the levels of Automation II and Automation III.
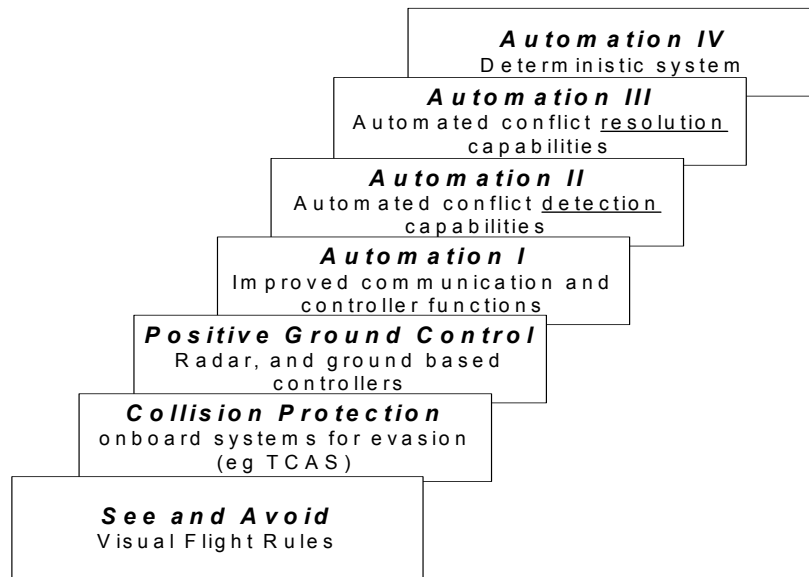
A Taxonomy of ATC Automation Evolution.

The broad topic of ATM decision aiding encompasses a number of goals, concepts, and time frames. The term can refer, for example, to such tactical aids as short term conflict alerting systems, which are currently in service and which provide controllers 2-3 minute look-ahead capability (Pélegrin, 1994). Strategic aids, on the

other hand, might provide as much as 30-45 minutes advance warning of predicted separation or scheduling irregularities. One goal of many current ATM automation development efforts is to incorporate "look ahead" capability on a strategic time frame (Meckliff & Gibbs, 1993). It is thought that this capability, if controllers were free to scroll ahead in fast time and observe the predicted outcomes of prospective control inputs, would greatly enhance the controller's ability to make accurate predictions, and assess the impact of prospective control inputs without the problems of real time (and real danger). This could also prove extremely useful for training purposes.

A major shift is afoot in ATM automation away from tactical monitoring aids, toward strategic planning functions. With the advent of automated planning, conflict prediction and problem solving tools, the hope is that much of the burden for strategic planning can be overseen by automation. In theory, such a system could not only obviate the need for the current planning controller (who makes up half of the typical planner-tactical controller team), but could also reduce the need for tactical control. Sound strategic planning, so the thinking goes, should minimize the need for tactical interventions.

Automation and Human Factors Concerns

The raison d'être of automation has traditionally been the reduction of operator workload and human error. Nonetheless, there is a wealth of anecdotal, empirical and theoretical evidence that improperly implemented automation invites many human performance costs. By effectively removing the human from the control loop, and relegating him/her to the role of passive monitor, automation can hinder the operator's ability to maintain an adequate mental representation of the system, as well as inhibit his/her ability to quickly reassume manual system control in case of emergency. Further, automation whose input/output relationships are not apparent to the operator, or whose actions are not consistent, increases the risk of human error. An in-depth discussion of the potential human factors concerns associated with automation is clearly beyond the scope of this paper. However, some of the (interrelated) problems that have been associated with the injudicious use of automation include the following:

- Reduced situation awareness;

- Increased monitoring demands;

- Mis-calibration of trust in automation (either excessive trust, termed "complacency," or, at the other extreme, mistrust of automation);

- Inability to reassume manual control;

- Degraded manual skills through lack of practice;

- The need for new selection and training procedures;

- Increased inter-operator coordination requirements;

- Increased workload management requirements, and

- Loss of motivation and job satisfaction.

THE NEED FOR USER ACCEPTANCE: EVIDENCE FROM THE CONTROLLER RESOLUTION ASSISTANT (CORA) PROJECT

EUROCONTROL's ongoing Conflict Resolution Assistant (CORA) project aims to define and develop operational requirements and prototypes for conflict resolution concepts, based around the introduction of

such computer-based ATM tools as Trajectory Planning, and Medium Term Conflict Detection (Eurocontrol, 2002). CORA is following a phased development, with CORA Phase 1 tools aimed at identifying conflicts or problems (and the aircraft involved), but stopping short of implementing a solution. Under CORA1, responsibility for devising and implementing a solution rests squarely with the human controller. CORA2, on the other hand, envisions a higher level of controller support with respect to conflict resolution and decision aiding. In short, CORA2 aims to provide resolution advisory assistance to the controller, either automatically or on-request, to help resolve any detected conflicts.

One of the greatest challenges currently facing the development and implementation of advanced ATM automation (such as conflict resolution tools) is to foster acceptance among those who must ultimately come to use the tools, namely controllers. It is an irony of advanced automation, that is increasingly capable of assuming control of higher (e.g. strategic) decision making functions, that the user must come to <u>trust</u> the system or tool, if it is to demonstrate any benefit. Before he is willing to trust a tool, however, the controller must be willing to <u>use</u> the tool. Given the nature of ATM (with its highly skilled performance routines) controllers can devise elaborate means of circumventing the use of a new tool, if they do not perceive its potential benefits.

Given the nature of the ATM domain (in which technical and operational advances are typically made in an evolutionary and incremental way), it is reasonable for controller acceptance to have been identified as a key factor in how fully the controller community will embrace advanced new forms of ATM automation. By identifying at an early stage the prevalent automation-related attitudes among a representative sample of the European ATM community (both controllers and managers), such issues can be better addressed through controller/manager training and information.

Controller Attitudes in General

It is a widely-held belief, both in ATM operations and among the research community, that air traffic controllers (ATCos) are reluctant to change, and that any attempts to introduce new forms of ATM automation will be met with great resistance from the control room floor. Various survey results over the years have portrayed controller attitudes as very positive, at least toward the overall ATM job (Kennholt & Bergstedt, 1971; Rajecki, 1990; Air Traffic Management, 1999).

The earliest systematic survey of controllers' automation attitudes (as opposed to those regarding the ATM job in general) appears to have been done at the University of Aston in Birmingham (UK) during the 1970s and 1980s. In two reports (Crawley, Spurgeon and Whitfield, 1980; Crawley, 1982), the group outlined the methods and results from a decade of research with British civil ATCos into attitudes toward (current and future) ATM automation. Perhaps not surprisingly, the researchers found that controllers tended to rate the less routine aspects of their job (e.g. identifying conflicts, deciding on solutions, and prioritising possible actions) more positively than the more mundane ones (e.g. preparing flight strips). Controllers, it seemed, preferred tasks that posed greater challenge, especially with respect to traffic handling skills.

Nonetheless, survey and anecdotal evidence has suggested aspects of the job with which controllers are less satisfied. Hopkin (1995) portrayed the current day controller as someone who, although very happy with the ATM job itself, expresses reservations about management, equipment, the media (which is seen as presenting an overly negative image of ATM), and some of their conditions of employment. Negative feelings toward ATM management, Hopkin noted, are heavily influenced by controllers' perception of management decisions as often ignorant of actual ATM tasks, particularly as such decisions relate to funding and equipment. Hopkin (1995) also observed that controllers' shared attitudes play an important role in group

culture, by helping to maintain professional norms and standards. Air Traffic Management magazine (1999) found that, while controllers did not report negative attitudes toward automation generally, they did appear reluctant to relinquish responsibility for decision making or control.

More recently, Air Traffic Management magazine (1999) concluded on the basis of surveys across Europe and North America that controllers tended to fear that future ATM automation will lead to greater monitoring demands, and greater risk of mistakes. Although most controllers denied mistrusting automation in general, controllers clearly wanted to remain in the position of making decisions. Controllers appeared accepting of an advisory automation (i.e. a system that presents alternative solutions), given that the authority of such automation stops short of actual implementation.

Method

As a first step in the current controller assessment project, a literature review was conducted, to survey past theoretical and empirical work on controller attitudes, and to establish what is currently known in this area. Literature were gathered on the subjects of controller attitudes toward advanced automation, controller attitudes generally (e.g. with respect to their job in general), and from other relevant domains. The resulting literature review formed the basis for the research hypotheses subsequently investigated through focus groups and surveys.

Site visits were conducted during the summer of 2000 to seven ATM centres across Europe. Together, these seven captured a reasonable cross-section of current European ATM operations, in terms of systems, work cultures, and traffic patterns. Of particular concern were attitudes related to future automation needs, system development issues and operational requirements. This assessment focused on prevalent attitudes toward automation, as they relate to perceptions of, for example: job security, safety, human performance issues, and management role changes. At each site, data collection centred on focus groups (FGs) with air traffic controllers (Hilburn & Flynn, 2001), as well as written surveys of both managers and controllers.

Results

Excerpted transcripts were supplemented with field notes, as well as summary comments from debriefings. Additional analysis of the focus group transcripts relied on content analysis – that is, computerised classification of textual material into relevant information (Weber, 1990). Computerised text analysis software was used to identify key words in context. Some of the major themes to emerge from the focus group sessions include the following[1]:

• Need to demonstrate benefits of new automation. Controllers reported that they would hold CORA-like automation to a high standard of reliability, despite the fact that they are currently relying on systems (e.g. STCA) and human colleagues of less than 100% reliability

• Automation-related human performance concerns

• Potential for automation to change role of Planner and Executive controllers

• Retraining for new skills and tools

• Job security / satisfaction concerns

• Need to involve controllers in system developments

---

[1] For a more complete overview of the methods and results of this study, refer to Hilburn and Flynn (2001).

One of the most notable themes to emerge from focus groups was the observation, phrased in various ways, that although controllers place high value in system reliability, they will accept systems of lower reliability under two conditions: First, the system in question is prone to false alarms rather than misses; Second, it must be easy to verify the (improper) functioning of such a system. The case of Short Term Conflict Alert (STCA) was often cited as an example in this regard. The purpose of STCA is to detect impending losses of separation between aircraft, and to display on screen for the controller a visual alert (usually via a change of data label colour). It is well-known that the algorithms underlying STCA vary in their sophistication. Some, for instance, fail to consider the planned or cleared route changes an aircraft is to make. Instead, they perform simple straight-line extrapolation of flight path. This can obviously result in a false alarm if an aircraft intends to level off directly under another.

It seems intuitive that, in most complex / time critical systems, automation misses (e.g. an STCA that failed to detect losses of separation) would be entirely unacceptable. The suggestion (consistent with some data from other domains) is however that false alarms may be tolerable, as long as the costs of verifying malfunction are not too high. In other words, controllers seem willing to accept a tool that gives false alerts, as long as they can at a glance see that the system (in this case, the air traffic pattern) is functioning properly.

Further, some controllers made a distinction between lower level tools such s STCA, and those such as CORA, which would be expected to assume control for higher level functions. Automation tasked with higher level functions, by definition, would be harder (at least, more time consuming) to check. The standard for reliability would therefore be much higher.

DECISION AIDING AUTOMATION IN ATM: RESULTS FROM CENTER TRACON AUTOMATION SYSTEM (CTAS) HUMAN-IN-THE-LOOP SIMULATIONS

Similarly to the CORA system, the system under study in this series of experiments aims to help the controller by providing resolution assistance. The chief goal of the current study was to establish whether such strategic decision aiding has the potential to benefit various aspects of human performance (primary among these is mental workload). As some have pointed out, this is not a foregone conclusion. It might be, for instance, that the human operator, if he / she is to remain in the control loop, must continuously compare the system's solution to their self-derived one. Carrying out this comparison might actually increase mental workload. Further, providing additional information beyond what the controller "needs" to derive a workable solution might invite overload. Kirlik (1993) demonstrated that an automated aid can (and should) go unused, if the costs of initiating the aid, considering its advice, generating one's own solution, and comparing solutions becomes too high.

The Center TRACON Automation System (CTAS) is a collection of automation tools designed to assist air traffic controllers in sequencing aircraft, maintaining arrival accuracy, and avoiding separation violations. The *Descent Advisor* (DA) component of CTAS, which was the focus of the current study, is a tool designed to help the en route controller meet scheduled arrival times (STAs), by generating a conflict-free continuous descent approach trajectory. Simulations were carried out with a number of goals in mind:

- To determine how and whether controllers used various levels of automation;

- To determine the workload (and other human performance effects) of various levels of decision aiding automation; and

- To demonstrate the effectiveness of various objective and subjective human performance metrics.

Method

In cooperation with NASA, NLR incorporated a research version of NASA's CTAS (including DA) into its NARSIM ATM simulator. Software modifications for this study provided three levels of experimenter-selectable automation assistance, defined as:

- *Traffic status* – Baseline configuration, in which the controller was presented only the traffic information (e.g., aircraft plots along with heading, speed, and aircraft type).

- *Conflict detection* – Additionally, the controller was alerted to both scheduling problems (discrepancies between STA and ETA) and planning conflicts (i.e., future separation conflicts), as calculated by the CTAS detection algorithm.

- *Conflict Resolution* – Various advisories (e.g. vectoring, speed and top-of-descent) were presented to the controller, in response to any detected conflicts. These advisories, if accepted, would resolve the detected conflict(s). The controller was free to accept or reject these advisories.

Dependent measures for this study included a variety of system and controller performance metrics, as well as measures of mental workload and visual scan pattern. The collected data included: Subjective workload assessments using two different instruments; heart rate variability (HRV) measures of controller mental workload; Controller acceptance survey responses; traffic awareness queries about the current and extrapolated future traffic pattern (Endsley & Rogers, 1994); eye gaze pattern, along with pupil diameter and fixation latencies. Mental workload was primarily assessed in terms of HRV, pupil diameter, and scanning randomness, or *entropy* (Harris, Glover, & Spady (1986)), as well as via the subjective NASA TLX instrument (Hart and Staveland, 1988).

Results

Since the primary intent of the following is to provide an overview of the techniques used to evaluate human performance, this section shall focus on only a subset of the data analysis from this series of experiments – namely workload results[2].

Several methods were used to evaluate workload: physiological (e.g. eye tracking and heart rate related measures) and subjective (TLX workload ratings). These will be discussed together. The following four figures show the HRV, pupil diameter, scanning randomness (entropy), and TLX results, respectively. Notice that all measures correlate positively with indicated workload, except for HRV. That is, HRV tends to decrease with increases in indicated workload. As a result, higher bars in the HRV data graph indicate <u>lower</u> indicated workload. HRV was assessed in terms of the 0.1 Hz component. HRV values were transformed into Z scores, by standardizing within each subject. What is noticeable from these data is the agreement among them on the influence of both traffic load (which was varied from low to high within subjects), as well as the influence of automation level. According to these results, automation benefited workload. This was in particular true for the highest "resolution" level of workload. Subjective workload, on the other hand, told an entirely different story. An automation by traffic load trend showed that subjective workload under high traffic conditions increased with automation condition, whereas under low traffic load, automation was associated with decreased workload.
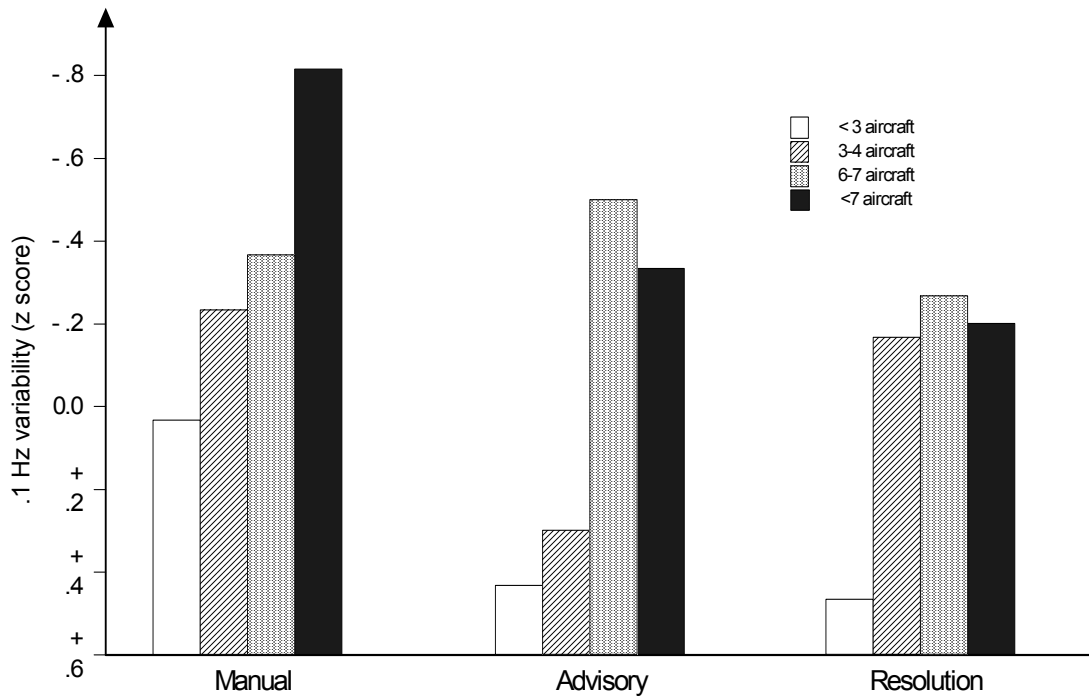
---

[2] For a more complete overview of the methods, and results, of this series of experiments, see Hilburn (1996).
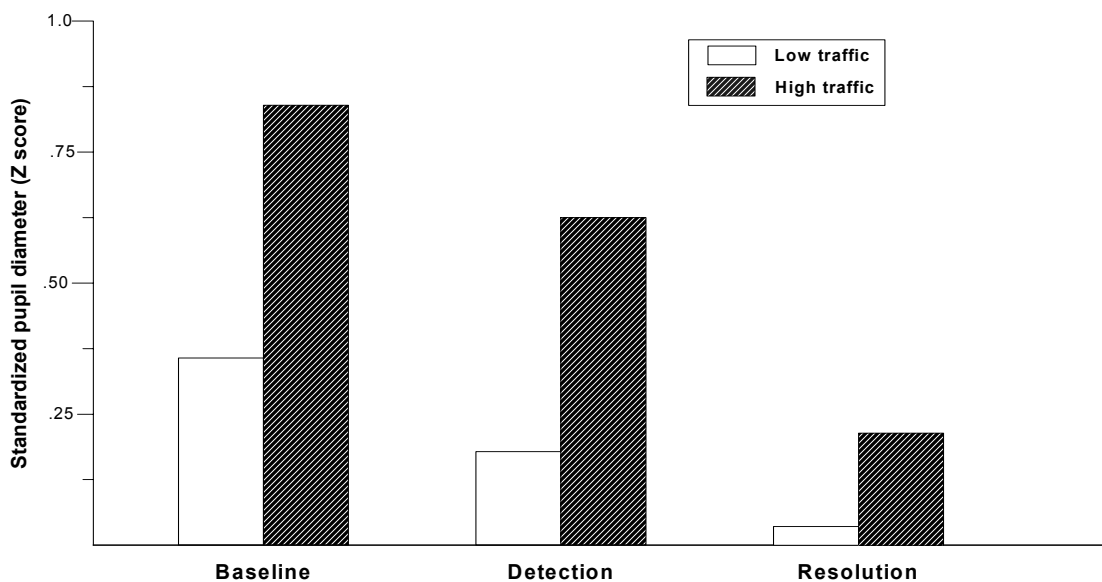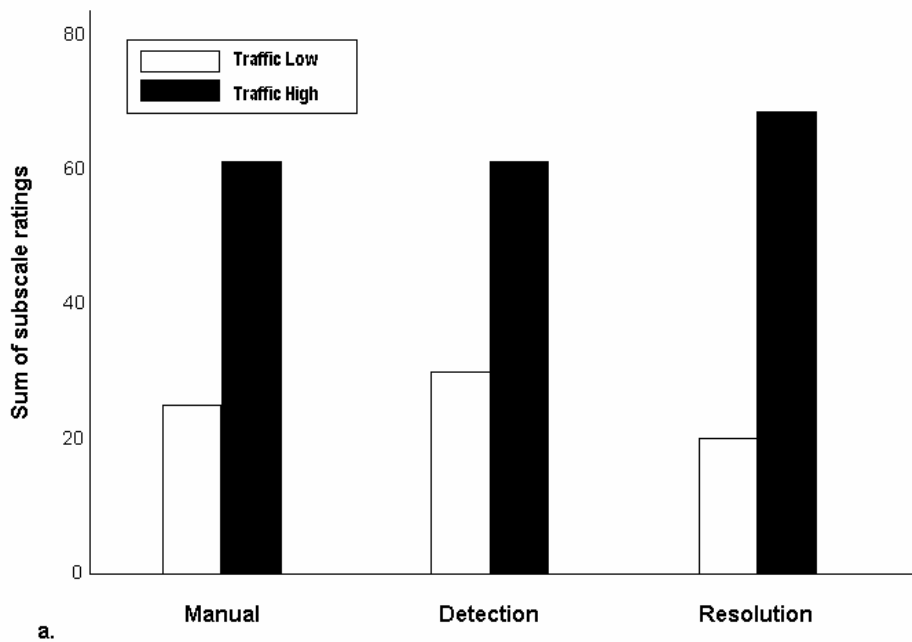
HRV (.1 Hz) as a Function of Automation Level and Workload.
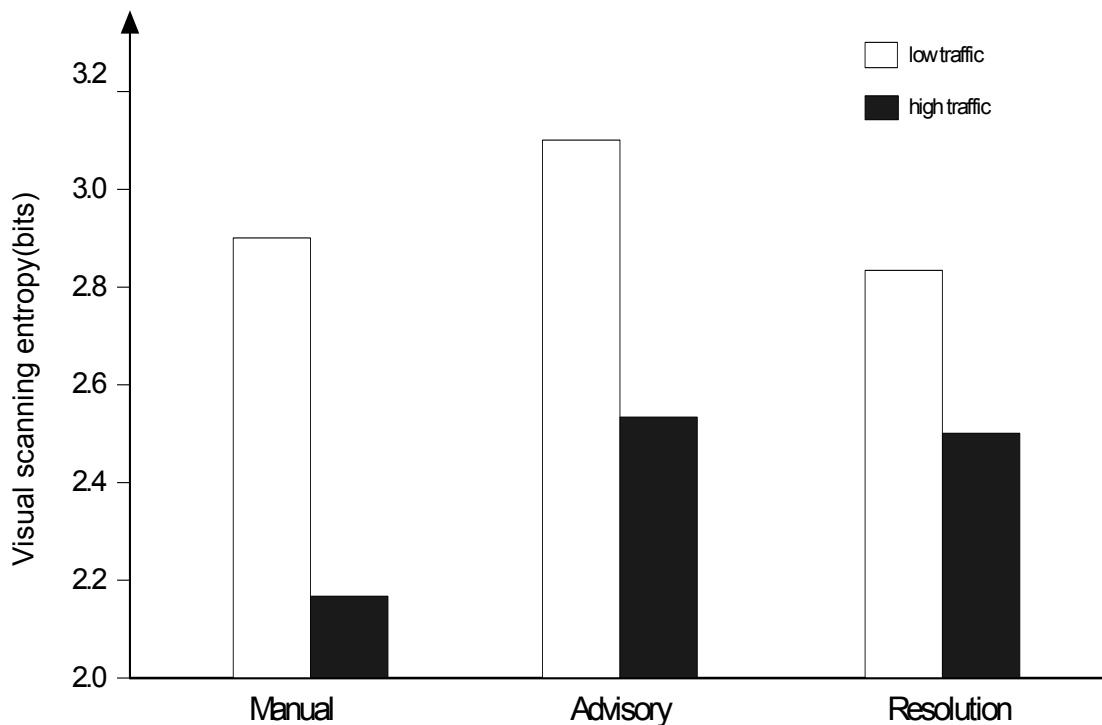
Pupil Diameter as a Function of Automation Level and Workload.

Scanning Entropy as a Function of Automation Level and Workload.

TLX Subjective Ratings as a Function of Automation Level and Workload.

DISCUSSION

The summaries of the two presented studies were intentionally abbreviated in this paper. Again, the interested reader is referred back to the sources cited. The preceding hopefully, however, made two clear points. First, the first study raised a number of potential issues in the design and development of strategic decision aiding for ATM tasks. As mentioned, user acceptance is heavily dependent not only on perceived reliability of the new system, but also on the nature of the reliability (e.g. is the system prone to false alarms? Misses?), and on the costs involved with verifying an automated system's functioning, as compared to the relative costs of for instance a miss/false alarm. Secondly, results of the second study underscore the dissociation that can often occur between subjective and objective measures, when studying human interaction with automated systems. Despite clear evidence from all physiological measures that the automated DA system benefited workload, subjective workload ratings were exactly opposite. Whether this dissociation is evidence of subjective bias, or that the two classes of metrics tap entirely different cognitive processes, is not clear. Nonetheless, it underscores the need for designers of future systems to bear in mind the potential for subjective experience to limit the initial acceptance of new technologies. If users are to rely on optional automated systems (such as "intelligent advisors"), they must perceive that the system provides benefits. Ironically, it is often not until they use such systems that they come to recognise the benefits.

On the basis of many years experience into evaluation of human-machine system interaction (primarily through human-in-the-loop simulations), a number of clear lessons can be drawn about how humans interact with advanced forms of automation. These lessons fall into two main categories, those concerning human-automation interaction per se, and those relating to the methods one can use to evaluate such interaction. Following are some of the major points.

The need for operator acceptance of new automation – more than ever before, automated systems are being introduced that are capable of taking over some of the highest level human functions (i.e. cognitive functions). One common trend across various automation projects is the notion of an "assistant" role for automation. As with a human assistant, however, an operator is free to ignore such a system if he/she does not recognise the benefits of its use. Ironically, such operators will only recognise the potential benefits of they use the system. It is clear to see, from projects such as CORA, for instance, how initial user acceptance has the potential to influence the introduction of new automation.

Training needs – are critical to ensure both technical skills and, at least as importantly, positive attitudes that foster acceptance, as just mentioned;

Familiarisation and trust – recognition of potential benefits, and operator acceptance, rely in large part on the operator's becoming familiar with the implicit "mental model" underlying a new system's algorithms. Knowing what information the system is using (and not using), and how it can be expected to behave in various situations, is critical for the development of operator trust in an automated "partner."

Differences in operator strategy – as with some other domains, the fundamental task of air traffic control presents a large "solution space" – as long as aircraft are moved through a volume of airspace in a safe and expeditious manner (i.e., pass quickly and do not contact one another), various control strategies can be accommodated. This has implications for systems that seek to provide strategic advice: Such advice should ideally not only fit with the task, but with the operator's preferred way of operating.

The need for comprehensive measurement techniques – given the changing nature of nature (again, increasingly tending toward assistant systems, and strategic decision aiding systems), it is essential to use

measurement techniques that can tap not only objective benefits (which are instructive in system design), but also subjective acceptance (which is increasingly essential for "optional-use" automation). Dissociations, when they occur, can be instructive – for instance, low acceptance even in the face of objectively demonstrable benefits can point toward the need for attitude training, to help foster user initial acceptance.

GLOSSARY

ATC	Air Traffic Control
ATCo	Air Traffic Controller
ATM	Air Traffic Management
CORA	Controller Resolution Assistant
CTAS	Center TRACON Automation System
DA	Descent Advisor
ETA	Estimated Time of Arrival
FG	Focus Group
HRV	Heart Rate Variability
NARSIM	NLR ATC Research Simulator
NLR	National Aerospace Laboratory of the Netherlands
RADAR	Radio Detection and Ranging
STA	Scheduled Time of Arrival
STCA	Short Term Conflict Alert
TRACON	Terminal RADAR Approach Control
VFR	Visual Flight Rules

REFERENCES

Air Traffic Management (1999). Attitude Adjustment: Controller Survey. ATM Special Insert: ATC Issues from the Controller's Perspective, March-April, 5-9.

Crawley, R. (1982). Predicting Air Traffic Controller Reaction to Computer Assistance: A Follow-up Study. AP Report 105. Birmingham, UK: Applied Psychology Department, University of Aston in Birmingham.

Endsley, M.R. & Rodgers, M.D. (1994). Situation Awareness Requirements for En Route Air Traffic Control. In *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting*. Nashville Tennessee, October 24-28.

Eurocontrol (2002). User Requirements Document for the CORA-2 Resolver. Eurocontrol document ASA.01.CORA.2.DEL07.URD. Bretigny, France: Eurocontrol Experimental Centre.

Gosling, G.D. & Hockaday, S.L.M. (1984). *Identification and Assessment of Artificial Intelligence Techniques in Air Traffic Control* (Research Report UCB-ITS-RR-84-14). Institute for Transportation Studies, University of California: Berkeley, California.

Harris, R.L., Glover, B.L., & Spady, A.A. (1986*). Analytic Techniques of Pilot Scanning Behavior and Their Application* (NASA Technical Paper 2525).

Hart, S.G. & Stavelend, L.E. (1988). Development of the NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P.A. Hancock & N. Meshkati (Eds.) *Human Mental Workload*. Holland: Elsevier.

Hilburn, B. & Flynn, M. (2001). An Evaluation of Air Traffic Controller Attitudes Toward Automation. In Proceedings of the Eleventh International Symposium on Aviation Psychology, March 5-9, Columbus, Ohio, USA.

Hopkin, V.D. (1995). Human Factors in Air Traffic Control. London: Taylor and Francis.

Hopkin, V.D. (1994). Human Performance Implications of Air Traffic Control Automation. In M. Mouloua & R. Parasuraman (Eds.) *Human Performance in Automated Systems: Current Research and Trends*. Hillsdale, New Jersey, USA: Erlbaum.

ICAO (1993). Increased ATC Automation may be Inevitable to Handle Increasing Traffic and Data. *ICAO Journal*, June 1993, pp. 16-20. Montreal, Quebec, Canada: International Civil Aviation Organization.

Kennholt, I. & Bergstedt, M. (1971). Attitudes Toward the Work and Working Conditions Among Air Traffic Control Personnel in the Aviation Administration. Stockholm: Swedish Personnel Administration Council.

Kirlik, A. (1993). Modeling Strategic Behavior in Human-Automation Interaction: Why an Aid Can and Should Go Unused. *Human Factors*, *35*, 221-242.

Meckliff, C. and Gibbs, P. (1993). PATs Problem Solver. In *PHARE Forum: Scientific Seminar of the Institute for Flight Guidance of DLR*. 6-8 October, 1993. Brussels: Eurocontrol.

Pélegrin, M. (1994). Electronic Aids to Controllers. In A. Benoît (Ed*.) On-Line Handling of Air Traffic*. AGARD document AG-321. Brussels: NATO.

Rajecki, D.W. (1990). Attitudes. Oxford: W.H. Freeman.

Weber, R.P. (1990). Basic Content Analysis (2nd edition). London: Sage.