

Evaluating safety and usability of ATM systems

Extended Abstract for ATM R&D Seminar2001

Patrizia Marti, Paola Lanzi, Francesco Pucci

University of Siena
Department of Communication Science
Via dei Termini 6
I-53100 Siena, Italy

[marti/lanzi/pucci@media.unisi.it]

Summary

In the paper we present methodologies and results of validation carried out within the ITI project, an innovative user interface for the En-Route and Approach Controller Working Positions for the new Ciampino ACC. The project developed as a co-operation between the Italian National Administration ENAV S.p.A. (Ente Nazionale Assistenza al Volo) and the Eurocontrol Experimental Centre (Bretigny, France). The evaluation was articulated along two dimensions, usability and safety, using a set of different methodologies. The assessment was based on the idea that critical situations are not due only to the availability of a certain information during the task execution but to the way in which different components in the process (software applications, organisational and cultural aspects, the physical layout, human operators) are balanced and interact to avoid or provoke breakdowns in the activity. The validation methodology we adopted allows to pro-actively assess which aspects of the system may impair or enhance safety after the introduction of new artefacts in the work setting.

Introduction

Evaluation of safety critical systems is a composite and articulated activity. Thanks to the availability of advanced technological tools, operators can demand routine tasks to the system and to concentrate on higher level mental operations. Therefore the activity of these operators evolves towards a flexible and context dependent process, where the knowledge that is daily produced is used to face new incoming situations. Indeed critical situations are not due only to the availability of a certain information in the execution of a procedure but to the way in which different components in the process (software applications, organisational and cultural aspects, the physical layout, human operators) are balanced and interact to avoid or provoke breakdowns in the activity. For this reason, the evaluation of complex safety critical systems requires an in depth analysis of the socio-technical context of the work, in order to assess the role that each component plays in the process. In this paper we present the experience we made evaluating an innovative user interface for the En-Route and Approach Controller Working Positions for the new

Ciampino ACC. In the project we adopted a multidimensional approach to the evaluation of safety and usability. More in detail we evaluated: 1) *technical usability*, that is the perceptual and physical aspects of the human computer interface such as display formatting as well as anthropometric characteristics of the object being worked with; 2) *domain suitability*, that refers to the appropriateness of the content of information and display representations; 3) *user-acceptability*, that is the ease of use and suitability of the system for supporting cognitive task requirements; 4) *safety* where we have adopted an approach that systematically assesses the components at stake in a process (software applications, organisational and cultural aspects, physical layout, human operators) and their interactions.

Usability aspects of the interface were evaluated applying user-centred methodologies, including heuristic evaluation, cognitive walkthrough and scenario-based evaluation. The reasons to use different methodologies is to address all the different dimensions of usability. Indeed heuristic evaluation covers issues related to the effectiveness and efficiency and can be used to guide a design decision or to critique a decision that has already been made. Cognitive walkthrough covers issues related to the effectiveness of the system, highlighting problems of action executions and feedback interpretation with respect to a specific goal. Scenarios were used both to evaluate usability and assess safety. Indeed what all of the recent literature on human error and new technology tells us is that human action, error and safety are inherently contextualised phenomena and cannot be analysed, evaluated or predicted without a model of context. The problem then becomes how to model context and scenarios can be a valid answer. In the following we describe in detail the methodology applied to the safety assessment.

Socio-technical approach to evaluation

The ITI Interface integrates basic features to support the controller's job. It is based on the EATCHIP Basic (i.e. SYSCO) and Advanced Functions (i.e. Safety Nets (SNET), Monitoring Aids (MONA)) specifically oriented to support some of the current controller's tasks. In ITI we have adopted an approach to evaluate usability and safety that systematically assesses the

components at stake in a process (software applications, organisational and cultural aspects, physical layout, human operators) and their interactions. The assessment is based on the idea that a criticality is a wrong distribution of resources among the components. The method allows to pro-actively assess which aspects of the system may impair or enhance safety after the introduction of new artefacts in the work setting.

This method is inspired by the well known conceptual model, named SHEL, developed by Elwyn Edwards (1972). The model describes the behaviour of interactive systems with special regard to human factors issues. SHEL is an acronym for Software, Hardware, Environment, and Liveware. Software refers not just to computer software but to the rules, procedures, practices that define the way in which the different components of the system interact among themselves and with the external environment. Hardware is used to refer to any physical and non-human component of the system such as vehicles, tools, manuals, signs and so on. Liveware refers to any human components of the system in the relational and communicational aspects. Environment refers to the socio-cultural and organisational environment in which the different components of the process interact.

The SHEL model concentrates on the interfaces among people and all system components including other Liveware resources. The important point about SHEL is that it offers a system view where humans cannot be considered as isolated from the other system components. This view is consistent with recent theoretical work in cognitive psychology including Distributed Cognition (Hutchins, 1995) and Activity Theory (Nardi, 1996) but is grounded in simple concepts that can be understood by system designers without this theoretical background. In particular, Activity Theory assumes that human behaviour is not a set of disembodied cognitive acts (e.g. decision making, classification, remembering). Rather conscious activity takes place in everyday practice and it is inextricably embedded in a social matrix of which every person is an organic part. In this respect the unit of analysis to take into account is wide and articulated. It consists of a subject (individual or group), an object or motive, artefacts (or tools) and socio-cultural rules and norms. Hence human activity should be considered as a socially and culturally organised ensemble where artefacts play a critical role in mediating human activity.

The SHEL model fits well with this theoretical framework since it considers any specific process as a combination of the three resources. This combination changes as soon as the process evolves and it is not exclusive in the sense that many combinations may occur during the process.

In ITI we applied the method in order to proactively assess and evaluate safety issues related to the impact that ITI tools can have in the context of Air Traffic Control.

The method develops in following three phases:

1 - Preparation of test material

- Identification of the basic Software (S), Hardware (H), Liveware (L) components that may affect the use of ITI. This phase of the method, as the following one, was possible thanks to an accurate activity analysis based on observation in the real operational context and to the analysis of official documents on operational procedures.
- Identification of safety issues. The following safety issues were identified: information visibility, consistency and integration of the information needed to perform the activity; conflict detection, coordination/transfer, hand-over procedure, monitoring.
- Scenario building. A set of scenarios were selected representing critical interactions among H,S,L components. Indeed the evaluation does not aim to sequentially test each single procedure as standing alone, but to create a simulated realistic operational context, in which non linear interactions among components could emerge. Scenarios are one of the possible representations of work processes. They are an interpretation and a reconstruction of the work processes. For this reason, scenarios focus on some aspects or features of the process and neglect some others. In this respect scenarios are fundamentally different from simple "traffic samples". They do not include only number and typology of traffic in a given unit of time but are realistic situations relevant in terms of safety (exceptional circumstances, ambiguous procedures, controllers' errors, communication misunderstandings). They are realistic since resulted from the analysis of the current work activity. It is important to point out that scenarios represent information coming from different sources (activity observations, documents, interviews, story telling) and different people with different knowledge and views (controllers, domain experts, human factors experts, developers).

In ITI scenarios were built in the following steps:

- 1- matching safety issues and ITI applications. We verified that the selected scenarios matched the identified safety issues and highlighted which ITI applications could impact on these safety issues.
- 2- Trying out scenarios on the simulation platform.
- 3- identifying SHEL components (H: all the applications implemented in ITI; S: all procedures needed to the process development; L: the actors involved in the simulation, couples of planner and executive Controllers, plus pseudo-pilots).
- 4- envisioning interactions among components. For each scenario we tried to identify which interactions between the operator and the other system components (L, S, H) could be safety critical using the ITI applications.

- 5- structuring scenario for test sessions to plan a complete and meaningful test. Elements of the structure included: rationale, estimated temporal duration, actors, goal (the objective of the scenario that the evaluators had to reach), initial condition (status of the interface), operational context, ITI applications involved, other external supports available to the controllers.
- 6- Preparation of the SHEL Question Table, that is a list of questions related to the coupling of H,S,L components of the selected scenarios.

2- Run the test

Each test was developed in three phases:

Warming up

The session started with a brief explanation of the simulation schedule and its objectives. Before starting the evaluation, the controllers were asked to familiarise with ITI even if they were already been trained in a previous pilot session.

Scenarios execution

The controllers received the scenario objective on a paper sheet. They were requested to execute the scenario simulating as much as possible the real operational conditions.

During the simulation the activity of controllers was observed and video recorded applying ethnographic methods. The objective of using ethnography to activity analysis is to understand the social context of the real work settings in which the activity takes place. The key benefit that ethnography offers to design and evaluation is a rich and detailed description of the complex features of the work setting. In ATM for example, what ethnography especially provides is a throughout insight into the subtleties involved in the work and in the routine interactions among members of the team work. The material which is possible to collect with this kind of analysis is related to the direct observation of an analyst who tries to understand the real work practices. In some way, ethnography is an alternative approach to analytical methodologies like task analysis (Kirwan and Ainsworth, 1992) and workflow. Indeed these methodologies do not focus on the social dimension of the work organisation, but aim at modelling more abstract and normative structure of tasks. The vital moment-by-moment mutual checking of “what is going on” by the various members of the team is missed in task analytic approaches to describe the work.

This kind of analysis is based on video and audio recording of the situation and notes taken by the observer. The use of audio and video analysis is not obvious (Hutchins, 1995). Some aspects of the setting are usually lost in the video and audio. The camera angle leaves some parts of the environment obscured, for example. However the data that can be collected are extremely rich. Once the situation has been recorded and transcribed, the data are

interpreted by the analyst. In order to avoid misinterpretations, the transcriptions and the related interpretations are discussed with users for a final check.

Post test: retrospective comments, focus group with the user.

After each scenario, the controllers were involved in a debriefing session based on the video recording of the test. The controllers were asked to freely comment their performance even if the designers drove the discussion on the assessment of safety issues. At the end of the debriefing the SHEL Question Table was filled out. The technique of Focus group allowed to obtain a wide variety of views from a range of people (controllers coming from different operational realities, system developers) who might have different but equally relevant perspectives about the use and the impact of the system. Moreover, due to the freeform nature of focus group, unexpected viewpoints were identified which may be otherwise overlooked if a more structured approach, such as questionnaire methodology, is taken alone. On the basis of video recordings, controllers were mainly asked to:

- discuss about the performance of the system (accuracy, representation, reliability etc.) also by asking explanations to the system developers;
- reason about their activity with the information provided by the new system;
- make a comparison among the activity carried out with or without the support of the system.

3- The data analysis

The data emerged from test sessions, controllers’ retrospective comments and focus group were analysed to evaluate the impact of ITI on safety issues and to proactively evaluate the possible occurrence of new safety issues due to the introduction of these tools. The findings of the analysis were based on answers to the SHEL Question Table, the discussion of safety issues with controllers and the observation of the activity during the simulation.

The results are provided in a synoptic view that highlights critical interactions among SHEL components and pro-active safety assessment. In the following we provide an extract of the outcomes we obtained from the application of the method.

Outcomes

This section describes one example of process (in this case the execution of a procedure) where multiple breakdowns have been observed. This example aims to show how a multidimensional approach can catch problematic interactions at different levels of the activity.

What characterises the ITI interface as an innovative system is the new basic HMI, the adaptation to a full stripless environment, the possibility to perform electronic co-ordinations and the introduction of

advanced functions like SNET and MONA. Although the operational concepts of these functionality are the real strength of the design philosophy, however sometimes the effectiveness of such concepts is reduced by the way in which they have been designed in the system (data presentation, interaction design mechanisms). The following extract mostly concerns the electronic coordination, with a particular emphasis on the stripless philosophy and the enhancement of aircraft (a/c) labels. The simulation, built around ad hoc scenarios, allowed us to observe and then analyse design defects that impact on usability and safety.

As a stripless system, the entry and exit flight levels can be coordinated directly through the a/c label. This allows controllers to perform their activity concentrating right on the radar tracks, without shifting their attention to peripheral windows or tools in general. This can reduce the time needed to perform the task and the risk of errors. The application of a multidimensional approach to validation allowed us to verify how different problems could combine together and provoke critical interactions within the same process.

The following example addresses one instance of the coordination procedure, i.e. when a controller proposes an exit (or entry) flight level and the other controller counter-proposes a different value. The system is designed to allow controllers to manage both proposal and counterproposal in a rapid and effective way. What happened during the simulation is that if the controller of a sector had several opened co-ordinations (both incoming and outgoing) he hardly realised the counterproposal received from the adjacent sector, since the appearance and position of proposed and counter-proposed values is the same as shown in figures 1A/B.

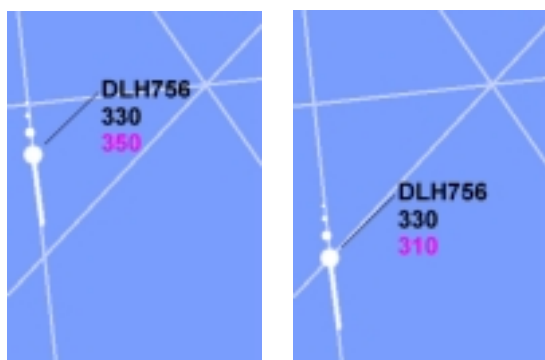


FIGURE 1A: PROPOSAL FIGURE 1B: COUNTERPROPOSAL

Just when the coordination had become in some way critical (that is the executive controller wanted to close the coordination in order to hand the a/c over and he couldn't), he realised the breakdown. He tried then to understand why the coordination was still open, asking the planner controller. At this time, the planner did not realise that a new value was counter proposed, since he did not notice that the magenta value was changed. Therefore, he asked the planner controller of the next sector to answer to the coordination proposal. In this

case the breakdown was solved and the coordination was closed. However the time needed to accomplish the procedure (and consequently the risk of errors) significantly increased.

The observation revealed that in later simulation sessions, when controllers became more familiar with the system functionality, the task was performed in more effective way. It was due to the fact ITI provides a tool that helps to discriminate between an outgoing coordination proposal and an incoming counterproposal: the Coordination IN and OUT windows. These windows list all incoming and outgoing coordination messages. Once the controllers understood the functionality of these windows, they used them to disambiguate the meaning of the magenta value in the a/c label simply checking whether the coordination message was present in the IN or OUT window.

Why didn't they use this strategy before? Actually a misinterpretation of the meaning of the Coordination IN and OUT windows occurred: instead of considering them two different lists for incoming and outgoing coordination messages, controllers interpreted them as lists for coordination of "inbound" and "outbound" flights.

To summarise, the event below is an example of breakdown occurred by the combination of different critical factors: attentive, wrong mental model, critical interaction between L, H, S. The use of different evaluation methodologies allows to detect and analyse the problems at the different levels.

The process and the breakdown are represented in the following figures 2A-C.

Figure 2A describes how the activity should be performed according to the system model. The Heuristic Evaluation revealed that the first breakdown (red square in Figure 2b) is a graphical interface design flaw, since different design principles were violated: provide clear feedback; make system status visible; allow recognition rather than recall.

Figure 2b and 2c. Applying the Cognitive Walkthrough it was possible to detect two main problems or cognitive distances. A cognitive distance refers to the human information processing needed to fill the gap between one system state and the next. The cognitive distance can be broken down into action or articulatory distance and evaluation or outcome distance. In the former case it refers to the cognitive effort required to translate the users' goal into actions understandable by the system. In the latter case it refers to the information processing required to interpret the system's output back in term of user's goals and intention.

In the evaluation we discovered two kinds of cognitive distances: semantic and inter-referential. Semantic distance is a specific kind of cognitive distance. For the output evaluation, it refers to the amount of human information processing needed to translate the meaning of the output of an action in the terms of the intention it serves (e.g. after obtaining a given result how close am

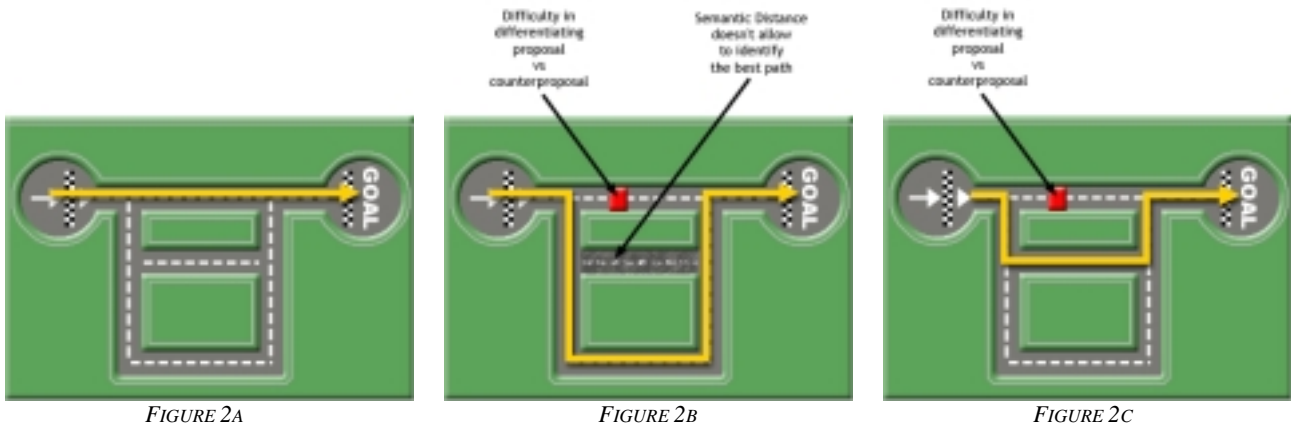


FIGURE 2A

FIGURE 2B

FIGURE 2C

I to the fulfilment of my intention?). In terms of action execution, it concerns the relationship between the user's intentions and the meaning of the actions that are possible in the interface language (e.g. is there any immediate way to map my intention in action that the system allows?).

The inter-referential distance is the cognitive processing needed to put in relationship the information processed in action execution and the information available as result of the action (e.g. where does the output of my action come out? which are the modalities of the feedback to my action?). These forms of distance allow to describe cognitively the relationship between the task the user has in mind and the way the task can be performed via the interface.

In our example the controllers' interpretation of Coordination Windows was conditioned by their previous experiences, since they are used to think in terms of flight movements, rather than in terms of exchange of system messages. This is the reason why the meaning of Coordination Windows was not easily understood and, consequently, their correct use required a considerable cognitive effort (semantic distance).

Moreover, even after training and familiarisation with interface, controllers still had problems in using the Coordination Windows, since the place of the input (label field) and of the output (Coordination Windows) are physically distant in the interface (inter-referential distance). This also contributed to a troublesome performance.

SHEL proactive assessment

The SHEL methodology provides a powerful framework for safety assessment. Its main assumptions can be resumed in the following statements:

- Each resource in a process is a stakeholder of knowledge needed to perform the process.
- The modification of a resource changes the interaction among SHEL components (i.e., the introduction of new tools in a work setting).
- Criticality is a wrong distribution of the resources among the components.
- Breakdown is a rupture in the interaction between the components.

- Any process can be executed with a different allocation of resources.
- Dynamic and complex system environments require flexible allocation of resources for process execution in order to deal with breakdowns of components and unpredictable situations.

Furthermore, SHEL allows also to proactively assess problems and critical interactions that were not observed during the simulation. Indeed simulation scenarios focus on specific accounts of activities and some critical interactions may not emerge during the observation. The SHEL analysis goes beyond the very observed activity allowing to proactively assess the impact of the new tools even on the entire process even if not directly observed. Indeed if we project an observed critical interaction to the development of the entire procedure, we can discover that the tools, in the current implementation, could put at risk also the interaction between other components of the process. In our coordination example, if we project the breakdown to the whole coordination procedure we can reasonably state that:

- Controller who counter-proposed a different value can be in trouble to understand why a coordination remains open;
- An open coordination can result in a critical situation both for the controller who hands the a/c over and for other controller who receives it. In this case the adopted strategy is to move the surrounding traffic away in order to not interfere with the problematic traffic;
- A/c that are involved in open coordinations are not aware of the problem, so any requests from their side can intrude with the activity.

Conclusions

In conclusion, the methodology we presented was successfully tried out in different contexts of safety critical applications (Rizzo et al, 2000). The application described in this paper confirms its potential to a proactive evaluation of the impact of new technological tools in real operational settings. In particular the method offers the following advantages:

- it allows to systematically detect critical interactions about system components and to infer new ones;

- it allows to overcome the limitation of scenarios that represent categories of single events;
- it provides the knowledge necessary to specify requirements and re-design defects. Indeed the method clearly detects at what level the problem occurs and which interactions among system components should be redesigned to solve it.

References

Edwards, E., 1972, Man and machine: Systems for safety, Proceedings of British Airline Pilots

Associations Technical Symposium (British Airline Pilots Associations, London), pp. 21-36.

Rizzo, A., Pasquini A., Di Nucci, P., Bagnara, S. (2000) SHELFs: Managing critical issues through experience feedback in railways. *Human Factors and Ergonomics in Manufacturing*, 10, 83-98.

Hutchins, E. (1995). *Cognition in the Wild*. MIT Press.

Kirwan, B. & Ainsworth, L.K., (1992) *A Guide to Task Analysis*. Taylor & Francis, London.

Nardi, B. (Ed) (1996). *Context and Consciousness: Activity Theory and Human-Computer Interaction* Cambridge: MIT Press.

Author Biographies

Patrizia Marti.

Patrizia Marti is lecturer in Educational Technologies at the Communication Science Department of the University of Siena (Italy) and in Cognitive Ergonomics at the Industrial Design Department of the Polytechnic, Milano (Italy). She has been involved in international research projects in the areas of nomadic systems, educational technologies and air traffic management. Her current research interests include the design of human activities in context (situated interaction) and human factors in safety critical applications. In the past, she has been involved in numerous projects related to the design and evaluation of ATM systems, in strict collaboration with Eurocontrol Experimental Centre (Bretigny, France) and ENAV S.p.A. (the Italian Aviation Authority).

Paola Lanzi

Degree in Communication Science at the University of Siena, Italy. Since August 2000 she has been collaborating with the Multimedia Communication Laboratory of the University of Siena working on international projects for the evaluation of safety critical applications and nomadic systems for art and entertainment. Her activity is currently focused on the analysis and evaluation of ATM systems. Her research area includes user studies, distributed cognition, real time simulations.

Francesco Pucci.

Born in 1974 in the south of Tuscany, in 1999 he gets his Master Degree (Laurea) in "COMMUNICATION SCIENCE - Management of Communication Technologies" at the University of Siena, Italy with a thesis in Human-Computer Interaction. Since October 1999 he works as a free-lance consultant for Butera e Partners (Rome-Milan, Italy) and as an HCI researcher at the Multimedia Communication Lab of the University of Siena. Involved in several international research projects, his areas of interest include the design of systems for education and cultural heritage, and activity analysis and system evaluation in the Air Traffic Management context. Current research addresses real time simulation, scenario based design and distributed cognition.