A Human Operator Function Reliability Statistical Modelling For Preliminary System Safety And Risk Assessment

Dr. Sergey Nesterov
Defence And Systems Institute University of South Australia (DASI, UniSA)
sergey.nesterov@unisa.edu.au

Abstract. This paper describes key concepts, approach and modelling details for preliminary safety (risk) assessment of human functions in complex systems controlled by a human operator or a group of operators. Human function consists of sequence of tasks and each task can be “abnormally” executed by human potentially leading to undesired events compromising system safety. Usually function reliability is measured in terms of function failure probability (likelihood) and the actual figure comes from a subject expert. A new approach has been developed to assist human factor and safety experts in human function reliability estimations and assessment. Monte-Carlo statistical modelling is used for human function assessment. In the model each function execution (as a set of tasks) generates an output - stochastic sequence of normal or failed tasks. This output is subject for a safety analysis, conducted by a system specific algorithm which generates output safety index value (representing levels of safety), for example, an accident, major or minor system mishap. Number of tasks in a function and individual task failure probabilities will determine number of iterations to produce credible safety index distribution. The prime focus is “spontaneous” human errors not complicated by unhelpful system events, like equipment failure, or other human performance shaping factors, like, fatigue. The approach gives an estimation of a “pure” human impact on function or operational procedure reliability. However, this model and the associated software tool can be used in the different contexts, for example, in estimating function reliability under fatigue by applying “fatigue” multiplier to each task failure probability. The heart of the model is human agent which takes input information: task, failure task nominal probability, nominal task safety level associated with the particular task (provided by safety expert) and produces stochastic output result: task failure or normal task execution. Human agent model behaviour has interesting features based on psychological observations, for example, subjective task failure probability correction, ability to notice and correct (mitigate) failed tasks. A new concept of “human error entropy” as a human performance factor influencing failure rate probability introduced in the human agent model will be discussed.

1. INTRODUCTION

Designers of safety critical systems in areas such as nuclear power, transport, chemical processing, aviation and military systems, have considerable interest in qualitative aspects of human reliability (performance). The DASI team undertakes research in the human performance area and is developing a simulation tool, called HOPS – Human Operator Performance Simulator to provide qualitative human (reliability) performance assessment and safety index evaluation in the operational context with the prime focus on aviation domain.

A fundamental part of the system design process involves the evaluation of potential human errors (or deviations in performance), their impact upon successful operation of the system and potential methods for recovering errors or mitigating their consequences. In order to perform comprehensive and cost-effective analyses, there is also a requirement to quantify the likelihood of potential failures revealed by the quantitative analyses [3].

HOPS is a software simulation tool containing models of human operator stochastic (probabilistic) behaviour which can be used to evaluate reliability (safety) of system operational procedures prior to more expensive and time consuming human subject experiments and detailed safety analysis. Simulation results can be used by systems engineering designers and safety engineers as a basis for preliminary system reliability and risk assessment.

Paper [1] presents cognitive task simulation approach for single pilot in aircraft operation and underlying simulation philosophy. Extended approach for two pilots (crew), operating in civil aviation, has been presented in paper [2]. It is accepted that modelling human crews presents major challenges. However, while crew error remains a major contributor to the majority of accidents, the need for prediction requires that this difficult problem be addressed and solved.

I strongly believe that proposed modeling techniques and modeling approach are generic and can be applicable in number of other domains. Purpose of this paper is to demonstrate generic applicability of the method for system reliability and safety preliminary analysis.

2. SIMULATION PHILOSOPHY OF HUMAN PERFORMANCE

Our approach to human performance (and error) is based on Hollnagel’s human performance and error “philosophy” [4]. Quoting Hollnagel, “Errors” or deviations have a positive side since they enable users – and systems – to learn about the nature of system behaviour and accidents. Indeed, deviations from the norm can have a distinct positive effect and be a source not only of learning but also of innovation. This requires that the system has sufficient resilience to...
withstand the consequences of the uncommon action and that it is possible for the users to see what has happened and how. Performance variability management captures this dual nature of performance deviations. This approach fully acknowledges that unwanted outcomes usually are the result of coincidences, which are the inevitable consequences of the natural variability of a system’s performance” [5]. In other words human operator performance always has variations from the “ideal” pattern. Some deviations may be useful while some deviations can be classified (by an observer) as “errors” and lead to an accident. Distinctions between different types of performance are usually based on an evaluation of the potential outcome. It is thus common to distinguish between accidents, incidents, near misses, and normal performance. In addition a human operator can correct “errors” or mitigate the effect of undesired deviations. Therefore, it is necessary to include in the model mechanisms for actions such as: outcome evaluation, error correction and human operator adaptability to changes in the system or in the environment.

The heart of the HOPS simulator is a stochastic human cognitive agent accepting tasks and “executing” them right with certain level of “failure” probability. If the task fails, then “error” (as a measure of deviation) is introduced into the system with a corresponding severity level. The severity level is an expert estimation of task failure impact on system safety, and measured as low, medium and high. The safety measure of the human performance deviation is based on the number, sequence and severity of all task failures introduced into the system during simulation.

This approach is very close to the Human Reliability Analysis (HRA) defined as the probability that a human will correctly perform some system-required activity during a given time period without performing any extraneous activity that can degrade the system. As Holnagell writes: “Two problems that need to be addressed by HRA methods: the problem of context and the problem of cognition. The actions of operators are not simply responses to external events, but are governed by their beliefs as to the current state of the system. Since operators make use of their knowledge and experience, their beliefs at any given point in time are influenced by the past sequence of events and by their earlier trains of thought.

Alltogether this means that human performance takes place in a context which consists both of the actual working conditions and the operator’s perception or understanding of them. It also means that the operator’s actions are a result of cognition and beliefs, rather than simple responses to events in the environment and that the beliefs may be shaped - and shared - by the group” [4].

Contemporary HRA includes quantitative performance prediction approach which has been used in the HOPS simulator. Quantification involves finding the probability that a specific action may go wrong and then modifying that by the aggregated effect of a more or less systematic set of performance shaping factors (PSFs) in a given context: scenario and human agent state. In other words, human reliability is an abstract measure of human performance. I believe that human performance variability is a natural phenomenon of stochastic nature. There is a distinction between variability in the real world and uncertainty in the models or uncertainty in our predictions of these phenomena. Human errors are regular and predictable consequences of variety of factors [5].

3. HUMAN RELIABILITY AND SAFETY PRELIMINARY ANALYSIS AS A STATISTICAL EXPERIMENT

I consider that preliminary system reliability and safety analysis have to be done at system conceptual design phase. For simplicity system reliability may be presented by two figures: system hardware reliability and reliability of system operation, represented as human (crew) reliability executing given operational procedure. The simplicity means that hardware and human operator failures are independent. It is not necessary true in some cases but acceptable assumption for a preliminary analysis.

![Figure 1: Reliability preliminary analysis model for systems with human in the loop.](image)

System engineers developed good understanding and number of well known techniques for hardware reliability assessment based on the notion of sequential and parallel hardware connections. Also there are widely available sources of hardware elements reliability, for example, electronic or mechanical components. Reliability of Operational Procedure (OP) is less known and there is no universal approach how to measure or calculate reliability figure, based on human factors analysis. It is even harder to make a safety analysis which includes and interprets in the context results of system reliability analysis.

To address this complex problem of OP reliability I think to develop a human performance simulator capable of evaluating the reliability (or safety) index of the given operational procedure. This operational
A task-oriented approach is based on a human (crew) cognitive model and a stochastic modelling approach, see figure 2.

The key conceptual elements of the model are:

- Event or task driven simulation approach, implemented in agent based programming techniques
- Stochastic dynamic modelling of a task sequence outcome, human behavior as a stochastic (Markov) process
- Focus on “pure” human failures under a normal (standard) operational conditions
- Focus on realism of a cognitive psychological human operator (crew) model
- Concept of a nominal task failure probability reflecting “pure” human failure under standard conditions
- Concept of failure probability adjustment depending on human “perceived hazard” level of a task failure
- Concept of operator “error related entropy” reflecting cognitive state of human. Error related entropy is a numeric parameter used for nominal probability adjustment for each task outcome modelling
- Concept of operational procedure reliability (safety) statistical index as a measure of hazard potential (or precondition) for accidents, incidents or safe flights.

Each simulation run accepts as input parameters:

- Number of simulated runs of an operational procedure – N, where N value should guarantee modeling outcome confidence
- Operation Procedure (OP) as a sequence of predefined operational tasks:
  \[ OP = \{ task_1, task_2, ..., task_m \} \]

where each task has associated properties: Perceived Hazard Level (or severity) of a failure of a task and associated nominal task failure probability (nPr):

\[ PHL = \{ \text{low} | \text{medium} | \text{high} \} \]

\[ nPr \] – nominal task failure probability value as an expert opinion or taken from valid sources

Therefore each task is a pair of values:

\[ Task_j = \{ \text{Severity}_j, nPr_j \} \]

Nominal probability concepts have been introduced in HRA techniques, these techniques fall generally into two categories [6]:

- Those that use a database (e.g. THERP, JHEDI, HEART). Use a collection of generic error probabilities. These probabilities are manipulated by the assessor to fit the context-related Performance Shaping Factors (PSF) in the scenario being assessed.
- Those that use expert opinions (e.g. APJ, PC, SLIM, IDM). Uses expert judgments of particular scenarios and PSFs to render information into Human Error Probabilities (HEPs)

For a predictive tool, the “database” technique has significant advantages:

- A new innovative task can be “approximated” by an existing generic task with associated failure probability
- Alternatively, a new generic task can be added as a unique generic task to the database and failure probability can be estimated.

Every single step of task execution takes input parameters associated with particular task and transform pilot state (“entropy”) producing stochastic task outcome success or failure.

Figure 2: Simulation Diagram.

Figure 3. Single step for task outcome modelling.
As an output for each run of an operational procedure the simulator produces probabilistic sequence of the tasks outcomes (successes or failures).

\[ Output = \{ \ldots \text{TaskOutcome}, \ldots \} \]

The second step of each is outcome sequence safety (reliability) analysis. During this analysis operational procedure Safety Outcome Value SOV will be produced based on the specific system routine (algorithm) developed under safety engineer guidelines. For example, outcome may be classified as catastrophic in case of double failures in the procedure modeling.

\[ SafetyOutput = SOV\{Output\} \]

For a given number of operational procedure traditions statistical distribution can be built for, let say catastrophic safety output events normalized to the total number of iterations. This index may be taken as reliability index for a operational procedure.

4. SINGLE OPERATOR RELIABILITY MODEL

There are a number of unique features in the human model providing a mechanism for task nominal failure probability adjustment in a dynamic context.

Firstly, the pilot model has two intermediate internal states related to a task execution after an initial task failure:

- Internal failure (or error) detection state
- Internal failure (or error) mitigation state

The core task outcome modelling algorithm for a single operator is shown below.

![Figure 4: Single task outcome modeling state machine.](image)

Therefore, task success outcome can be achieved in two ways:

- In one single step – a successful task execution
- In two additional sub-steps (if the first task execution attempt fails) - pilot notices a failure and after that he successfully mitigates a failed task.

Secondly, it is important to introduce two mechanisms for a nominal task failure probability adjustment addressing two observed psychological phenomena. The first phenomenon states that a failure probability for the task depends on the criticality of task (or severity impact on the system safety). In other words, similar tasks would have different rates for failure depending on the subjective pilot perception of the task criticality, the more critical the task – the less the failure probability.

The second phenomenon relates to a stressor (performance shaping factor) affecting cognitive performance. It is likely the best option to call this stressor as “entropy”\(^1\). I assume that if a pilot “posses” an optimal entropy then under this condition a pilot has the best performance equivalent to nominal failure probability for a given task. If this entropy deviates from the optimal, the task failure probability is rising.

Entropy is rising on different levels (degrees) in the situations (see Figure 4):

- when a task failure happened
- when a pilot notices a failure
- when a pilot fails to mitigate a failure
- when a pilot successfully mitigates a failure

Entropy is going down when a task has been executed successfully\(^2\).

This entropy mechanism leads to the generation of failure clusters (burst of errors), that are often observed in reality. At some degree of generalisation it is possible to consider entropy similar to a stress concept. Further study is required to investigate the nature of the phenomenon and its impact on performance. I believe that entropy impact on performance is not linear and may be similar to the stress impact curve\(^3\).

Individual tasks and associated nominal probabilities can be expressed as domain specific or can be generic depending on the task data availability.

Below, there are examples of cognitive and non-cognitive generic “primitive” tasks:

- Motor Action
- Communication
- Monitoring
- Procedure Recall
- Information Recall
- Situation Comprehension
- Situation Assessment
- Decision-Making

5. CREW (TWO OPERATORS) MODEL

A two-pilot crew flying a commercial aircraft has been selected as an example and represents a single cognitive

---

1. In earlier papers I called this stressor as a ‘temperature’, but later decided to use entropy instead to avoid confusion.
2. Real time is also a “cooling” factor, but I do not count real time as a factor in event driven simulation.
3. Error entropy mechanism can be blocked in simulation if needed.
work system which is not easily decomposable into independent elements. In this model two pilots, with common goals and objectives, have specific roles and functions and mission requires some form of dependency. In addition, such cognitive work includes dynamic interactions across a set of factors. A simplified crew model has been proposed, which can be used for a performance (and safety) prediction.

The main questions I try to address in the crew model development are:

- How is crew functioning?
- What are crew performance (reliability) indicators and their impact on reliability (safety)?
- How does one single operator performance and conditions influence the other pilot?

In the crew model the first pilot module (Pilot 1) is functioning exactly under the same rules as a pilot in the single pilot model, described earlier. The second operator (pilot) is acting as a “supervisor” to check executed by the first operator task for failure. Therefore, a single pilot failure in execution of a particular task does not necessary lead to a final crew task failure, unless the other pilot does not notice a “root” failure.

![Figure 5: Crew task processing model.](image)

The proposed functional model has important underlying assumptions:

- pilots cross supervise each other’s task execution
- the model implements the most efficient crew performance mechanism

By analogy with the one pilot model a crew performance is measured in terms of crew task failures and associated pilot’s states of entropy.

Preliminary analysis and simulations of a two pilot scenario for Boeing 737-NG start takeoff procedure highlight areas for further investigation.

The most important issue is pilot’s interdependence in aircraft handling. For example, cross supervision performance has the best results if both pilot behaviors are independent. Some tasks are allocated for both pilots in the SOP, e.g. “verify increasing N2 RPM”. That “ambiguous” task allocation raises the question: do they really execute task independently in parallel without cross-check? Alternatively, “both” can mean: the first available pilot picks up a task? Therefore careful task decomposition would be required.

I believe that crew represents a single cognitive work system and certain level of pilot’s interdependence exists and influences crew performance: task failures and mitigation. In our case that interdependence can be modelled through pilot entropy mechanism and subject for further study.

For example:

- strong leadership (in crew) could mean that the leader entropy influences the second pilot
- equal partnership (in crew) could mean pilot’s entropy averaging after each task

6. FUTURE MODEL ENHANCEMENTS

It would be beneficial to include a number of performance shaping factors (PFS) affecting operators performance, for example, fatigue and operator’ recency factors in the model. Our particular interest is performance shaping factors “close to the cockpit”, like these, as well as experience, and environmental factors, directly affecting human cognitive performance, like cabin (work place) environmental parameters, such as, temperature and vibration. The final list of important PFS variables and their impact on a operator performance (task failure) are subjects for further investigation.

The full scale model implementation requires a significant amount of intellectual and programming effort. However, I can envisage two additional developed phases of the model.

- Event driven model of two pilots with extended model for pilot to pilot and pilot to ATC communication, paper [7] presents a modelling approach on ATC control tasks. It will be the next level of simulation complexity which also will include modelling of the pilot’s roles, knowledge, and workload sharing. There are a number of issues which have to be addressed:
  - How pilots communicate: linguistic factors
  - What pilots communicate: content-coding factors [9]
- Indicators of the communication performance
- Event driven with real time constraints for crew operation. This brings additional complexity in the model introducing time

---

4 Other types of crew configuration are available, for example, without cross supervision.
constraints on the decision making and procedure execution processes.

6.1 Sources Of Empirical Data For Model Calibration And Verification

Empirical data required for the simulation are:

- Failure probabilities associated with the tasks
- Task severity values (criticality)
- Performance shaping factors and their impact on human performance

Some of these data can be found in the literature, whereas other data can be derived from the existing databases collected in the aviation industry. However, severity values associated with the particular task must be obtained from experts, mainly experienced pilots.

In civil aviation domain data sets collected under the LOSA program [10] (world wide program) are a credible source of information to be used in the required procedure database. By analysing raw data it is possible to identify the type of task and estimate task failure probability (attributed to the particular aircraft model). At the next step, it is possible to make a rating of the procedure failure severity providing experienced pilots’ opinion and provide statistical rating validation.

7. CONCLUSION

It is a great advantage to have a predictive simulation tool which gives an opportunity to be proactive in system design and analysis, as well as, in system operational stage in order to evaluate proposed system changes. For example, the proposed simulator evaluates for designers the reliability (or safety index) of the particular standard operation procedure as a part of advanced system risk analysis.

It is accepted that modelling human crews presents major challenges. However, while crew error remains a major contributor to the majority of system malfunctions or accidents, the need for prediction requires that this difficult problem be addressed and solved. I believe that modern, agent-based modelling techniques offer an opportunity to achieve desired objectives.

REFERENCES


4. Erik Hollnagel. www.ida.liu.se/~eriho/


