The LABY Microworld: A Platform for Research, Training and System Engineering

Jean-Paul Imbert[1], Helen M. Hodgetts[2, 3], Robert Parise[4], François Vachon[2], & Sébastien Tremblay[2]

[1]Laboratoire d'informatique interactive, ENAC, Toulouse, France
[2]École de psychologie, Université Laval, Québec, Canada
[3]Department of Applied Psychology, Cardiff Metropolitan University, Cardiff, UK
[4]DSNA, DTI R&D/EgisAvia, Toulouse, France

The LABY microworld, a functional simulation of Air Traffic Control (ATC), captures the underlying processes involved in electronic air traffic management with a simplified version of the operational human-machine interface. LABY is a computer-based human-in-the-loop dynamic environment whereby a controller must issue directional commands to guide aircraft along a predetermined route, while avoiding potential conflicts and dealing concurrently with other incoming information. It can be used for human factors research or system engineering purposes, or configured specifically for use with expert controllers for the training of non-technical skills in ATC. We present a use case of LABY, comparing the efficiency of input devices for ATC: Input times using the mouse were quicker than with the stylus, but error was not greater. We discuss the potential of LABY for system engineering, training and research purposes.

INTRODUCTION

Microworlds are interactive and dynamic simulation environments that recreate the functional relations of a task, providing a high degree of external realism compared to laboratory studies, yet retaining a level of control over the manipulation and measurement of task variables that is not possible in the field (Brehmer & Dörner, 1993; Gonzalez, Vanyukov & Martin, 2005; Gray, 2002). From a Cognitive System Engineering perspective (e.g., Rassmussen, Pejtersen, & Goodstein, 1994), microworlds sit at a mid-point on the continuum between initial laboratory work and studies conducted in operational settings. Semantically-rich microworlds create an immersive experience, and share certain key characteristics with their real-world counterparts (Brehmer, 1992) in that they are dynamic (evolving autonomously in real time), complex (involve interacting components and conflicting goals), and opaque (relationships between variables must be inferred). Examples cover a range of domains, such as military settings (Hodgetts, Vachon, & Tremblay, 2014), crisis management (Tremblay, Vachon, Lafond, & Kramer, 2012), and emergency response (Gagnon, Couderc, Rivest, & Tremblay, 2011). In the current paper we present the LABY microworld which simulates an air-traffic control (ATC) task (e.g., reading instructions, monitoring movement, anticipating trajectories). It can be used for system engineering and evaluation, training, or research purposes within ATC specifically, or within the context of supervisory and monitoring tasks more broadly.

PRACTICE INNOVATION

The LABY Microworld

LABY is based upon the main task of guiding *n* plane(s) around a predetermined route, indicated by a green path. Participants must input numerical values such as heading, flight level, speed, etc., in order to direct flights around the trajectory and avoid any obstacles which may occur; for example, aircraft managed automatically by the system can be used to create conflict situations, and obstacles can be created along the route which require users to issue heading changes to avoid them. LABY is dynamic in that the situation changes both autonomously (system-controlled aircraft) and as a direct result of actions made by the participant (guided aircraft). It is a complex multitasking environment, which involves controlling multiple aircraft and engaging in other concurrent tasks depending on the particular configuration, such as responding to system alerts. Scenarios can be created that are opaque in the sense that the participant needs to build up an understanding of the impact of different variables on the aircraft under their control, e.g., the speed or heading correction needed to avoid a prohibited area or a potential conflict between aircraft.

Construction of the experimental task and the manipulation of objects is performed and controlled using an editor, *LABYedit* (see Figure 1). This editor provides the experimenter with graphic elements to build a route, for example, path segments (straight lines or curved), constraint areas (where the operator is required to issue a command, such as a change in speed), zones triggering an interfering flight, or other obstacles. Workload can be altered according to how many aircraft the participant must control, the number and type of clearances required, and the number/trajectory of other interfering flights; while time pressure can be augmented by increasing aircraft speed. Critical events (e.g., conflict with other aircraft), or adverse conditions (e.g., a prohibited area) can be implemented when a guided aircraft enters a particular zone. Other features of the system can also be altered for research or training purposes, such as the size or number of screens, information input modes/methods, and the presence or nature of alerts/reminders (visual/auditory). The current list of graphics is not limited, and other types of instructions or constraints could be implemented.

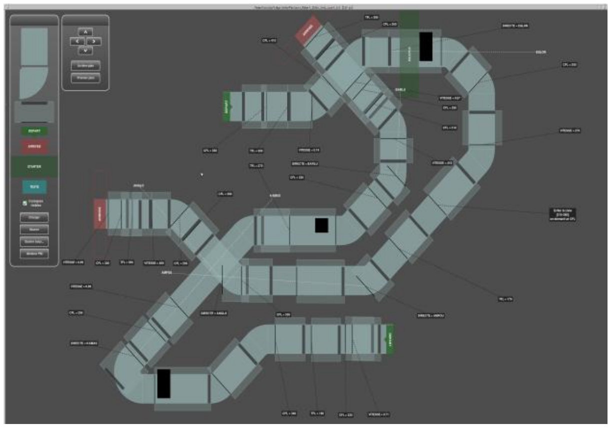


Figure 1. The course editor *LABYedit*.

System engineering

With the increasing amount of information to be displayed – often across multiple screens – the cognitive demands of ATC can sometimes be overwhelming. LABY offers a platform with which to assess prototype support systems within a risk-free simulated environment. For example, we present later a use case whereby LABY was used to compare different ATC information input devices (mouse vs. a stylus with handwriting recognition). While more basic experimental platforms may not provide a high enough level of realism – e.g., controllers may only show certain mental activities in complex, dynamic, and time-pressured situations – microworlds such as LABY allow human-in-the-loop experiments that are considered sufficiently engaging for experimental results to be generalized for actual use (see, e.g., DiFonzo, Hantula & Bordia, 1998). It is often impossible to fully anticipate all options and eventualities when testing a new system, but LABY allows researchers an initial environment to evaluate human interaction with new technological solutions before moving on to more developed simulations. It offers a controlled environment for the iterative process with the flexibility to change dynamics, modify design features, and examine alternatives. Further along the continuum, simulator equipment or actual working set-ups can be used, but these tend to be more complex and expensive, and require the use of expert controllers.

LABY supports the use of a holistic approach, by allowing researchers to examine a range of variables both in isolation and in combination, to gain a comprehensive understanding of the impact of a particular variable on performance as a whole. For example, a gain along one dimension (e.g., accuracy) may be accompanied by a loss in another (e.g., reduced reaction time), and it is important that any support tool does not compromise performance in another domain (e.g., Vachon, Lafond, Vallières, Rousseau, & Tremblay, 2011).

LABY also collects and stores raw data from the eye tracker (Parise, Imbert, Marais, & Alonso, 2012) which can be used to infer the processes underlying performance when evaluating different system designs. LABY allows the online tracking of eye gaze on any object on the screen, whether static or moving, and is not limited to pre-defined areas of interest. Using eye movement data, a process tracing approach could be useful for extracting how operators work with a new system, identifying the optimal sequence of information seeking, requests, decisions and actions using a particular interface configuration (Lafond et al., 2009).

LABY has already been used to examine the effectiveness of different information notification designs in a multitasking environment (Imbert, Hodgetts, Parise, Vachon, & Tremblay, submitted; Parise et al., 2012). In ATC, static red text is used for a low level warning and blinking red text is used for a higher level alert; results showed that participants were no quicker to respond to the animated alert than the static one, and both elicited slower response times compared to the three other prototype designs tested. Furthermore, these two operational alerts occasionally went unnoticed.

In addition to qualitative data such as interviews with controllers, LABY provides objective behavioral measures for evaluation of new technologies. Subjective reports and introspection are fraught with difficulties because often the participant is unaware of subtle differences in his/her performance under certain conditions. Indeed, controllers believed that the short term conflict alert – blinking red text – currently used in operational HMI was sufficient to capture attention, but using objective performance measures in this study we demonstrated incidents of missed alerts and slower reaction times than other notification designs.

Training

LABY can also be used as a training device; either to train novice operators in the skills necessary for air traffic management, or to provide established controllers with practice on a new operating system. In French ATC, the paper flight progress strips are gradually being phased out to eventually be replaced by an entirely electronic environment; as such, this move brings about significant changes in the methods of working and requires operators to learn the new system. To facilitate this transition, LABY has been used as a ‘serious game’ to strengthen expertise at the functional level of the task before progressing to training with the actual operating system. Superior to table-top exercises which cannot recreate the pressures of a real-life situation, this immersive microworld environment is engaging and authentic. Because time is compressed within the serious game, training with different conditions and situations can be accelerated relative to training in the field. As well as being more cost-effective than real world simulations, LABY provides a safe environment in which novices can learn from their errors, and receive feedback during and/or after the task as an intelligent tutor system (e.g., Lafond, Tremblay, DuCharme, St-Louis, & Gagnon, 2012).

Within the context of man-machine interfaces, Rasmussen (1983) distinguishes three levels of human performance: skill-, rule-, and knowledge-based performance. First, LABY can be used to train underlying cognitive non-technical skills in ATC. Computer-based training materials have been created using LABY to bridge the gap between table top and operational training. These materials allow the acquisition of basic skills – such as learning how to handle the interface and interact with the system, how to enter commands, and how to find required information – with the aim of providing a more complete knowledge of the system through interaction with a simplified version. Furthermore, it is possible to configure scripts in LABY for the training of novice controllers and to test specific rule-based performance. For example, operators can be trained in how to manage specific control situations, such as crossing planes of different angles and speeds; the action to be taken in a potential conflict situation; the minimum separation distance between aircraft in different conditions; avoiding obstacles (e.g., simulated hazardous areas such as cumulonimbus clouds, or prohibited areas like parachuting); how to sequence aircraft in a stream; or how to integrate multiple streams into a single stream. Finally, Rasmussen's knowledge level of performance would generally be acquired from first-hand experience in the field, in order to build up schemas and mental models associated with that particular operational system. However, LABY can also assist with integrating operators into new control centers by training them in the specific configurations necessary for that particular airspace, e.g., constraints associated with existing routes, or recurrent situations with the air traffic in that area.

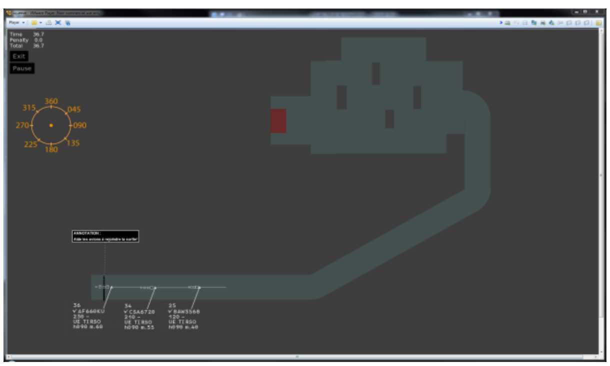


Figure 2. Training a landing sequence.

As a training tool, LABY can be modified to reproduce the design and the main aircraft interaction requirements of the particular operational interface. For training purposes, we have created an organization of routes geared towards ATC problems of workflow management, crossing flight paths, conflict management, and training landing sequences (Figure 2), guided by an operational representation of the tasks. LABY has now been made available in a control center classroom, and sessions are scheduled to follow completion of the theoretical course on learning the new input method but before practice in the operational simulator. Currently, expert controllers oversee the novices across each of the different stages of training. This set-up has worked well, but in the longer term it is envisaged that learners will be able to use LABY in classrooms without supervision. To date, LABY has been used by around 100 controllers in control centers near Brest. Although we have no formal feedback back as yet, comments from trainers indicate that LABY is greatly appreciated as a training device. LABY has also been used in the context of a EUROCONTROL SESAR nextgen system project (WP 4.7.2), to gradually train controllers in the skills necessary for the new electronic flight strip system. The training process comprised three steps. The first stage involved a laboratory-based task, and was designed to familiarize controllers with the interactive pen device by manipulating circles and moving them to a specific area on the interface. At the second stage, LABY was used as a 'stand alone' training tool (the interface was completely different to the operational one). Using LABY it was possible to monitor and evaluate the learning progress of each controller, and provide timely corrective feedback. Finally, the third stage involved a complex simulation with the SESAR human-machine interface which includes a minimum of two controllers at their work stations and one or more pseudo-pilots.

Research

The task of guiding a plane along a course uses general characteristics of a command and control activity (e.g., surveillance, attention, anticipation, dynamic decision making, executing a sequence of actions, cognitive/perceptual load) that are relevant to a number of other domains such as driving, security surveillance, surveillance for cyber attacks, crisis management, and emergency response.

The complex and dynamic nature of microworld tasks, as well as the way in which specific variables can be manipulated and measured, makes LABY a useful platform for a wide range of human factors and applied psychology domains. For example, interruptions, background sound, information overload, psychological, and temporal pressure are all stressors that may impact performance on a supervisory task; using LABY to achieve a greater understanding of those processes operating at a conceptual level could later inform the development of support devices for such conditions.

LABY provides a research platform to better characterize the cognitive processes and limitations to human cognition in multitasking situations by drawing upon a variety of performance measures and process indicators, e.g., decision times/reaction times, performance accuracy, online situation awareness/workload questionnaires and eye movements. Eye tracking data can facilitate a process tracing approach, whereby the processes leading up to a particular decision can be inferred unobtrusively from eye movements (Lafond et al., 2009). LABY can also be used in conjunction with EEG technology to understand more about the brain's electrical activity under particular circumstances. Giraudet, Bérenger, Imbert, Tremblay, and Causse (2014) used EEG in conjunction with LABY to suggest that the P300 amplitude is a valid physiological indicator of vulnerability to inattentional deafness in complex environments. Using the range of measures possible with LABY will help researchers to understand interactive effects between variables, and contribute to conceptual developments in cognitive psychology.

FINDINGS: A USE CASE

The current move away from paper flight progress strips towards an entirely electronic representation of aircraft requires research to determine optimal methods for interacting with the radar screen to input the necessary flight information (e.g., speed, altitude). It has been shown that touch-screen tablet computers can be used for selecting aircraft labels on the radar screen (Alonso, Terrier, Parise, & Cellier, 2010), and a free head-motion eye-tracking system has also been tested as an input device in ATC (Alonso et al., 2013).

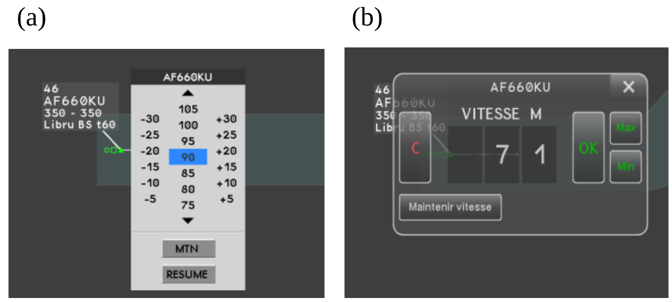
Figure 3. The LABY microworld set-up, including interactive input stylus.

Here, we compared three different input methods for air traffic controllers entering aircraft instructions. The first uses a mouse to click on the precise field on the aircraft label that needs changing (e.g., heading), which then opens a drop-down menu for the value to be selected from a list (Figure 4a). This is the current operational method. The other two methods use an interactive pressure-sensitive stylus (Wacom pen; Figure 3). For one, the pen is used to point on the precise field on the aircraft label; this then opens a window in which the value is written and interpreted by a handwriting recognition system (Figure 4b). For the other, the participant simply points on the lower part of the aircraft label to open a pie-menu from which the precise field can be selected (Figure 4c). The value is then written and interpreted via the recognition system as before. This last condition entails less gestural precision than the previous two, but it does require an additional step in selecting the instruction in the pie-menu.

Method

1. Participants. Twenty-six participants, all subject matter experts (SME) – nine controllers with over 15 years of operational experience and 17 controllers in training at the École Nationale de l'Aviation Civile (ENAC, France) – were familiar with the use of mouse but not with the stylus.

2. Design. We compared three input devices: mouse; stylus with precision pointing on the label; and stylus with pie-menu. In addition, we compared the different types of instruction: heading, cleared flight level (CFL), transfer flight level (TFL), direct clearance to a waypoint, and speed. Input times and errors in input were recorded.



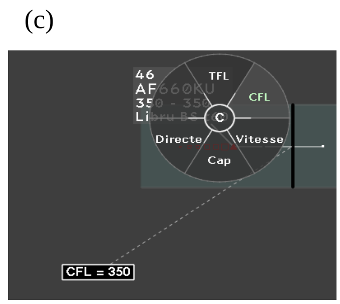


Figure 4. The three input devices (a) mouse and scrollable list (b) stylus label pointing and hand-writing recognition (c) stylus and pie-menu.

3. Procedure. Participants first read standardized instructions informing them of the task requirements. A familiarization phase involved two 10-min exercises to gain familiarity with the stylus. The experimental phase comprised three 20-min scenarios, one in each of the three experimental conditions (counterbalanced). Finally, participants completed a debriefing questionnaire in which they were able to express a subjective preference for the input devices used.

Results

We compared between the three input methods in terms of total time to input control instructions (this included the time to access the menu and to input the necessary value). Input times were further analysed according to the five different types of instruction (Figure 5).

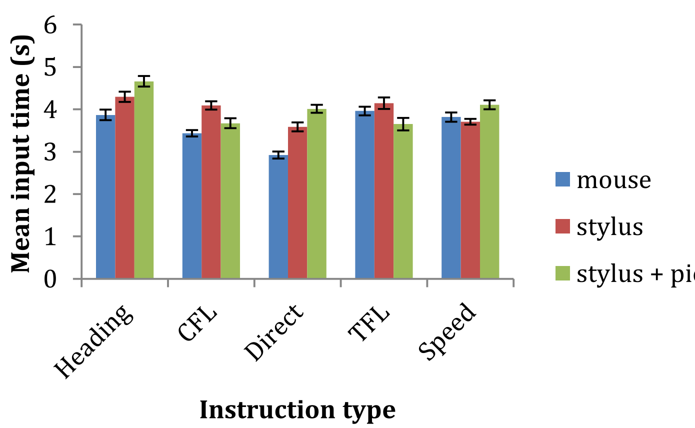


Figure 5. Mean total time to input control instructions according to instruction type.

A 3 (device) × 5 (instruction type) repeated-measures ANOVA revealed a significant main effect of device, $F(2, 46) = 17.33$, $p < .01$, whereby there was no difference between the two stylus input methods, but the mouse was significantly quicker than both. There was a significant effect of instruction, $F(4, 92) = 32.40$, $p < .01$, and also a significant interaction, $F(8, 184) = 21.42$, $p < .01$. Pairwise comparisons showed that the three types of device were each significantly different from each other when entering heading, TFL or direct clearances (the mouse was always faster, because of the time needed to write the numbers in the stylus conditions), but the stylus was comparable to the mouse for entering TFL or speed instructions (when the scrollable lists became more complex).

In fact, the mean time to select the type of instruction was comparable across conditions – mouse (1.33 s), stylus (1.28 s), stylus and pie-menu (1.19 s), despite the extra step required with the pie-menu. The differences reported in Figure 5 relate to the time to input the actual values, which one might expect to be greater with the stylus: mouse (2.37s), stylus (2.82s), stylus and pie-menu (2.91s). Those instruction types for which the stylus was slower (e.g., heading) may be linked to poorer detection of +/- signs by the recognition tool. Familiarity with using the stylus is likely to reduce input times.

Input errors were very low and comparable across conditions ($F < 1$), suggesting that stylus input is promising and does not increase error compared to using the mouse. The stylus and pie-menu are being integrated into a European SESAR project. Although using a stylus takes slightly longer than the mouse, we suggest that the act of writing instructions may bring other benefits such as better memory for information inputted, and therefore improved situation awareness.

DISCUSSION

Microworlds are intended to approximate the dynamic coupling between a task and operator in target situations. Within an ATC environment, LABY provides an intermediate step between lab and field studies by preserving a part of the complexity of an operational setting while having greater control over the studied variables.

One limitation to LABY is that at present it is a stand-alone program that involves no interaction with pilots or other controllers. Verbal communication is an important part of ATC, so it would be interesting to introduce controller dyads (e.g., a radar controller and a planning controller) who must coordinate air traffic together within a single sector; or even just a pseudo pilot. Although the LABY microworld can incorporate many aspects of flight task performance, it should be taken as just one stage on the CSE continuum (Rassmussen et al., 1994). This simplified ATC environment can help with understanding underlying task functions and interactions, but results will then need to be validated in more realistic simulators. Similarly, LABY can help with early stages of controller training, but experience with the actual system is still necessary to learn the specific technical aspects of the job.

LABY is a useful tool for human factors and applied cognitive research since its generic nature and controlled environment enables findings to be generalized to other supervisory and monitoring tasks. LABY is also a platform for system design and has already been used to objectively evaluate input methods and notification designs in terms of usability. It facilitates the transition to a more complex testing simulation, avoiding implementing experimental protocols that are too complex, allowing difficulties/issues arising from the prototypes to be modified at an early stage, and making short iteration cycles possible. LABY is emerging as a high fidelity but simple training environment to introduce student controllers to the initial notions of control, and to familiarize more established controllers with the processes required in future operational systems without the need to use a complex simulation platform.

REFERENCES

Alonso, R., Causse, M., Vachon, F., Parise, R., Dehais, F., & Terrier, P. (2013). Evaluation of head-free eye tracking as an input device for air traffic control. *Ergonomics*, *56*, 246-255.

Alonso, R., Terrier, P., Parise, R., & Cellier, J.M. (2010). Effet d'une de corrélation visuo-motrice sur le mouvement de pointage. *Le Travail Humain*, *72*, 245–66.

Brehmer, B. (1992). Dynamic decision making: Human control of complex systems. *Acta Psychologica*, *81*, 211-241.

Brehmer, B., & Dörner, D. (1993). Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of the field study. *Computers in Human Behavior*, *9*, 171–184.

DiFonzo, N., Hantula, D., & Bordia, P. (1998). Microworlds for experimental research: Having your (control and collection) cake, and realism too. *Behavior Research Methods, Instruments, and Computers*, *30*, 278-286.

Gagnon, J., Couderc, F., Rivest, M., & Tremblay, S. (2011). Capturing the Task Model of Experts in Emergency Response using SYnRGY. *Proceedings of the 9th International ISCRAM Conference, Vancouver.*

Giraudet, L., Bérenger, M., Imbert, J-P., Tremblay, S., & Causse, M. (2014). Inattentional Deafness in Simulated Air Traffic Control Tasks: A Behavioral and P300 Analysis. *Proceedings of the 5th International AHFE Conference, Krakow, Poland.*

Gonzalez, C., Vanyukov, P., & Martin, M. K. (2005). The use of microworlds to study dynamic decision making. *Computers in Human Behavior*, *21*, 273–286.

Gray, W. D. (2002). Simulated task environments: The role of high-fidelity simulations, scaled worlds, synthetic environments, and laboratory tasks in basic and applied cognitive research. *Cognitive Science Quarterly*, *2*, 205–227.

Hodgetts, H.M., Vachon, F., & Tremblay, S. (2014). Background sound impairs interruption recovery in dynamic tasks: Procedural conflict? *Applied Cognitive Psychology*, *28*, 10 – 21.

Imbert, J.-P, Hodgetts, H. M., Parise, R., Vachon, F., & Tremblay, S. (submitted). Attentional costs and failures in air traffic control notifications. *Ergonomics.*

Lafond, D., Champagne, J., Hervet, G., Gagnon, J., Tremblay, S., & Rousseau, R. (2009). Decision Analysis Using Policy Capturing and Process Tracing Techniques in a Simulated Naval Air-Defense Task. *Proceedings of the 53rd HFES Annual Meeting* (pp. 1220-1224).

Lafond, D., Tremblay, S., DuCharme, M. B., St-Louis, M. E., & Gagnon, J. F. (2012). Assessment and Refinement of an Intelligent Tutor for Complex Decision Making. *Proceedings in 34th Annual Conference of the Cognitive Science Society*. Sapporo, Japan: Cognitive Science Society.

Parise, R., Imbert, J-P., Marais, F., & Alonso, R. (2012). LABY : Un support d'aide à l'évaluation de choix conception d'IHM pour le contrôle aérien. *Ergo'IHM Proceedings.*

Rasmussen, J. (1983). Skills, Rules, and Knowledge; Signals, Signs, and Symbols, and Other Distinctions in Human Performance Models. *IEEE Transactions on Systems, Man, and Cybernetics*, *13*, 257-266.

Rasmussen, J., Pejtersen, A. M., & Goodstein, L. P. (1994). *Cognitive Systems Engineering*. Wiley.

Tremblay, S., Vachon, F., Lafond, D., & Kramer, C. (2012). Dealing With Task Interruptions in Complex Dynamic Environments Are Two Heads Better Than One? *Human Factors*, *54*, 70-83.

Vachon, F., Lafond, D., Vallières, B. R., Rousseau, R., & Tremblay, S. (2011). Supporting situation awareness: A tradeoff between benefits and overhead. In *Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), 2011 IEEE First International Multi-Disciplinary Conference on* (pp. 284-291).