Human factors in ATC alarms and notifications design: an experimental evaluation

Sylvie Athènes, Stéphane Chatty, Alexandre Bustico
Centre d'Etudes de la Navigation Aérienne
7, avenue Edouard Belin
31055 Toulouse cedex, France
{athenes, chatty, bustico}@cena.fr

ABSTRACT

With the growing use of computerised working position, alarm design on air traffic control displays is a concern as events to be notified increase in number and diversity. Safety requires visual notifications that can be efficiently detected and understood. It also requires that no information on the radar display is obscured by the visual notifications. We also need to design hierarchies of notifications, from most severe to benign. Taking advantage of the current graphical capabilities of computers, we have identified and explored various dimensions in visual alarm design. We present in this paper an experimental evaluation, in terms of detection time and precision, of several of those dimensions: opacity, size, temporal profile of animation and signal frequency. From our results, we conclude that opacity, size and temporal profile of animation are well suited to introduce some nuances in the way we convey notifications on visual displays. We also show that detection of a given signal will measurably vary according to the amount of attention available at that time.

INTRODUCTION

When designing user interfaces for command and control systems such as air traffic control workstations, the safety of the device itself and the adequacy and efficiency of users' actions are not the only critical issues for safety. The way information can be perceived and interpreted by users is also important to safety, especially when it comes to notifying alarms or other asynchronous events. With aircraft flying one nautical mile in ten seconds, one does not want alarms to be noticed in a matter of minutes, but rather in a matter of seconds. And in any case, one does not want alarms to go unnoticed.

Over the last years, the increasing use of computing systems has led to two major evolutions in the design of user interfaces for air traffic control systems: more information to be conveyed, including alarms, and more sophisticated displays.

On the one hand, as the underlying systems are growing more complex, new notifications have to be conveyed to users. There are new kinds of alarms associated to new types of breakdowns or problem detection, such as a ground collision risk detection. But there also are less serious events that nonetheless need to be notified, such as asynchronous communications or changes in the status of aircraft. For instance, incoming requests from aircraft through an air-ground data link have to be responded to at some point, and thus need an appropriate notification. This is new for interface designers, who so far had only to provide for a handful of severe alarms and could rely on blinking labels. A wider range of signals is now required.

On the other hand, the evolution of computing systems has provided designers with an extremely wide range of visual techniques and allowed the design of visually richer displays. Between 1985 and today, air traffic control engineers have explored many ways of displaying new pieces of information on graphical radar screens: multi-windowing, coloured backgrounds for geographical and meteorological information, enriched flight labels with colour coding, etc. Avoiding visual clutter has become very difficult. If no real visual design is exercised, adding the necessary range of information (see above) to such a display can lead to disastrous results. Displayed messages may obscure other pieces of information, or may go unnoticed, or both. For instance, at the CENA we have had examples of realistic simulations where controllers focusing on an area of their large screen did not notice an alarm based on colour change for as long as 20 minutes!

In order to ensure alarm perception, designers tend to add sounds to notify information. Sound is indeed an efficient medium, but it cannot be used in every contexts. Air traffic controllers are usually reluctant to use sound, and prefer to keep it for emergency alarms. Therefore, other means have to be found for lower priority notifications. We believe that the graphical capabilities of computers can be used to design such solutions.

We have looked for potential solutions in other domains. The following work put together findings from two domains: in the neurophysiology of perception, it has been known for a long time that peripheral vision is very sensitive to movement whereas central vision is very accurate. Therefore, in order to be really efficient, signals that do not appear exactly where a person is looking should be animated. In the art domain, sets of rules have been devised by painters and graphic designers, in order to influence the way the eye analyses an image. Modern computers provide a significant subset of the visual techniques that are available to painters or cartoon authors. But surprisingly, very few of those capabilities have been used in air traffic control: in contrast, video games for instance make a far more extensive use of them. Among these visual techniques, most interesting is the possibility to design animations that efficiently raises attention without obscuring background information. The goal of our research is to identify salient design dimensions and to explore the mechanism of animation efficiency. This should allow us to provide engineers with hierarchies of visual signals to be mapped on hierarchies of alarms. In this article, we identify some dimensions, and report an experiment that provides preliminary results on their compared salience.

NOTIFYING INFORMATION

Basically, a good alarm should do two things: raise attention and provide some information about the situation it is signalling (location, nature, degree of urgency, what kind of action is required, to name but a few). These two aspects of alarm design have been mostly treated separately.

On the one hand, there are classical results about visual alarm design that can efficiently raise attention (for a review, see [1]). However, a vast majority of these results is based on light displays and the parameters tested pertain mostly to colour, shape or location. For example, the guidelines in “Human Factors in the Design and Evaluation of ATC Systems”[2] specify colour (red for warning/danger, yellow for caution and green for normal/ready status), location (the central 15 degrees of the area where the controller normally looks) and blinking (no more than 2 levels, rate between 2 and 3 Hz).

On the other hand, HMI researchers have stepped further away from the characteristics of visual perception *per se* and focused on the cognitive mechanisms of alarm awareness [3, 9]. Wickens [10], for example, defined a set of basic principles allowing for a better readability of the signals and their relationship to the situation at hand. Particularly efficient are the following principles: sequencing, where it should be possible to recover the order in which a series of signals progressed; grouping and prioritising, where the spatial organisation of the signals on the display should be closely related to the functional organisation of the system; colour, where the conventional meaning of colours should be congruent with their use (e.g. red for danger); informativeness, where each signal should be designed to provide higher-level combined information.

Lately, research work in the so-called ecological perspective has been advocating for a more integrated representation of information. In this approach, the fundamental unit of analysis is the human-environment system (or the human-machine system, as the case may be). Typically, these studies question the relevance and format of the information presented to the user [4]. Their designs attempt to exploit the large capacity of the human perceptuo-motor system as it is found in natural environments by supplying a complex, but transparent information environment. A good example is provided by head-up displays for aircraft pilots where design has focused on representation of those features of the optical array that support perception of speed, height, and flight direction [6].

DESIGNING ATTENTION GETTERS

Resting on accumulated results from traditional research, user interface design guides tend to be conservative and address mostly the issue of colour, symbol, font, location and blinking rate. We believe that it is time to go beyond a mere transposition of classical results. Today's graphical screens allow production of complex graphical effects that surpass the capacities of light signals. Their endless variations allow the design of signals for a whole range of events of different priorities. As the number of dialogues with the machine increases, it becomes important that the user’s attention not be grabbed all the time, but rather attracted discerningly. Furthermore, as the amount of information to be conveyed increases, it becomes crucial to keep fluidity of work and to avoid upstaging of the main task [5].

We are thus exploring a series of graphical attributes or effects that could support a large range of signals with distinctive attention "grabbing" capabilities. Among the possible attributes are classical parameters like the shape, size and dominant colour of the graphical signal. As we are looking for signals that do not clutter the display or obscure other pieces of information, transparency as provided by modern graphical systems is also an interesting dimension. The transparency level of the visual signal can be used as well as the transparency gradient that determines its contrast with the background. Finally, animation provides a whole set of dimensions. The type of animation is probably meaningful though difficult to characterise: pulsating objects, moving objects, etc. Then for periodic animations as for any other periodic signals, classical variables are signal frequency and profile, as illustrated in figure 1.

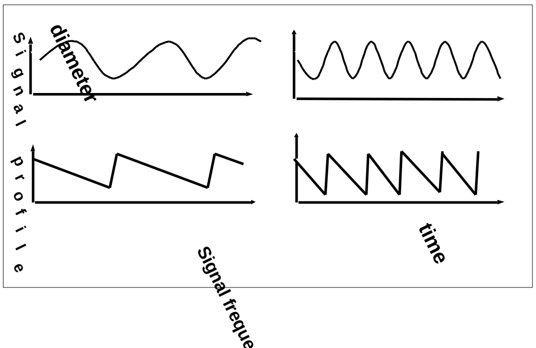

Figure 1. Variations in the way the signal evolves over time. Top to bottom, variation on the temporal profile. Left to right, variation on frequency.

The present paper reports on a subset of the above-mentioned parameters. Because both shape and colour have been extensively explored, we decided to keep these two parameters as neutral (or all-encompassing) as possible and to use only one shape (a circle) and one colour (yellow). We chose a circle because of its relative symbolic neutrality (as opposed to the "caution" meaning of the triangle in some cultures for example) and the colour yellow because it is the colour with the largest perception field. We manipulated the transparency and the animation of the signals. Among the animation parameters, we chose frequency and temporal profile. We used pulsating signals, with two variations: "local" signals were small pulsating disks, whereas "non-local" or "global" signals were ripples that encompassed the whole screen, with the goal of catching the eye wherever it rests.

EXPERIMENT

We have carried out an experiment in order to assess the influence of the signals we have listed. The experiment consisted in a series of random signals that subjects had to detect and designate on a touchscreen (see Figure 2). As an indicator of a signal capacity to raise the attention, we collected the reaction time to a signal (i.e. the time needed for subjects to react to a signal while attending the Stroop task). We also computed the pointing accuracy (i.e. the distance in cm between the signal and the point where the subject's finger landed on the touchscreen) in order to assess the accuracy with which a given signal would be perceived.

Figure 2. Experimental setup.

Design and procedure

Six right-handed subjects took part in the experiment. They were seated in front of a 21" screen equipped with a touch input layer, as shown on Figure 2. The eyes-screen distance was 60 to 70cm (24 to 28 in.), depending on the length of the subject's arm.

To evaluate the capacity of the computed signals to raise attention, we used a dual-task paradigm: signals appeared randomly at the periphery of the touchscreen, while the subjects attended a Stroop Colour-Word test displayed in the screen central part (see Figure 3). This test, which has now been in existence for over 60 years, yields highly reliable and stable measures of individuals' performance. In this experiment, the underlying assumption is that the more difficult the task, the more attention it requires. Using the Stroop test provided us with a tool to grade the amount of attention required from the subjects and a wealth of published results to check the reliability of our data (for a review, see [6, 7]).

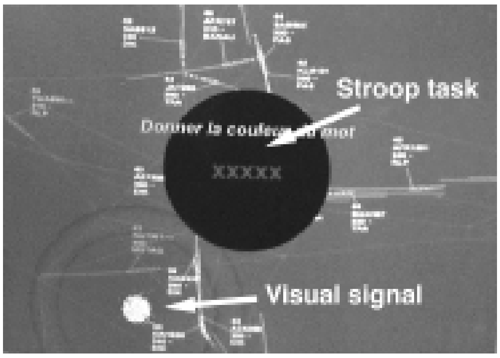

Figure 3. The experimental display: the Stroop task in the middle, and visual signals at the periphery. Here, the signal is a non-local stylised ripple converging toward a small disk.

Detection task. The signals were pulsing yellow circles whose diameters increased and decreased between 10 and 25mm. In addition, they were varied systematically along 4 dimensions: transparency (18, 30 and 50% of opacity), signal

Stroop task. The Stroop colour-word test consisted of words or non-words displayed in black or in colour in the central part of the touch-screen. The test entailed three levels of difficulty, from the easiest to the most difficult: subjects were instructed to, either read aloud colour names printed in black (e.g., the word 'red' printed in black: correct answer is red), or name the ink colour of non-words (e.g.: 'xxx' printed in green: correct answer is green), or name the ink colour of colour names while ignoring the word (e.g., the word 'red' printed in green: correct answer is green).

Subjects were instructed to keep their attention focused at all time on the Stroop test and to perform that task as quickly as possible while keeping a minimal error rate.

The experiment was divided into three one-hour sessions (one per day over three days). During each session, corresponding to one of the three conditions of the Stroop task, subjects were presented with ten repetitions of each signal configuration (i.e. 10x36), randomised over the session.

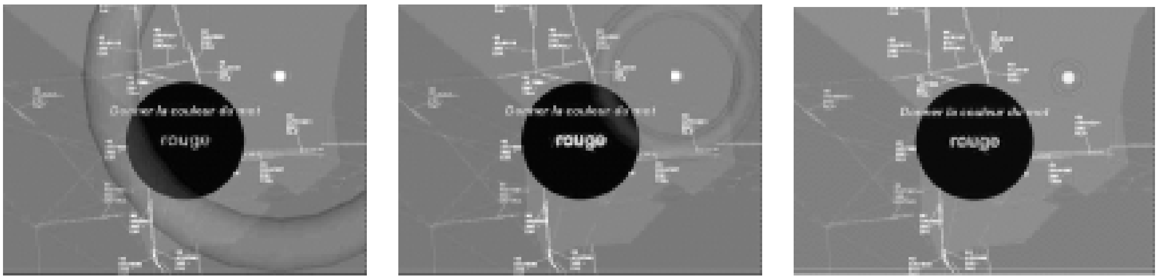

Figure 4. A non-local animation: a ripple crosses the whole screen and converges toward the point of interest.

frequency (0.75, 1.50 and 2.50 Hz), temporal profile (wave or step, see Figure 1), and the local/global aspect of the signal display (the yellow circle, alone or surrounded by animated rings/waves (see Figures 3 and 4). There were 10 repetitions for each of the 36 signal configurations.

In the course of a session, each signal could appear anywhere on the screen with the exception of the central part dedicated to the Stroop task. Subjects were instructed to point towards the signals as soon as they were detected. However, they were forbidden to visually explore the screen. Reaction time and pointing accuracy on the touch-screen were measured.

RESULTS AND DISCUSSION

We shall report results, first pertaining to reaction time (i.e. the time needed for subjects to react to a signal while attending the Stroop task) and, second, pertaining to pointing accuracy (i.e. the distance in cm between the signal and the point where the subject's finger landed on the touchscreen).

For the reaction time, we performed an ANOVA with a within-subject design.

Firstly, as expected, attention load (as influenced by the level of difficulty of the Stroop task) does indeed have an effect on the reaction time to a signal ($F = 5.74$; $p < .02$): reaction times to identical signals increase with the level of difficulty of the Stroop task. We thus confirm that the amount of attention required by a given level of difficulty of the Stroop task is indeed reflected in reaction time to a secondary task.

The ANOVA shows a significant main effect for subjects ($F = 29.51$; $p < .0001$): reaction time means range from 470ms for the fastest subject to 639ms for the slowest subject. We also observe a significant interaction between subjects and level of difficulty of the Stroop task ($F = 7.20$; $p < .0001$). However, this is due to a learning effect for the signal detection task: regardless of the level of difficulty of the concurring Stroop test, signal detection during the first session yielded the longest reaction times. It is worth noting that this result has no consequence on the preceding main effects as the experimental conditions have been balanced across subjects.

Additionally, we observe a significant main effect for three signal parameters:

- transparency ($F = 97.36$; $p < .0001$): more transparent signals yield longer reaction times;
- temporal shape ($F = 6.59$; $p < .02$): slow increase of the signal surface, as opposed to a step-like increase, entails a longer reaction time;
- local/global aspect ($F = 51.94$; $p < .0001$): local signals yield longer reaction times.

Some interaction effects allow for a richer description of the observed behaviour.

- global signals decrease reaction time, but only for the very transparent signal ($F = 24.55$; $p < .0001$).
- step-like increase of the signal surface decreases reaction time, but the difference is most dramatic (125ms) for the very transparent signal ($F = 3.26$; $p < .05$).

In order to find out if the position of the signal on the screen influences reaction time, we performed correlations between reaction time and, either the x and y coordinates, or the polar coordinates, of the signal. We found no significant results, the value of the correlations ranging from 0.031 to 0.135. In our experiment, signal position on the screen does not influence reaction time.

Lastly, we performed a within-subject design ANOVA with pointing accuracy as dependent variable. We found a main subject effect ($F = 55.09$; $p < .0001$) reflecting the fact that some subjects were more accurate than others (mean accuracy ranges from 10mm to 16mm). There are additional main effects: pointing is more accurate when the concurring Stroop test condition has the highest level of difficulty ($F = 20.77$; $p < .0001$); more transparent signal leads to less accurate pointing ($F = 5.44$; $p < .01$); lower signal frequency entails lower pointing accuracy ($F = 5.91$; $p < .01$); pointing is less accurate for global signals ($F = 16.23$; $p < .0001$). Even though the above effects are significant, the magnitude of the differences is very small, about 1mm. It would be far fetched to argue that such a difference could be of interest at the scale of an air traffic control position.

To sum up, we showed that perception of signals is measurably influenced by the amount of attention available at any given time. This result should caution us when designing signals for air traffic controllers whose workload, and therefore attentional load, can vary widely over time. We also found that, among our variables, three are worth manipulating: the temporal shape of the animation, the transparency, and the local/global aspect of the signal. The differences in reaction time depend on their combination. Considering that the nature of the task was to expect and react quickly to signals, observed differences between signal configurations are likely to be higher in a real life situation.

Transparency is the trickiest, but the most promising of the parameters: although one should probably aim for opacity levels below 20% and be careful to fine-tune the levels appropriate to the required effects, the interaction of this parameter with most of the other parameters makes it a good candidate for notifications and suchlike low priority information. The “enhancement effect” of the transparency parameter upon the other tested parameters hints towards specific usage where transparency could be a mode to be chosen as opposed to a set parameter. This feature could be useful when increasing workloads lead controllers to mentally postpone or tune out a subset of events. Preliminary results on the error rate indicate that the most transparent signals are more frequently missed. Therefore, such signals can be appropriate for notifying events of very minor urgency, because they are to be noticed only when the user is not immersed in resolving a conflict and scans the screen.

CONCLUSION

This paper is a first step toward the elaboration of design principles based on the use of graphical transparency and animation. This first experiment allowed us to determine which dimensions, within the explored set of parameters, had the most important effect on visual detection time. It would be very interesting to test if our findings hold when the attentional field is not restricted to a 21" screen. Indeed, air traffic controllers spread their attention across a much wider visual field, comprising minimally a radar screen, a strip board and various small screens and/or button displays. Another question is raised about the transposability of our results when several signals appear at the same time. Do we observe a hierarchy of detection following the distribution of the reaction times observed in our study? We are at the present conducting an experiment designed to confirm our findings using a different paradigm in a manner akin to some video games. This will unable us to propose hierarchies of visual signals rated in terms of detection time and probability. Further work also includes the introduction of sound, first as a way to shorten even further the signal detection times, and then as a mean to reinforce or enrich the effect of visual signals through multimodal design.

REFERENCES

1. Boff, K.R. & Lincoln, J.E. (Eds.): Engineering Data Compendium, Human Perception and Performance. Harry G. Armstrong Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio, 1988.
2. Cardosi, K.M. & Murphy, E.D. (Eds.): Human Factors in the Design and Evaluation of Air Traffic Control Systems. Federal Aviation Administration, Final Report, April 1995.
3. Edworthy, J. & Adams, A.: Warning Design: A Research Prospective. Taylor & Francis, 1996.
4. Flach, J., Hancock, P., Caird, J. & Vicente, K.J. (Eds.): Global Perspectives on the Ecology of Human-Machine Systems. Lawrence Erlbaum Associates, Hillsdale, 1995.
5. Harrison, B.L. & Vicente, K.J.: An experimental evaluation of transparent menu usage. Proc. of CHI'96, Vancouver, Canada..
6. Jensen, A.R. & Rohwer, W.D.: The Stroop Color-Word test: a review. Acta Psychologica (25), 36-93, 1966.
7. MacLeod, C.M.: Half a century of research on the Stroop effect: an integrative review. Psychological Bulletin, Vol. 109, No.2, 163-203, 1991.
8. Rasmussen, J. & Pejtersen, A.M.:Virtual ecology of work. In "Global Perspectives on the Ecology of Human-Machine Systems", Flach, J., Hancock, P., Caird, J. & Vicente, K. (Eds.), Lawrence Erlbaum Associates, Hillsdale, 1995.
9. Stanton, N. (Ed.): Human Factors in Alarm Design. Taylor & Francis, 1994.
10. Wickens, C.D.: Engineering Psychology and Human Performance, HarperCollins, 1992.