The Neuroergonomic Evaluation of Human Machine Interface Design in Air Traffic Control using behavioral and EEG/ERP measures

Louise Giraudet, Jean-Paul Imbert, M Bérenger, Sebastien Tremblay, M Causse

The Neuroergonomic Evaluation of Human Machine Interface Design in Air Traffic Control using behavioral and EGG/ERP measures

ARTICLE *in* BEHAVIOURAL BRAIN RESEARCH · JULY 2015

Impact Factor: 3.39 · DOI: 10.1016/j.bbr.2015.07.041 · Source: PubMed

5 AUTHORS, INCLUDING:

Louise Giraudet
Institut Supérieur de l'Aéronautique et de l'E…
2 PUBLICATIONS **0** CITATIONS
SEE PROFILE

Jean-Paul Imbert
Ecole Nationale de l'Aviation Civile
13 PUBLICATIONS **8** CITATIONS
SEE PROFILE

Sébastien Tremblay
Laval University
131 PUBLICATIONS **1,301** CITATIONS
SEE PROFILE

Mickael Causse
Institut Supérieur de l'Aéronautique et de l'E…
34 PUBLICATIONS **122** CITATIONS
SEE PROFILE

Title: The Neuroergonomic Evaluation of Human Machine Interface Design in Air Traffic Control using behavioral and EGG/ERP measures

Author: Giraudet L. Imbert J-P. Bérenger M. Tremblay S. Causse M.

PII: S0166-4328(15)30112-1
DOI: http://dx.doi.org/doi:10.1016/j.bbr.2015.07.041
Reference: BBR 9729

To appear in: *Behavioural Brain Research*

Received date: 5-2-2015
Revised date: 15-7-2015
Accepted date: 16-7-2015

Please cite this article as: Giraudet L, Imbert J-P, Bérenger M, Tremblay S, Causse M.The Neuroergonomic Evaluation of Human Machine Interface Design in Air Traffic Control using behavioral and EGG/ERP measures.*Behavioural Brain Research* http://dx.doi.org/10.1016/j.bbr.2015.07.041

The Neuroergonomic Evaluation of Human Machine Interface Design in Air Traffic Control using behavioral and EGG/ERP measures

Giraudet, L.[1], Imbert, J-P.[2], Bérenger, M.[1], Tremblay, S.[3], Causse, M.[1,3]

[1]Institut Supérieur de l'Aéronautique et de l'Espace, 10 avenue Edouard Belin, 31055 Toulouse, France

[2]Laboratoire d'Informatique Interactive, Ecole Nationale de l'Aviation Civile, Université de Toulouse, 31055 Toulouse, FRANCE

[3]School of Psychology, Université Laval, Québec, QC G1V 0A6, CANADA

Corresponding author : GIRAUDET Louise, ISAE service DCAS, 10 avenue Edouard Belin, 31055 Toulouse Cedex 4, France
e-mail : louise.giraudet@isae.fr
Tel : 0033 683947003

Highlights

- We propose a neuroergonomic approach to evaluate notification designs
- Participants performed an Air Traffic Control task with two different visual designs
- The more salient visual design globally enhanced the performance to the task
- Cerebral response to auditory alarms was enhanced thanks to the salient design
- Results have implications in the evaluation of human machine interface design

Abstract

The Air Traffic Control (ATC) environment is complex and safety-critical. Whilst exchanging information with pilots, controllers must also be alert to visual notifications displayed on the radar screen (e.g. warning which indicates a loss of minimum separation between aircraft). Under the assumption that attentional resources are shared between vision and hearing, the visual interface design may also impact the ability to process these auditory stimuli. Using a simulated ATC task, we compared the behavioral and neural responses to two different visual notification designs - the operational alarm that involves blinking

colored “ALRT” displayed around the label of the notified plane (“Color-Blink”), and the more salient alarm involving the same blinking text plus four moving yellow chevrons (“Box-Animation”). Participants performed a concurrent auditory task with the requirement to react to rare pitch tones. P300 from the occurrence of the tones was taken as an indicator of remaining attentional resources. Participants who were presented with the more salient visual design showed better accuracy than the group with the suboptimal operational design. On a physiological level, auditory P300 amplitude in the former group was greater than that observed in the latter group. One potential explanation is that the enhanced visual design freed up attentional resources which, in turn, improved the cerebral processing of the auditory stimuli. These results suggest that P300 amplitude can be used as a valid estimation of the efficiency of interface designs, and of cognitive load more generally.

Keywords
Air Traffic Control, attentional resources, ERP, Human Machine Interface evaluation, Neuroergonomics

1. Introduction

Within safety-critical, continuously-evolving, and visually-rich environments such as air traffic control, supervisory control of emergency response, and security surveillance, operators must deal with dynamic and cognitively demanding tasks whilst confronted with temporal pressure, stress, and high-risk decision-making situations. In the case of Air Traffic Control (ATC), the main task is to guide aircraft through controlled airspace with the safety requirements of maintaining a minimal distance and an altitude of separation between them while optimizing their trajectories. Each controller is responsible for an airspace volume that is represented on a radar visualization system where numerous aircraft positions are displayed. They also must be vigilant and responsive to the occurrence of various on-screen visual notifications triggered by safety nets. In the present study, within a simulated-ATC task, we used one key safety-critical visual notification that serves to indicate an impending loss of separation between aircraft.

The auditory channel is also essential for ATC as controllers also need to exchange information with pilots and other controllers through radio and phone communications. Auditory warnings such as ground collision avoidance alerts or area infringement warnings have been increasingly integrated into ATC workstations. This recent introduction of auditory alerts raises new human factors issues, as several theories have indicated that a high cognitive load context can lead to a neglect of auditory alerts. One could argue that the high perceptual and cognitive load typical of ATC operations may consume a large proportion of attentional resources – especially when sub-optimal visual designs are used – which in turn can reduce the available attentional capacity for processing the task at hand, as well as for additional

unexpected events. Indeed, according to perceptual load theory [1-3], tasks involving high perceptual load can consume most of attentional capacity, leaving little remaining for processing information that is not directly related to the focal task, such as unexpected alarms [4-7]. In this sense, several researches have shown that attentional resources are shared between vision and hearing [8-11]. Some authors also postulate that tasks with high cognitive load (e.g., load in working memory) can lead to a reduced openness to additional stimuli such as auditory distractors [12-14]. In line with these theories, we suggest that introducing efficient and salient visual designs that can reduce the perceptual and cognitive load is important not only to improve performance of the ATC task itself, but to also help preserve attentional resources that may potentially be required by other information channels.

Several studies have demonstrated that salient stimuli promote fast and effortless processing of information (see [15] for review). This automatic and preattentive process has been explained by salience map models; two-dimensional maps that encode locations to be processed in priority according to their salience. This is supported by recent work concerning the brain structures that might contain such salience maps [16]. Nardo et al. [17] showed the efficacy of a bottom-up signal for the orienting of spatial attention in a complex and dynamic environment. By using a more salient visual design for the critical visual notifications occurring in ATC, the allocation of visual spatial attention should be directed foremost toward those stimuli, sparing controllers a costly visual search in terms of attentional resources.

Concerning the evaluation of cognitive load, the use of the oddball paradigm together with event-related brain potentials (ERP) has been proposed as a valid cognitive load index in various realistic tasks such as simulated flight missions [18, 19], gauge monitoring [20] or video games [21]. However, to the best of our knowledge, very few authors have explicitly used such paradigms to measure the cognitive load elicited by various human machine interface (HMI) designs. P300, usually measured between 300 and 500 ms post-stimulus [22] is one of the most commonly studied ERPs and is known to be observed during oddball paradigms. In this paradigm, participants are instructed to detect targets among non-targets (series of standard to-be-ignored stimuli; see [22]). The oddball paradigm is a well-known example that incorporates cognitive and attentional processes for stimulus recognition and attention allocation [23]. When attentional focus deviates from the target detection task (e.g., in a dual task paradigm), the P300 amplitude decreases significantly [12, 24, 25]. P300 is also modulated by the load of the concurrent task as increases in memory load reduce P300 component size because task processing demands increase [26, 27]. Importantly, it is generally accepted that a distinction can be made between two subcomponents of the P300, the P3a and the P3b. The P3a seems to be more specifically related to the novelty of deviant auditory stimuli [28], independently of task-relevance. It has a shorter latency, a fronto-central scalp distribution and its generation involves the frontal lobe and the hippocampus. The P3a amplitude decreases with repetition and habituates rapidly. It is sensitive to variations in top-down monitoring by

frontal attention mechanisms engaged to evaluate incoming stimuli and is related to the orienting response [22]. In contrast, the P3b potential, partially generated in the medial temporal lobe, has a more posterior-parietal scalp distribution, a somewhat longer latency and is less sensitive to habituation, than P3a. Several studies also suggests that the locus coeruleus-norepinephrine (LC-NE) system underlies P3b generation for a target detection task [29], which is consonant with attentional resource allocation and arousal-related effects in humans. The P3b has been thought to reflect such processes as memory access, memory storage and response initiation that are evoked by the evaluation of stimuli in tasks that require some form of action like a covert or overt response. In summary, P3a is produced in response to the processing of sensory stimuli with frontal lobe activation from attention-driven working memory changes; conversely, P3b is produced as a result of temporal/parietal lobe activation from memory and context updating operations and subsequent memory storage. In this paper, the term P300 will be used to refer to P3b, as our oddball task was task-relevant and required an open response. The high cognitive load involved in ATC should solicit the temporal lobe for sensory processing and memory operations, therefore affecting those functions and limiting auditory target processing.

Our study is based on a neuroergonomic approach [30-33] which merges knowledge and methods from cognitive psychology, system engineering, and neurosciences. This approach aims to improve the system safety and efficiency at the workplace by considering human brain functioning. We used an ATC-like synthetic environment called Laby [34] which simulates key features of a dynamic visual monitoring radar task. Participants had to acknowledge notifications displayed close to aircraft located in peripheral vision, which simulated a collision avoidance alarm. Two notification designs have previously been shown to elicit a difference in performance in this environment [34]. Box-Animation (BA), a very salient visual notification, with brackets pulsing around the notified aircraft, is extremely well detected by the controllers. On the contrary, the Color-Blink (CB) notification — similar to the classical operational design of the critical notification indicating a loss of minimum separation between aircraft — is a much less salient design that causes a lower detection rate. The Box-Animation design is very noticeable and does not require a sustained visual search to be perceived; on the other hand, the Color-Blink notifications can sometimes go unnoticed if the controller is not actively monitoring the radar screen.

2. Objectives and hypotheses

Two groups of participants were recruited. One group performed the ATC task with Box-Animation and the other with Color-Blink notifications. To further improve the level of realism, each participant performed the task according to two levels of cognitive load (tempo, i.e. the number of events per unit of time) with various numbers of aircraft in the visual scene (between 5 and 21). Simultaneously with the ATC task, participants were asked to respond to the occurrence of low probability tones and to ignore high

probability tones. P300 auditory-evoked potentials were recorded from the occurrence of the tones both in parallel with the ATC task and in two control conditions (tones alone without the ATC task), as an indicator of remaining attentional resources. Measuring P300 amplitude variations will indicate if the variations in HMI design affected attentional processes and response initiation.

We predicted that the introduction of the ATC task would reduce ERPs amplitude to the rare target tones in comparison to the baseline condition, in which the ATC task is not administered. This might demonstrate a reduced availability of the attentional resources for processing the auditory stimuli. According to the initial study comparing the two notification types [34], we also hypothesized that the ATC task would consume fewer attentional resources when performed with Box-Animation compared to Color-Blink notifications. Consequently a lower subjective mental load, a better detection rate and higher ERPs amplitude should be observed with Box-Animation than Color-Blink notifications.

3. Method

3.1. Participants

42 volunteers, all students of Université Laval between 19 and 46 years old, were recruited for this study. None had a history of neurological disease, psychiatric disturbance, substance abuse, or took psychoactive medications. They all received full information on the experiment protocol, signed an informed consent and received compensation for their participation in the study. All participants had a STAI Y-B score below 55 (average anxiety). Six participants were removed from the analysis due to a lack of compliance with instructions and/or data acquisition problems. The 36 remaining participants (M = 24.1 years, SD = 5.8) were divided into two groups of 18. The first group was associated with the classic ATC type of visual notifications called Color-Blink, and the second group with the newly-developed type of notification called Box-Animation.

3.2. The Laby microworld and the auditory oddball task

3.2.1. The ATC Task

The Laby microworld is a functional simulation of ATC, built on the main task of guiding an aircraft around a route shown on the center part of the screen (Figure 1). Participants had to regularly modify the flight path and altitude of an aircraft using drop-down menus. The instructions were given via a pop-up window close to the aircraft (cf. Figure 2).

Figure 1. Screenshot of the Laby microworld simulation. On the top, an example with 5 static peripheral aircraft positioned around the corridor. Below, an example with 21 peripheral aircraft. The radar labels of the

peripheral aircraft are always displayed. In both images, the main aircraft navigates through the corridor. An altitude instruction is displayed on its left (the radar label of the main aircraft appears only in this case).

Figure 2. Zoom on the Laby interface. Participants had to select the altitude of the central aircraft according to the instruction given on the black window above the aircraft.

In addition to the central aircraft, participants had to monitor a set of static aircraft located around the main aircraft corridor (Figure 1). Visual notifications were displayed in or around the radar label located in the vicinity of these peripheral aircraft, either the Color-Blink type for the first group, or the Box-Animation type for the second group (Figure 3). Color-Blink is colored text with the word “ALRT” which blinks at a rate of 800 ms on/200 ms off (see Figure 3, 1a-1b). It is used in ATC operational radar visualization for high-priority short-term conflict alerts. Box-Animation involves the same colored text “ALRT” but also four yellow chevrons placed around the label of the notified plane (Figure 3, 2a-2b). These chevrons move outwards from the label by 60 pixels following a slow in/slow out animation cycle of 1 Hz. It corresponds to a radar display prototype being used in a previous study [35].

Figure 3. The two types of visual notifications inspired from the one triggered in ATC radar screen when minimum separation between aircraft is lost. In the Color-Blink notification, the text ALRT switches from white (1a) to red (1b) at a rate of 200 ms white on/800 ms red. In the Box-Animation notification, the text ALRT is displayed in red (2b) and four yellow chevrons placed around the label (2a) move outward from the label (2b) by 60 pixels following a slow in/slow out animation cycle of 1Hz.

Participants had to acknowledge the notifications by clicking on the associated aircraft. The notified aircraft was randomly selected among the static aircraft, and only one notification was issued at a time. The notification disappeared as soon as the participant clicked on the aircraft. If the participant did not react within a given time (depending on the speed condition), the notification disappeared. Thirty-four visual notifications were displayed in each scenario.

In order to engage the participant in the ATC-like simulation, a score was displayed on the top left of the screen. The score decreased for the following three reasons: first, when a participant led the aircraft outside of the corridor, second when he/she gave an incorrect instruction, third when he/she failed to click on a peripheral notification in the time limit. The simulation ended as soon as the aircraft reached the arrival area, colored in red, at the end of the corridor.

To further improve the level of realism, the participants performed the simulated ATC-like tasks within the Laby microworld software in four different scenarios: two with low cognitive load and two with high cognitive load. The cognitive load was manipulated by the speed of the task. In the low

cognitive load condition, the central aircraft moved to 0.6 velocity units and peripheral aircraft notifications were displayed every 17 seconds on average. In the high cognitive load condition, the central aircraft moved to 0.99 velocity units and peripheral aircraft notifications were displayed every 12 seconds on average. In addition, the number of aircraft in the visual scene varied, from 5 in two scenarios to 21 in the two other scenarios. We only considered the effect of the speed. The number of paths, altitude instructions and visual notifications were the same in each of the four scenarios, and the order of the four scenarios was counterbalanced among participants.

3.2.2. Auditory oddball task

In parallel to the ATC task, participants had to perform an auditory alarm detection task. Standard tones (1000 Hz, 52.5 dB, 500 ms long, probability = 0.8) and deviant tones (2000 Hz, 52.5 dB, 500 ms long, probability = 0.2) were randomly played. The tones were not representative of the auditory alerts recently integrated in ATC operations. The frequencies were chosen from the study of P300 components conducted by Kolev et al. [36]. The mean time window between successive tones depended on the speed of the scenario (slow = 4.2; fast = 2.6 mean time window in seconds between two tones). Participants were told to consider the deviant tones as auditory warnings and to report them as fast as possible by pressing a specific button. The auditory oddball detection task had no impact on the score. The number of auditory alarms (n = 20) was the same in each of the four scenarios.

In order to determine individual baseline P300 amplitudes, participants were asked to perform two auditory oddball control tasks. These oddball control tasks were similar to the auditory oddball task administered in parallel to the ATC task, the only difference was that a white cross was displayed at the center of the screen instead of the ATC task. One auditory control task was performed in the slow speed condition, and another one was performed in the high speed condition. The order was counterbalanced among participants. These two oddball control tasks were completed after the four ATC scenarios. Importantly, after having checked the lack of significant effect of speed on N100 and P300 components, we merged these two oddball control tasks into a “baseline condition”. A 42 dB white noise was played continuously during each ATC scenario and during the oddball control tasks.

3.3. Procedure

The whole procedure lasted about 2.5 hours. First, participants had to fill out two behavioral questionnaires: the Pichot Fatigue questionnaire [37] and the State-Trait Anxiety Inventory (STAI Form Y-B, [38]). Next, participants were seated comfortably at 60 cm from the 30 inch screen in a sound-attenuated room with their right hand on the computer mouse and their left hand on the auditory alarm button. Second, they completed a training phase to familiarize with the Laby microworld software, i.e.

enter correctly path and altitude instructions by the drop-down menus, acknowledge visual notifications, and report deviant sounds. After the training, electrodes were placed on the participants' scalps before they completed the four counterbalanced ATC scenarios. Between each scenario, participants filled out the NASA Task Load Index (NASA TLX, see [39]). Finally, participants performed the two control oddball tasks in the two speed conditions.

3.3.1. EEG recordings and data processing

Continuous EEG recordings were performed with a ProComp Infinity™ encoder (Thought Technology Ltd) during the four ATC scenarios and the two control tasks. Prior to the four scenarios, three electrodes were placed for bipolar measurements: the positive electrode on the Pz site (parietal lobe), the reference electrode on the left side of the forehead and the ground ear-clip electrode on the right ear lobe. The EEG signal was recorded at a sampling rate of 256Hz.

EEG data analysis was performed using EEGLAB 11.0.3.1b [40] running under MATLAB 7.1 (The Mathworks). The EEG signals were filtered with a 0.5Hz high-pass filter and 20Hz low-pass filter, and then segmented into epochs around the auditory stimulus (from 200 ms before stimulus onset to 1000 ms after stimulus onset). The amplitude of the P300 was defined as the average amplitude within 364 to 464 ms post-stimulus. These windows were determined from a 100 ms wide time window around the peak latency for deviant tones (414 ms post-stimulus) among participants during the control task (oddball alone).

3.3.2. Statistical analysis

Mean detection rates of peripheral visual notifications were calculated for the four scenarios. ERP amplitudes were computed for the four Laby scenarios and for the two oddball control tasks. Statistical analyses were performed using Statistica 7.1 (StatSoft ©). Differences between the experimental conditions were investigated with the use of ANOVA followed by post hoc testing (Tukey's honestly significant difference, Tukey HSD).

4. RESULTS

4.1. Subjective results

We performed 2 * 2 ANOVAs with "group" (notification type) as a categorical variable and within-subject factor "speed" (cognitive load) to investigate the effect of the notification design and the task speed on the "mental demand" and "temporal demand" dimensions (NASA TLX). Although the mean scores for mental and temporal demands were lower in the Box-Animation vs. Color-Blink groups (mental demand: 57.42 vs. 48.06; temporal demand: 54.24 vs. 49.31) there was no significant main effect of the

group (mental demand: $F(1, 34) = 1.86$, $p = .18$, $\eta^2p = .05$; temporal demand: $F(1, 34) = .55$, $p = .46$, $\eta^2p = .02$). These results reveal that participants did not feel a significantly lower effort with the noticeable Box-Animation design compared to the basic Color-Blink design. There was a main effect of the speed on both mental demand ($F(1, 34) = 30.84$, $p < .001$, $\eta^2p = .48$) and temporal demand ($F(1, 34) = 31.77$, $p < .001$, $\eta^2p = .48$), and no interaction. The effect of the speed shows that the increase in speed was perceived by participants as an increase in difficulty (mental and temporal demands).

4.2. Behavioral results

4.2.1. Peripheral notifications detection rate

We performed a 2 * 2 ANOVA with "group" (notification type) as a categorical variable and within-subject factor "speed" (cognitive load) to investigate the effect of the notification design and the task speed on the peripheral notifications detection rate. Importantly, we found a main effect of the group ($F(1, 34) = 20.14$, $p < .001$, $\eta^2p = .37$). As expected, participants had a higher notification detection rate in the Box-Animation group (mean M = 99.83, standard deviation SD = 0.33) than in the Color-Blink group (M = 95.70, SD = 3.89). We also found a main effect of speed ($F(1, 34) = 14.78$, $p < .001$, $\eta^2p = .30$). Significantly fewer visual notifications were reported under the fast condition (M = 96.41, SD = 6.48) than under the slow condition (M = 99.11, SD = 2.42). Interestingly, there was a significant interaction between speed and group ($F(1, 34) = 11.24$, $p = .002$, $\eta^2p = .25$). Tukey HSD post-hoc analysis revealed that increasing simulation speed significantly decreased the detection rate of peripheral notifications for the participants that used the Color-Blink design ($p < .001$) while the detection rate of the participants that used the Box-Animation notifications was unaffected by the higher level of speed ($p = .99$). Box Animation design seems to ease the detection task up to a point where speed increases did not affect detection rate.

4.2.2. Accuracy to the central aircraft guiding and the oddball task

As a supplementary analysis, we examined the effects of group and speed on the accuracy rate to the central aircraft guiding task with a 2 * 2 ANOVA with group as categorical variable. There was a significant effect of the speed ($F(1, 34) = 44.71$, $p < .001$, $\eta^2p = .57$) on the accuracy for altitude instructions, no effect of the group ($F(1, 34) = 1.39$, $p = .25$, $\eta^2p = .04$) and no interaction ($F(1, 34) = .69$, $p = .41$, $\eta^2p = .02$). We also performed a 2 * 2 ANOVA with group as categorical variable on the rare tones detection for the ATC scenarios. There was no effect of the speed ($F(1, 34) = .95$, $p = .34$, $\eta^2p = .03$), no effect of the group ($F(1, 34) = 30.84$, $p = .52$, $\eta^2p = .01$) and no interaction ($F(1, 34) = 2.96$, $p = .095$, $\eta^2p = .08$). Interestingly, though not statistically significant, there was a numerical difference between the means performance to the oddball task for this interaction. In the slow condition, participants

had very similar tone detection rates (Box-Animation group: M = 95.62 %, SD = 6.39; Color-Blink group: M = 96.25 %, SD = 5.19), while in the fast condition, they had a slightly better performance in the Box-Animation group (M = 96.46 %, SD = 3.06) than in the Color-Blink group (M = 93.26 %, SD = 10.50). These results may suggest that the better accuracy to peripheral notification detection allowed by the Box-Animation design was not detrimental to auditory detection or guiding the central aircraft, and could even marginally improve the performance in auditory detection.

4.3. EEG results

4.3.1. Averaging of the two oddball control tasks into one baseline condition

We first compared the two oddball control tasks (slow and fast) before merging them into a single baseline condition, to exclude potential effects of the speed on the auditory P300 amplitude. The 2 * 2 ANOVA with within-subject factors “speed” and “type of sound” showed no effect of the speed ($F(1, 35) = 1.72$, $p = .20$, $\eta^2p = .05$), a classic significant effect of the type of sound ($F(1, 35) = 68.78$, $p < .001$, $\eta^2p = .66$), with a higher P300 for target deviant tones (M = 4.57 µV, SD = 3.37 µV) than for standard tones (M = -.13 µV, SD = 1.52). There was no significant interaction ($F(1, 35) = .12$, $p = .73$, $\eta^2p = .003$). Speed of the oddball control task having no impact P300 amplitude, the two control tasks were averaged to create the baseline condition. For the following analyses, we only focused on the deviant tones.

4.3.2. P300 results

We compared the deviant tones P300 response in the baseline condition to the Laby scenarios with a one-way ANOVA with within-subjects factor “task” (baseline condition vs. the four Laby scenarios averaged). There was a significant main effect of the introduction of the ATC task ($F(1, 35) = 13.20$, $p = .001$, $\eta^2p = .27$), with a lower P300 when the ATC task was performed (M = 2.64 µV, SD = 2.35 µV) than in the baseline condition (M = 4.57 µV, SD = 3.37 µV). This result is coherent with Kramer, Trejo [41] findings that showed that P300 amplitude is sensitive to the mental workload generated by the introduction of a radar-monitoring task vs. a baseline condition (tones alone) [42].

Finally, a 2 * 2 ANOVA on the Laby scenarios with within-subject factor “speed” and categorical variable “group” revealed a significant main effect of the group ($F(1, 34) = 4.20$, $p = .048$, $\eta^2p = .11$), no main effect of the speed ($F(1, 34) = 1.29$, $p = .04$) and no interaction ($F(1, 34) = 2.41$, $p = .13$, $\eta^2p = .07$). These results revealed that Box-Animation notifications elicited a higher P300 for deviant tones (M = 3.41 µV, SD = 2.70 µV) in comparison to Color-Blink notifications (M = 1.87 µV, SD = 1.68 µV), as shown in Figure 4. As expected, it suggests an enhanced processing of auditory deviant target tones allowed by a release of attentional resources when the better HMI design was used.

Figure 4. ERPs for the BA (red) and CB (blue) and Control (black) conditions, for alarm tones, on the Pz electrode. The horizontal axis denotes time in ms, and the vertical axis denotes amplitude in μV. P300 amplitude is significantly higher in BA group than in CB group.

5. Discussion

The current study used EEG techniques to assess the impact of cognitive load during a simulated ATC task that also required responding to auditory targets. The visual notification detection aspect of the ATC task was performed either with the Box-Animation design, a very noticeable visual notification, or with the Color-Blink design, a much less perceptible notification. The main objective of this study was to investigate if an enhanced visual design can improve the cerebral processing of supplementary auditory stimuli during the ATC task.

Behavioral results showed that participants who used the ATC interface with the Box-Animation design were more accurate in the detection of peripheral notifications compared to those who were presented with the Color-Blink notifications. In addition, we found that those participants in the Box-Animation group were also less affected – or even unaffected – by an increase in speed. It is essential in multi-task situations to evaluate performance across all tasks to ensure that any new design does not just improve performance on one particular task while degrading performance on others [43]. Accordingly, the lack of impact of the notification design on the concurrent tasks (guiding the central aircraft and detecting auditory targets) shows the efficiency of the Box-Animation design to draw participant's attention towards peripheral notifications, without causing undesirable interference with other critical tasks. This pattern of results may be taken to suggest that there is a release of attentional resources due to the Box-Animation design and not only a trade-off between visual notifications and the concurrent tasks.

However, the subjective questionnaires (NASA TLX) revealed that participants did not perceive a lower mental demand with the noticeable Box-Animation design compared to the basic Color-Blink design. Yet, the effect sizes calculated using the partial eta squared showed that the behavioral impact (peripheral notification rates) of the notification type was higher than the simulation speed (respectively, $\eta^2p = .37$ and $\eta^2p = .30$). This inconsistency between the subjective assessment and the objective behavioral performance demonstrates the importance of considering both subjective and behavioral objective metrics in design evaluation. Nevertheless the subjective judgment should not be dismissed; indeed, the compliance of the operators of critical systems is essential, especially as it can jeopardize the use and acceptability of the system.

In the ATC-like Laby simulation, participants were instructed to focus on the main task of guiding an aircraft around a given route, in accordance with centrally displayed instructions, always appearing next to the central aircraft. The validation of visual notifications displayed in the periphery of the screen at

random locations was regarded as secondary to the main guiding task. According to the NSEEV's model [44], these latter notifications are more likely to be missed, especially under high cognitive load, because of their greater eccentricity (higher effort needed to direct attention toward the item of interest), their occurrence in a random location (lower expectancy of the event to appear in a particular location), and the fact that they are seen as secondary to the main guiding task (lower value of the item). However the current experiment demonstrates that this effect may be compensated by a more salient design. The Box-Animation design is larger than color-blink (static salience, see [45]), and the slow in/slow out animation involves greater dynamic salience [46]. While both notifications involve a repeated animation cycle, the 'popping' motion of the chevrons in Box-Animation creates a 'deviant' quality that is more likely to capture attention analogous to deviance in the auditory modality, [47]. As such, detection of the Box-Animation notifications required fewer attentional resources than Color-Blink notifications, meaning that detection was achieved with greater ease and was less vulnerable to increases in workload. Accordingly, when the difficulty of the ATC task increased (speed), participants' performance remained unaffected in the Box-Animation condition while it declined with Color-Blink.

The analysis of physiological results showed a higher P300 in the baseline condition – which required participants to simply detect the oddball sounds – than when performing the simulated ATC-like set of tasks. This analysis also revealed that, as hypothesized, the better performance in the Box-Animation group was concomitant with a greater auditory P300 amplitude for deviant tones compared to that found in the Color-Blink group. It seems that when the simulated ATC-like tasks were performed with the Box-Animation design, the P300 amplitude observed for the deviant tones (3.41 μV) was much closer to the amplitude observed in the baseline condition (4.57 μV) in comparison with the P300 amplitude observed in the Color-Blink group (1.87 μV). This pattern of results suggests that there is a lower depletion of attentional resources from working with the Box-Animation notification design. Taken together, these results support the idea that P300 amplitude can serve as a reliable cognitive load index in ecological settings [18, 20, 21, 41].

As the auditory P300 reflects the cerebral response to an auditory stimulus, it seems likely that P300 reduction may also indicate a decline in the probability of detecting an auditory stimulus. This is supported by the smaller auditory detection rate (though not statistically significant) in the Color-Blink group for the high speed condition, compared to the Box-Animation group. The P300 decrease in the Color-Blink can also be seen as a precursory effect of the diminishing attentional and perceptual processing resources, which would eventually lead to a decrease in performance on the auditory detection task if cognitive resources were completely exhausted. According to several authors [48, 49], the problem of missed alarms occurs frequently across a range of flight environments and extends to ATC since the development of auditory notifications and warnings is increasingly integrated within ATC workstations

[50]. In addition to their well-documented limitations (e.g. stress, cry-wolf effect, cf. [51], auditory alarms sometimes fail to be perceived, especially in critical situations. This propensity to remain unaware of fully audible stimuli under high workload conditions is referred to as inattentional deafness (e.g. [52]. Consequently, using more salient notification designs to restore P300 could help prevent inattentional deafness in high multitasking situations such as ATC and piloting.

6. Conclusion

The analysis of behavioral, subjective and physiological results, i.e. the neuroergonomic approach, gives us a more complete understanding of the complex impact of changes in interface design. The benefits of the Box-Animation design were better understood given its impact on the subjective perception of participants, on their behavioral performance and their cerebral response, the latter revealing an otherwise invisible effect on available attentional resources. The neuroergonomic approach offers a more complete and objective way to evaluate HMI design. Our study corroborates the assumption that an enhanced HMI design liberates attentional resources of an operator, making him/her more efficient to process other additional critical stimuli such as auditory alarms. We also confirmed that P300 amplitude represent a reliable cognitive load index in ecological settings. The investigation of the relationship between HMI design in an ATC context and ERPs amplitude is the first step towards real-time monitoring of operators for adaptive intelligent systems.

7. Acknowledgements

We would like to thank Danny Lebel for his technical support, and Anaït Bagramian and Marie-Pierre B. Tremblay for their help running the experiment. We are also grateful to Zarrin Chua, Julia Behrend and Helen Hodgetts for proofreading earlier versions of the paper. Financial support was provided by a Discovery grant from the National Sciences and Engineering Research Council of Canada (NSERC), awarded to Sébastien Tremblay [grant number CG073877], by the Institut Supérieur de l'Aéronautique et de l'Espace to in the form of an operating grant to M. Causse, and also by the Direction Générale de l'Armement in the form of a scholarship to L. Giraudet.

8. References

[1] Lavie N. Perceptual load as a necessary condition for selective attention. Journal of Experimental Psychology-Human Perception and Performance. 1995;21:451-68.
[2] Lavie N, Hirst A, de Fockert JW, Viding E. Load theory of selective attention and cognitive control. Journal of Experimental Psychology: General. 2004;133:339.
[3] Lavie N, Tsal Y. Perceptual load as a major determinant of the locus of selection in visual attention. Perception & Psychophysics. 1994;56:183-97.
[4] Hughes RW, Hurlstone MJ, Marsh JE, Vachon F, Jones DM. Cognitive control of auditory distraction: Impact of

task difficulty, foreknowledge, and working memory capacity supports duplex-mechanism account. Journal of Experimental Psychology: Human Perception and Performance. 2013;39:539.
[5] Moore DR, Fuchs PA, Rees A, Palmer A, Plack CJ. The Oxford Handbook of Auditory Science: The Auditory Brain: Oxford University Press; 2010.
[6] Spence C, Read L. Speech shadowing while driving on the difficulty of splitting attention between eye and ear. Psychological science. 2003;14:251-6.
[7] Wood NL, Cowan N. The cocktail party phenomenon revisited: Attention and memory in the classic selective listening procedure of Cherry (1953). Journal of Experimental Psychology: General. 1995;124:243.
[8] Brand-D'Abrescia M, Lavie N. Task coordination between and within sensory modalities: Effects on distraction. Perception & psychophysics. 2008;70:508-15.
[9] Macdonald JS, Lavie N. Visual perceptual load induces inattentional deafness. Attention, Perception, & Psychophysics. 2011;73:1780-9.
[10] Santangelo V, Olivetti Belardinelli M, Spence C. The suppression of reflexive visual and auditory orienting when attention is otherwise engaged. Journal of experimental psychology Human perception and performance. 2007;33:137.
[11] Sinnett S, Costa A, Soto-Faraco S. Manipulating inattentional blindness within and across sensory modalities. The quarterly journal of experimental psychology. 2006;59:1425-42.
[12] Berti S. The role of auditory transient and deviance processing in distraction of task performance: a combined behavioral and event-related brain potential study. Frontiers in human neuroscience. 2013;7.
[13] Berti S, Schröger E. Working memory controls involuntary attention switching: evidence from an auditory distraction paradigm. European Journal of Neuroscience. 2003;17:1119–22.
[14] Munka L, Berti S. Examining task-dependencies of different attentional processes as reflected in the P3a and reorienting negativity components of the human event-related brain potential. Neuroscience letters. 2006;396:177–81.
[15] Santangelo V. Forced to remember: When memory is biased by salient information. Behavioural brain research. 2015.
[16] Thompson KG, Bichot NP. A visual salience map in the primate frontal eye field. Progress in brain research. 2005;147:249-62.
[17] Nardo D, Santangelo V, Macaluso E. Stimulus-driven orienting of visuo-spatial attention in complex dynamic environments. Neuron. 2011;69:1015-28.
[18] Fowler B. P300 as a measure of workload during a simulated aircraft landing task. Human Factors: The Journal of the Human Factors and Ergonomics Society. 1994;36:670–83.
[19] Kramer AF, Sirevaag EJ, Braune R. A psychophysiological assessment of operator workload during simulated flight missions. Human Factors: The Journal of the Human Factors and Ergonomics Society. 1987;29:145–60.
[20] Ullsperger P, Freude G, Erdmann U. Auditory probe sensitivity to mental workload changes–an event-related potential study. International Journal of Psychophysiology. 2001;40:201–9.
[21] Allison BZ, Polich J. Workload assessment of computer gaming using a single-stimulus event-related potential paradigm. Biological psychology. 2008;77:277–83.
[22] Polich J. Updating P300: an integrative theory of P3a and P3b. Clinical neurophysiology: official journal of the International Federation of Clinical Neurophysiology. 2007;118:2128.
[23] Polich J, Kok A. Cognitive and biological determinants of P300: an integrative review. Biological psychology. 1995;41:103–46.
[24] Munka L, Berti S. Examining task-dependencies of different attentional processes as reflected in the P3a and reorienting negativity components of the human event-related brain potential. Neuroscience letters. 2006;396:177-81.
[25] Singhal A, Doerfling P, Fowler B. Effects of a dual task on the N100–P200 complex and the early and late Nd attention waveforms. Psychophysiology. 2002;39:236-45.
[26] Kok A. On the utility of P3 amplitude as a measure of processing capacity. Psychophysiology. 2001;38:557–77.
[27] Wijers AA, Otten LJ, Feenstra S, Mulder G, Mulder LJ. Brain potentials during selective attention, memory search, and mental rotation. Psychophysiology. 1989;26:452-67.
[28] Squires NK, Squires KC, Hillyard SA. Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. Electroencephalography and clinical neurophysiology. 1975;38:387–401.
[29] Nieuwenhuis S, Aston-Jones G, Cohen JD. Decision making, the P3, and the locus coeruleus--norepinephrine system. Psychological bulletin. 2005;131:510.

[30] Causse M, Dehais F, Péran P, Sabatini U, Pastor J. The effects of emotion on pilot decision-making: A neuroergonomic approach to aviation safety. Transportation Research Part C: Emerging Technologies. 2013;33:272-81.
[31] Parasuraman R, Rizzo M. Neuroergonomics: The brain at work: Oxford University Press, USA; 2007.
[32] Durantin G, Gagnon J-F, Tremblay S, Dehais F. Using near infrared spectroscopy and heart rate variability to detect mental overload. Behavioural brain research. 2014;259:16-23.
[33] Gagnon J-F, Durantin G, Vachon F, Causse M, Tremblay S, Dehais F. Anticipating human error before it happens: Towards a psychophysiological model for online prediction of mental workload. Human Factors and Ergonomics Society Chapter Europe Proceedings2012.
[34] Imbert J-P, Hodgetts H, Parise R, Vachon F, Tremblay S. The LABY microworld: A Platform for Research, Training and System Engineering. 2014.
[35] Imbert J-P, Hodgetts HM, Parise R, Vachon F, Dehais F, Tremblay S. Attentional costs and failures in air traffic control notifications. Ergonomics. 2014;57:1817-32.
[36] Kolev V, Demiralp T, Yordanova J, Ademoglu A, Isoglu-Alkaç Ü. Time–frequency analysis reveals multiple functional components during oddball P300. NeuroReport. 1997;8:2061-5.
[37] Pichot P, Brun J-P. Questionnaire bref d'auto-évaluation des dimensions dépressive, asthénique et anxieuse.: Elsevier Science; 1984.
[38] Spielberger CD. State-Trait Anxiety Inventory: Wiley Online Library; 2010.
[39] Cao A, Chintamani KK, Pandya AK, Ellis RD. NASA TLX: Software for assessing subjective mental workload. Behavior research methods. 2009;41:113–7.
[40] Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of neuroscience methods. 2004;134:9–21.
[41] Kramer AF, Trejo LJ, Humphrey D. Assessment of mental workload with task-irrelevant auditory probes. Biological Psychology. 1995;40:83–100.
[42] Giraudet L, Saint-Louis M, Causse M. Electrophysiological correlates of inattentional deafness: No hearing without listening. Human Factors and Ergonomics Society Europe Chapter Conference, Toulouse, France2012.
[43] Lafond D, Vachon F, Rousseau R, Tremblay S. A cognitive and holistic approach to developing metrics for decision support in command and control. Advances in cognitive ergonomics. 2010:65–73.
[44] Steelman KS, McCarley JS, Wickens CD. Modeling the control of attention in visual workspaces. Human Factors: The Journal of the Human Factors and Ergonomics Society. 2011;53:142–53.
[45] Itti L, Koch C. A saliency-based search mechanism for overt and covert shifts of visual attention. Vision research. 2000;40:1489-506.
[46] Yantis S, Jonides J. Abrupt visual onsets and selective attention: voluntary versus automatic allocation. Journal of Experimental Psychology: Human perception and performance. 1990;16:121.
[47] Hughes RW, Vachon F, Jones DM. Auditory attentional capture during serial recall: Violations at encoding of an algorithm-based neural model? Journal of Experimental Psychology: Learning, Memory, and Cognition. 2005;31:736.
[48] Bliss JP. Investigation of alarm-related accidents and incidents in aviation. The International Journal of Aviation Psychology. 2003;13:249-68.
[49] Dehais F, Causse M, Vachon F, Régis N, Menant E, Tremblay S. Failure to Detect Critical Auditory Alerts in the Cockpit Evidence for Inattentional Deafness. Human Factors: The Journal of the Human Factors and Ergonomics Society. 2014;56:631–44.
[50] Cabrera D, Ferguson S, Laing G. Development of auditory alerts for air traffic control consoles. Audio Engineering Society; 2005.
[51] Edworthy J, Loxley S, Dennis I. Improving auditory warning design: Relationship between warning sound parameters and perceived urgency. Human Factors: The Journal of the Human Factors and Ergonomics Society. 1991;33:205-31.
[52] Koreimann S, Strauß S, Vitouch O. Inattentional deafness under dynamic musical conditions. 2009.

figure 1 .

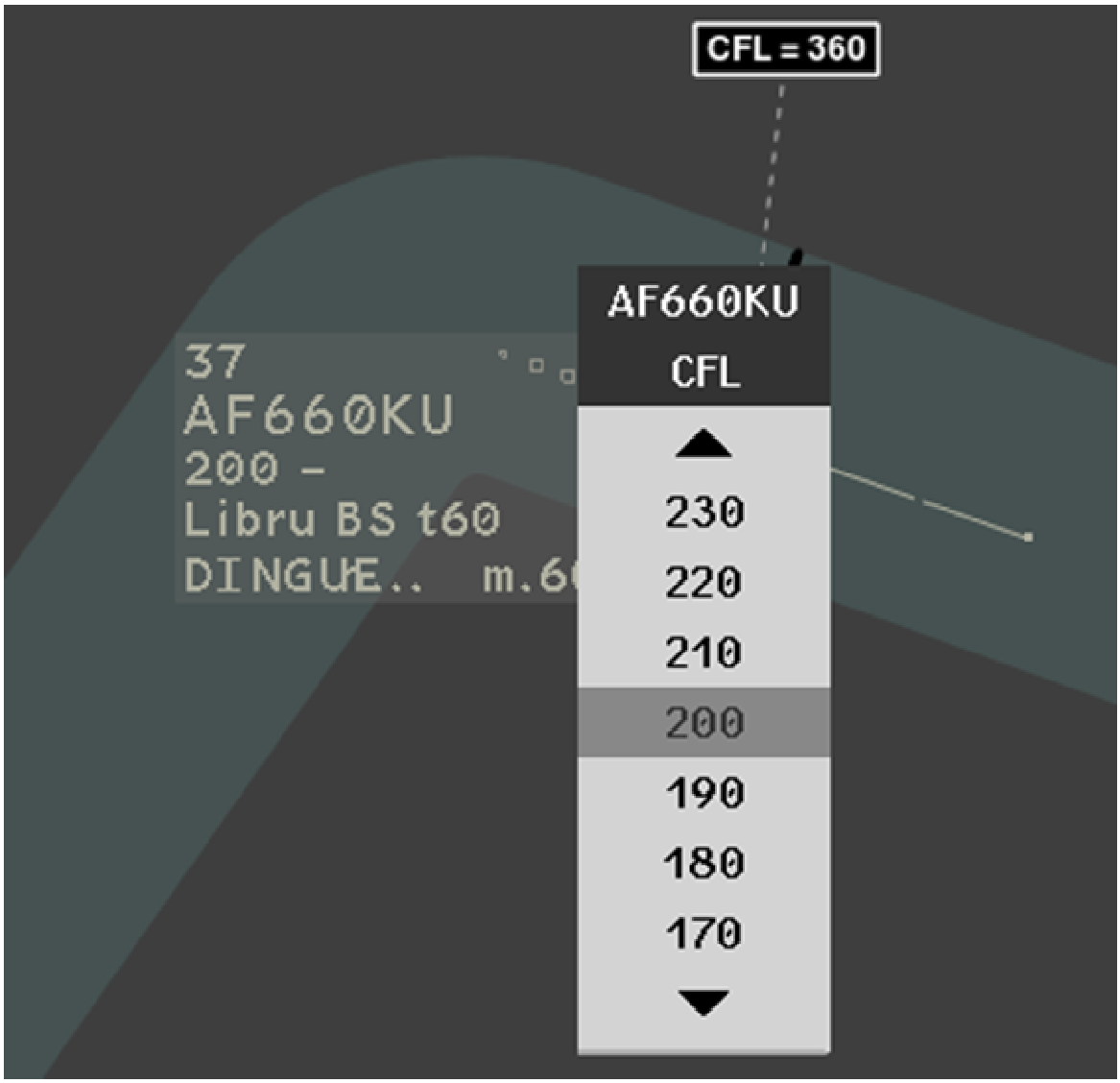

figure 2 .

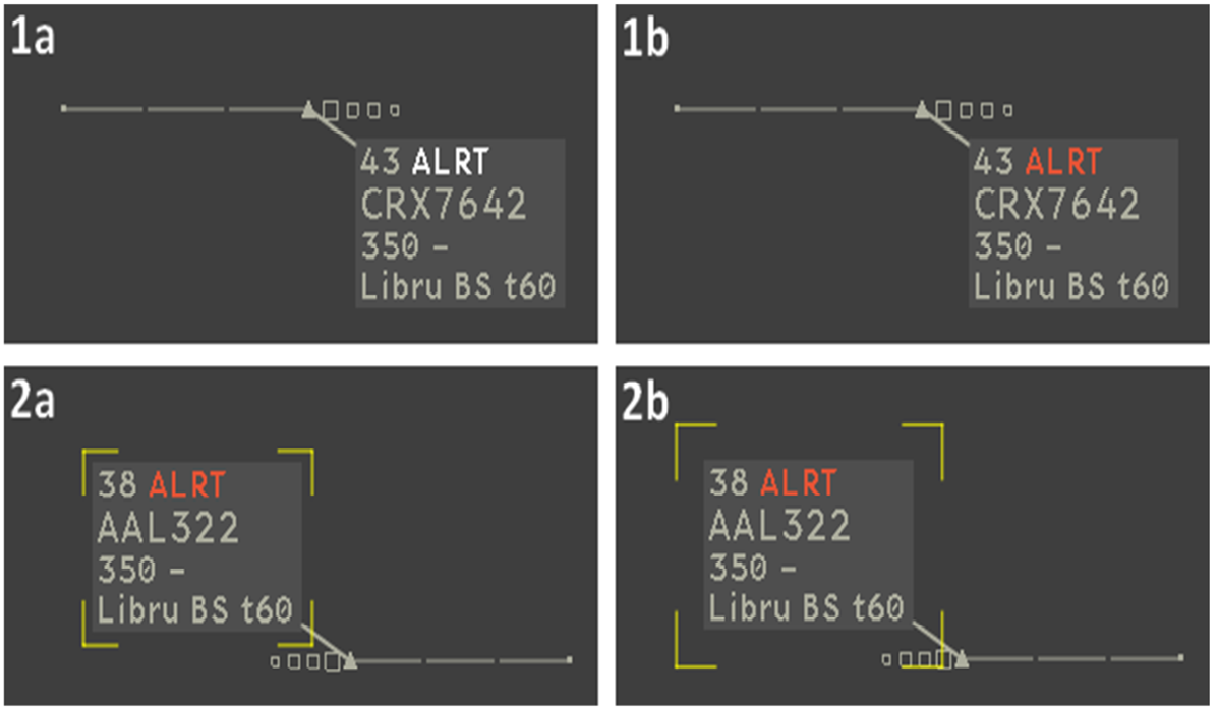

figure 3 .

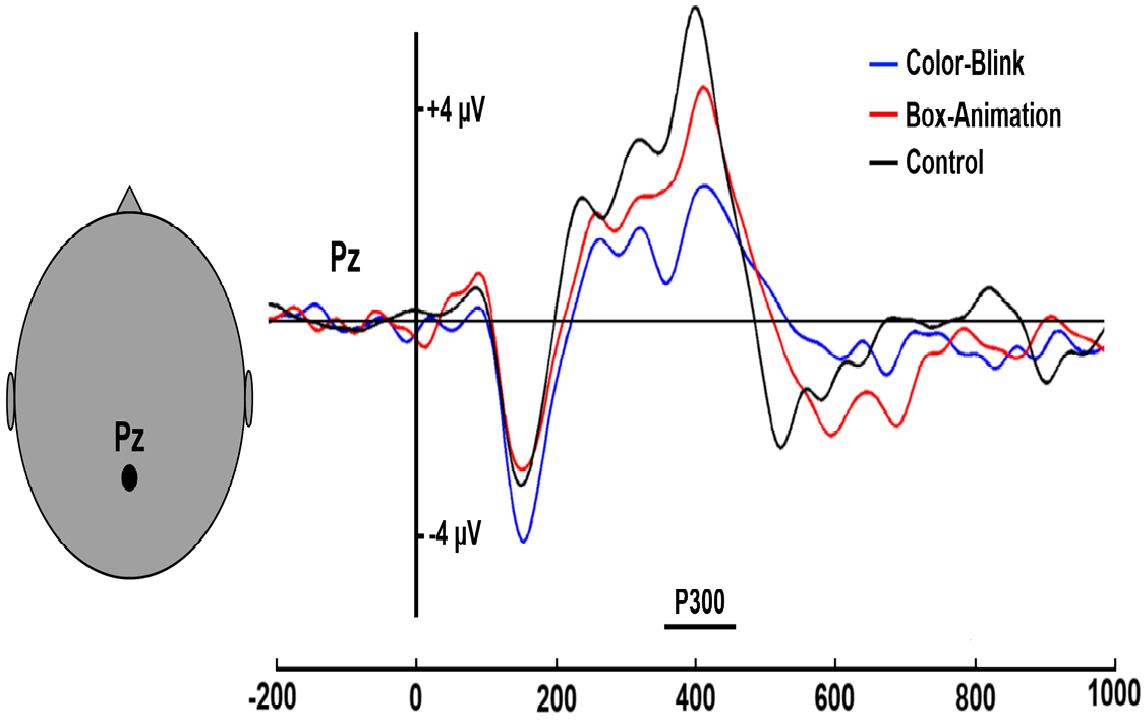

figure 4 .