The Benefits and the Costs of Using Auditory Warning Messages in Dynamic Decision Making Settings

Cindy Chamberland[1], Helen M. Hodgetts[1,2], Benoît R. Vallières[1], François Vachon[1], & Sébastien Tremblay[1]

[1] Université Laval, Québec, Canada

[2] Cardiff Metropolitan University, UK

Address correspondence to Cindy Chamberland, École de psychologie, Pavillon Félix-Antoine-Savard, 2325 rue des Bibliothèques, Université Laval, Québec (QC) Canada, G1V 0A6; cindy.chamberland@psy.ulaval.ca

Abstract

The failure to notice critical changes in both visual and auditory scenes may have important consequences for performance in complex dynamic environments, especially those related to security such as aviation, surveillance during major events, and command and control of emergency response. Previous work has shown that a significant number of situation changes remain undetected by operators in such environments. In the current study, we examined the impact of using auditory warning messages to support the detection of critical situation changes and to a broader extent the decision making required by the environment. Twenty-two participants performed a radar operator task involving multiple subtasks while detecting critical task-related events that were cued by a specific type of audio message. Results showed that about 22% of the critical changes remained undetected by participants, a percentage similar to that found in previous work using visual cues to support change detection. However, we found that audio messages tended to bias threat evaluation towards perceiving objects as more threatening than they were in reality. Such findings revealed both benefits and costs associated with using audio messages to support change detection in complex dynamic environments.

WORDCOUNT (max = 200): 188
KEYWORDS: Auditory Alarms; Change Detection; Dynamic Decision Making; Decision Support System; Eye tracking; Pupillometry.

INTRODUCTION

In complex and dynamic work environments such as aviation, surveillance, and emergency response command and control (C2), system operators are exposed to a high load of information from multiple sources and sensory modalities, which must be continually processed, filtered, and integrated with the task at hand. The capability of the operator to discern significant objects and events in his/her environment is crucial to maintaining performance in these often uncertain and time-pressured environments (Durlach, 2004). However, in situations of high workload, when attention must be divided between multiple sources (different modalities, dealing with unexpected interruptions, etc.), significant events may go unnoticed (St. John & Smallman, 2008). In one study, over 13% of critical changes on a visual interface remained undetected by operators in a simulated C2 environment (e.g., Vachon, Vallières, Jones, & Tremblay, 2012). This so-called inattentional blindness (i.e., the inability to detect unexpected and obvious events in the visual scene; Mack & Rock, 1998) is believed to manifest by a failure to direct attention towards a change or a failure of attentional processes to extract and process enough visual information regarding the critical change for it to affect subsequent actions. One way to reduce the number of undetected events in such an environment is to reduce the visual load of the display by exploiting other sensory modalities. According to the multiple resource theory of attention (Wickens, 2008), using the same modality to perform multiple tasks leads to interference, which may be reduced if different tasks use different modalities. In contrast, it has been shown that integrating information cross-modality is more demanding (Penney, 1989), especially given that the transient nature of sound might force participants to remember the auditory information to the detriment of the visual input, therefore placing a greater burden on working memory. The present study explored the benefits and the costs of using auditory warning messages to support change detection in complex and dynamic work environments.

In the current work, we use a microworld that provides a simplified simulation of above-water C2 warfare to assess decision making and the ability to detect critical events in complex and dynamic settings (Hodgetts, Vachon, & Tremblay, 2014; Hodgetts, Tremblay, Vallières, & Vachon, 2015; Vachon, Vallières, et al., 2012; Vallières, Hodgetts, Vachon, & Tremblay, 2016). This microworld requires participants to monitor a radar screen representing the airspace around the ship, evaluate the threat level of every airborne aircraft moving in the vicinity of the ship based on a visual list of parameters, and take retaliatory defensive measures against hostile aircraft. Critical changes consist of an aircraft changing from a non-hostile to a hostile threat level. Participants must detect such changes and perform a retaliatory defensive action since hostile aircraft are programed to attack the ship. In previous studies using this microworld, each critical change was accompanied by a visual change on the radar screen (i.e., a change in the direction and/or the speed of the aircraft) that made it visually noticeable (e.g., Hodgetts et al., 2014; Vachon, Vallières, et al., 2012). The detection of visual changes can be improved by automatic change detection software (St. John, Smallman, & Manes, 2005), although care must be taken that such decision aids fit within the available workload capacity of the user (Vallières et al., 2016). Another possibility to support change detection might be to reduce the visual load of the task by using the auditory rather than the visual modality to announce the critical changes. Accordingly, in the current study, critical changes were cued by auditory warning messages rather than by visual changes on the radar. A change in some characteristics of the audio messages (i.e., voice and content) signaled that an aircraft’s threat level had turned to hostile and thus required further action.

There are good reasons to believe that the presence of such auditory cues may be helpful in alerting the operator to the presence of a new threat. The use of auditory warnings as a means of improving safety and performance is now relatively common in applications such as hospital equipment or aircraft navigation systems (Stanton & Edworthy, 1999). More recently, new in-car technologies have led to an increasing number of sound interfaces, ranging from information related to satellite navigation to warning signals used to alert drivers to potential danger. Ho and Spence (2005) demonstrated that semantically meaningful auditory warning signals may provide an effective means of capturing attention in a simulated driving task and facilitate the detection of potentially urgent visual driving events (e.g., the rapid approach of a car in front or from behind) when compared to auditory cues that are not semantically associated with the signalled emergency. In complex and dynamic environments, auditory alerts are sometimes regarded as preferable to visual alarms since it is widely accepted that processing of an auditory stimulus is obligatory without the need for explicit head/eye movements (Edworthy, Loxley, & Dennis, 1991), and auditory alarms are often superior to their visual counterparts in terms of detection speed (Morris & Montano, 1996) and compliance (Wogalter & Young, 1991). Moreover, unexpected changes in an auditory stream, such as voice changes (Hughes, Vachon, & Jones, 2007; Vachon, Hughes, & Jones, 2012), are known to involuntarily attract attention, even when the stimulus is irrelevant to the task at hand. However, the auditory presentation of information may come at a cost whereby the background sound impairs other facets of the ongoing task. For instance, the presence of irrelevant radio messages can impair the ability to process navigational information (Banbury, Fricker, Tremblay, & Emery, 2003). There is also evidence that the content of background sound can interfere with ongoing semantic processing (e.g., Marsh, Hughes, & Jones, 2008).

In addition to the potential effects of sound on cognitive processing, the use of the auditory modality to detect critical changes may be vulnerable to inattentional deafness (e.g., Vitevitch, 2003), the failure to detect (often obvious) changes in the auditory scene (Bregman, 1994). This phenomenon has been observed with various types of auditory changes (e.g., voice, location, rhythm) and is highly influenced by the allocation of attention (Eramudugolla, Irvine, McAnally, Martin, & Mattingley, 2005). Recently, Dalton and Fraenkel (2012) demonstrated sustained inattentional deafness in a dynamic setting over an extended period of time, which provided a direct analogy to the famous gorilla study (Simons & Chabris, 1999) in which half the participants counting passing between basketball players in a short video failed to notice a gorilla walking amongst the players. In this auditory equivalent, participants were presented with an auditory scene recorded binaurally using a dummy head. When participants were asked to focus on a specific conversation within the auditory scene, 70% of people failed to notice an 'auditory gorilla' present for 19 s in the recording (i.e., a male character entering from the back of the auditory scene and walking through the scene continually repeating the sentence "I'm a gorilla"). The phenomenon of inattentional deafness is relatively less well researched than its visual counterpart, particularly in complex C2 environments that are likely to be susceptible to such an occurrence. The vulnerability to overlook important events in the auditory domain thus appears similar in many ways to its visual counterpart and as such, poses a considerable risk for complex work environments that use the auditory modality to warn operators of emergency situations in the belief that they are less fallible (e.g., Dehais et al., 2014).

The present study

In the current study, we explored the benefits and the costs of using auditory warning messages to support change detection in a realistic C2 environment. The effectiveness of the auditory warning messages was evaluated by contrasting the percentage of actions performed on the corresponding aircraft following each type of audio message as well as by their impact—positive or negative—on threat-evaluation performance.

We were also interested in recording eye movements during the simulation to examine the possible interaction between attentional resources available during critical change detection performance. According to Multiple Resource Theory (Wickens, 2008), separate pools of attentional resources are drawn upon according to sensory modality. While the critical messages were presented in the auditory modality, we were interested to see whether detection may also be influenced by the direction of attention in the visual modality, a finding which may inform the argument regarding whether attentional resources are separate or shared. This investigation was first conducted by looking at the pupillary response following the presentation of auditory warning messages according to whether or not the change was detected. Pupil size has been studied in relation to many cognitive processes and states (see Beatty, 1982; Kahneman, 1973; Marshall, 2007; Wang, 2011) but most importantly, it has been demonstrated that under conditions of constant illumination, pupillary size increased linearly with attentional effort and could therefore serve as an index of attentional effort or focused attention (Hoeks & Levelt, 1993; Kahneman, 1973). Exploiting pupillometry in the context of change detection can provide important inferences as to why some events remain undetected even though they were fixated. For instance, Vachon, Vallières, et al. (2012; cf. Privitera, Renninger, Carney, Klein, & Aguilar, 2010) were able to delineate two sources of detection failures in the visual domain based on pupillometry. A first source derived from a lack of attention on the critical event (i.e., not fixating the aircraft that visibly changed in speed and/or heading) and was not accompanied by any pupil dilation. A second source, on the other hand, derived from a failure of attentional resources despite fixating the critical event; the undetected critical event triggered pupil dilation, but to a lesser degree than the pupil response triggered when the event was actually detected.

In the auditory modality, one might presume that the first source of detection failure may not apply; due to the pervasive nature of sound, detection is not dependent on the specific orientation of the operator’s head at the critical event moment. Auditory information, however, still appears vulnerable to the second source, and seemingly obvious events or changes can go unnoticed in an auditory scene (e.g., Dalton & Fraenkel, 2012; Greg & Samuel, 2008; Vitevitch, 2003). Contrasting the change in pupil size for detected and undetected auditory warning messages might provide empirical evidence that non-detection is the result of a failure of attentional processes (cf. Vachon, Vallières, et al., 2012).

As a second step to our investigation, we were interested in testing whether auditory change detection performance could be influenced by the participant’s gaze position on the radar, a potential proxy for visual attention (e.g., McCarley & Kramer, 2008; Rayner, 2009). Previous work conducted on change detection has shown that visual changes were more likely to be detected when occurring close to a fixated position (Hollingworth, Schrock, & Henderson, 2001; O'Regan, Deubel, Clark, & Rensink, 2000; Vachon, Vallières, et al., 2012), indicating that gaze position plays an important role in the detection of visual changes. In the present study, critical changes were cued via the auditory modality and so gaze position would not be expected to facilitate detection in the same manner. Nevertheless, since auditory warning messages conveyed information about a particular aircraft on the radar, change detection performance might be improved when gaze position happens to be close to the mentioned aircraft. For example, the closer the gaze position, the less time auditory information must be stored for, and the more likely the aircraft in question is recognized. We tested this hypothesis by examining whether change detection performance was influenced by participants' gaze position on the visual scene at the exact moment of the onset of the auditory warning message. Another important finding of Vachon, Vallières, and colleagues (2012) was that visual change detection was facilitated when the critical object was fixated just before the change, showing the contribution of attentional resources in the capacity to detect visual changes. Based on that finding, we were also interested in examining whether allocating visual attention toward an aircraft just before it becomes hostile could facilitate auditory detection of the critical event due to attentional pre-processing of that aircraft. This question was addressed by comparing the percentage of change detection for aircraft that received at least one fixation 5 s before the change with those that were not fixated.

METHOD

Participants

Twenty-two students from Université Laval (12 men, 10 women, *M* = 25.72 years, *SD* = 6.90) participated in a single 2-hour experimental session and received CAD $20 compensation for their time. All reported normal or corrected-to-normal vision and normal hearing.

Apparatus/materials

The experiment used the Simulated Combat Control System (S-CCS) microworld (see Hodgetts et al., 2014; Vachon, Vallières, et al, 2012) run on a PC. This microworld provides a functional simulation of threat evaluation and combat power management processes (i.e., response planning, execution, and monitoring) that can also be generalized to other C2 situations. The visual interface includes three parts: (a) a black radar screen; (b) a list of parameters relating to the aircraft selected; and (c) a set of action buttons (Figure 1). At the center of the screen is the ownship with multiple aircraft moving in the vicinity in real time. An aircraft is represented by a white dot surrounded by a green square with a line attached; this line indicates the direction in which the aircraft is moving, and its length is proportional to the aircraft speed. Each scenario lasted 4 min and involved 27 aircraft in total, starting with five and increasing to a maximum of 10 at any one time.

Task. Participants were asked to perform four subtasks concurrently throughout the entire scenario: (a) threat-level classification; (b) threat-immediacy classification; (c) threat neutralization; and (d) critical change detection. For the threat-level classification subtask, participants were required to classify all aircraft on the radar according to their threat level (non-hostile, uncertain, hostile). Clicking with the mouse on an aircraft icon would turn the surrounding square red and display a list of parameters relating to that aircraft (see Figure 1). Five of these parameters, which could take either a threatening or a non-threatening value, must be considered to determine the threat level of the selected aircraft; (a) country of origin (ADRK, WEIV, CBOR; ADRK = threatening); (b) altitude (low, high; low = threatening); (c) identification friend or foe (IFF) (friend, neutral, foe; foe = threatening); (d) detection of weapons (yes, no; yes = threatening); and (e) military electronic emissions (yes, no; yes = threatening). Based on a pre-set classification rule, participants were asked to classify the threat level of aircraft as either non-hostile (0 or 1 threatening parameters), uncertain (2 or 3 threatening parameters), or hostile (4 or 5 threatening parameters) by clicking on the corresponding action button on the interface. Once an aircraft had been classified, the white dot changed color according to the threat level assigned to it: green (non-hostile), yellow (uncertain), or red (hostile). The threat-level classification subtask was crucial because it determined subsequent behaviors. For aircraft classified as hostile, immediate actions were required given that they were programed to hit the ownship.

For the threat-immediacy classification subtask, participants were asked to classify the level of threat immediacy of any aircraft classified as hostile based on its temporal proximity from the ownship. Temporal proximity (in seconds) could be determined by summing up the value of two parameters presented on the visual interface: i) The Time to Closest Point of Approach (TCPA), corresponding to the point at which, if the aircraft continues on the same trajectory, it will be closest to the ownship, and ii) the Closest Point of Approach by Units of Time (CPAUT), defined as the aircraft's distance from the ownship at the closest point of approach divided by its speed. The addition of these two parameters gave the overall "time before hit" (Roy, Paradis, & Allouche, 2002). Based on a pre-set classification rule, threat immediacy could be high (< 15 s from hitting the ownship), moderate (15-30 s), or low (> 30 s). Responses were made by clicking on the corresponding immediacy button on the interface (1 = high, 2 = moderate, 3 = low).

For the threat neutralization subtask, participants should then choose to launch an antiaircraft missile in defense towards hostile aircraft. Clicking on the 'engage' button launched a missile with a 2-s delay, and only one could be airborne at any one time. The critical change detection subtask consisted of detecting critical changes regarding the threat level of classified aircraft. When an aircraft appeared on the radar (either at the beginning or during an ongoing scenario), its threat level was either non-hostile or uncertain. However, its parameters (and therefore its threat level) could change over time: An aircraft's threat level could turn from non-hostile to uncertain, from non-hostile to hostile, and from uncertain to hostile, and so it was necessary to regularly check at the five relevant parameters of classified aircraft in order to reassess threat level according to the same pre-set classification rule. A total of 33 changes in aircraft parameters were presented during a scenario: 25 were non-critical (i.e., the aircraft's threat level did not change, or it did change to uncertain) and 8 were critical (i.e., the aircraft's threat level changed to hostile). More than one change occurred for some aircraft during the ongoing scenario, but an aircraft could not be presented with more than one critical change. Critical changes occurred unexpectedly in each scenario and were separated by a minimum of 15 s. Of the total 27 aircraft presented in a scenario, 11 ended with a non-hostile threat level, 8 with an uncertain threat level, and 8 with a hostile threat level (corresponding to the 8 critical changes).

Eye tracking. Eye movements were recorded with a Tobii TX300 eye tracker (Tobii Technology, 2010), integrated into a 23-in widescreen monitor with a resolution of 1280 × 800 and with a sampling rate of 300 Hz. Eye movements were calibrated before each test session to prevent data loss caused by participants movements or sitting position between test sessions. Tobii studio 3.0 software was used to analyze eye movement data. The threshold to detect an eye fixation was set at 70 ms and the fixation field corresponded to a circle with a 50-pixel radius.

<Insert Figure 1>

Manipulations

Additional information about aircraft visible on the radar was provided throughout the scenario via two auditory channels, one for each ear. Audio messages included the identification number of a particular aircraft (e.g., *Track 132*) and information about a parameter that was not displayed on the visual interface: Either communication channel or flying pattern. Such information is known to be indicative of hostility in naval C2 environments and can be transmitted through auditory channels in the form of intelligence updates from another ship (Chalmers, Webb, & Keeble, 2002). In a similar manner to the parameters presented visually, each of the two parameters presented in the auditory modality could take either a threatening or a non-threatening value. Audio messages presented to the left channel conveyed information on whether the communication channel was open (non-threatening) or closed (threatening), for example, *"Track 383, communication open"*. Audio messages presented to the right ear indicated whether the flying pattern was showing no intent (non-threatening) or deception (threatening), e.g., *"Track 888, pattern deception"*. Each message was edited to last 2,750 ms and an interval of at least 4,000 ms separated each message in order to prevent overlaps within or between auditory channels. Most audio messages were delivered in a neutral fashion by a female voice (72% of the audio messages): Messages presented to the left channel were conveyed by Female A, while messages presented to the right channel were spoken by Female B. Occasionally, audio messages were delivered by a male voice in either of the two channels (28% of the audio messages).

These combinations yielded four types of audio messages: (a) messages with a threatening content conveyed by the male voice; (b) messages with a threatening content conveyed by a female voice; (c) messages with a non-threatening content conveyed by the male voice; and (d) messages with a non-threatening content conveyed by a female voice. During the simulation, all critical changes were accompanied by a message with a threatening content and spoken by the male voice (hereinafter referred to as an auditory warning message), acting as an auditory cue to warn participants that the aircraft's threat level had turned to hostile. Such auditory warning messages were exclusively presented when critical changes occurred, which made them perfectly reliable cues. The other types of audio messages were to be ignored by participants and were included as auditory distractors to replicate a realistic complex and dynamic environment in which operators have to differentiate tens, even hundreds, of auditory alarms that could share common features (Edworthy & Hellier, 2000; see also Momtahan, Hétu, & Tansley, 1993). These auditory distractors were not associated with non-critical changes or any particular event in the scenario. Audio messages conveyed information about two new parameters that were not presented in the visual parameters list and for which there were no visual representations on the radar. Auditory warning messages were specifically introduced in the microworld to warn participants about a critical change without having to detect it by applying the pre-set classification rule on aircraft parameters presented on the visual interface. These auditory cues were therefore designed to increase the speed and the probability to detect critical changes. They also provided information that was consistent with that presented in the parameter list (i.e., if participants clicked on the aircraft mentioned in the auditory warning message, they would also find its threat level being hostile if they applied the pre-set classification rule).

Auditory warning messages used in the present study were based on a specific combination of two message characteristics: a male voice *and* a threatening content. Unlike searching for a single attribute that could be performed by using automatic processes, searching for a conjunction of features requires focused attention in order to synthesize the incoming auditory information and identify the specific target (Treisman & Gelade, 1980). Given the fact that only 28% of audio messages were conveyed by the male voice, participants were required to notice both a change of voice and a threatening content in order to detect and respond to auditory warning messages. A male voice was used to convey auditory warning messages since neutral male voices are perceived as more urgent than neutral female voices (Edworthy, Hellier, Walters, Clift-Mathews, & Crowther, 2003). Audio messages parameters and aircraft threat level associated to each type of audio messages are summarized in Table 1. Overall, each scenario included the presentation of 8 auditory warning messages, 8 messages with a non-threatening content conveyed by the male voice, 14 messages with a threatening content conveyed by a female voice, and 26 messages with a non-threatening content conveyed by a female voice.

<Insert Table 1>

A critical change was considered detected if the aircraft mentioned in the auditory warning message was selected or classified within a 10-s interval following its presentation. This temporal window was selected to make sure only one critical change was presented during the selected time frame, and to ensure that most detections were captured, given the time required to hear and analyze the message, and to select the aircraft on the interface (or to classify the aircraft if it had already been selected when the critical change occurred). No visual cues indicating a critical change were presented on the radar within this temporal window. Of course, the heading of hostile aircraft changed at some point to hit the ownship, but this visual cue occurred at least 10 s after the presentation of the auditory warning message to ensure that our measure of change detection was not influenced by the presentation of any visual cue.

Procedure

Participants were presented with a PowerPoint tutorial, which they read through at their own pace, explaining the context of the simulation and providing instructions to complete the task. The four types of audio messages were explicitly presented to participants, along with their particular relevance for the task. They were told that they should be alert to auditory warning messages as they always indicated that a critical change had occurred and required further action, but that the other types of audio messages could be ignored. To check understanding of all the instructions before starting the experiment, participants were presented with three static screenshots from the microworld task and asked to perform the threat-level classification task and the threat-immediacy task, after which they familiarized themselves with the microworld's dynamic environment through two training sessions. Each training session lasted 16 minutes and comprised four scenarios. Participants were then asked to complete four test sessions presented in a counterbalanced fashion across participants and separated by a 5-min rest period. Each test sessions lasted 16 min and comprised four different scenarios (e.g., different parameter values and different trajectories) that were similar in terms of design and difficulty (e.g., number of audio messages, number of critical changes) and presented in a random order. Overall, a participant had to detect 128 critical changes (32 critical changes in each test session) among 400 non-critical changes (100 non-critical changes in each test session). A summary of instructions was presented on the screen at the beginning of each test session, and participants clicked a 'Continue' button to initiate the first scenario. After each test session, participants were administered the mental and the temporal demand subscales of the NASA-TLX subjective workload questionnaire (Hart & Staveland, 1988). More precisely, they were asked to answer two questions aloud on a 10-point Likert-type scale: i) "How mentally demanding was the task?"; and ii) "How hurried or rushed was the pace of the task?". The experimental session lasted about two hours.

RESULTS

The influence of auditory warning messages on change detection and decision making was investigated by using three categories of measures: i) critical change detection; ii) eye movement; and iii) threat-level classification. In all analyses, the alpha level was set to .05, and the Greenhouse-Geisser correction was applied when the sphericity criterion was not met. When post hoc analyses were performed, a Bonferroni correction was applied to prevent an increase of the familywise error rate. Additionally, normality was examined for all variables by assessing the kurtosis and skewness of the data. Following recommendations by Leech, Barrett, and Morgan (2005), the distribution of the variable was accepted as approximately normal if the absolute value of the statistic divided by the respective standard error was 2.5 or less. Based on that criterion, most variables satisfied the assumption of normality and were analyzed using parametric analyses. Three variables did not meet the normality assumption and were analyzed using non-parametric analyses.

Results of the present study were computed by aggregating data from the four test sessions. In order to ensure that they were similar in difficulty, we conducted a series of analyses to test whether there was any difference between test sessions regarding subjective (subscales of the NASA-TLX) and objective measures of difficulty (percentage of ship hits, percentage of correct threat-level classification). Table 2 reports mean and standard deviation associated with each measure as a function of test session. One-way repeated-measures analyses of variance (ANOVAs) revealed no difference between test sessions regarding perceived mental demand and perceived temporal demand (*F*s < 1). Similarly, Friedman tests showed no effect of test sessions on the percentage of ship hits, F_r (3, N = 22) = 2.78, p = .43, and the percentage of correct threat-level classification, F_r (3, N = 22) = 2.67, p = .45. These results confirmed that test sessions were similar in difficulty.

<Insert Table 2>

Change detection performance

Overall, 78% of critical changes were detected by participants, indicating that in the majority of cases, participants selected or classified (if already selected) the aircraft concerned within the 10-s interval following the auditory warning message. Given the still high percentage of undetected changes (22%), we were interested in determining whether the 10-s interval was sufficient to capture most detections. To this end, we computed change-detection performance using a 15-s interval, which is the longest post-change interval possible to avoid a potential overlap between two critical changes. The percentage of detected changes slightly increased when using a longer temporal window (from 78% to 85.23%), but most importantly, results showed that mean detection time was 5.5 s (*SD* = 0.57 s), and that among all critical changes that were detected, 81.3% of them were detected within the 10-s period following a critical change. In light of these results, a 10-s interval was considered to be an appropriate temporal window to detect critical changes. Given that the heading of hostile aircraft changed at some point following the 10-s interval to hit the ownship (introducing a visual cue on the radar), increasing the temporal window to 15-s would have introduced a confounding variable. It should be noted that change-detection performance was not interpreted in the present study per se. It was rather compared between experimental conditions.

Additional analyses were performed in order to determine whether participants really used auditory warning messages to perform the critical change-detection task. Given that critical changes were not followed by any visual cue on the radar during the 10s-interval following their presentation, participants who neglected to use auditory information to detect critical changes would have to regularly select all aircraft on the radar and reassess their threat level by applying the pre-set classification rule. The utilization of audio messages during the simulation was assessed by computing the percentage of audio messages that were followed within a 10-s interval by a selection or a classification on the corresponding aircraft. If participants correctly used audio messages to perform the change-detection task, then the percentage of actions (either a selection or a classification) performed on the corresponding aircraft should then be greater following male threatening messages (auditory warning messages) than to any other type of audio messages. In contrast, if participants did not use audio messages to perform the change-detection task, the percentage of actions performed on the corresponding aircraft should then be similar following any type of message. This hypothesis was tested by conducting a 2 (Voice: Female, Male) × 2 (Content: Non-threatening, threatening) repeated-measures ANOVA on the percentage of actions (selection or classification) performed on the corresponding aircraft within the 10-s interval following the audio message. Means and standard deviations are reported in Table 3. The analysis revealed that the percentage of actions was greater following messages spoken by the male voice than by a female voice, $F(1, 21) = 233.24$, $p < .001$, $\eta_p^2 = .92$, and greater following messages with a threatening content than messages with a non-threatening content, $F(1, 21) = 316.96$, $p < .001$, $\eta_p^2 = .94$. The two-way interaction was also significant, $F(1, 21) = 357.34$, $p < .001$, $\eta_p^2 = .94$. In line with our hypothesis, paired-samples *t*-tests revealed that actions performed on an aircraft following a male threatening message (auditory warning messages) were much more frequent than those following messages with a non-threatening content spoken by the male voice, $t(21) = 18.99$, $p < .001$, messages with a threatening content spoken by a female voice, $t(21) = 18.84$, $p < .001$, and messages with a non-threatening content spoken by a female voice, $t(21) = 19.61$, $p < .001$. Such results suggest that participants most likely used auditory cues to perform the threat-evaluation task and that they were able to respond differently to warning and distracting audio messages.

<Insert Table 3>

Change detection and oculometry

Pupil size and gaze position were measured in the current study to assess the possible interaction between attentional resources available during critical changes and the change-detection performance. This investigation was first conducted by computing the change in pupil size following the presentation of audio messages. The key analyses consisted in examining whether the pupillary response to audio messages varied according to their parameters (female/male voice and non-threatening/threatening content) and according to whether or not the critical change was detected by participants (in the case of an auditory warning message). A second step in this investigation consisted of testing if change detection was improved when participants' gaze position was close to the position of the hostile aircraft during the presentation of the auditory warning message. As a third and final step, we evaluated whether hostile aircraft were more likely to be detected if they were fixated (i.e. attended) just before the critical change than if they were not.

Pupillometry. The influence of audio messages on the pupil size was assessed by measuring the percentage of change in pupil size (PCPS) evoked by each type of audio message (see Beatty, 1982). PCPS was computed as the average pupil size within the 5-s interval following the onset of the audio message minus the average pupil size within the 5-s interval preceding the audio message (i.e., the baseline), divided by this same baseline. A temporal window of 5 s was used for our analysis to ensure that the difference in the average pupil size mostly reflected the processing of the auditory message rather than subsequent actions performed after the analysis of the message (e.g., threat-level classification, threat-immediacy classification, threat neutralization). Furthermore, this time frame was determined based on the fact that an audio message lasted 2,750 ms and that participants most likely needed additional time to analyze its content and determine whether it was threatening or not. Messages with a threatening content spoken by the male voice (auditory warning messages) were further divided into two categories (detected and undetected) in order to determine the influence of change-detection performance on the PCPS. The PCPS evoked by each type of audio message is depicted in Figure 2. A first 2 (Voice: Female, Male) × 2 (Content: Non-threatening, Threatening) repeated-measures ANOVA was performed on the PCPS using only male threatening messages that were detected. The analysis revealed main effects of Voice, $F(1, 21) = 42.82$, $p < .001$, $h_p^2 = .67$, but the main effect of Content was not significant, $F(1, 21) = 4.05$, $p = .06$, $h_p^2 = .16$. Most importantly, the two-way interaction was significant, $F(1, 21) = 6.92$, $p = .02$, $h_p^2 = .25$, indicating that messages with a threatening content produced larger PCPS than messages with a non-threatening content when they were spoken by the male voice, $t(21) = -2.51$, $p = .02$, but this difference was absent when the message was spoken by a female voice, $t(21) = 0.55$, $p = .59$. A second 2 × 2 repeated-measures ANOVA was conducted on the PCPS, but this time only male threatening messages that were undetected by participants were selected. Results showed that messages conveyed by the male voice produced a larger PCPS than those spoken by female voices, $F(1, 21) = 26.06$, $p < .001$, $h_p^2 = .55$. However, there was no significant effect of Content, $F < 1$, and no significant interaction between Voice and Content, $F < 1$. Taken together, these results showed that the male voice produced a greater PCPS than female voices regardless of message content, and that male threatening messages (auditory warning messages) produced a larger PCPS than any other type of message when detected by participants, but a similar PCPS to messages with a non-threatening content spoken by the male voice when not detected.

<Insert Figure 2>

Gaze position at the time of critical change. For all critical changes, we measured the distance separating participants' gaze position from the position of the hostile aircraft on the radar when the critical change occurred (also corresponding to the onset of the auditory warning message). The distance was then recoded as a categorical variable that could take three possible values (i.e., 0 to 199 pixels, 200 to 699 pixels, and 700 or more pixels). As shown in Figure 3, the percentage of undetected changes increased with the distance interval separating gaze position from the position of the hostile aircraft on the radar. A repeated-measures ANOVA confirmed the main effect of distance interval, $F(2, 42) = 8.28$, $p = .001$, $h_p^2 = .28$. Pairwise comparisons revealed that the percentage of undetected changes was lower when the distance interval was 0-199 pixels compared to 200-699 pixels ($p < .001$) or 700+ pixels ($p = .02$). However, there was no significant difference in the percentage of undetected changes between the 200-699 and the 700+ distance intervals ($p = 1$). Such results indicated that change detection was improved when the distance between gaze position and the position of the hostile aircraft was lower than 200 pixels during the critical change.

<Insert Figure 3>

Fixations on the aircraft before the critical change. Results showed that among all hostile aircraft that were fixated within 5 s before the critical changes, 85.3% were detected by participants, whereas this percentage dropped to 71.4% when hostile aircraft were not fixated before the critical changes. A chi-square of independence revealed a significant relationship between change detection and aircraft fixation prior to the change, $\chi^2(1, N = 2{,}816) = 79.04$, $p < .001$. A change was 1.19 times more likely to be detected if the hostile aircraft was fixated just before the critical change than if it was not fixated.

Audio messages and threat-level classification

A final objective of the study was to examine whether auditory warning messages were efficient in improving decision making in dynamic settings. One of the key tasks required by the participants during the simulation was to determine the threat level (non-hostile, uncertain, hostile) of all aircraft on the radar screen. Any further action performed on an aircraft depended directly on its threat level. Overall, we found that the percentage of correct threat-level classification was greater for hostile (95.8%; $SD = 1.4\%$) than for uncertain (79.0%; $SD = 16.2\%$) and non-hostile aircraft (88.3%; $SD = 16.8\%$). A Friedman test was conducted to assess differences in threat level since the normality assumption was not met. Results showed a main effect of threat level, $F_r(2, N = 22) = 22.18$, $p < .001$. Pairwise comparisons revealed that the percentage of correct classification was lower for uncertain aircraft than for hostile ($p < .001$) and non-hostile aircraft ($p < .001$). However, there was no significant difference between hostile and non-hostile aircraft ($p = .76$).

Based on these results, we were particularly interested in examining whether the percentage of correct classification for non-hostile and uncertain aircraft was influenced by the presentation of auditory distractors that could have produced a certain number of false alarms during the simulation. More precisely, we looked at whether participants were more likely to misclassify an aircraft as hostile (compared to non-hostile or uncertain) when that specific aircraft was mentioned in an auditory distractor within a 10-s period preceding the threat evaluation task. To this end, we examined the distribution of threat-level classification errors for non-hostile aircraft and found that 4.4% of these aircraft were misclassified as hostile aircraft when threat-level classification was not preceded by an auditory distractor related to the aircraft (the other 95.6% were misclassified as uncertain aircraft). This percentage increased to 10.1% when threat-level classification was preceded by an auditory distractor related to the aircraft (although such messages were not auditory warning messages). A non-hostile aircraft was therefore 2.30 times more likely to be misclassified as hostile aircraft if threat-level classification was preceded by an auditory distractors than if it was not, $\chi^2(1, N = 466) = 5.64$, $p = .02$. The same analysis was performed on uncertain aircraft and revealed that 50.7% of aircraft were misclassified as hostile when threat-level classification was not preceded by an auditory distractor (the other 49.3% were misclassified as non-hostile aircraft). This percentage increased to 64.0% when threat-level classification was preceded by an auditory distractor related to the aircraft. An uncertain aircraft was 1.26 times more likely to be misclassified as hostile if threat-level classification was preceded by an auditory distractor than if it was not, $\chi^2(1, N = 652) =$

10.91, $p = .001$. Taken together, these results showed that the very act of hearing an audio message related to a particular aircraft (albeit an auditory distractor) tended to bias threat evaluation towards perceiving the aircraft threat level as more hostile than it was in reality.

GENERAL DISCUSSION

The purpose of this study was to explore the benefits and the costs of using auditory warning messages to support change detection and decision making in dynamic settings. Although participants were successful in discriminating auditory warning messages and auditory distractors, results showed that a significant percentage of auditory warning messages remained undetected during the simulation (22%). Pupillometry and eye tracking were used in the current study to assess the possible relationship between change detection and attentional capacities. Our results revealed a greater pupillary response to auditory warning messages when critical changes were detected than when they were not, indicating that change detection most likely relies on attentional resources available during the critical change. We furthermore demonstrated that gaze position plays an important role in the detection of auditory events. Results showed that auditory change detection was facilitated when the mentioned aircraft was fixated just before the presentation of the auditory warning message or when gaze position happened to be close to the mentioned aircraft. The present study also provides evidence that using the auditory modality to support change detection can come with potential performance costs. When analyzing classification errors on the threat-evaluation task, we found that auditory distractors tended to bias threat evaluation towards perceiving aircraft as more hostile than they actually were.

The source of the inability to detect auditory information

Oculometric data were able to provide new insights into the nature of the attentional processes involved in auditory change detection. Pupillometry showed that messages presented by the threatening voice (the male voice in this experiment)—regardless of content—produced an orientation response (pupil dilation) similar to that found when presenting deviant background sounds (i.e., occasional sounds that are incongruent with the rest of the auditory stream and thus appear more salient, see, e.g., Maher & Furedy, 1979; Qiyuan, Richer, Wagoner, & Beatty, 1985; Steiner & Barry, 2011). Previous research in cognitive psychology has shown that when participants are asked to perform a visual serial recall task in the presence of background sound, recall performance is better when the task-irrelevant auditory stream comprises a single voice (e.g., male), compared to a stream that includes an embedded deviant voice (e.g., a unexpected female voice; Hughes et al., 2007; see also Vachon, Hughes, & Jones, 2012). The auditory deviant is thought to capture attention, drawing attentional resources away from the to-be-recalled material, thereby impairing recall performance. This explanation has received support from several studies showing that deviant sounds produce a pupillary response greater than repeated sounds (e.g., Liao, Kidani, Yoneya, Kashino, & Furukawa, 2016; Liao, Yoneya, Kidani, Kashino, & Furukawa, 2016; Steiner & Barry, 2011; Wetzel, Buttelmann, Schieler, & Widmann, 2015). In the current study, the probability of hearing a message conveyed by the male voice is quite low (28%) compared to that of a female voice (72%). Consequently, it is likely that messages conveyed by the male voice acted as deviant sounds, capturing participants' attention and producing a pupillary response. However, our results suggest that this automatic processing of the male voice was not necessarily sufficient for the critical changes to be reported; while critical changes did tend to generate a pupillary response, this pupil dilation was significantly greater in cases when the changes were actually reported than in cases that remained unreported. Based on Vachon, Vallières, and colleagues (2012), we propose that the critical voice generated a 'call for attention' for further processing (cf. Näätänen, 1990); however, this automatic call for attention remains sometimes unfulfilled in complex dynamic environments because attentional processes are overloaded. Non-detection of auditory warning messages would then be attributed to a failure of attentional processes. This interpretation is furthermore supported by our analysis of gaze position prior to critical changes showing that the probability of detection is greater when critical objects are fixated, and thus receive attentional processing, in the 5-s period prior to the auditory warning messages. This pre-processing of critical objects combined with the orientation response evoked by the critical voice was sufficient, in most cases, to enable a detection of critical changes.

The failure to notice the presence of auditory tones or alarms while performing a primary task has been studied as the inattentional deafness phenomenon. Such a phenomenon has been investigated mostly in static laboratory settings (e.g., MacDonald & Lavie, 2011; Raveh & Lavie, 2015), but it has been recently replicated in more realistic environments (Dehais et al., 2014; Giraudet, St-Louis, Scannella, & Causse, 2015). These studies suggest that the inability to detect auditory tones or alarms might be related to attentional capacities. There is a growing body of evidence showing that inattentional deafness is more likely to occur while performing tasks involving a high perceptual or cognitive load that consume most of attentional capacities, leaving few or none to consciously detect the alarms (e.g., Dehais et al., 2014; Giraudet et al., 2015, MacDonald & Lavie, 2011; Raveh & Lavie, 2015). The contribution of attentional processes to the inattentional deafness phenomenon was also investigated by Giraudet and colleagues (2015) by measuring event-related potentials during the presentation of auditory alarms. The authors reported a negative relationship between the individual number rate of undetected tones and the individual P300 amplitude (more specifically the P3b), a component associated with voluntary orientation of attention. In addition to showing that operators are vulnerable to such a phenomenon in complex dynamic environments, through the use of pupillometry, the current study provides further evidence that the inability to detect auditory information is most likely due to a failure of attentional processes.

Similarities between visual and auditory change detection

Based on our results, we were interested in examining whether the percentage of non-detection was similar to that found in exactly the same setting when the critical changes were promoted by visual cues on the radar screen (i.e., a marked increase in aircraft speed and/or a change in aircraft heading toward the ownship) rather than by audio messages (see Vachon, Vallières, et al., 2012). In order to compare the efficiency of auditory warning messages relative to visual cues to support change detection, we reanalyzed Vachon, Vallières, and colleagues' (2012) data using the same change detection criterion as in the current study (the authors used a 15-s interval). When contrasting the percentage of undetected changes in both studies with a Mann-Whitney test, we found that the percentage of non-detection observed in the current study (*Mdn* = 16.41) was not significantly different from that calculated from Vachon, Vallières, and colleagues' data (*Mdn* = 21.09), $U = 171.00$, $p = .32$. This result indicated that the inability to detect critical changes was equivalent in both experiments, regardless of whether they were cued by visual or auditory information.

In the visual domain, there is accumulating evidence that undetected changes might result from two distinct sources: i) a lack of attention, whereby the event is not detected because the operator is not attending; perhaps by looking elsewhere in the visual scene at the critical moment, and ii) a failure of attentional process, whereby information may still not reach conscious awareness even if the eye focuses directly on it (Vachon, Vallières, et al., 2012). Given the omnipresent nature of audio information, the first source of non-detection cannot be extended to the auditory domain since auditory detection is not dependent on the specific orientation of the participant's head during the critical change. Auditory information, however, appeared vulnerable to the second source of non-detection, that is, participants sometimes failed to adequately process and act upon the auditory warning message, even if it was clearly audible and their attention was captured by the male voice (as demonstrated by increased pupillary response). A further parallel between the current eye-tracking data and that obtained by Vachon, Vallières, et al. (2012) is in terms of gaze position. Critical changes (visual or auditory) are more likely detected if the participant fixates the aircraft concerned immediately prior to that change; this attentional pre-processing of the aircraft appears to circumvent that second source of detection failure and makes it more likely that change information will be adequately processed and actively reported.

According to the multiple resource theory of attention (Wickens, 2008), reducing the visual load on display by using auditory rather than visual cues to support change detection should increase the amount of attentional resources available for detecting critical changes, and therefore decrease the incidence of non-detection. Instead, our pattern of results points more toward a shared attentional capacity across vision and hearing (cf. Strayer & Johnson, 2001). That is, regardless of the modality of the cue, there was no difference in terms of change detection performance, and auditory presentation showed the same pattern of eye movement during detection, and the same source of non-detection (pupillometry data) as previously demonstrated in the visual modality. It is more likely that performing the threat-evaluation and weapon-assignment tasks in addition to the monitoring of the airspace consumed most of attentional capacity, leaving insufficient resources for detecting critical events, regardless of whether they were cued by visual or auditory information. The failure to detect critical events would then reflect a central limit of human cognition that is amodal in nature. This theory is supported by a growing body of research showing that the ability to detect an auditory tone depends on the perceptual load of the visual task. For example, it has been shown that participants engaged in low and high load versions of a visual discrimination task fail to notice the presence of an auditory tone significantly more often under the conditions of high workload (MacDonald & Lavie, 2011; Raveh & Lavie, 2015). These studies demonstrate load-induced cross-modal effects and provide evidence for a shared attentional capacity across vision and hearing.

Improving the efficiency of auditory alarms

In order to improve the detection rate of auditory messages, one possibility for future research might be to increase the urgency with which the message is spoken. Warning words spoken in an ‘urgent’ voice are perceived as being more believable and more demanding of attention than their ‘non-urgent’ counterparts (e.g., Edworthy et al., 2003). In addition, words spoken in an urgent voice show a higher attention-grabbing power than non-urgently spoken words (e.g., Ljungberg, Parmentier, Hughes, Macken, & Jones, 2012). In fact, Ljungberg and colleagues (2012) demonstrated that it is the urgency of the voice rather than the spoken content that determines attentional capture. While a subjective perception of urgency may have its benefits in improving detection rates, this demand for attention may also be a double-edged sword and have a negative impact on other aspects of the task. For example, one issue with using an urgent voice might be that participants in a decision-making task perceive degree of ‘urgency’ as necessarily equating to the degree of threat, resulting in a decision bias. In the current study, we found a similar decision bias even though auditory warning messages were not spoken in an urgent manner. Indeed, threatening value (and therefore perhaps higher perceived urgency) associated with the male voice may have led participants to perceive the aircraft concerned as more hostile than it actually was (when the content was non-threatening), thus leading to inaccurate hostile classifications. Of course, further research taking subjective ratings of urgency would be necessary to verify if the voice with a threatening value (the male voice in the current experiment) was necessarily equated with an increase in perceived urgency. If the mere presentation of an audio message spoken in a neutral voice can bias threat evaluation, it is plausible that such misperception of threat would increase if audio messages were spoken urgently. If one perceives a threat, there is then the tendency to search for confirmatory information in relation to it (Fischer et al., 2011), which in turn can influence decision making (Josephs, Larrick, Steele, & Nisbett, 1992). As such, switching to a confirmatory information search mode may bias one's subsequent decisions regarding perception of threat, and may result in more false alarms.

While verbal warnings may prove to be effective in supporting change detection under levels of low workload, in high workload situations—when attention must be divided between multiple sources—such warnings may be less than optimal if operators need to dedicate significant cognitive resources to the evaluation of the semantic content and voice of the messages. In the current experiment, the content of most of the tasks to perform by the operator (visual or auditory) were verbal in nature (e.g., aircraft parameters, aircraft identification, audio message content), what could have contributed to the lack of benefit of the auditory warning messages. A more appropriate design could reduce the cognitive load required by the environment by using, for example, non-verbal sounds. Although abstract tones and chimes may not provide the richness of information of speech, an alternative could be auditory icons, i.e. non-speech sounds that are ecologically associated with their referent processes (McKeown &

Isherwood, 2007). For example, the sound of liquid pouring would be semantically associated with low fuel, and may provide an alert that is processed more directly than voice messages. A recent study of alert types in self-driving cars found that both auditory icons and speech alerts resulted in self-reports of lower effort compared to visual notifications (Ness, Helbein, & Porter, 2016).

Gaze position appears to be another important factor in determining the efficiency of auditory alarms since detection rates improved when the target was fixated just before a critical change or when gaze was in close proximity to the target. While gaze position is determined to a certain extent by the user, it can also be influenced by the interface display and guided towards a given area; as such, the current work has implications for the development of adaptive systems based on eye-tracking to help reduce attentional errors. Attention aware systems assess a user's current attentional capacity—through gaze tracking or physiological measures (e.g., Chen & Vertegaal, 2004; Lieberman, Kramer, Montain, & Niro, 2007)—to optimize the timing/manner in which information is displayed (e.g., Roda & Thomas, 2006). If certain cognitive states are detected (e.g., attentional tunneling, inattentional blindness/deafness) then adaptions to the interface can be made to control what the user sees at critical points. For example, if an operator is experiencing attentional tunnelling and focuses on one aspect of the visual scene to the neglect of others, then the information upon which the operator is inappropriately focusing could be frozen or temporarily removed in order to break his/her perseverative behavior and redirect attention towards the most relevant activity at that time, potentially circumventing attentional errors (see, e.g., Dehais, Causse, & Tremblay, 2011). Promising results have identified physiological markers for attentional tunnelling (Dehais et al., 2011) and the level of fun in video games (Chamberland et al., 2015), as well as methods to approximate mental models of decision makers (Lafond, Tremblay, & Banbury, 2013) that are the core of adaptive systems and individually-tailored cognitive assistants.

Conclusion

The current study has revealed both benefits and costs associated with using audio messages to support change detection in a dynamic C2 environment. Our results clearly showed that a notable percentage of critical events (22%) might go unnoticed in such an environment, most likely attributed to a failure of attentional processes. While the non-detection of auditory warning messages are characterized by the absence of a behavioral response (the critical event went unreported), physiological measures demonstrated that such messages are nevertheless associated with a pupillary response (although to a lesser extent than a detected change). Thus in some cases the critical event underwent a degree of processing that may have generated a phenomenal awareness, but due to high load, there were insufficient attentional resources for the further sensemaking required to actively detect and respond to the event. Spreading attentional resources across modalities did not ease the burden on the visual domain and improve detection performance. While offloading to the auditory modality may benefit the first source of detection failure, the fact that auditory changes are susceptible to the second source of non-detection means that no improvement in the detection rate was observed relative to visual changes (Vachon, Vallières, et al., 2012). That is, attentional failure prevents the detection of changes, regardless of their sensory modality. Given the high percentage of undetected changes and the false alarms observed in the current study, using verbal audio messages might not be the optimal solution to support change detection in dynamic and complex environments.

ACKNOWLEDGMENTS

This research was supported by a collaborative research and development grant to Sébastien Tremblay and François Vachon from the National Sciences and Engineering Research Council of Canada (NSERC) in collaboration with Thales Canada, Defense and Security Division. Thanks are due to Marianne Couture and Marie-Hélène Bédard for assistance in data collection.

REFERENCES

Banbury, S., Fricker, L., Tremblay, S., & Emery, L. (2003). Using auditory streaming to reduce disruption to serial memory by extraneous auditory warnings. *Journal of Experimental Psychology: Applied, 9*, 12–22. doi:10.1037/1076-898X.9.1.12

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*, 276–292. doi:10.1037/0033-2909.91.2.276

Bregman, A. S. (1990). *Auditory Scene Analysis.* MIT Press: Cambridge, MA.

Chalmers, B. A., Webb, R. D., & Keeble, R. (2002). Modeling shipboard tactical picture compilation. *Proceedings of the Fifth International Conference on Information Fusion*, *2*, 1292-1299. doi:10.1109/ICIF.2002.1020962

Chamberland, C., Grégoire, M., Michon, P. E., Gagnon, J. C., Jackson, P. L., & Tremblay, S. (2015). A cognitive and affective neuroergonomics approach to game design. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *59*(1), 1075-1079.

Chen, D., & Vertegaal, R. (2004). Using mental load for managing interruptions in physiologically attentive user interfaces. In *CHI'04 extended abstracts on Human Factors in Computing Systems* (pp. 1513-1516). ACM.

Dalton, P., & Fraenkel, N. (2012). Gorillas we have missed: Sustained inattentional deafness for dynamic events. *Cognition, 124*, 367–372. doi:10.1016/j.cognition.2012.05.012

Dehais, F., Causse, M., & Tremblay, S. (2011). Mitigation of conflicts with automation. *Human Factors, 53*, 448–460. doi: 10.1177/0018720811418635

Dehais, F., Causse, M., Vachon, F., Regis, N., Menant, E., & Tremblay, S. (2014). Failure to detect critical auditory alerts in the cockpit: Evidence for inattentional deafness. *Human Factors*, *56*, 631–644. doi:10.1177/0018720813510735

Durlach, P. J. (2004). Change blindness and its implications for complex monitoring and control systems design and operator training. *Human Computer Interaction*, *19*, 423–451. doi:10.1207/s15327051hci1904

Edworthy, J., & Hellier, E. (2000). Auditory warnings in noisy environments. *Noise & Health, 2,* 27-39.

Edworthy, J., Hellier, E., Walters, K., Clift-Mathews, W., & Crowther, M. (2003). Acoustic, semantic and phonetic influences in spoken warning signal words. *Applied Cognitive Psychology*, *17*, 915–933. doi:10.1002/acp.927

Edworthy, J., Loxley, S., & Dennis, I. (1991). Improving auditory warning design: Relationship between warning sound parameters and perceived urgency. *Human Factors*, *33*, 205–231. doi:10.1177/001872089103300206

Eramudugolla, R., Irvine, D. R. F., McAnally, K. I., Martin, R. L., & Mattingley, J. B. (2005). Directed attention eliminates “change deafness” in complex auditory scenes. *Current Biology*, *15*, 1108–1113. doi:10.1016/j.cub.2005.05.051

Fischer, P., Kastenmüller, A., Greitemeyer, T., Fischer, J., Frey, D., & Crelley, D. (2011). Threat and selective exposure: The moderating role of threat and decision context on confirmatory information search after decisions. *Journal of Experimental Psychology. General*, *140*, 51–62. doi:10.1037/a0021595

Giraudet, L., St-Louis, M.-E., Scannella, S., & Causse, M. (2015). P300 event-related potential as an indicator of inattentional deafness? *PLoS One*, *10*. doi:10.1371/journal.pone.0118556

Gregg, M. K., & Samuel, A. G. (2008). Change deafness and the organizational properties of sounds. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 974–991. doi:10.1037/0096-1523.34.4.974

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds), *Advances in Psychology*, *52*. Human mental workload (pp. 139-183). doi:10.1016/S0166-4115(08)62386-9

Ho, C., & Spence, C. (2005). Assessing the effectiveness of various auditory cues in capturing a driver's visual attention. *Journal of Experimental Psychology: Applied*, *11*, 157–174. doi:10.1037/1076-898X.11.3.157

Hodgetts, H. M., Tremblay, S., Vallières, B. R., & Vachon, F. (2015). Decision support and vulnerability to interruption in a dynamic multitasking environment. *International Journal of Human-Computer Studies*, *79*, 106–117. doi:10.1016/j.ijhcs.2015.01.009

Hodgetts, H. M., Vachon, F., & Tremblay, S. (2014). Background sound impairs interruption recovery in dynamic task situations: Procedural conflict? *Applied Cognitive Psychology*, *28*, 10–21. doi:10.1002/acp.2952

Hoeks, B., & Levelt, W. J. (1993). Pupillary dilation as a measure of attention: A quantitative system analysis. *Behavior Research Methods, Instruments & Computers*, *25*, 16–26. doi:10.3758/BF03204445

Hollingworth, A., Schrock, G., & Henderson, J. M. (2001). Change detection in the flicker paradigm: The role of fixation position within the scene. *Memory & Cognition*, *29*, 296-304. doi:10.3758/BF03194923

Hughes, R. W., Vachon, F., & Jones, D. M. (2007). Disruption of short-term memory by changing and deviant sounds: Support for a duplex-mechanism account of auditory distraction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 1050–1061. doi:10.1037/0278-7393.33.6.1050

Josephs, R. A., Larrick, R. P., Steele, C. M., & Nisbett, R. E. (1992). Protecting the self from the negative consequences of risky decisions. *Journal of Personality and Social Psychology*, *62*, 26–37. doi:10.1037/0022-3514.62.1.26

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Lafond, D., Tremblay, S., & Banbury, S. (2013). Cognitive shadow: A policy capturing tool to support naturalistic decision making. In *Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), 2013 IEEE International Multi-Disciplinary Conference on* (pp. 139-142). IEEE.

Leech, N. L., Barrett, K. C., & Morgan, G. A. (2005). *SPSS for intermediate statistics: Use and interpretation*. Mahwah, NJ: Lawrence Erlbaum Associates.

Liao, H.-I., Kidani, S., Yoneya, M., Kashino, M., & Furukawa, S. (2016). Correspondences among pupillary dilation response, subjective salience of sounds, and loudness. *Psychonomic Bulletin & Review*, *23*, 412-425. doi:10.3758/s13423-015-0898-0

Liao, H.-I., Yoneya, M., Kidani, S., Kashino, M., & Furukawa, S. (2016). Human pupillary dilation response to deviant auditory stimuli: Effects of stimulus properties and voluntary attention. *Frontiers in Neuroscience*, *10*. doi:10.3389/fnins.2016.00043

Lieberman, H. R., Kramer, F. M., Montain, S. J., & Niro, P. (2007). Field assessment and enhancement of cognitive performance: Development of an ambulatory vigilance monitor. *Aviation, Space, and Environmental Medicine, 78,* B268-B275.

Ljungberg, J. K., Parmentier, F. B. R., Hughes, R. W., Macken, W. J., & Jones, D. M. (2012). Listen out! Behavioural and subjective responses to verbal warnings. *Applied Cognitive Psychology, 26*, 451–461. doi:10.1002/acp.2818

Macdonald, J. S. P., & Lavie, N. (2011). Visual perceptual load induces inattentional deafness. *Attention, Perception & Psychophysics*, *73*, 1780–1789. doi:10.3758/s13414-011-0144-4

Mack, A., & Rock, I. (1998). Inattentional blindness: Perception without attention. In *Visual Attention. Vancouver studies in cognitive science, Vol. 8* (pp. 55–76). New York, NY: Oxford University Press.

Maher, T. F., & Furedy, J. J. (1979). A comparison of the pupillary and electrodermal components of the orienting reflex in sensitivity to initial stimulus presentation, repetition, and change. In H. D. Kimmel, E. H. van Olst, & J. F. Orlebeke (Eds.), *The orienting reflex in humans* (pp. 381-391). Hillsdale, NJ: Erlbaum.

Marsh, J. E., Hughes, R. W., & Jones, D. M. (2008). Auditory distraction in semantic memory: A process-based approach. *Journal of Memory and Language*, *58*, 682–700. doi:10.1016/j.jml.2007.05.002

Marshall, S. P. (2007). Identifying cognitive state from eye metrics. *Aviation, Space, and Environmental Medicine, 78*(5), B165-175.

McCarley, J. S., & Kramer, A. F. (2008). Eye movements as a window on perception and cognition. In R. Parasuraman & M. Rizzo (Eds.), *Neuroergonomics: The brain at work* (pp. 95–112). New York, NY: Oxford University Press.

McKeown, D., & Isherwood, S. (2007). Mapping candidate within vehicle auditory displays to their referents. *Human Factors, 49*, 417–428. doi:10.1518/001872007X200067

Momtahan, K., Hétu, R., & Tansley, B. (1993). Audibility and identification of auditory alarms in the operating room and intensive care unit. *Ergonomics*, *36*, 1159-1176. doi:10.1080/00140139308967986

Morris, R. W., & Montano, S. R. (1996). Response times to visual and auditory alarms during anaesthesia. *Anaesthesia and Intensive Care*, *24*(6), 682–684.

Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related and other brain measures of cognitive function. *Behavioral and Brain Sciences*, *13*, 201–233. doi:10.1017/S0140525X00078407

Ness, M. A., Helbein, B. & Porter, A. (2016). Speech auditory alerts promote memory for alerted events in a video-simulated self-driving car ride. *Human Factors*, *58*, 416-426. doi:10.1177/0018720816629279

O'Regan, J. K., Deubel, H., Clark, J. J., & Rensink, R. A. (2000). Picture changes during blinks: Looking without seeing and seeing without looking. *Visual Cognition*, *7*, 191-211. doi:10.1080/135062800394766

Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition*, *17*, 398–422. doi:10.3758/BF03202613

Privitera, C. M., Renninger, L. W., Carney, T., Klein, S., & Aguilar, M. (2010). Pupil dilation during visual target detection. *Journal of Vision*, *10*, 1-14. doi:10.1167/10.10.3

Qiyuan, J., Richer, F., Wagoner, B. L., & Beatty, J. (1985). The pupil and stimulus probability. *Psychophysiology*, *22*, 530-534. doi:10.1111/j.1469-8986.1985.tb01645.x

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*, 1457–1506. doi:10.1080/17470210902816461

Raveh, D., & Lavie, N. (2015). Load-induced inattentional deafness. *Attention, Perception & Psychophysics*, *77*, 483–492. doi:10.3758/s13414-014-0776-2

Roda, C. & Thomas, J. (2006). Attention aware systems: Theories, applications, and research agenda. *Computers in Human Behavior*, *22*, 557–587. doi:10.1016/j.chb.2005.12.005

Roy, J., Paradis, S., & Allouche, M. (2002). Threat evaluation for impact assessment in situation analysis systems. In I. Kadar (Ed.), *Signal processing, sensor fusion, and target recognition XI; Vol. 4729. Proceedings of SPIE* (pp. 329-341). Orlando, FL. SPIE Press. doi:10.1117/12.477618

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inatentionnal blindness for dynamic events. *Perception*, *28*, 1059–1074. doi:10.1068/p281059

Stanton, N. A., & Edworthy, J. (1999). Auditory warnings and displays: An overview. In N. A. Stanton & E. Edworthy (Eds.), *Human factors in auditory warnings* (pp. 3–30). Aldershot, UK: Ashgate.

Steiner, G. Z., & Barry, R. J. (2011). Pupillary responses and event- related potentials as indices of the orienting reflex. *Psychophysiology*, *48*, 1648-1655. doi:10.1111/j.1469-8986.2011.01271.x

St. John, M., & Smallman, H. S. (2008). Staying up to speed: Four design principles for maintaining and recovering situation awareness. *Human Factors*, *2*, 118–139. doi:10.1518/155534308X284381

St. John, M., Smallman, H. S., & Manes, D. I. (2005). Recovery from interruptions to a dynamic monitoring task: The beguiling utility of instant replay. *Proceedings from the 49th Annual Meeting of Human Factors and Ergonomics Society* (pp. 473-477). Santa Monica, CA: Human Factors and Ergonomics Society. doi:10.1518/001872005774860014

Strayer, D. L., & Johnson, W. A. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone. *Psychological Science, 12*, 462-466. doi:10.1111/1467-9280.00386

Treisman, A. & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97-136. doi: 10.1016/0010-0285(80)90005-5

Vachon, F., Hughes, R. W., & Jones, D. M. (2012). Broken expectations: Violation of expectancies, not novelty, captures auditory attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 164-177. doi:10.1037/a0025054

Vachon, F., Vallières, B. R., Jones, D. M., & Tremblay, S. (2012). Nonexplicit change detection in complex dynamic settings: What eye movements reveal. *Human Factors*, *54*, 996–1007. doi:10.1177/0018720812443066

Vallières, B. R., Hodgetts, H. M., Vachon, F., & Tremblay, S. (2016). Support dynamic change detection: Using the right tool for the task. *Cognitive Research: Principles and Implications, 1*. doi:10.1186/s41235-016-0033-4

Vitevitch, M. S. (2003). Change deafness: The inability to detect changes between two voices. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 333–342. doi:10.1037/0096-1523.29.2.333

Wang, J. T.-Y. (2011). Pupil dilatation and eye tracking. In M. Schulte-Mecklenbeck, A. Kühberger, & R. Ranyard (Eds.), *A handbook of process tracing methods for decision research: A critical review and user's guide* (pp. 185–204). New York, NY: Psychology Press.

Wetzel, N., Buttelmann, D., Schieler, A., & Widmann, A. (2015). Infant and adult pupil dilation in response to unexpected sounds. *Developmental Psychobiology*, *58*, 382–392. doi:10.1002/dev.21377

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors*, *50*, 449–455. doi:10.1518/001872008X288394

Wogalter, M. S., & Young, S. L. (1991). Behavioural compliance to voice and print warnings. *Ergonomics*, *34*, 79–89. doi:10.1080/00140139108967290

AUTHOR BIOGRAPHIES

Cindy Chamberland is a Postdoctoral Fellow in cognitive and social neuroscience at Université Laval. She previously completed postdoctoral fellowships in human factors at the Institut Supérieur de l'Aéronautique et de l'Espace (2014), in applied cognitive psychology at Université Laval (2012-2013), and in cognitive sciences at University of Moncton (2011). She was the recipient of the 2013 *Mitacs Post-Doctoral Award for Outstanding Innovation.*

Helen M. Hodgetts is a Lecturer and Research Associate in experimental psychology/human factors at the School of Applied Psychology, Cardiff Metropolitan University, UK. She received her Ph.D. in cognitive psychology from Cardiff University in 2005 and was a postdoctoral researcher at Cardiff University (2005 – 2009; 2011 – 2012) and Université Laval (2009 – 2010).

Benoît R. Vallières is a Postdoctoral fellow in applied cognitive psychology at Université Laval. He received his Ph.D. in cognitive psychology from Université Laval in 2015. He also completed a research internship at Advanced Distributed Learning Co-Laboratories in Orlando in 2013. His research interests include change blindness, decision support system, and cognitive training.

François Vachon is an Associate Professor at the School of Psychology, Université Laval Québec, Canada. He received his Ph.D. in cognitive psychology from Université Laval in 2007. He then completed postdoctoral fellowships in cognitive psychology at Cardiff University (2007), in cognitive neuroscience at Université de Montréal, Canada (2008-2009), and in applied cognitive psychology at Université Laval (2009-2011).

Sébastien Tremblay is a Full Professor at the School of Psychology, Université Laval. He is also Honorary Professor of Cardiff University (UK) and scientific director of the Urban

Science Joint Research Unit. He has expertise in a wide range of cognitive human factors issues. He holds a Ph.D. in Psychology (1999, Cardiff University).

Table 1.
Audio message parameters and aircraft threat level associated to each type of audio messages.

Type of audio messages	**Audio message parameters**		**Aircraft threat level**
	Voice	Content	
Auditory warning messages	Male	Threatening (communication closed or pattern deception)	Hostile
Auditory distractors	Male	Non-threatening (communication open or pattern no intent)	Non-hostile/uncertain
Auditory distractors	Female	Threatening (communication closed or pattern deception)	Non-hostile/uncertain
Auditory distractors	Female	Non-threatening (communication open or pattern no intent)	Non-hostile/uncertain

Table 2.
Mean (+ standard deviation) associated with perceived mental demand (max = 10), perceived temporal demand (max = 10), percentage of ship hits, and percentage of correct threat-level classification as a function of test session.

	Test session			
Measures	1	2	3	4
	Subjective measures			
Perceived mental demand	6.01 (1.88)	6.14 (1.39)	5.99 (1.87)	5.98 (1.57)
Perceived temporal demand	5.90 (1.64)	5.97 (1.28)	5.80 (1.65)	5.91 (1.43)
	Objective measures			
Percentage of ship hits	5.97 (8.18)	5.82 (9.21)	4.83 (7.00)	6.11 (6.92)
Percentage of correct classification	86.69 (11.42)	87.96 (10.12)	87.60 (11.66)	87.92 (10.49)

Table 3.
Mean (+ standard deviation) associated with the percentage of audio messages that was followed within a 10-s interval by any action on the corresponding aircraft (selection or classification) as a function of message voice (female, male) and message content (non-threatening, threatening).

	Voice	
Content	Female	Male
Non-threatening	18.0 (4.6)	27.5 (10.1)
Threatening	19.7 (4.7)	78.0 (14.6)

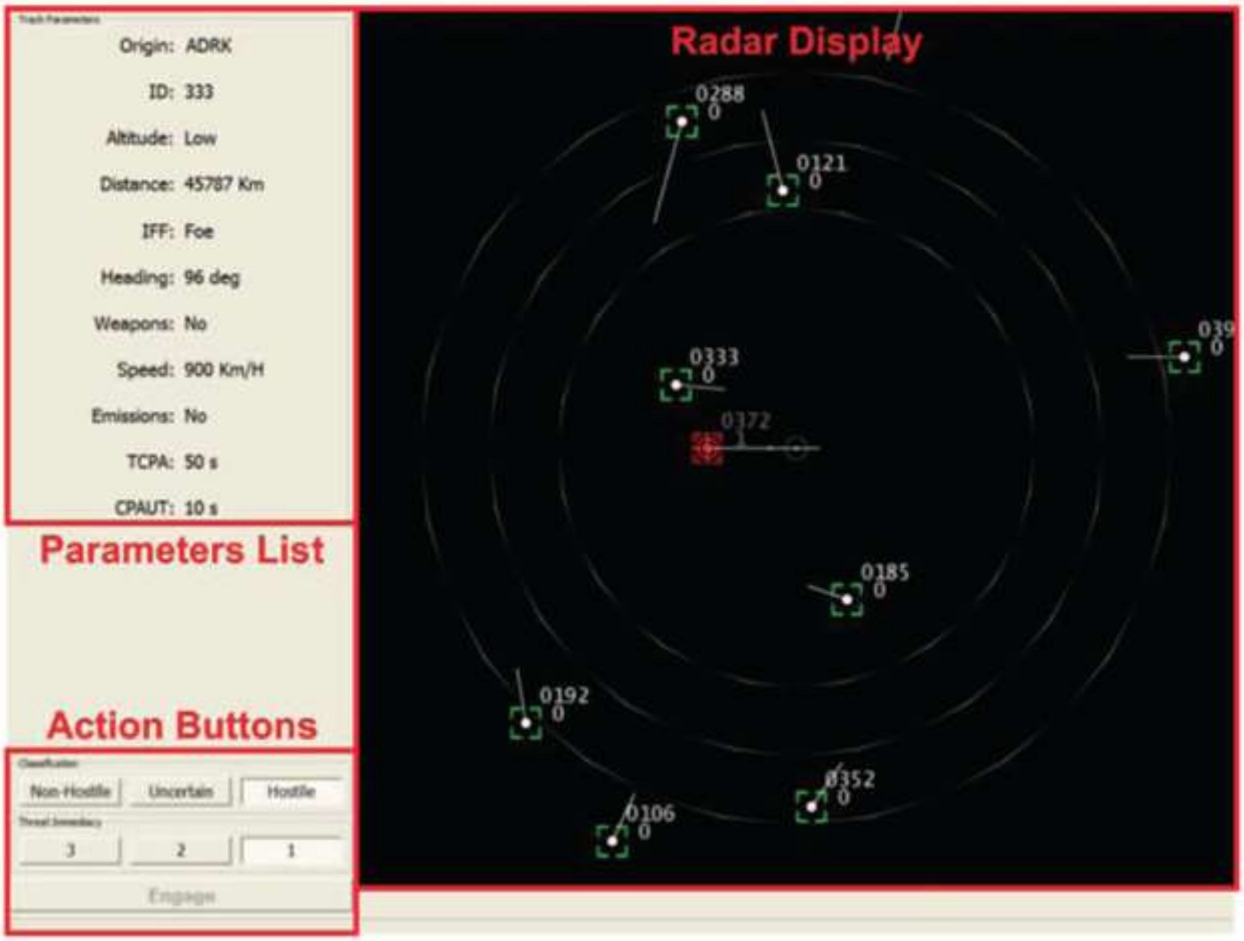

Figure 1. Screenshot of the Simulated Combat Control System (S-CCS) microworld visual interface. This interface can be divided into three parts: (a) a radar display depicting in real time all aircraft (represented by a white dot surrounded by a green square) moving at various speeds and trajectories around the ship (represented by the central point); (b) a list of parameters providing information about the selected aircraft; and (c) a set of action buttons allowing the participant to allocate threat level to an aircraft and to engage with missile fire a hostile aircraft.

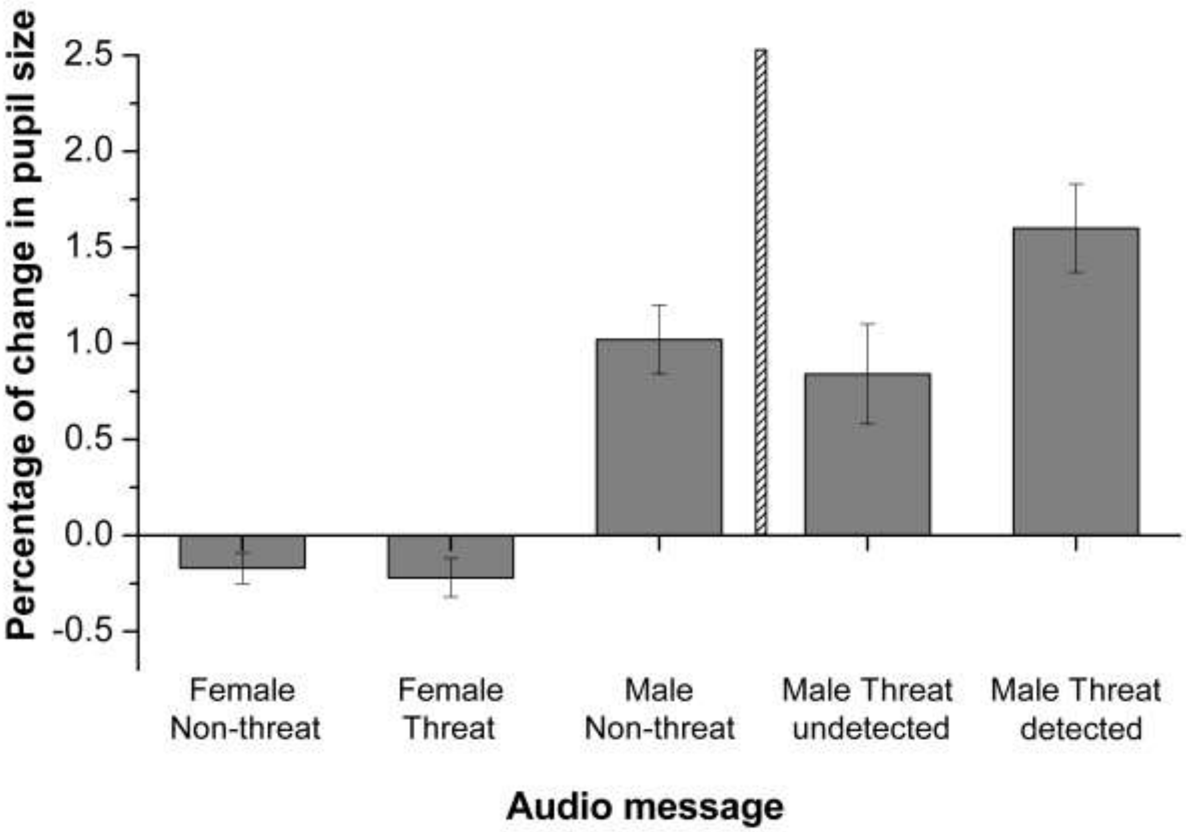

Figure 2. Average percentage of change in pupil size (PCPS) evoked by each type of audio messages. The PCPS is computed as the average pupil size within the 5-s interval after the onset of the audio message minus the average pupil size within the 5-s interval preceding the audio message (i.e., the baseline), divided by this same baseline. Error bars represent mean standard errors.

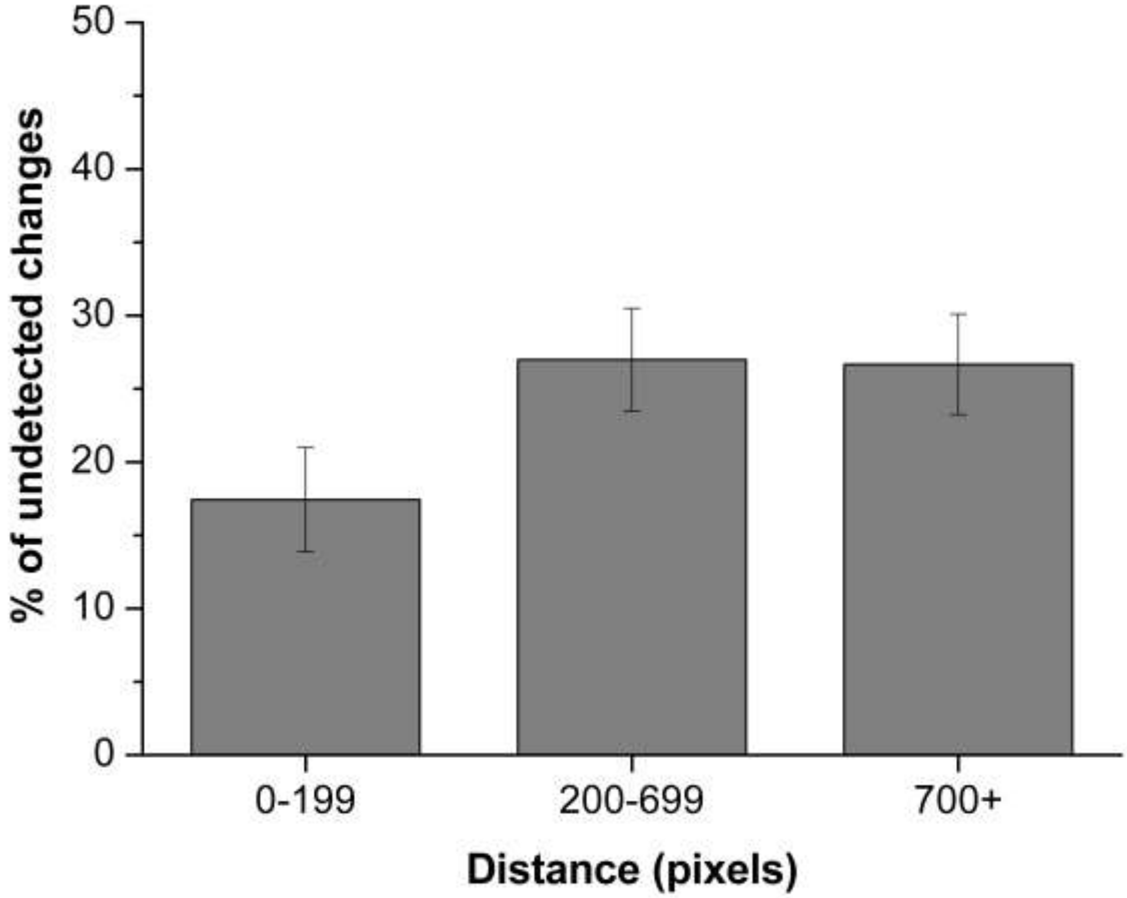

Figure 3. Percentage of undetected changes as a function of the distance interval separating gaze position from the position of the hostile aircraft during critical changes (0 to 199, 200 to 699, and 700 or more pixels). Error bars represent mean standard errors.