Fast & Scrupulous: Gesture-Based Alarms Improve Accuracy and Reaction Times Under Various Mental Workload Levels. An ERSP Study.

Eve Floriane Fabre [1], Bertille Somon [2], Valeria Baragona [1], Quentin Uhl [1] & Mickaël Causse [1]

[1] ISAE-SUPAERO, Neuroergonomics and Human Factors Research group, DCAS, Toulouse University, France

[2] ONERA, The French Aerospace Lab, Information Processing and Systems Department, Salon Cedex Air, France

*** Correspondence: Eve Floriane Fabre**

Keywords: Gesture alarms; Emergency situations; Operators; Safety; EEG; Time-Frequency Analysis.

Author Note

This work was supported by a Chair grant from Dassault Aviation ("CASAC", holder: Prof. Mickaël Causse) and a postdoctoral grant from the AXA Research Fund (awarded to Eve Fabre).

Abstract

In high-risk environments, fast and accurate responses to warning systems are essential to efficiently handle emergency situations. The aim of the present study was twofold: 1) investigating whether hand action videos (i.e., gesture alarms) trigger faster and more accurate responses than text alarm messages (i.e., written alarms), especially when mental workload (MWL) is high; and 2) investigating the brain activity in response to both types of alarms as a function of MWL. Regardless of MWL, participants (N = 28) were found to be both faster and more accurate when responding to gesture alarms than to written alarms. Brain electrophysiological results suggest that this greater efficiency might be due to a facilitation of the action execution, reflected by the decrease in mu and beta power observed around the response time window. These results suggest that gesture alarms may improve operators' performances in emergency situations.

1. INTRODUCTION

1.1 Reacting to emergency situations

In high-risk environments, such as transportation or nuclear power plants, fast and accurate reactions are essential to efficiently handle emergency situations and prevent potential catastrophic events. Since alarm systems play a critical role in triggering timely and appropriate responses in operators, particular care must be taken to ensure that their design maximizes the efficacy of the operator's response in the event of an emergency (e.g., Brown et al., 2000; Torralba et al., 2007; Wu & Li, 2018).

Alarm systems are expected to distract operators from the task currently at hand and provide them with crucial information such as: 1) the gravity of the failure, 2) its cause, 3) its location and even sometimes 4) how to respond to it (Guillaume, 2011). Because stimulating two or more sensory channels of information (e.g., visual, auditory and/or haptic) facilitates both the detection and the reaction to alarms, emergency alarm systems are almost always multimodal usually composed of an auditory component and a visual component (e.g., Alirezaee et al., 2017; Chan & Chan, 2006; Child & Wendt, 1938; Doyle & Snowden, 2001; Hughes et al., 1994; Liu, 2001; Schröger & Widmann, 1998). On the one hand, auditory alarms are very effective to capture operators' attention because they are omnidirectional and can be detected without having to gaze a specific area. On the other hand, visual alarms allow to provide more information to operators compared to auditory alarms (Guillaume, 2011; Stevens et al., 2009).

Emergency alarm systems generally combine symbolic (e.g., flashing lights, blares, or bips) and speech alarms (Guillaume, 2011). Symbolic alarms are very effective to attract the operator's attention, even though their use implies a thorough and repeated training of operators to create and sufficiently consolidate the association between a symbolic alarm and the particular event it signals to guarantee satisfactory levels of retention and detection (e.g.,

Momtahan, et al., 1993; Stevens et al., 2009). While compared to symbolic alarms, speech alarms tend to trigger slower reactions in operators (e.g., Wheale, 1982), they allow to convey complex information (Whittemore & Woods, 2021). They can either be *simple awareness warnings*, such as the Stall alarm, that indicate the nature of the emergency (e.g., in this case that the aircraft is about to enter a stall), but not how to respond to it (i.e., decreasing thrust and performing a nose-down pitch action; IATA, 2018a; Whittemore & Woods, 2021) or *pro-active instructions*, such as the Ground Proximity Warning System (GPWS; EASA, 2007), that indicate how to respond to the emergency (i.e., increasing thrust and performing a nose-up pitch action called "Pull Up") without explicitly stating its nature (i.e., the aircraft is about to hit the ground; IATA, 2018b; Moroze & Snow, 1999).

1.2 Why operators sometimes fail to appropriately react to emergency situations

Despite the important efforts that have been made to improve emergency alarm systems, failures to effectively respond to emergency alarms are still responsible for a significant number of accidents (Bliss, 2003; Kelly & Efthymiou, 2019; Scoot, 1996; Whittemore & Woods, 2021). For instance, pilots' delayed or inadequate reactions to the Stall or the GPSW alarms were found to be an important contributive factor in a significant part of respectively Controlled Flight Into Stall (CFIS; IATA, 2018a; Oliver et al., 2019; Sherry & Mauro, 2014) and Controlled Flight Into Terrain (CFIT; Cooper, 1995; Moroze & Snow, 1999; Shappell & Wiegmann, 2001; 2003) accidents (i.e., when an airworthy aircraft under the complete control of the crew is inadvertently flown into respectively the ground or a stall), accounting for many fatalities in commercial aviation (IATA, 2021; ICAO, 2019).

Ergonomics flaws are one of the reasons why operators do not always properly respond to emergency alarms (Scoot, 1996; Whittemore & Woods, 2021). For instance, symbolic auditory alarms become hard to recognize for operators when many of them are played concurrently (e.g., Causse et al., 2022; Momtahan, et al., 1993), and speech auditory alarms

can blend with operators' communications and not be detected or properly understood (Guillaume, 2011). Moreover, visual alarms can be difficult to detect due to their small size and the important amount of concurrent visual information (Corwin, 1995).

Mental workload (MWL) can also affect operators' response to emergency alarms (e.g., Bliss & Dunn, 2000; Bliss, 2003; Simonson et al., 2022). When performing a high MWL task, operators have to dedicate an important part of their cognitive resources to it, leaving few resources available to allocate to an incoming emergency alarm (Wickens, 2008; 2020). In these conditions, operators may not be able to detect and properly process an incoming alarm (Causse, Imbert et al., 2016; Causse et al., 2022; Giraudet et al., 2015), a phenomenon referred to as *inattentional deafness* for auditory stimuli (e.g., Dehais et al., 2019; Zhu et al., 2022) and *inattentional blindness* for visual stimuli (e.g., Lavie et al., 2014; Simons, 2000; White & O'Hare, 2022). Perceptual load was also found to affect the detection and processing of incoming alarms in a similar way as MWL (Jaquess et al., 2017; Lavie et al., 2014; MacDonald & Lavie, 2011; Molloy et al., 2015; Raveh & Lavie, 2015). The deleterious effects of perceptual load and MWL are likely to combine in high MWL situations as the latter are usually characterized by elevated levels of visual (e.g., numerous displays to monitor) and auditory (e.g., auditory alerts, communications) perceptual load (Causse et al., 2022) with potential catastrophic consequences (see for instance Eastern Air Lines Flight 401: NTSB, 1973)

Even though an alarm has been successfully detected and processed by an operator, different factors, such as stress or MWL, can still impair the cognitive ability and psychomotor skills of operators and disrupt the response to the alarm (Martin et al., 2015, 2016; Schmidt et al., 2008; Whittemore & Woods, 2021). The sudden and intense stress generated by emergency alarms can trigger inappropriate reactions in operators, such as freezing reactions (i.e., a defensive reaction consisting in a state of attentive immobility and

body tension; Roelofs, 2017) that can slow down or even inhibit operators' response to the alarm (DeCelles, 1991; Heaslip et al., 1991; Martin et al., 2012; Air Canada Flight 189: TSB, 1998), or startle reactions (i.e., "a simple reflex action that generally commences with an eye blink and develops into an aversive movement away from the stimulus"; Martin et al., 2015, 2016; Air France Flight 447: BEA, 2012; Colgan Air Flight 3407: NTSB, 2010). High levels of MWL (e.g., competing task) can also affect the execution of the response to emergency alarms (Bustamente et al., 2007; Shaw et al., 2018), with potential catastrophic consequences (Kuchar, 2001; NTSB, 2000).

1.3 Improving operators' reactions with gesture alarms?

Because operators have limited attention and working memory resources (Jaquess et al., 2017; Wickens, 2020), they should be presented with concise and easy-to-understand information (Endsley & Jones, 2004) that can spread out the workload and allow optimal responses (Bliss & Gilson, 1998). More specifically, Whittemore and Woods (2021) recommend using proactive instruction alarms (e.g., the "Pull Up" of the GPSW alarm) in emergency situations, instead of simple awareness warnings (e.g., Stall alarm).

Speech alarms are not the only way to communicate proactive instructions to the operators. Videos of the action to be performed in response to a specific emergency could also be used. Fabre and colleagues (2021) have recently proposed that, because action videos both strongly attract attention (Abrams & Christ, 2003) and preactivate the motor cortex (Rizzolatti et al., 1996), they might (compared to simple written speech alarms) facilitate both the detection of emergencies and the initiation of the expected motor responses in operators (Brucker et al., 2015; Iacoboni et al., 1999; Jeannerod, 1994; Meltzoff & Prinz, 2002).

To test this assumption, Fabre and colleagues (2021) conducted a two-experiment study investigating both 1) the response performances to a "Pull Up" hand action video (hereinafter referred to as gesture alarm) and the current GPWS written inscription and 2) the associated

brain electrophysiological responses to these two types of stimuli. Faster reaction times and greater mu and beta desynchronizations were found in response to gesture alarms than to the current GPWS written inscription, suggesting that the faster reaction to gesture alarms might result from a greater motor preparation and pre-activation of "Pull Up" recovery maneuver (Fadiga et al., 1995). No difference in response accuracy was found between the two types of alarms. The authors attributed this result to the fact that participants always had to execute the same action (i.e., Pull Up) in response to the two types of alarms, which might have triggered automatic responses and erased a potential difference in accuracy.

Emergency situations are usually associated with a high MWL, which is known to affect both the detection and the response to emergency alarms (Simonson et al., 2022). It is therefore essential to evaluate whether and to what extent high levels of MWL might affect the response to new emergency systems before they are implemented in operational environments. Even though the preliminary results of Fabre and colleagues are encouraging, the impact of MWL on operators' response performance (i.e., accuracy and reaction times) to gesture alarms compared to written inscription alarms remains unknown.

1.4 The present study

The present study aimed at investigating the efficiency of gesture alarms (compared to written alarms) under variable levels of MWL. Participants had to perform an *alarm-response task* drawn from the alarms (i.e., Stall and GPWS alarms) and the expected responses of two aviation emergency situations (i.e., incoming CFIS or CFIT). Participants had to respond to two opposite *gesture alarms* – i.e., videos of a hand either pushing or pulling a joystick – and the corresponding *written alarms* – i.e., written inscriptions of either the verb *"push"* or the verb *"pull"*. Participants had to respond to the alarms while performing: 1) no other competing task (no MWL), 2) a competing mental calculation task generating a low level of MWL and 3) a high level of MWL. Based on the results of Fabre et al. (2021), we predicted

to observe faster reaction times to gesture alarms than to written alarms and no difference in response accuracy between the two types of alarms in the no MWL run. We also predicted that 1) increasing participants' MWL would overall slowdown and negatively affect the accuracy of their responses to both types of alarms and that 2) participants would be faster and more accurate in responding to gesture alarms than to the written alarms as the level of MWL increased.

The present study also aimed at investigating the event-related spectral perturbations (ERSPs) observed in response to the gesture alarms and the written alarms as a function of the level of MWL (e.g., Behmer & Fournier, 2014). The high temporal and frequential resolution of ERSPs allows to investigate the brain processes associated with the processing of motor stimuli (Pfurtscheller & Da Silva,1999). Mu and beta desynchronization were repeatedly found at parietal electrodes (i.e., C3 and C4) when participants had to observe and/or imitate an action (e.g., Fox et al., 2016; Meyer et al., 2020; Pfurtscheller et al., 1997; Pineda, 2005; Quandt et al., 2012), or read action language (e.g. Moreno et al., 2013; Niccolai, et al., 2014), reflecting the preparation and/or the execution of actions and sometimes interpreted as reflecting the activity of mirror neuron system. We predicted that, similarly to when participants only had to observe the alarms (Fabre et al., 2021), greater mu and beta power decrease would arise in response to gesture alarms than to written alarms, reflecting the greater facilitation of the response preparation and/or execution triggered by gesture alarms (compared to written alarms). We also made the assumption that these differences in mu and beta power between the two types of alarms would vary as a function of the level of MWL (e.g., Behmer & Fournier, 2014; Jenson et al., 2019; Muthukumaraswamy & Singh, 2008).

2 MATERIAL AND METHODS

2.1 Participants

Based on the literature in the field, reporting in general between 17 and 37 participants for this type of experimental design with EEG measures (Mollo et al., 2016; Quandt et al., 2012), 28 participants (*M*age = 26, *SD* = 8, age range: 20 □ 48 years old; 12 females) took part in the present study. They were recruited thanks to paper ads that were disposed in various locations in the University campus. They were engineering students, PhD students or postdoctoral fellows. All had normal or corrected-to-normal vision and none of them reported a history of prior neurological disorder. All participants were informed of their rights and gave written informed consent for participation in the study.

2.2 Ethics Statement

This study was carried out in accordance with both the Declaration of Helsinki and the recommendations of the Research Ethics Committee of the University of Toulouse in France (CERNI no. 2018-107).

2.3 Material

2.3.1 Alarms

Six different stimuli were presented to the participants. The stimuli of interest were two *written alarms* consisting in a red written inscription of the verb "PULL" or "PUSH" (respectively Figures 1.A and 1.C.) and two *gesture alarms* consisting in videos of a hand pulling or pushing a joystick (respectively Figures 1.D & 1.F).

Insert Figure 1 here

To prevent automatic responses to the four alarms of interest, two filler alarms to which no response was expected were created. They consisted in a red written inscription of the verb

“STAY” (i.e., written filler) and a video of a motionless hand on a joystick (i.e., gesture filler; respectively Figures 1.B. & 1.E.).

2.3.2 Experimental apparatus

The experimental paradigm was presented using E-Prime 3 (Psychology Software Tools, Inc., Pittsburgh, PA) on a computer. Participants were seated in a chair and placed in front of a 19” monitor and had to respond to the alarms using a Logitech © Extreme 3D Pro joystick.

2.3.3 Electroencephalogram (EEG)

The EEG signals were recorded and amplified with the Biosemi© system (Amsterdam, The Netherlands; https://www.biosemi.com/products.htm) using 64 Ag-AgCl active scalp electrodes (Fp1, AF7, AF3, F1, F3, F5, F7, FT7, FC5, FC3, FC1, C2, C3, C5, T7, TP7, CP5, CP3, CP1, P1, P3, P5, P7, P9, PO7, PO3, O1, Iz, Oz, POz, Pz, CPz, Fpz, Fp2, AF8, AF4, AFz, Fz, F2, F4, F6, F8, FT8, FC6, FC4, FC2, FCz, Cz, C2, C4, C6, T8, TP8, CP6, CP4, CP2, P2, P4, P6, P8, P10, PO8, PO4, O2) arranged according to the international 10-20 system. Two electrodes were placed on the mastoids. Four external electrodes were placed around the eyes to record the vertical (two electrodes) and horizontal (two electrodes) electro-oculogram (EOG). Two additional electrodes placed close to Cz, the Common Mode Sense [CMS] active electrode and the Driven Right Leg [DRL] passive electrode, were used to form the feedback loop that drives the average potential of the participant as close as possible to the AD-box reference potential (Metting van Rijn, et al., 1990). Skin-electrode contact, obtained using electro-conductive gel, was monitored, keeping voltage offset from the CMS below 20 mV for each measurement site as recommended by the manufacturer. All the signals were (DC) amplified and digitized continuously with a sampling rate of 512 Hz with an anti-aliasing filter with 3 dB point at 104 Hz (fifth order sinc filter), no high pass filtering was applied online. The triggers indicating the onsets of the alarms were recorded on additional digital channels.

2.4 Tasks

2.4.1 Alarm response task

In the *alarm response task*, participants were presented with the four alarms of interest (i.e., push gesture, push written, pull gesture, pull written alarms) that were played 15 times each, and the two filler alarms (i.e., stay gesture, stay written alarms) that were played 30 times each. The alarms and the fillers were randomly presented to the participants for 4 seconds, with a variable inter-trial interval lasting between 4 and 6 seconds. Participants' task consisted in 1) responding to the alarms of interest according to the action they described (i.e., pushing the joystick in response the push gesture and push written alarms and pulling the joystick in response the pull gesture and pull written alarms), and 2) remaining still when filler alarms were displayed (i.e., stay gesture and stay written alarm).

2.4.2 Mental arithmetic task

The *mental arithmetic task* used in the present study was a subtraction task drawn from the Trier Social Stress Test (TSST; Birkett, 2011). Participants' task consisted in repeatedly subtracting either 5 in the low MWL run or 13 in the high MWL run from a fixed starting number (1022). Participants had to orally state their response and continue their subtraction loop until 1) they reached the closest number to 0; or 2) they made a mistake. In the former case (i.e., participants made no mistake), they were asked to perform the same task a second time (and sometimes even a third time) from a different starting number (i.e., 1023 the second time and 1024 the third time) until the end of the run (see section 2.6. for further detail). In the latter case (i.e., participants made a mistake), the experimenter notified the participants of their mistake, and the participants were required to start all over again from the initial number (e.g., 1022).

2.5 Data measures

2.5.1 Behavioral measurements

Alarm response task. Performances to the *alarm response task* were assessed in terms of accuracy (i.e., erroneous responses) and reaction times to correct responses. During the pre-tests, we found that the E-Prime default detection threshold of the responses was too sensitive for our experimental paradigm (i.e., micro-movements of the hand on the joystick were counted as responses). We determined that only considering the joystick inputs characterized by a vertical amplitude superior to 100 pixels above (i.e., pull) and below (i.e., push) the center of the screen (i.e., the default position of the joystick) allowed an optimal detection level of the responses on our 1920 x 1080 screen. As a result, only the joystick inputs superior to +/- 100 pixels on the vertical axis were considered to be responses to the alarms. Erroneous responses were computed as the sum of non-responses (i.e. when the participant did not respond to an alarm of interest) and inadequate responses (i.e., pulling in response to a push alarm and pushing in response to a pull alarm). Reaction times were defined as the time period between the onset of the alarm and the beginning of the participants' response. The reaction times to erroneous responses and those above or below two standard deviations around the mean were eliminated (Ratcliff, 1993).

Mental arithmetic task. Performances to the modified TSST in the low MWL run and the high MWL run were assessed in terms of 1) number of mathematical operations, 2) number of mathematical errors and 3) error ratio, which was computed as the number of mathematical errors divided by the total number of operations.

2.5.2 Subjective measures

At the end of each run, participants were asked to subjectively evaluate the difficulty of the run (i.e., performing both tasks irrespective of the type of alarm) they just completed by filling the six sub-scales of the NASA-TLX questionnaire which assess the perceived mental demand, the physical demand, temporal demand, performance, effort and frustration of participants (Hart & Staveland, 1988). Afterwards, they were asked to evaluate the extent to

which they found it difficult to respond to 1) the written alarms and 2) the gesture alarms on two different 7-point Likert scales (1 -> *not difficult at all* to 7 -> *extremely difficult*).

2.5.3 Brain electrophysiological measures

The Event-Related Spectral Perturbations (ERSPs) time-locked to the onset of written alarms and gesture alarms were investigated in all three levels of MWL (i.e., runs). To compute the latter, EEG data were analyzed using EEGLAB v.13.6.5b open-source software (Delorme & Makeig, 2004) on Matlab 2017a. First, EEG signals were band-pass filtered ([0.1-30] Hz) using a basic FIR filter. An independent component analysis was performed to isolate eye-blinks and movement-related artifacts that were removed from the signal (Jung et al., 1998). A visual inspection of the data was performed to reject residual artifacts intervals. Data were segmented into epochs from - 1000 to 2000 ms around the stimuli onset. Four participants were removed from the analysis due to the low quality of their EEG signal (less than 10 trials per condition). For the remaining 24 participants, epoched data were separated for each experimental condition according to the run (no MWL, low MWL and high MWL) and to the event type (gesture or written alarm). ERSPs were computed with Morlet wavelets as performed by EEGLAB, time-locked to the event display. They were computed with 30 frequencies in the 3-30 Hz window, three wavelet cycles at the lowest frequency, an increasing factor of 0.8 and 350 time-points in the - 441 to 1441 ms time-window (with 0 corresponding to the onset of the stimuli). Finally, as recommended ERSPs were baseline-corrected with a common ERSP baseline across factors and averaged separately for each condition across participants.

2.6 Procedure

After signing the inform consent form, participants were asked to take place in the experimental room. They sat in a comfortable seat in front of the 1920 x 1080 computer

screen. Participants were asked to read the instructions of the experiment, while the experimenter placed a 64-electrode Biosemi EEG cap on their head.

Participants were told that they would have to perform the *alarm response task* under three different levels of MWL: once without any competing task (i.e., no MWL run), once simultaneously to a moderately demanding *mental arithmetic task* (i.e., low MWL run), and once simultaneously to a highly demanding *mental arithmetic task* (i.e., high MWL run). Participants were explained that they were expected 1) to respond as accurately and as fast as possible − when a response was expected − to different alarms in the *alarm response task* and 2) to make as many operations as possible without committing any calculation error in the *mental arithmetic task*. It was specified to the participants that the two tasks were of equal importance. In each run, participants were presented with 60 stimuli of interest and 60 fillers.

At the end of each run (lasting approximately 14 minutes), participants were asked to fill the NASA-TLX questionnaire (Hart & Staveland, 1988) and the two 7-point Likert scales evaluating the difficulty to react to respectively gesture and written alarms during the run. In order to prevent potential biases, the order of the runs was counterbalanced across participants and the instructions for the *mental arithmetic task* were provided to the participants just before the beginning of each run. The experimenter was sitting next to the participants during the three runs (but not when participants were filling the scales), evaluating the correctness of the calculations that were being made. The proximity of the experimenter aimed at both simulating the presence of a second operator and increasing participants' MWL as a result of being watched (Belletier et al., 2015). At the end of the experiment, participants were debriefed by the experimenter and thanked for their participation in the study before they left the experimental room.

2.7 Data Analysis

2.7.1 Behavioral data analysis

Mental arithmetic task. Three 2 [MWL (low MWL, high MWL)] paired t-tests were conducted on 1) the number of mathematical operations, 2) the number of mathematical errors and 3) the error ratio (*number of errors/number of operations*).

Alarm response task. As none of the variables was normally distributed (Shapiro-Wilk tests: $p_s < .05$), a 3 x 2 [MWL (no MWL, low MWL, high MWL) x Alarm (written, gesture)] binary logistic regression was performed on the errors made in response to the alarms of interest (with 0 = correct response and 1 = erroneous response). A 3 x 2 [MWL (no MWL, low MWL, high MWL) x Alarm (written, gesture)] repeated-measures ANOVA analysis was conducted on the corrected log-transformed reaction times.

2.7.2 Subjective data analysis

NASA-TLX. A 3 [MWL (no MWL, low MWL, high MWL)] one-way repeated measures ANOVA was conducted on the raw NASA-TLX data in order to assess the perceived task difficulty in the three experimental runs (Hart, 2006). The NASA-TLX sub-scales were also analyzed separately. As most variables were not normally distributed (Shapiro-Wilk tests: $p_s < .05$), six 3 [MWL (no MWL, low MWL, high MWL)] one-factor Friedman's ANOVAs were conducted on the NASA-TLX sub-scales. Wilcoxon signed-rank tests with a Bonferroni-adjusted alpha level of .017 (.05 / 3 comparisons) were conducted as post-hoc tests.

Difficulty to respond to the alarms. A 3 x 2 [MWL (no MWL, low MWL, high MWL) x Alarm (written, gesture)] ordinal logistic regression was performed on the ratings of difficulty to respond to the alarm from 1 (easy) to 7 (difficult), as none of the variables (with the exception of the difficulty to respond to the written alarms in the high MWL, $p = .11$) was normally distributed (Shapiro-Wilk tests: $p_s < .05$). A manual stepwise analysis was performed to remove non-significant interactions from the model.

2.7.3 EEG data analysis

Mean ERSPs were analyzed with permutation statistics (N = 2000 permutations) and a False Discovery Rate (FDR) correction for multiple comparisons with the run (i.e., no MWL, low MWL and high MWL) and event type (i.e., gesture alarm, written alarm) as within-subject factors. Decibel-corrected averages associated with power increase (positive values) or decrease (negative values) were then extracted for each frequency band i.e., delta [3 – 4[Hz, theta [4 – 8[Hz, mu [8 – 12[Hz, low beta [12 – 15[Hz and medium/high beta [15 – 30[Hz according to the time-window of significance identified by the permutation test at the C3 and C4 electrodes. ANOVA and Tukey post-hoc analyses were performed on the average power across participants in the frequency bands and time-windows in which the EEGLAB permutation test graphs revealed a significant main effect of MWL (see Figure 7.C & D). Paired t-test analyses were performed on the average power across participants in the frequency bands and time-windows in which the EEGLAB permutation test graphs revealed a significant main effect of the type of alarm (see Figure 7.C & D).

3 RESULTS

3.1 Behavioral results

3.1.1 Mental arithmetic task

Participants made a significantly greater number of operations in the low MWL run [t (23) = 13.94, p < .001, $CI_{95\%}$ (225.15, 303.60; M = 406.12, SD = 112.23] than in the high MWL run (M = 141.75, SD = 66.44; see Figure 2.A.). They also made a significantly greater number of mathematical errors in the high MWL run made [t (23) = - 3.31, p = .003, $CI_{95\%}$ (- 6.10, - 1.41); M = 7.79, SD = 4.70] than in the low MWL run (M = 4.04, SD = 3.02; see Figure 2.B.). Finally, their error ratio was significantly greater in the high MWL run [t (23) = - 6.22, p < .001, $CI_{95\%}$ (- .07, - .04); M = .06, SD = .04] than in the low MWL run (M = .01, SD =.01; see Figure 2.C.).

Insert Figure 2 here

3.1.2 Alarm response task

Erroneous responses to the alarms. The analysis revealed a significant main effect of alarm [*B* (SE) = 2.833 (1.106), CI $_{(95\%)}$ = (.666, 5.001), Wald χ^2 (1) = 6.565, p = .010; see Figure 3.A.], with gesture alarms (*M* = 3.50 %, *SD* = 6.08) predicting for a lower error rate than written alarms (*M* = 6.40 %, *SD* = 7.78). The analysis also revealed a significant main effect of MWL [no MWL: *B* (SE) = - 4.679 (1.270), CI $_{(95\%)}$ = (- 7.169, - 2.189), Wald χ^2 (1) = 13.563, p < .001; low MWL*: B* (SE) = - 3.377 (1.362), CI $_{(95\%)}$ = (- 6.048, - .707), Wald χ^2 (1) = 6.143, p = .013; with high MWL as dummy; see Figure 3.B.]. The high MWL run (*M* = 7.60 %, *SD* = 8.94) predicted for a lower accuracy than both the low MWL run (*M* = 4.34 %, *SD* = 5.50, p = .013) and the no MWL run (*M* = 2.91 %, *SD* = 5.64, p < .001), but no significant difference was found between the no MWL run and the low MWL run (p = .056).

Insert Figure 3 here

Reaction times to the alarms. The analysis revealed a significant main effect of alarm [*F* (1, 23) = 34.56, p < .001, ηp^2 = .60; see Figure 4.A.] with faster reaction times to gesture alarms (*M* = 930 ms, *SD* = 259) than to written alarms (*M* = 1140 ms, *SD* = 302). A significant main effect of MWL was also found [*F* (2, 46) = 46.75, p < .001, ηp^2 = .67; Figure 4.B.], with longer reaction times in the high MWL run (*M* = 1228 ms, *SD* = 330, p < .001) and in the low MWL run (*M* = 1013 ms, *SD* = 236, p < .001) than in the no MWL run (*M* = 864 ms, *SD* = 200); and in the high MWL run than in the low MWL run (p < .001). The MWL x Alarm interaction did not reach significance [*F* (2, 46) = 1.06, p = .35, ηp^2 = .04].

Insert Figure 4 here

3.2 Subjective results

3.2.1 NASA-TLX questionnaire.

The analysis revealed a significant main effect of MWL [F (2, 46) = 109.89, $p < .001$, ηp^2 = 1.00; see Figure 5], with greater task difficulty perceived in high MWL run (M = 11.41, SD = 2.17, $p_s < .001$) than in both the low MWL run (M = 9.89, SD = 2.28) and the no MWL run (M = 6.01, SD = 1.43); and in the low MWL run than in the no MWL run ($p < .001$).

Insert Table 1 here

For the sake of clarity, the results of the statistical analysis conducted on the NASA-TLX subscales, the means and standard deviations are reported in Table 1.

Insert Figure 5 here

3.2.2 Difficulty to respond to the alarm.

The analysis revealed a significant main effect of MWL [no MWL: B (SE) = - 1.625 (.244), CI $_{(95\%)}$ = (- 2.103, - 1.147), Wald χ^2 (1) = 44.371, $p < .001$; low MWL: B (SE) = - .833 (.244), CI $_{(95\%)}$ = (- 1.311, - .355), Wald χ^2 (1) = 11.67, $p < .001$; with high MWL as dummy; Figure 6A], with the no MWL run (M = 2.23, SD = 1.13) predicting for a lower difficulty to respond to the alarm than both the low MWL run (M = 3.02, SD = 1.49, $p < .001$) and the high MWL run (M = 3.85, SD = 1.48, $p < .001$); and the low MWL run than the high MWL

run ($p < .001$). The analysis also revealed a significant main effect of type of alarm [B (SE) = 1.325 (.157), CI $_{(95\%)}$ = (1.016, 1.634), Wald χ^2 (1) = 70.735, $p < .001$; Figure 6B] with gesture alarms (M = 2.37, SD = 1.41) predicting for a lower difficulty to respond than written alarms (M = 3.70, SD = 1.34).

Insert Figure 6 here

3.3 Brain electrophysiological results

For the sake of clarity, the results of the effect of MWL and the type of alarm are reported respectively in Table 2 and Table 3, as well as displayed on Figure 7.

3.3.1 Effect of the MWL

Permutation tests revealed a significant main effect of MWL at both the C3 and C4 electrodes in the low and high beta frequency bands (all F_s (2, 46) > 11; all p_s <.001 and all ηp^2 > .34; see Figure 7.C & D). Post-hoc analyses revealed that at both electrodes, the average power was significantly lower in the no MWL run than in both the low MWL and high MWL runs (all p_s < .001; see Table 2). No difference was observed between the low MWL and high MWL conditions.

Insert Tables 2 & 3 here

3.3.2 Effect of the type of alarm

Permutation tests revealed a significant main effect of the type of alarm at both C3 and C4 in the delta, theta and mu frequency bands. Post-hoc analyses revealed an increase in the delta

band, whereas we observed a decrease in both the theta and mu frequency bands for gesture compared to written alarms (all $p_s < .05$; see Table 3 for the respective values and time windows). Finally, permutation tests also revealed a significant main effect of alarm in the low beta frequency band, but only at the C3 electrodes, with a lower activity for gesture compared to written alarms ($p < .05$; see Table 3 for the values and time window).

Insert Figure 7 here

4 DISCUSSION

The aim of the present experiment was to investigate the efficiency of gesture alarms (compared to written inscriptions) under different levels of MWL at the subjective, behavioral and physiological levels. Participants performed an alarm response task concurrently to a mental subtraction task inducing several MWL levels.

4.1 Manipulation check of the mental arithmetic task

In line with literature, participants' performance decreased (i.e., less mathematical operations, more calculation errors and higher error ratio) in the high MWL run compared to the low MWL run (Birkett, 2011; Linares et al., 2020). The analysis of the NASA-TLX ratings (Hart, 2006) also revealed that participants reported significantly higher levels of perceived task difficulty during the high MWL run than during both the low MWL and the no MWL runs; and during the low MWL run than the no MWL run. Taken together, these results serve as a manipulation check and confirm that the level of MWL significantly increased between the no, the low and the high MWL runs (Feltman et al., 2022; Wickens, 2020).

4.2 Impact of MWL on the response to the alarms

While participants took significantly longer and found it more difficult to respond to the alarms in the low MWL run than in the no MWL run, no difference in error rates to the alarm-

response task was found between these two runs. These results suggest that even though the increase in MWL between the no MWL run and the low MWL run slowed participants' reactions, enough attentional and cognitive resources were still available in the low MWL run to maintain a similar level of response accuracy as in the no MWL run (De Sanctis et al., 2014; Wickens, 2008). In the high MWL run, participants took significantly longer, made more erroneous responses, and found it harder to respond to both types of alarm than in both the no MWL run and the low MWL run. First, the general decrease in performance to both the *mental arithmetic task* (i.e., number of operations and number of errors) and the *alarm-response task* (i.e., both accuracy and speed; in line with literature; e.g., Causse et al., 2022; Giraudet et al., 2015) show that participants followed the instructions they were given, that is to consider the two tasks as of equal importance and not prioritize one over the other. Second, these results also suggest that the reserve of attentional and cognitive resources was not sufficient to enable the completion of the *alarm-response task* in the high MWL run with the same level of performance, as in the no MWL and low MWL runs (Jaquess et al., 2017; Wickens, 2008; 2020), resulting in a significant drop in response accuracy and longer reactions times (in line with De Sanctis et al., 2014; Diekfuss et al., 2017; Shaw et al., 2018).

At the brain electrophysiological level, greater beta activity was found in response to the alarms in general in the low MWL and high MWL runs than in the no MWL run, along the whole epoch. Greater beta activity was found to reflect increased task demand/difficulty, (mental) workload and/or (mental) fatigue (e.g., Kumar & Kumar, 2016; Michels et al., 2010; Morales et al., 2019; Okogbaa et al., 1994; Rietschel et al., 2012). Therefore, the greater beta activity observed in response to the alarms in both the low and high MWL runs may reflect the increase in task demand triggered by the mental arithmetic task. The fact that the difference in beta activity was found on the whole epoch (even before the alarms were presented) suggests that this increase in task demand might have been sustained along the

whole run, confirming once again the efficiency of this mental arithmetic task to trigger MWL in participants.

4.3 Gesture alarms versus written alarms

Participants were found to be both faster and more accurate when responding to the gesture alarms than to the written alarms (in line with Fabre et al., 2021), regardless of the level of MWL. They also reported that they found it easier to respond to the gesture alarms than to the written alarms, once again regardless of the level of MWL. Taken together, these results suggested that gesture alarms might be more efficient to trigger fast and accurate responses than simple written instructions, even in high MWL situations.

In line with our predictions and the previous results of Fabre and colleagues (2021), greater mu and low beta power decreases were found in response to gesture alarms than to written alarms. Mu and beta power desynchronizations are known to reflect the preparation and the execution of actions (e.g., Brinkman et al., 2016; Fox et al., 2016; Meyer et al., 2020; Pfurtscheller et al., 1997; Pineda, 2005; Quandt et al., 2012). The results of the present study indicate that, compared to written alarms, gesture alarms might facilitate the preparation and/or the execution of the response. A greater power decrease in the mu frequency band was found for gesture alarms than for written alarms in time-window of the response to the alarms (i.e., on average 930ms for the gesture alarms and 1140ms for the written alarms). These results suggest that the greater decrease in power observed in response to gesture alarms is more likely to reflect the facilitation of the action execution, rather than the facilitation of its preparation. However, since we were not able to investigate the signal associated with the onset of the response observed for both gesture and written alarms (i.e., no triggers were sent when participants performed the action), we cannot conclude with certainty on this point.

The results also suggest a greater high delta (3-4 Hz) synchronizations in response to the gesture alarms than to the written alarms. Increased delta activities were found to underpin

action selection and/or execution in cognitively demanding situations (Beste et al., 2012) and to be correlated to performance at action selection (Hamel-Thibault et al., 2018). In the present study, the greater high delta activities observed in response to gesture alarms than to written alarms occurring before the execution of the action could reflect the facilitation of the action selection in response to gesture alarms. However, this result should be taken with caution because the present study was not designed to investigate the delta rhythm. First, the time-frequency tradeoff intrinsic to the definition of Morlet wavelets did not create a sufficient precision in the frequency domain to analyze very low frequency activities with three cycles. Second, this small number of cycles which is required to reach a sufficient time-domain accuracy at low frequencies induces a frequency leakage from close-by frequencies when transposed to the frequency domain (Cohen, 2014). Consequently, the delta power measured by the EEG could have been polluted by surrounding (e.g., theta) frequency activities. Further research is therefore necessary to confirm this potential difference in delta synchronization between gesture and written alarms.

4.4 Applications, limitations and future works

Compared to written instruction alarms, gesture alarms were found to facilitate the initiation of the expected response and to trigger faster and more accurate responses under all three MWL levels. The results of the present study therefore suggest that gesture alarms could be a suitable solution to improve both the visual modality of emergency alarms and the operators' performances, even under high levels of MWL (Bliss & Gilson, 1998; Endsley & Jones, 2004).

In the context of aviation, gesture alarms could be used to improve the GPWS alarm, and more importantly the Stall alarm (Whittemore & Woods, 2021). While the maneuver that has to be performed in reaction to the Stall alarm is both stressful and counter-intuitive for pilots (BEA, 2012; IATA, 2018a; NTSB, 2010), the current Stall alarm consists in a simple

awareness warning (i.e., only indicating the dangerous deviation of a parameter; Ewbank, et al., 2016). Replacing the visual modality of the Stall alarm by a gesture alarm (i.e., hand video pushing on the stick) might significantly enhance its design and improve pilots' response performance to it.

The implementation of gesture alarm systems might also help improve the reactions of operators in many other fields where fast and accurate responses are needed. Gesture alarms may be particularly adapted to nuclear powerplants and transportation in general (i.e., air, rail, or road transport), to improve emergency shutdown or collision avoidance maneuvers for instance. They might also be used in the medical field (Alizezaee et al., 2017; Edworthy, 2013).

However, further research is needed before gesture alarms could be implemented in real operational environments. First, while most alarm systems are multimodal, we focused exclusively on the visual component of the alarms. While, it is very likely that adding an auditory component will improve even more the operators' response performances to gesture alarms (Alirezaee et al., 2017; Child & Wendt, 1938; Hughes et al., 1994; Liu, 2001), it remains to be demonstrated. More specifically, new studies should be conducted to determine which auditory component among symbolic sounds, simple auditory awareness warnings (e.g., "stall"), auditory pro-active instructions (e.g., "push down") or a combination of these auditory alarms (Whittemore & Woods, 2021) would maximize the efficiency of gesture alarms. Second, the fact that gesture alarms were not tested on real-life operators in their usual work environment is another important limitation of the present study. Before gesture alarms could be implemented in work environments, they should be thoroughly tested on trained operators in the different operational contexts in which they could be implemented.

Conclusion

The behavioral results of the present study demonstrate that compared to written instructions gesture alarms trigger faster and more accurate responses, regardless of the operators' level of MWL. The brain electrophysiological results suggest that this greater efficiency might be due to the fact that (compared to written alarms) gesture alarms facilitate the execution (and also possibly the preparation) of the associated action. While further research is necessary to confirm the greater efficiency of gesture alarms in real operational environments, our study provides a step forward towards the implementation of such gesture-based warnings to improve operators' performances in emergency situations.

REFERENCES

Abrams, R. A., & Christ, S. E. (2003). Motion onset captures attention. *Psychological Science*, *14*(5), 427–432. https://doi.org/10.1111/1467-9280.01458.

Alirezaee, P., Girgis, R., Kim, T., Schlesinger, J. J., & Cooperstock, J. R. (2017). Did you feel that? Developing novel multimodal alarms for high consequence clinical environments. *The 23rd International Conference on Auditory Display*, 175 – 181. https://doi.org/10.21785/icad2017.066.

BEA (2012). Final Report: On the accident on 1st June 2009 to the Airbus A330-203 registered F-GZCP operated by Air France flight AF 447 Rio de Janeiro – Paris. https://www.bea.aero/docspa/2009/f-cp090601.en/pdf/f-cp090601.en.pdf

Behmer Jr, L. P., & Fournier, L. R. (2014). Working memory modulates neural efficiency over motor components during a novel action planning task: An EEG study. *Behavioural brain research*, *260*, 1-7. https://doi.org/10.1016/j.bbr.2013.11.031

Belletier, C., Davranche, K., Tellier, I. S., Dumas, F., Vidal, F., Hasbroucq, T., & Huguet, P. (2015). Choking under monitoring pressure: being watched by the experimenter reduces executive attention. *Psychonomic bulletin & review*, *22*(5), 1410-1416. https://doi.org/10.3758/s13423-015-0804-9

Beste, C., Ness, V., Lukas, C., Hoffmann, R., Stüwe, S., Falkenstein, M., & Saft, C. (2012). Mechanisms mediating parallel action monitoring in fronto-striatal circuits. *Neuroimage*, *62*(1), 137-146. https://doi.org/10.1016/j.neuroimage.2012.05.019

Birkett, M. A. (2011). The Trier Social Stress Test protocol for inducing psychological stress. *JoVE (Journal of Visualized Experiments)*, (56), e3238. https://doi.org/10.3791/3238.

Bliss, J. P. (2003). Investigation of alarm-related accidents and incidents in aviation. *The International Journal of Aviation Psychology*, *13*(3), 249–268. https://doi.org/10.1207/S1532 7108IJAP13 03_04.

Bliss, J. P., & Dunn, M. C. (2000). Behavioural implications of alarm mistrust as a function of task workload. *Ergonomics*, *43*(9), 1283-1300. https://doi.org/10.1080/001401300421743

Bliss, J. P., & Gilson, R. D. (1998). Emergency signal failure: Implications and recommendations. *Ergonomics*, *41*(1), 57-72. https://doi.org/10.1080/001401398187323

Brinkman, L., Stolk, A., Marshall, T. R., Esterer, S., Sharp, P., Dijkerman, H. C., ... & Toni, I. (2016). Independent causal contributions of alpha-and beta-band oscillations during movement selection. *Journal of Neuroscience*, *36*(33), 8726-8733. https://doi.org/10.1523/JNEUROSCI.0868-16.2016

Brown, W. S., O'Hara, J. M., & Higgins, J. C. (2000). *Advanced alarm systems: Revision of guidance and its technical basis* (Rep. BNLNUREG-52593). Upton, NY, Brookhaven National Lab.

Brucker, B., Ehlis, A. C., Häußinger, F. B., Fallgatter, A. J., & Gerjets, P. (2015). Watching corresponding gestures facilitates learning with animations by activating human mirror-neurons: An fNIRS study. *Learning and Instruction*, *36*, 27-37. https://doi.org/10.1016/j.learninstruc.2014.11.003

Bustamante, E. A., Bliss, J. P., & Anderson, B. L. (2007). Effects of varying the threshold of alarm systems and workload on human performance. *Ergonomics*, *50*(7), 1127-1147. https://doi.org/10.1080/00140130701237345

Causse, M., Parmentier, F. B., Mouratille, D., Thibaut, D., Kisselenko, M., & Fabre, E. (2022). Busy and confused? High risk of missed alerts in the cockpit: an electrophysiological study. *Brain research*, *1793*, 148035. https://doi.org/10.1016/j.brainres.2022.148035

Causse, M., Imbert, J. P., Giraudet, L., Jouffrais, C., & Tremblay, S. (2016). The role of cognitive and perceptual loads in inattentional deafness. *Frontiers in human neuroscience*, *10*, 344. https://doi.org/10.3389/fnhum.2016.00344

Causse, M., Peysakhovich, V., & Fabre, E. F. (2016). High working memory load impairs language processing during a simulated piloting task: an ERP and pupillometry study. *Frontiers in human neuroscience*, *10*, 240. https://doi.org/10.3389/fnhum.2016.00240.

Chan, A. H., & Chan, K. W. (2006). Synchronous and asynchronous presentations of auditory and visual signals: Implications for control console design. *Applied Ergonomics*, *37*(2), 131-140. https://doi.org/10.1016/j.apergo.2005.06.006

Child, I. L., & Wendt, G. R. (1938). The temporal course of the influence of visual stimulation upon the auditory threshold. *Journal of Experimental Psychology*, *23*(2), 109. https://doi.org/10.1037/h0059792

Cohen, M. X. (2014). *Analyzing neural time series data: theory and practice*. MIT press. ISBN: 9780262019873

Cooper, J. (1995). Controlled Flight into terrain. *Aerospace (UK)*, 305–831.

Corwin, W. H. (1995). Controlled flight into terrain avoidance- Why the ground proximity warning system is too little, too late. *Human Factors in Aviation Operations*, 155–160.

De Sanctis, P., Butler, J. S., Malcolm, B. R., & Foxe, J. J. (2014). Recalibration of inhibitory control systems during walking-related dual-task interference: a mobile brain-body imaging (MOBI) study. *Neuroimage*, *94*, 55-64. https://doi.org/10.1016/j.neuroimage.2014.03.016

DeCelles, J. L. (1991). The delayed GPWS response syndrome. *Aviation Research and Education Foundation. Herndon, VA. July*.

Dehais, F., Roy, R. N., & Scannella, S. (2019). Inattentional deafness to auditory alarms: Inter-individual differences, electrophysiological signature and single trial classification. *Behavioural brain research*, *360*, 51-59. https://doi.org/10.1016/j.bbr.2018.11.045

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, *134*(1), 9-21. https://doi.org/10.1016/j.jneumeth.2003.10.009.

Diekfuss, J. A., Ward, P., & Raisbeck, L. D. (2017). Attention, workload, and performance: A dual-task simulated shooting study. *International Journal of Sport and Exercise Psychology*, *15*(4), 423-437. https://doi.org/10.1080/1612197X.2015.1121508

Doyle, M. C., & Snowden, R. J. (2001). Identification of visual stimuli is improved by accompanying auditory stimuli: The role of eye movements and sound location. *Perception*, *30*(7), 795-810. https://doi.org/10.1068/p3126

Edworthy, J. (2013). Medical audible alarms: a review. *Journal of the American Medical Informatics Association*, *20*(3), 584-589. https://doi.org/10.1136/amiajnl-2012-001061.

EASA. (2007). *Decision n° 2007/017/R of the executive director of the european aviation safety agency* (2007/017/R). https://www.easa.europa.eu/system/files/dfu/ED%2520Decision%25202007-017-R.pdf

Endsley, M. R., & Jones, D. G. (2004). *Designing for situation awareness: An approach to user-centered design (2nd ed.).* Boca Raton, FL: CRC Press, Taylor & Francis Group.

Ewbank, C. E., Mumaw, R. J., & Snow, M. P. (2016, September). Development of the enhanced bank angle warning. In *2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)* (pp. 1-9). IEEE. https://doi.org/10.1109/DASC.2016.7778061

Fabre, E. F., Jahanpour, E. S., & Causse, M. (2021). Human Mirror Neuron System Based Alarms in the Cockpit: A Neuroergonomic Evaluation. *Applied Psychophysiology and Biofeedback*, 1-14. https://doi.org/10.1007/s10484-020-09481-0.

Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: A magnetic stimulation study. *Journal of Neurophysiology*, *73*(6), 2608–2611. https://doi.org/10.1152/jn.1995.73.6.2608.

Feltman, K. A., Randles, L., Goldie, C., & Bernhardt, K. A. (2022). A new approach to manipulating aviator workload: A case for individualizing workload manipulations. *Applied Ergonomics*, *102*, 103736. https://doi.org/10.1016/j.apergo.2022.103736

Fox, N. A., Bakermans-Kranenburg, M. J., Yoo, K. H., Bowman, L. C., Cannon, E. N., Vanderwert, R. E., Ferrari, P. F., & van IJzendoorn, M. H. (2016). Assessing human mirror activity with EEG mu rhythm: A meta-analysis. *Psychological Bulletin, 142*(3), 291–313. https://doi.org/10.1037/bul0000031

Guillaume, A. (2011). Intelligent auditory alarms. In T. Hermann, A. Hunt, J. G. Neuhoff (Eds.). *The Sonification Handbook* (pp. 493–508). Logos Publishing House, Berlin, Germany. https://sonification.de/handbook/download/TheSonificationHandbook-chapter19.pdf

Giraudet, L., St-Louis, M.-E., Scannella, S., & Causse, M. (2015). P300 event-related potential as an indicator of inattentional deafness? *PLoS ONE, 10*, e0118556–e0118556. https://doi.org/10.1371/journal.pone.0118556.

Hamel-Thibault, A., Thénault, F., Whittingstall, K., & Bernier, P. M. (2018). Delta-band oscillations in motor regions predict hand selection for reaching. *Cerebral cortex, 28*(2), 574-584. https://doi.org/10.1093/cercor/bhw392

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *In Advances in psychology* (Vol. 52, pp. 139-183). North-Holland. https://doi.org/10.1016/S0166-4115(08)62386-9.

Hart, S. G. (2006, October). NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 50, No. 9, pp. 904-908). Sage CA: Los Angeles, CA: Sage Publications.

Heaslip, T. W., Hull, N., McLeod, R. K., & VERMIJ, M. (1991). The frozen pilot syndrome. In *International Symposium on Aviation Psychology, 6 th, Columbus, OH* (pp. 782-787).

Hughes, H. C., Reuter-Lorenz, P. A., Nozawa, G., & Fendrich, R. (1994). Visual-auditory interactions in sensorimotor processing: saccades versus manual responses. *Journal of Experimental Psychology: Human Perception and Performance, 20*(1), 131. https://doi.org/10.1037/0096-1523.20.1.131.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science, 286*(5449), 2526–2528. https://doi.org/10.1126/science.286.5449.2526.

IATA (2015). *Loss of control in-flight accident analysis report*. Montreal-Geneva: International Air Transport Association, 1st Edition. https://flightsafety.org/wp-content/uploads/2017/07/IATA-LOC-I-1st-Ed-2015.pdf

IATA (2018a). *Guidance material and best practices for the implementation of upset prevention and recovery training.* Montreal-Geneva: International Air Transport Association, 2nd Edition.

https://www.iata.org/contentassets/b6eb2adc248c484192101edd1ed36015/gmbp_uprt.pdf

IATA (2018b). *IATA Controlled Flight Into Terrain Accident Analysis Report—2008–2017 Data* (2018 Edition). https://www.iata.org/contentassets/06377898f60c46028a4dd38f13f979ad/cfit-report.pdf

IATA (2021). *Safety Report 2020*. https://www.iata.org/en/publications/safety-report/

ICAO (2019). *State of Global Aviation Safety Report*. https://www.icao.int/safety/Documents/ICAO_SR_2019_final_web.pdf

Jaquess, K. J., Gentili, R. J., Lo, L. C., Oh, H., Zhang, J., Rietschel, J. C., ... & Hatfield, B. D. (2017). Empirical evidence for the relationship between cognitive workload and attentional reserve. *International Journal of Psychophysiology*, *121*, 46-55. https://doi.org/10.1016/j.ijpsycho.2017.09.007

Jeannerod, M. (1994). The representing brain: Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences*, *17*(2), 187–202. https://doi.org/10.1017/S0140525X0003402.

Jenson, D., Thornton, D., Harkrider, A. W., & Saltuklaroglu, T. (2019). Influences of cognitive load on sensorimotor contributions to working memory: an EEG investigation of mu rhythm activity during speech discrimination. *Neurobiology of Learning and Memory*, *166*, 107098. https://doi.org/10.1016/j.nlm.2019.107098

Jung, T. P., Humphries, C., Lee, T. W., Makeig, S., McKeown, M. J., Iragui, V., & Sejnowski, T. J. (1998). Extended ICA removes artifacts from electroencephalographic recordings. *Advances in neural information processing systems*, 894-900.

Kelly, D., & Efthymiou, M. (2019). An analysis of human factors in fifty controlled flight into terrain aviation accidents from 2007 to 2017. *Journal of safety research*, *69*, 155-165. https://doi.org/10.1016/j.jsr.2019.03.009.

Kuchar, J. K. (2001). Markov model of terrain for evaluation of ground proximity warning system thresholds. *Journal of Guidance, Control, and Dynamics*, *24*(3), 428-435. https://doi.org/10.2514/2.4748

Kumar, N., & Kumar, J. (2016). Measurement of cognitive load in HCI systems using EEG power spectrum: an experimental study. *Procedia Computer Science*, *84*, 70-78. https://doi.org/10.1016/j.procs.2016.04.068

Lavie, N., Beck, D. M., & Konstantinou, N. (2014). Blinded by the load: attention, awareness and the role of perceptual load. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1641), 20130205. https://doi.org/10.1098/rstb.2013.0205

Linares, N. N., Charron, V., Ouimet, A. J., Labelle, P. R., & Plamondon, H. (2020). A systematic review of the Trier Social Stress Test methodology: Issues in promoting study comparison and replicable research. *Neurobiology of stress*, *13*, 100235. https://doi.org/10.1016/j.ynstr.2020.100235

Liu, Y. C. (2001). Comparative study of the effects of auditory, visual and multimodality displays on drivers' performance in advanced traveller information systems. *Ergonomics*, *44*(4), 425-442. https://doi.org/10.1080/00140130010011369

Macdonald, J. S., & Lavie, N. (2011). Visual perceptual load induces inattentional deafness. *Attention, Perception, & Psychophysics*, *73*(6), 1780-1789. https://doi.org/10.3758/s13414-011-0144-4

Martin, W. L., Murray, P. S., Bates, P. R., & Lee, P. S. Y. (2016). A Flight Simulator Study of the Impairment Effects of Startle on Pilots During Unexpected Critical Events. *Aviation Psychology and Applied Human Factors*, *6*(1), 24–32. https://doi.org/10.1027/2192-0923/a000092

Martin, W. L., Murray, P. S., Bates, P. R., & Lee, P. S. (2015). Fear-potentiated startle: A review from an aviation perspective. *The International Journal of Aviation Psychology*, *25*(2), 97-107. https://doi.org/10.1080/10508414.2015.1128293

Martin, W., Murray, P., & Bates, P. (2012). Startle, freeze and denial: An analysis of pathological pilot reactions during unexpected events. In *Proceedings of the 65th International Air Safety Seminar* (pp. 140-5).

Meltzoff, A. N., and Prinz, W. (2002). *The imitative mind: Development, evolution and brain bases* (Vol. 6). Cambridge University Press.

Metting van Rijn, A. C., Peper, A., & Grimbergen, C. A. (1990). High-quality recording of bioelectric events. *Medical and Biological Engineering and Computing*, *28*(5), 389–397. https://doi.org/10.1007/BF02441666

Meyer, M., Endedijk, H. M., & Hunnius, S. (2020). Intention to imitate: top-down effects on 4-year-olds' neural processing of others' actions. *Developmental cognitive neuroscience*, *45*, 100851. https://doi.org/10.1016/j.dcn.2020.100851

Michels, L., Bucher, K., Lüchinger, R., Klaver, P., Martin, E., Jeanmonod, D., & Brandeis, D. (2010). Simultaneous EEG-fMRI during a working memory task: modulations in low and high frequency bands. *PloS one*, *5*(4), e10298. https://doi.org/10.1371/journal.pone.0010298

Momtahan, K., Hetu, R., & Tansley, B. (1993). Audibility and identification of auditory alarms in the operating room and intensive care unit. *Ergonomics*, *36*(10), 1159-1176. https://doi.org/10.1080/00140139308967986

Mollo, G., Pulvermüller, F., & Hauk, O. (2016). Movement priming of EEG/MEG brain responses for action-words characterizes the link between language and action. *cortex*, *74*, 262-276. https://doi.org/10.1016/j.cortex.2015.10.021

Molloy, K., Griffiths, T. D., Chait, M., & Lavie, N. (2015). Inattentional deafness: visual load leads to time-specific suppression of auditory evoked responses. *Journal of Neuroscience*, *35*(49), 16046-16054. https://doi.org/10.1523/JNEUROSCI.2931-15.2015

Morales, J. M., Ruiz-Rabelo, J. F., Diaz-Piedra, C., & Di Stasi, L. L. (2019). Detecting mental workload in surgical teams using a wearable single-channel electroencephalographic device. *Journal of surgical education*, *76*(4), 1107-1115. https://doi.org/10.1016/j.jsurg.2019.01.005

Moreno, I., de Vega, M., & León, I. (2013). Understanding action language modulates oscillatory mu and beta rhythms in the same way as observing actions. *Brain and cognition*, *82*(3), 236-242. https://doi.org/10.1016/j.bandc.2013.04.010

Moroze, M. L., and Snow, M. P. (1999). Causes and Remedies of Controlled Flight into Terrain in Military and Civil Aviation. *10th International Aviation Psychology Symposium*, 1143–1149.

Muthukumaraswamy, S. D., & Singh, K. D. (2008). Modulation of the human mirror neuron system during cognitive activity. *Psychophysiology*, *45*(6), 896-905. https://doi.org/10.1111/j.1469-8986.2008.00711.x

Niccolai, V., Klepp, A., Weissler, H., Hoogenboom, N., Schnitzler, A., & Biermann-Ruben, K. (2014). Grasping hand verbs: oscillatory beta and alpha correlates of action-word processing. *PloS one*, *9*(9), e108059. https://doi.org/10.1371/journal.pone.0108059

NTSB (1973). *Aircraft Accident Report Eastern AirLines, Inc. L-1011,N310EA Miami, Florida, December 29, 1972* [Aircraft Accident Report NTSB/AAR-73/14]. https://www.ntsb.gov/investigations/AccidentReports/Reports/AAR7314.pdf

NTSB (2000). *Controlled Flight Into Terrain, Korean Air Flight 801, Boeing 747–300, HL7468, Nimitz Hill, Guam, August 6, 1997* [Aircraft Accident Report NTSB/AAR-00/01]. https://www.ntsb.gov/investigations/AccidentReports/Pages/AAR0001.aspx

NTSB (2010). *Loss of Control on Approach, Colgan Air, Inc., Operating as Continental Connection Flight 3407, Bombardier DHC-8-400, N200WQ, Clarence Center, New York, February 12, 2009. NTSB/AAR-10/01. Washington, DC.* https://www.ntsb.gov/investigations/AccidentReports/Reports/AAR1001.pdf

Okogbaa, O. G., Shell, R. L., & Filipusic, D. (1994). On the investigation of the neurophysiological correlates of knowledge worker mental fatigue using the EEG

signal. *Applied Ergonomics*, *25*(6), 355-365. https://doi.org/10.1016/0003-6870(94)90054-X

Oliver, N., Calvard, T., & Potočnik, K. (2019). Safe limits, mindful organizing and loss of control in commercial aviation. *Safety Science*, *120*, 772-780. https://doi.org/10.1016/j.ssci.2019.08.018

Pfurtscheller, G., Neuper, C., Andrew, C., & Edlinger, G. (1997). Foot and hand area mu rhythms. *International Journal of Psychophysiology*, *26*(1-3), 121-135. https://doi.org/10.1016/S0167-8760(97)00760-5

Pfurtscheller, G., & Da Silva, F. L. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical neurophysiology*, *110*(11), 1842-1857. https://doi.org/10.1016/S1388-2457(99)00141-8

Pineda, J. A. (2005). The functional significance of mu rhythms: translating "seeing" and "hearing" into "doing". *Brain research reviews*, *50*(1), 57-68. https://doi.org/10.1016/j.brainresrev.2005.04.005

Quandt, L. C., Marshall, P. J., Shipley, T. F., Beilock, S. L., & Goldin-Meadow, S. (2012). Sensitivity of alpha and beta oscillations to sensorimotor characteristics of action: An EEG study of action production and gesture observation. *Neuropsychologia*, *50*(12), 2745-2751. https://doi.org/10.1016/j.neuropsychologia.2012.08.005.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological bulletin*, *114*(3), 510. https://doi.org/10.1037/0033-2909.114.3.510.

Raveh, D., & Lavie, N. (2015). Load-induced inattentional deafness. *Attention, Perception, & Psychophysics*, *77*(2), 483-49. https://doi.org/10.3758/s13414-014-0776-2

Rietschel, J. C., Miller, M. W., Gentili, R. J., Goodman, R. N., McDonald, C. G., & Hatfield, B. D. (2012). Cerebral-cortical networking and activation increase as a function of cognitive-motor task difficulty. *Biological psychology*, *90*(2), 127-133. https://doi.org/10.1016/j.biopsycho.2012.02.022

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, *3*(2), 131–141. https://doi.org/10.1016/0926-6410(95)00038-0.

Roelofs, K. (2017). Freeze for action: neurobiological mechanisms in animal and human freezing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1718), 20160206. https://doi.org/10.1098/rstb.2016.0206

Schröger, E., & Widmann, A. (1998). Speeded responses to audiovisual signal changes result from bimodal integration. *Psychophysiology*, *35*(6), 755-759. https://doi.org/10.1111/1469-8986.3560755

Schmidt, N. B., Richey, J. A., Zvolensky, M. J., & Maner, J. K. (2008). Exploring human freeze responses to a threat stressor. *Journal of Behavior Therapy and Experimental Psychiatry*, *39*(3), 292–304. https://doi.org/10.1016/j.jbtep.2007.08.002.

Scoot, W. B. (1996). New Research Identifies Causes of CFIT. *Aviation Week & Space Technology*, *144*(25), 40–71.

Shappell, S. A., & Wiegmann, D. A. (2001). Unraveling the mystery of general aviation controlled flight into terrain accidents using HFACS. *Proceedings of the 11th Symposium on Aviation Psychology, 1-6.* https://www.faa.gov/about/initiatives/maintenance_hf/library/documents/media/human_factors_maintenance/unraveling_the_mystery_of_general_aviation_controlled_flight_into_terrain_accidents_using_hfacs.pdf

Shappell, S. A., & Wiegmann, D. A. (2003). *A human error analysis of general aviation controlled flight into terrain accidents occurring between 1990-1998*. DOT/FAA/AM-03/4. Office of Aerospace Medicine Washington, DC 20591

Shaw, E. P., Rietschel, J. C., Hendershot, B. D., Pruziner, A. L., Miller, M. W., Hatfield, B. D., & Gentili, R. J. (2018). Measurement of attentional reserve and mental effort for cognitive workload assessment under various task demands during dual-task walking. *Biological psychology*, *134*, 39-51. https://doi.org/10.1016/j.biopsycho.2018.01.009

Sherry, L., & Mauro, R. (2014, April). Controlled Flight into Stall (CFIS): Functional complexity failures and automation surprises. In *2014 Integrated Communications, Navigation and Surveillance Conference (ICNS) Conference Proceedings* (pp. D1-1). IEEE. https://doi.org/10.1109/ICNSurv.2014.6819980

Simons, D. J. (2000). Attentional capture and inattentional blindness. *Trends in cognitive sciences*, *4*(4), 147-155. https://doi.org/10.1016/S1364-6613(00)01455-8

Simonson, R. J., Keebler, J. R., Blickensderfer, E. L., & Besuijen, R. (2022). Impact of alarm management and automation on abnormal operations: A human-in-the-loop simulation study. *Applied Ergonomics*, *100*, 103670. https://doi.org/10.1016/j.apergo.2021.103670

Stevens, C. J., Brennan, D., Petocz, A., & Howell, C. (2009). Designing informative warning signals: Effects of indicator type, modality, and task demand on recognition speed and accuracy. *Advances in Cognitive Psychology*, *5*, 84. https://doi.org/10.2478/v10053-008-0064-6

TSB (2014). Aviation Investigation Report A11H0002. Controlled flight into terrain Bradley Air Services Limited (First Air) Boeing 737-210C, C-GNWN Resolute Bay, Nunavut 20 August 2011. https://www.tsb.gc.ca/eng/rapports-reports/aviation/2011/a11h0002/a11h0002.pdf

TSB (1998). Aviation Occurrence Report A97H0011. Loss Of Control On Go-Around (Rejected Landing) Air Canada Canadair Cl-600-2b19 C-FSKI Fredericton Airport, New Brunswick 16 December 1997. https://www.tsb.gc.ca/eng/rapports-reports/aviation/1997/a97h0011/a97h0011.pdf

Torralba, B., Martinez-Arias, R., & Sola, R. (2007). *An experimental study of operator activity using an advanced alarm system in nuclear power plant control room.* Human Factors and Power Plants and HPRCT 13th Annual Meeting, 2007 IEEE 8th (pp. 92–97). Monterey, CA: IEEE.

Wheale, J. (1982). Decrements in performance on a primary task associated with the reaction to voice warning messages. *British Journal of Audiology*, *16*(4), 265-272. https://doi.org/10.3109/03005368209081471

Wickens, C. D. (2008). Multiple resources and mental workload. *Human factors*, *50*(3), 449-455. https://doi.org/10.1518/001872008X28839

Wickens, C. D. (2020). Processing resources and attention. In *Multiple-task performance* (pp. 3-34). CRC Press.

White, A., & O'Hare, D. (2022). In plane sight: Inattentional blindness affects visual detection of external targets in simulated flight. *Applied Ergonomics*, *98*, 103578. https://doi.org/10.1016/j.apergo.2021.103578

Whittemore, T. J., & Woods, S. (2021). Analysis on the Negative Emotional, Physiological, and Cognitive Responses Elicited from of the Activation of a Stall Alarm. *Journal of Aviation/Aerospace Education & Research*, *30*(1). https://doi.org/10.15394/jaaer.2021.1881

Wu, X., & Li, Z. (2018). A review of alarm system design for advanced control rooms of nuclear power plants. *International Journal of Human–Computer Interaction*, *34*(6), 477-490. https://doi.org/10.1080/10447318.2017.1371950.

Zhu, R., Wang, Z., Ma, X., & You, X. (2022). High expectancy influences the role of cognitive load in inattentional deafness during landing decision-making. *Applied Ergonomics*, *99*, 103629. https://doi.org/10.1016/j.apergo.2021.103629

NASA-TLX	No MWL		Low MWL		High MWL		Friedman ANOVA		No vs. Low MWL			No vs. High MWL			Low vs. High MWL		
	M	*SD*	*M*	*SD*	*M*	*SD*	*X²F (2)*	*p*	*T*	z	*P*	*T*	z	*p*	*T*	z	*p*
Mental Demand	3.79	2.81	13.13	4.09	17.17	2.53	45.66	**< 0.001***	0.00	4.29	**< 0.001***	0.00	4.29	**< 0.001***	2.00	4.14	**< 0.001***
Physical Demand	3.21	1.86	3.58	2.15	4.58	2.84	11.13	**0.004***	38.00	0.91	0.360	44.5	2.03	**0.042***	12.00	2.54	**0.011***
Temporal Demand	5.00	3.95	9.88	3.39	11.54	5.19	29.34	**< 0.001***	17.50	3.79	**< 0.001***	2.50	4.12	**< 0.001***	52.50	2.18	**0.029***
Performance	3.79	1.74	8.87	3.33	11.92	4.65	40.06	**< 0.001***	0.00	4.20	**< 0.001***	0.00	4.20	**< 0.001***	19.00	3.49	**< 0.001***
Effort	4.62	3.46	12.29	4.76	15.37	4.13	41.93	**< 0.001***	0.00	4.29	**< 0.001***	0.00	4.20	**< 0.001***	6.00	3.81	**< 0.001***
Frustration	3.25	3.08	9.33	4.72	11.71	5.34	28.57	**< 0.001***	8.00	3.95	**< 0.001***	2.50	4.12	**< 0.001***	49.00	2.31	**0.021***

Table 1. Summary of the means, standard deviations and statistical analysis performed on the ratings of NASA-TLX subscales.

Workload Main Effect		No MWL		Low MWL		High MWL		ANOVA			Tukey post-hoc tests (*p*)		
		M	*SD*	*M*	*SD*	*M*	*SD*	*F (2,46)*	*p*	*ηp²*	*NWL vs. LWL*	*NWL vs LWL.*	*LWL vs. HWL*
C3	Beta [12 - 30 Hz[	-2.420	1.882	0.331	1.178	0.215	1.106	23.910	**< 0.001***	0.509	**< 0.001***	**< 0.001***	0.964
	Low Beta [12 - 15 Hz[	-1.329	1.570	0.313	1.173	0.208	1.094	11.865	**< 0.001***	0.340	**< 0.001***	**< 0.001***	0.958
	High Beta [15 - 30 Hz[	-2.784	2.071	0.337	1.232	0.217	1.184	26.374	**< 0.001***	0.534	**< 0.001***	**< 0.001***	0.967
C4	Beta [12 - 30 Hz[	-2.447	1.973	0.527	1.018	0.148	0.817	28.570	**< 0.001***	0.554	**< 0.001***	**< 0.001***	0.653
	Low Beta [12 - 15 Hz[	-1.316	1.643	0.469	1.032	0.227	0.907	13.654	**< 0.001***	0.372	**< 0.001***	**< 0.001***	0.791
	High Beta [15 - 30 Hz[	-2.824	2.185	0.546	1.057	0.122	0.867	31.205	**< 0.001***	0.576	**< 0.001***	**< 0.001***	0.635

Table 2. Summary of the means, standard deviations and statistical analysis performed on the total beta, low beta and high beta powers (dB) measured at C3 and C4 in response to the alarms in the no MWL (NWL), the low MWL (LWL) and the high MWL (HWL) runs in the 0 – 1441 ms time window after the onset of the alarms.

Alarm Main Effect		Time window	Gesture Alarm		Written Alarm	
			M	*SD*	*M*	*SD*
C3	Delta [3 - 4 Hz[	660 - 720 ms	1.361	1.151	0.500	0.974
	Theta [4 - 8 Hz[	1000 -1440 ms	-0.893	1.151	-0.159	1.026
	Mu [8 - 12 Hz[	1020 - 1350 ms	-1.523	1.463	-0.867	1.568
	Low Beta [12 - 15Hz[	1170 - 1230 ms	-0.691	0.945	0.162	1.180
C4	Delta [3 - 4 Hz[	590 - 820 ms	1.511	1.156	0.430	0.940
	Theta [4 - 8 Hz[	970 - 1240 ms	-0.836	1.045	-0.171	1.065
	Mu [8 - 12 Hz[	950 - 1100 ms	-1.850	1.391	-1.040	1.640
	Low Beta [12 - 15Hz[	/	/	/	/	/

Table 3. Average power and standard deviations (dB) measured at C3 and C4 in response to both gesture and written alarms in the delta, theta [4 - 8 Hz[, mu [8 - 12 Hz[and low beta [12 - 15Hz[frequency bands in significant time windows. Time-windows were selected according to the results of the permutation tests with FDR correction for multiple comparison (see Methods section).

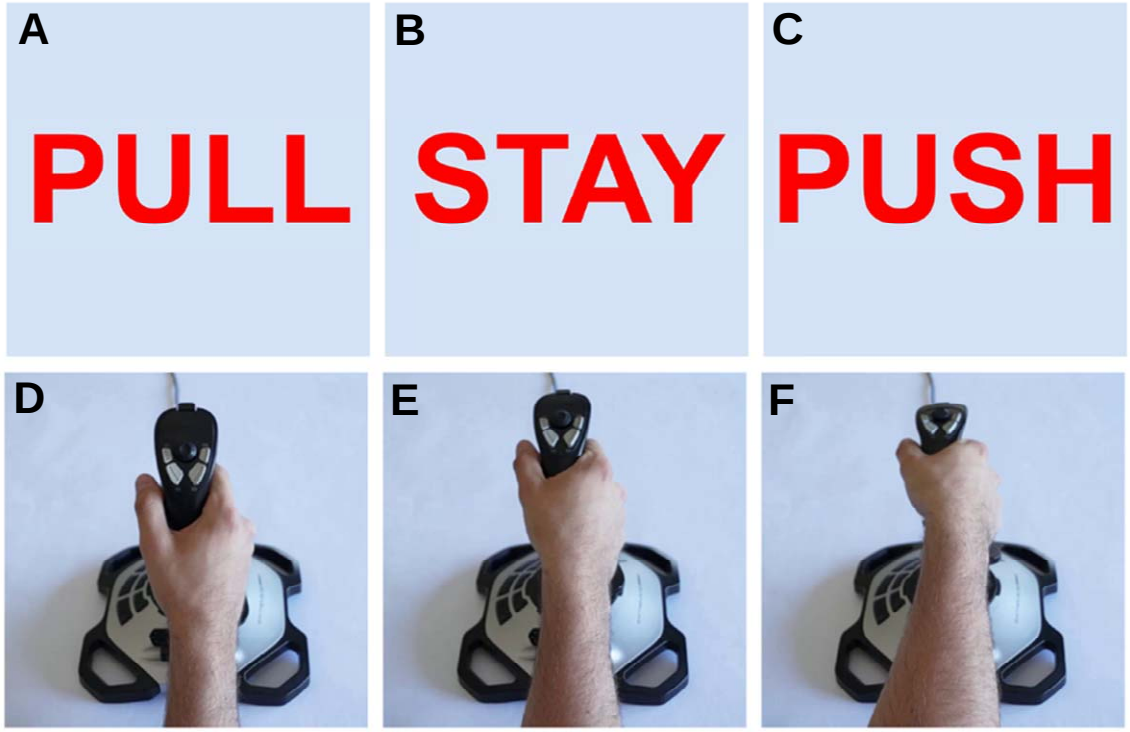


Figure 1. Illustration of the two (A) PULL and (C) PUSH written alarms and the (B) STAY written filler; and the corresponding (D) PULL and (F) PUSH gesture alarms and (E) gesture filler. For the written alarms and written filler, the verb was displayed twice for 1800ms followed by a grey background for 200ms each time. For the gesture alarms and the gesture filler, the video lasted 2000ms (the hand gesture lasted 1800ms and the hand remained still for 200ms at the end of the gesture) and was played twice.

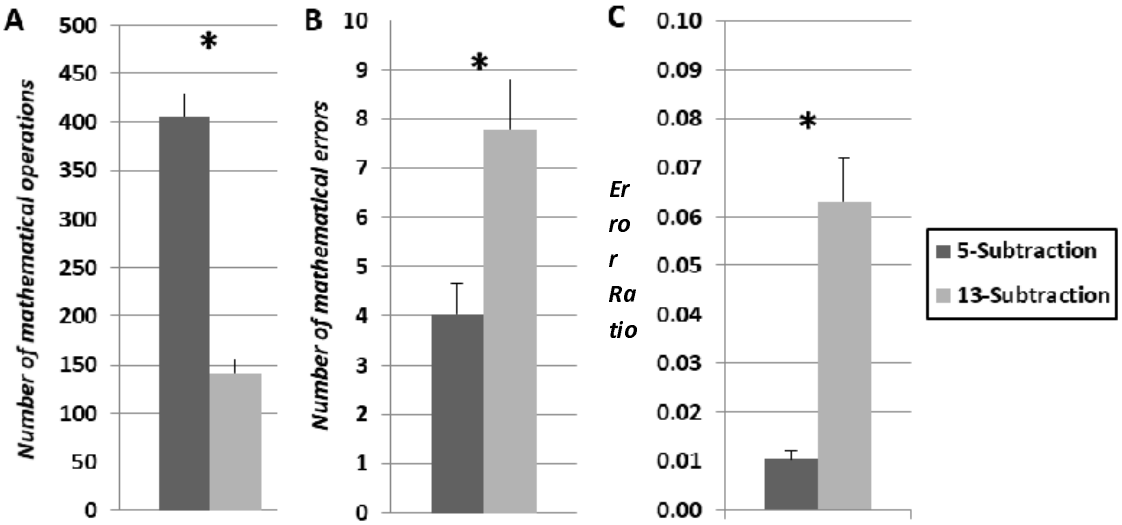
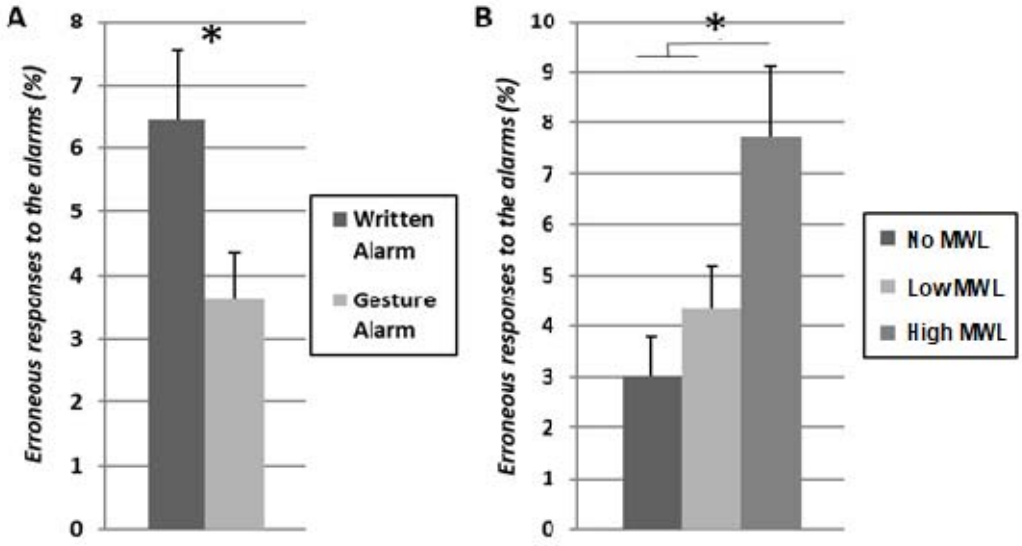


Figure 2. Illustration of (A) the number of mathematical operations, (B) the number of mathematical errors and (C) the error ratio in the low MWL run (dark grey) and in the high MWL run (light grey). * $p < .01$

Figure 3. Illustration of the erroneous responses to the alarm (%) as a function of (A) the type of alarms with written alarms (dark grey) and gesture alarms (light grey); and (B) the level of MWL with no-MWL (dark grey), low MWL (light grey) and high MWL (mid grey). *: $p < .05$



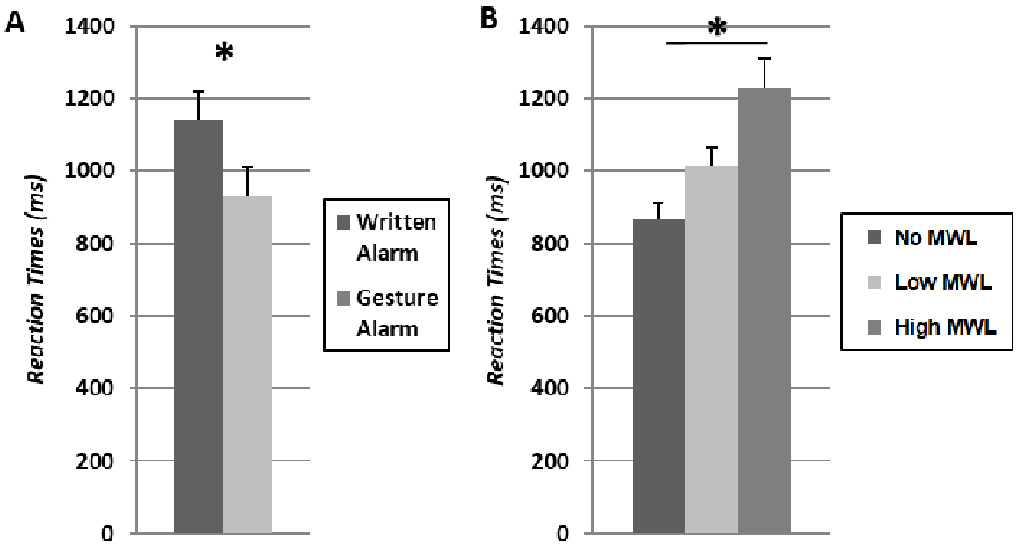
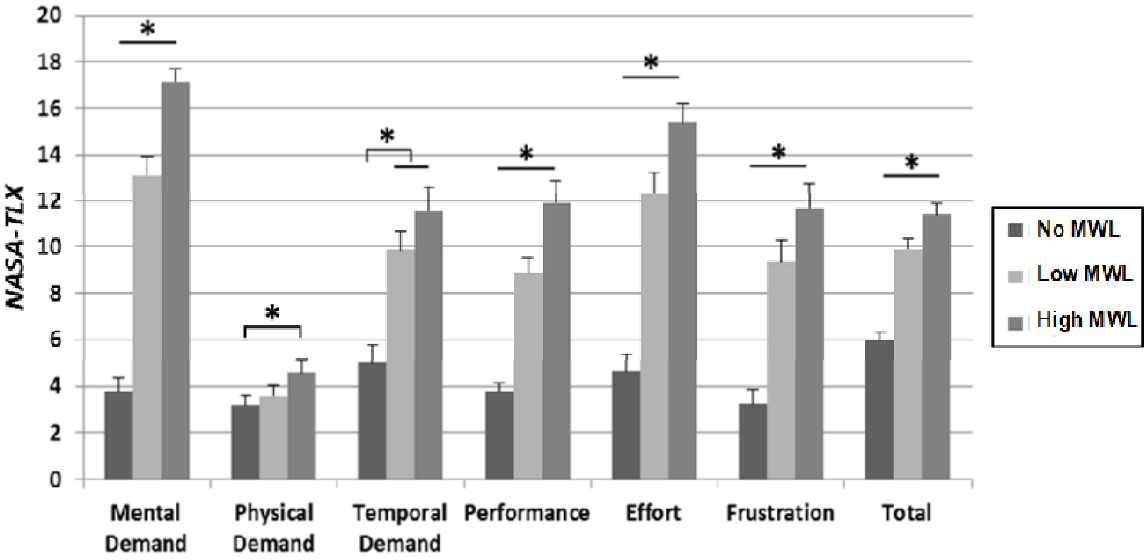


Figure 4. Illustration of the reaction times to the alarm (ms) as a function of (A) the type of alarms with written alarms (dark grey) and gesture alarms (light grey); and (B) the level of MWL with no MWL (dark grey), low MWL (light grey) and high MWL (mid grey). *: $p < .05$

Figure 5. Illustration of the NASA-TLX ratings observed in the no MWL run (dark grey), the low MWL run (light grey) and the high MWL run (middle grey). *: $p < .05$



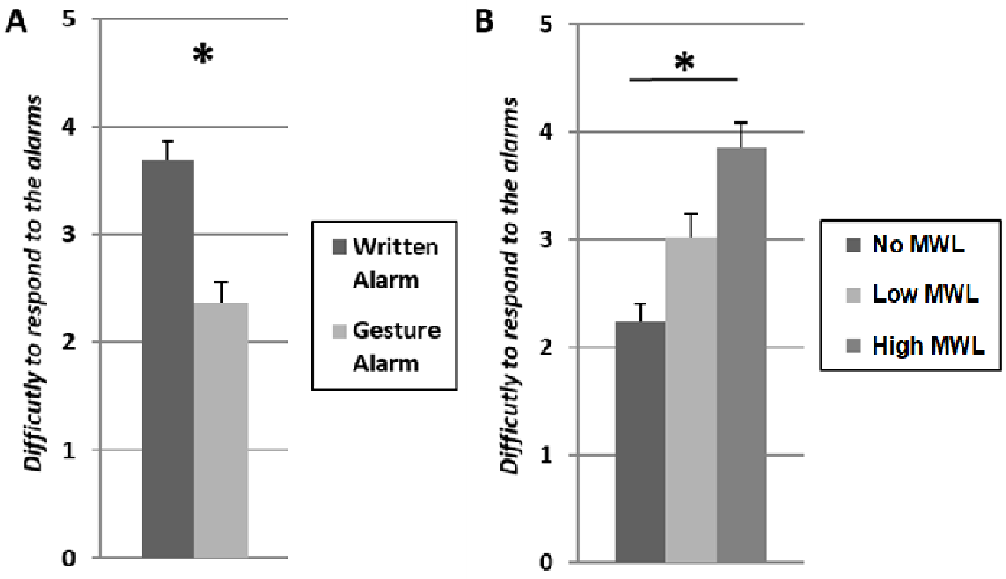


Figure 6. Illustration of the difficulty to respond to the alarms as a function of (A) the type of alarms with written alarms (dark grey) and gesture alarms (light grey); and (B) the level of MWL with no MWL (dark grey), low MWL (light grey) and high MWL (mid grey). *: $p < .001$.

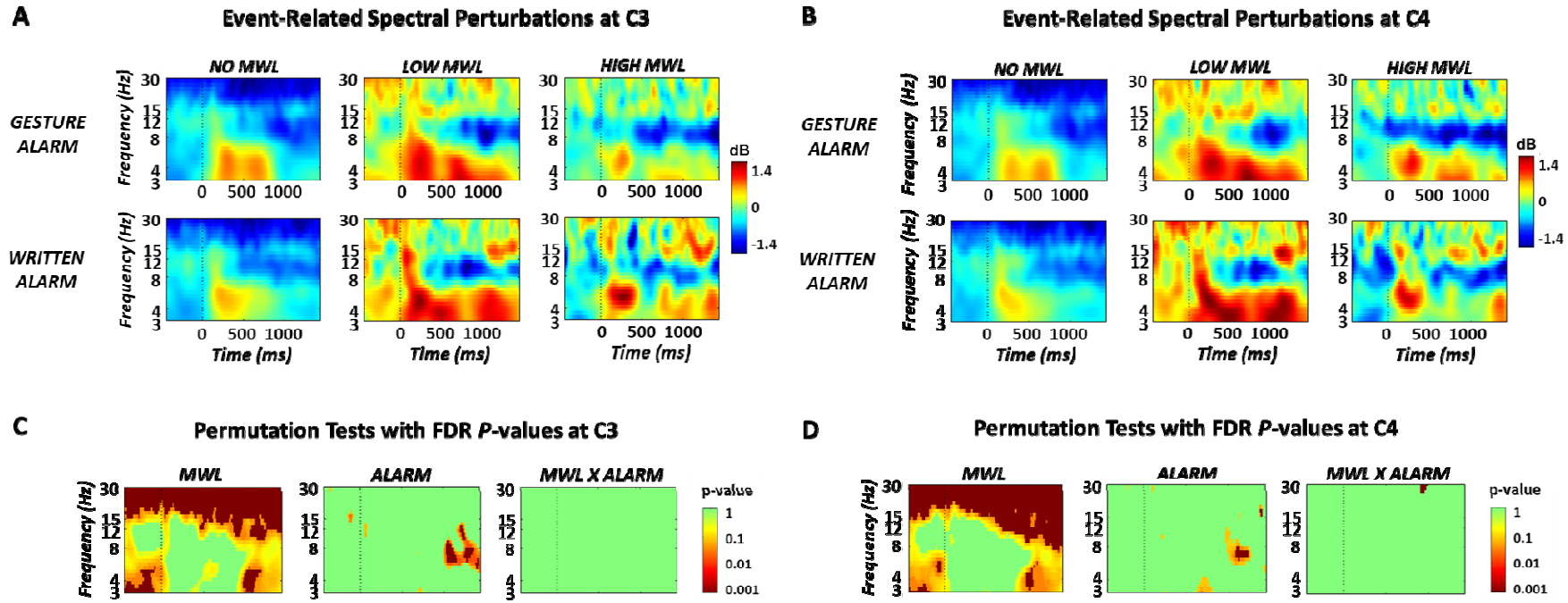


Figure 7. Event Related Spectral Perturbations (ERSPs) showing increases (positive values – red) and decreases (negative values – blue) in dB-corrected power in the three MWL levels (no MWL – left, low MWL – middle, and high MWL – right) for the two types of alarms (Gesture – top and written – bottom) at the C3 (A) and C4 (B) electrodes. Graphs C and D show results of the permutation tests for the MWL main effect (no MWL vs. low MWL vs. high MWL – right hand side top graph), the alarm main effect (gesture vs. written – bottom central graph) and the MWL x Alarm interaction (right hand side bottom graph). P-values are displayed with a log-scale from green (nonsignificant) to dark red (highly significant).