Fast & Scrupulous: Gesture-Based Alarms Improve Performances Under Several Mental Workloads. An ERSP Study.

Eve Floriane Fabre, Bertille Somon, Valeria Baragona, Quentin Uhl & Mickaël Causse

ISAE-SUPAERO, Université de Toulouse, France

*** Correspondence: Eve Floriane Fabre**

Keywords: Emergency situations; Operators; Safety; Take Over; EEG; Time-Frequency Analysis.

Author Note

This work was supported by a chair grant from Dassault Aviation ("CASAC", holder: Prof. Mickaël Causse).

Abstract

In high-risk environments, fast and accurate responses to warning systems are essential to efficiently handle emergency situations and prevent potential catastrophic events. The aim of the present experiment was twofold: 1) investigating whether hand action videos, referred as to gesture alarms, are more efficient to trigger fast and accurate responses than text alarm messages (referred as written alarms), especially when the mental workload is high; and 2) investigating the brain activity in response to both types of alarms as a function of the level of mental workload. Regardless the level of mental workload, participants were found to be both faster and more accurate when responding to the gesture alarms than to the written alarms. Brain electrophysiological results suggest that this greater efficiency might be due to the fact that (compared to written alarms) gesture alarms facilitate the action execution reflected by the mu and beta power decrease observed around the response time window. Taken together, these results demonstrate that gesture alarm systems may help improving 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 the appropriate response 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).

In aviation, one of the most critical emergency alarm systems is probably the "Ground Proximity Warning Systems" (GPWS). This alarm system (and its later version Enhanced-GPWS; EASA, 2007) is designed to prevent one of the deadliest categories of accidents (IATA, 2018a, 2021): the Controlled Flight Into Terrain (CFIT). A CFIT occurs when an airworthy aircraft under the complete control of the crew that is inadvertently flown into terrain, water, or an obstacle (Moroze & Snow, 1999; Shappell & Wiegmann, 2001; 2003). Because stimulating two or more sensory channels of information (e.g., visual, auditory and/or haptic) facilitates the detection of an alarm, the GPWS is − like most emergency alarm systems – multimodal with a visual and an auditory modality (Alirezaee et al., 2017; Child & Wendt, 1938; Hughes et al., 1994; Liu, 2001). The visual modality of the GPWS consists of a red-colored "Pull Up" inscription flashing on the Primary Flight Display (PFD; i.e., the display located in front of the pilots that presents the primary flight parameters). Its auditory modality consists of a combination of non-speech (i.e., bips) and speech (i.e., "Pull Up" pronounced by a synthetic voice) alarms. Despite the fact that the compulsory implementation of the GPWS alarm in all large turbine and turbojet airplanes decided by the FAA in 1974 led to a significant drop in CFIT accidents, these accidents remain the second cause of fatalities in commercial aviation (IATA, 2021; ICAO, 2019) with three out of four CFIT accidents

resulting in the death of all passengers and crew members (Moroze & Snow 1999; Scoot, 1996).

1.2. Alarm design, workload and stress

Respectively 20% and 70% of CFIT accidents occur during the cruise phase and the approach/landing phases (Scott, 1996). While automation has significantly reduced pilots' workload during the majority of the flight (e.g., cruise phase), pilots' activity is particularly intense during the approach and landing phases (Cooper, 1995; Corwin, 1995; Kelly & Efthymiou, 2019) − especially in peak traffic periods, when the visibility is low or/and in adverse weather conditions (Wiggins & O'Hare, 1995). In these conditions, pilots can sometimes channelize their attention or struggle to handle the workload, which is likely to affect their situation awareness (i.e., "the perception, understanding, and ability to forecast the factors affecting the aircraft at any moment in time"; Moroze & Snow, 1999), partly explaining why an important part of CFITs occur during the approach and the landing phases.

In many CFITs involving aircraft with a functioning GPWS alarm, the latter did not allow the pilots to avoid the accident either because it did not warn the pilots at all (24% of the cases) or warned them of the impending impact with an important delay that did not leave enough time to respond to the alert (38% of the cases; DeCelles, 1991; IATA, 2018a; Kuchar, 2001). However, in 38% of these accidents the delayed or inadequate reaction of pilots to the GPWS alarm was found to be an important contributive factor (see for instance Korean Air Flight 801: NTSB, 2000; or First Air Flight 6560: TSB, 2014). There are various reasons why pilots sometimes fail to properly respond to the GPWS alarm (Scoot, 1996). A startle effect following the unexpected and stressful "*Pull Up*" alarm may leave the pilot unable to act appropriately during a few seconds (Martin et al., 2015, 2016). A high mental workload (MWL) does not only impair pilots' situation awareness, but can also affect their capacity to detect and/or appropriately react to alarms (e.g., Bliss, 2003; Causse, Peysakhovich, et al.

2016; Dehais et al. 2014; Giraudet et al., 2015; Martin et al., 2016; NTSB, 2000; Schmidt et al., 2008). Moreover, the important MWL observed during the approach and landing phases adds up to the already elevated visual (e.g., numerous displays, cockpit windows) and auditory (e.g., auditory alerts, communications) perceptual load, which is likely to leave pilots with few attentional and cognitive resources to allocate to the incoming GPWS alert (e.g., Causse, Imbert et al., 2016; Jaquess et al., 2017; Lavie et al., 2014). Finally, the design of the GPWS alarm might not be optimal to trigger fast and accurate reactions in pilots (Fabre et al., 2021). First, its visual features can be difficult to detect due to its small size (Corwin 1995). Second, MWL was found to affect language processing (e.g., Causse, Peysakhovich, et al., 2016), suggesting that the linguistic features of the GPWS alarm (i.e., spoken and written "Pull Up") may not be particularly adapted to trigger fast and accurate reactions in pilots under high levels of MWL.

1.3. Improving the GPWS alarm using a hand action video

One way to improve airline pilots' reactions to the GPWS alarm could be to enhance its visual feature (Fabre et al., 2021). Hand action videos were found to both strongly attract attention (Abrams & Christ, 2003) and preactivate the motor cortex (Rizzolatti et al., 1996), suggesting that using hand action videos as visual alarms might facilitate the initiation of the associated motor response in operators (Brucker et al., 2015; Iacoboni et al., 1999; Jeannerod, 1994; Meltzoff & Prinz, 2002). Fabre and colleagues (2021) recently conducted a two-experiment study to investigate whether using a hand action video (referred as to gesture alarm) as visual feature of the GPWS alarm could improve pilots' reactions compared to the current written inscription. Experiment 1 was conducted in laboratory conditions to guarantee a good quality of the signal and aimed at comparing the brain signatures associated with the observation of two types of "Pull Up" warnings: a red-inked "Pull Up" inscription and a video of the action to perform in response to the "Pull Up" alarm (i.e., *gesture alarm*). Experiment 2

aimed at investigating whether enhancing the visual feature of the GPWS alarm − using a "Pull Up" hand action video instead of a "Pull Up" inscription – could improve pilots' response to the GPWS alarm. To this purpose, participants were placed in a A320 flight simulator and asked to apply a Pull-Up maneuver as soon as they detected the "Pull Up" alarms (i.e., "Pull Up" inscription and "Pull Up" hand action video), while monitoring the external environment (i.e., through the cockpit windows) for a potential air traffic conflict in conditions of reduced visibility. The results of experiment 1 revealed that gesture alarms triggered greater mu and beta desynchronizations than written alarms, suggesting that the gesture alarm elicits greater motor preparation and pre-activation of "Pull Up" recovery maneuver compared to the inscription alarm, at least when no motor response is required (e.g., Fadiga et al., 1995). However, the results of experiment 2 revealed no difference in reaction times between the two types of alarms in the flight simulator. The authors assumed that the flight scenario that was used might not have been adapted to investigate potential differences between the two types of alarms.

1.4. The present study

The great majority of CFIT accidents occur during the cruise, the approach and the final approach/landing phases (Scoot, 1996). When the GPSW alarm sounds in the cockpit during the cruise phase, reacting to this emergency situation is usually one of the few tasks pilots have to perform as the cruise phase is associated with the lowest MWL of the flight (e.g., Lin et al., 2012; Rouwhorst, et al., 2017). However, when the GPSW alarm sounds in the cockpit during the approach or the final approach/landing phase pilots are required to react to it while they are already performing competing tasks associated with a moderate (i.e., approach) and a high (i.e., final approach/landing) MWL (Rouwhorst, et al., 2017). The present study aimed at investigating 1) the efficiency of gesture alarms (compared to written alarms) under variable levels of MWL simulating pilots' activity during the cruise, the

approach and the final approach/landing phases; and 2) the time-frequency brain activity (i.e., event-related spectral perturbations or ERSPs) time-locked to both gesture and written alarms as a function of the level of MWL. In all three MWL levels, participants had to perform an *alarm-response task*, in which they were presented with both *gesture alarms* – i.e., videos of a hand either pushing (i.e., push gesture alarm) or pulling (i.e., pull gesture alarm) a joystick – and the corresponding *written alarms* – i.e., written inscriptions of either the verb *"push"* (i.e., push written alarm) or the verb *"pull"* (i.e., pull written alarm). Depending on the alarm, they were required to either pull or push the joystick, as fast and accurately as possible. In order to prevent automatic responses to these alarms, participants were also presented half of the time with "stay" filler alarms (i.e., either a fixed gesture video or a written inscription) to which no response was expected. Participants had to respond to the alarms (i.e., alarm response task) while performing: 1) no other competing task to simulate the cruise phase (i.e., no arithmetic task), 2) a competing task generating a low level of MWL to simulate the approach phase (i.e., moderately difficult arithmetic task), and 3) a competing task generating a high level of MWL to simulate the final approach/landing phase (i.e., difficult arithmetic task). The arithmetic task was drawn from the Trier Social Stress Test, an experimental paradigm known to trigger MWL and stress (Birkett, 2011). This task consists in making complex calculations and in verbally providing the answers to the experimenter, simulating the pilots' activity (i.e., making complex calculations to operate the aircraft and verbally communicating with the crew and the air traffic control).

Based on the results of Fabre et al. (2021), we predicted that we might find no difference in response accuracy and reaction times between gesture alarms and written alarms under low MWL conditions. We also predicted that increasing participants' MWL would slow down their reactions and negatively affect the accuracy of their responses to both types of alarms and thatparticipants would be faster and more accurate to respond to gesture alarms

than to the written alarms under high levels of MWL. Regarding the brain electrophysiological activity, many studies have shown that mu and beta desynchronization are found when participants have 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). Therefore, 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 in the present study, where participants were expected to respond to the alarms. We also made the assumption that these differences in mu and beta power decrease between the two types of alarms might vary depending on 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. All had normal or corrected-to-normal vision. None of the participants 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 participants. Two were *gesture alarms* consisting in 1) a video of a hand pushing (i.e., push gesture alarm; see Figure 1.F.) and 2) a video of a hand pulling (i.e., pull gesture alarm; see Figure 1.D.) an aircraft joystick. Two were *written alarms* consisting in red written inscriptions of 1) the verb "PUSH" (i.e., push written alarm; see Figure 1.C) and 2) the verb "PULL" (i.e., pull written alarm; see Figure 1.A). Participants were expected to respond as fast as possible to these four alarms according to the action described by it (i.e., either pulling or pushing the joystick).

Insert Figure 1 here

To prevent automatic responses to the four alarms of interest described above, two filler alarms to which no response was expected were created. The written version of the filler consisted in a red written inscription of the verb "STAY" (see Figure 1.B.), while its gesture version consisted in a video of a motionless hand on a joystick (see Figure 1.E.). For the written alarms and written filler, the verb was displayed twice for 1800 ms followed by a grey background for 200 ms each time. For the gesture alarms and the gesture filler, the video lasted 2000 ms (the hand gesture lasted 1800 ms and the hand remained still for 200 ms at the end of the gesture) and was played twice.

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. Description of the tasks

2.4.1. Alarm response task

In the *alarm response task*, participants were presented with 1) the four alarms of interest (i.e., push gesture, push written, pull gesture, pull written alarms) that were played 15 times each, and 2) the two filler alarms (i.e., stay gesture and stay written alarms) that were played 30 times each. The alarms and the fillers were randomly presented to the participants for 4 seconds (see 2.3.1. Alarms), 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). The performance to this task was assessed in terms of 1) number of erroneous responses and 2) reaction times to the alarms of interest.

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 (i.e., low MWL run) or 13 (i.e., high MWL run) from a starting number (e.g., 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 2.5. 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). The performance to this task was assessed in terms of 1) number of mathematical operations, 2) number of mathematical errors and 3) error ratio (*number of errors/number of operations*).

2.5. 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 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* (see section 2.4.1.) 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; see section 2.4.2.). 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. 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.

At the end of each run, participants were asked to subjectively evaluate 1) the task difficulty of the run they just completed by filling the French validation of the NASA-TLX (Hart & Staveland, 1988) and 2) the extent to which they found it difficult to respond to the gesture alarms and the written alarms on a 7-point Likert scale (1 -> *not difficult at all* to 7 -> *extremely difficult*). 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 whole experiment, evaluating the correctness of the calculations that were being made. The proximity of the experimenter aimed at both simulating the presence of a second pilot and increasing the 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.

2.6. Data Analysis

2.6.1. Behavioral data analysis

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 (*number of errors/number of operations*) observed for the two runs. Three paired t-tests were conducted on these three types of data.

Alarm response task. Performances to the *alarm response task* were assessed in terms of reaction times and accuracy (i.e., erroneous responses). Reaction times above or below two standard deviations around the mean were eliminated (Ratcliff, 1993). 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. A response was considered as erroneous when a participant did not respond to an alarm of interest or responded to it in the inadequate way (i.e., pulling in response to a push alarm and pushing in response to a pull alarm). Because the data were not normally distributed (Shapiro-Wilk tests: $p_s < .05$), a 3 x 2 [MWL (no MWL, low MWL, high MWL) x Alarm (written, gesture)] linear logistic regression was performed on the errors made in response to the alarms of interest.

2.6.2. Subjective data analysis

NASA-TLX. A 3 levels (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 (Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, Frustration) were also analyzed separately. As the data 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)] logistic regression was performed on the ratings of difficulty to respond to the alarm from 1 (easy) to 7 (difficult), as the data were not normally distributed (Shapiro-Wilk tests: $p_s < .05$). A manual stepwise analysis was performed to remove non-significant interactions from the model.

2.6.3. EEG data analysis

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 between 0.1 and 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 done 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 averaged for each experimental condition separately according to the run (no MWL, low MWL and high MWL) and to the event type (type of alarm: gesture or written alarm). Event-related spectral perturbations (ERSPs) were computed time-locked to the event display. ERSPs were computed with Morlet wavelets as performed by EEGLAB 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). ERSPs were baseline-corrected with a common ERSP baseline across factors and 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 average associated with power increase (positive values) or decrease (negative values) was 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 a significant main effect of MWL was found. Paired t-test analyses were performed on the average power across participants in the frequency bands and time-windows in which a significant main effect of the type of alarm was found.

3. RESULTS

3.1. Behavioral results

3.1.1. Mental arithmetic task

The paired t-test conducted on the number of mathematical operations [t (23) = 13.94, p < .001, $CI_{95\%}$ (225.15, 303.60); see Figure 2.A.] revealed that participants made a significantly greater number of operations in the low MWL run (M = 406.12, SD = 112.23) than in the high MWL run (M = 141.75, SD = 66.44). The paired t-test conducted on the number of mathematical errors [t (23) = - 3.31, p = .003, $CI_{95\%}$ (- 6.10, - 1.41); see Figure 2.B.] revealed that participants made a significantly greater number of mathematical errors in the high MWL run (M = 7.79, SD = 4.70) than in the low MWL run (M = 4.04, SD = 3.02). Finally, the paired t-test performed on the error ratio (*number of errors/number of operation*) [t (23) = - 6.22, p < .001, $CI_{95\%}$ (- .07, - .04); see Figure 2.C.] revealed that participants had a significantly greater error ratio in the high MWL run (M = .06, SD = .04) than in the low MWL run (M = .01, SD =.01).

Insert Figure 2 here

3.1.2. Alarm response task

Erroneous responses to the alarms. The 3 x 2 [MWL (no MWL, low MWL, high MWL) x Alarm (written, gesture)] logistic regression analysis revealed a 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 3 x 2 [MWL (no MWL, low MWL, high MWL) x Alarm (written, gesture)] repeated-measures ANOVA analysis performed on the corrected log-transformed reaction times revealed a 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). The analysis also revealed a main effect MWL [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, $p = .35$, $\eta p^2 = .04$].

Insert Figure 4 here

3.2. Subjective results

3.2.1. NASA-TLX questionnaire.

The 3 (MWL [no MWL, low MWL, high MWL]) one-way ANOVA conducted on the NASA-TLX raw data revealed a main effect of MWL [F (2, 46) = 109.89, $p < .001$, $\eta p^2 = 1.00$; 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 the standard deviations are reported in Table 1.

Insert Figure 5 here

3.2.2. Difficulty to respond to the alarm.

The 3 x 2 [MWL (no MWL, low MWL, high MWL) x Alarm (written, gesture)] logistic regression 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 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

3.3.1. Effect of the MWL

Statistical analyses (permutation tests) revealed an across-trial significant main effect of MWL on high frequency bands (> 12Hz; see Figure 7). To further investigate these effects, three 3 [MWL (no MWL, low MWL, high MWL)] one-way repeated-measures ANOVAs were conducted on the average power measured in the 0 – 1441 ms time-window after the onset of the alarms measured for 1) the whole beta frequency band [12 - 30 Hz[, 2) the low beta frequency band [12 - 15 Hz[and 3) the high beta frequency band [15 - 30 Hz[. Tukey post-hoc tests were performed. For the sake of clarity, these results are reported in Table 2.

Insert Table 2 here

3.3.2. Effect of the type of alarm

When considering the effect of the type of alarm (gesture alarms versus written alarms), we observed a statistically significant difference in the delta frequency band at both C3 and C4 electrodes respectively in the 660 − 720 ms, and the 590 − 820 ms time windows after the alarm onset. At both electrodes, we observed a higher power increase in the delta band for the gesture alarm compared the written alarm (see Table 3). Oppositely, we observed a decreased activity in the theta and mu frequency bands for gesture compared to written alarms (more negative power) at both C3 and C4 electrodes. In the theta band, this decrease was statistically significant in the 1000 − 1440 ms time window at C3 and in the 970 − 1240 ms time window after the onset of the alarms C4 for gesture and written alarm respectively (see Table 3).

Insert Figure 7 here

In the mu frequency band, the activity was also statistically lower for gesture compared to written alarms in the 1020 − 1350 ms and the 950 − 1100 ms time windows after the onset of the alarms for C3 and C4 respectively. Finally, a significant decrease was also observed in the low beta [12-15 Hz[frequency band, but only at the C3 electrode in the 1170 − 1230 ms time window, with the power for gesture alarms significantly lower than the one associated with written alarms (see Figure 7).

Insert Table 3 here

4. DISCUSSION

The aim of the present experiment was twofold: 1) investigating whether gesture alarms are more efficient to trigger fast and accurate responses than written alarms, especially when the MWL is high; 2) investigating the ERSPs observed in response to both types of alarms as a function of the level of MWL. Participants had to perform an alarm-response task where participants had to respond accurately to gesture alarms and written alarms under three

different levels of MWL − induced by subtraction tasks drawn from the Trier Social Stress Test (Birkett, 2011).

4.1. MWL manipulation check

In line with the literature, participants made significantly less mathematical operations, more calculation errors, and were overall less performant (i.e., higher *number of errors/number of operations* ratio) in the 13-subtraction task (i.e., high MWL run) than in the 5-subtraction task (i.e., low MWL run). 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 the low MWL run than the no MWL run. Taken together, these results serve as a manipulation check and confirm that participants' level of MWL was significantly higher in the low MWL run than in the no MWL run; and in the high MWL run than in the low MWL run.

4.2. Impact of MWL on the response to the alarms

The results showed that participants took significantly longer, made more erroneous responses, and reported that they found it more difficult to respond to the alarms (i.e., both written and gesture alarms) in the high MWL run than in both the no MWL run and the low MWL run. 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. Taken together, these results suggest that while participants experienced an increase in MWL in the low MWL run compared to no MWL run (i.e., longer reactions times observed in the low MWL run), they might have been able to recruit enough cognitive resources to prevent a decrease in response accuracy. In the high MWL run however, the MWL appears to have been so intense that despite the increase in reaction times participants were no longer able to maintain the response accuracy observed

in the two other runs, resulting in a significant decrease in general performance (i.e., both accuracy and rapidity) to the *alarm-response task*.

At the brain electrophysiological level, greater beta activity was found in response to the alarms in general (i.e., both written and gesture alarms) 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 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 arithmetic task to trigger MWL.

4.3. Gesture alarms versus written alarms

While we predicted that we would find differences in both reaction times and response accuracy between the gesture alarms and the written alarms only under high levels of MWL, participants were found to be both faster and more accurate when responding to the gesture alarms than to the written alarms, regardless of the level of MWL. Moreover, participants reported that they found it less difficult to respond to gesture alarms than to written alarms, once again regardless of the level of MWL. Taken together, these results demonstrate that, independently of the level of MWL, gesture alarms were associated with a higher performance to the alarm response task (i.e., faster and more accurate responses) than written alarms.

In line with our predictions and the results of our previous study (Fabre et al., 2021), greater mu and low beta power decreases were also found in response to the gesture alarms

than to the written alarms when participants had to respond to the alarms. Mu and beta power decrease, but also 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). Therefore, the results of the present study indicate that, compared to written alarms, gesture alarms might facilitate the preparation and/or the execution of the action. Significant power decrease in the mu frequency band for gesture compared to written alarms were found in the 950 − 1110 ms (C3) and the 1020 − 1350 ms (C4) time windows, while the difference in low power decrease was significant in the 1170 − 1230 ms time window (C3) after the onset of the alarm. In other words, these power differences were significant in the time window of the response to the alarms (i.e., on average 930 ms for the gesture alarms and 1140 ms for the written alarms). These results suggest that the greater mu power and the beta power decrease observed in response to gesture alarms are more likely to reflect the facilitation of the action execution triggered by this type of alarm (rather than the facilitation of its preparation). However, since we were not able to compare the signal associated with the onset of the response observed for both gesture and written alarms (i.e., triggers were not sent on the EEG signal when participants performed the action), we cannot conclude with certainty on this point. Further research is therefore necessary to confirm this assumption. The results also suggest a greater high delta (3-4 Hz) synchronizations in response to the gesture alarms than to the written alarms in the 660 − 720 ms (C3) and the 590 − 820 ms (C4) time windows after the onset of the alarms. Increased delta activities were found to underpin action selection and/or execution in cognitively demanding situations (Beste et al., 2012), and were also found to be correlated to performance at action selection (Hamel-Thibault, et al., 2018). In the present study, participants had to inhibit their responses half of the time and had to accurately select and execute one of two possible actions (i.e., push or pull). 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, the present study was not designed to investigate the delta rhythm. Indeed, 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. In addition, this number of cycles which is required to reach a sufficient time-domain accuracy at low frequencies induces a frequency leakage from close frequencies activity (Cohen, 2014). Therefore, this result should be taken with caution. Further research is necessary to confirm this potential difference in delta synchronization between gesture and written alarms.

4.4. Applications and future works

The behavioral results of the present study demonstrate the greater efficiency of gesture alarms to trigger fast and 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. The results also suggest that it might be possible to safely implement different gesture alarms to improve reaction in various emergency situations.

Besides preventing CFIT accidents, gesture alarms could also help preventing Loss Of Control In-Flight (LOC-I) accidents, which are by far the main cause of fatalities and costs in commercial aviation (IATA, 2021). The latter refer to accidents in which the flight crew was unable to maintain control of the aircraft in flight, which resulted in an unrecoverable deviation from the intended flight path (IATA, 2015). LOC-I accidents usually result from a failure to prevent and/or recover from a stall (i.e., an aerodynamic loss of lift caused by exceeding the critical angle of attack) or an upset i.e., an undesired state of the aircraft characterized by unintentional divergences from parameters normally experienced during

operations (e.g., a pitch attitude greater than 25° nose up or than 10° nose down, a bank angle greater than 45° or inappropriate airspeeds; IATA, 2018b). While LOC-I accidents result from multiple events and conditions which interact in unpredictable and unforeseen ways, inappropriate crew actions are one of the leading causes of these accidents (Sherry & Mauro, 2014; Oliver et al., 2019). This is partially due to the fact that the maneuvers to perform to prevent and/or recover from a stall or an upset (e.g., reducing thrust and performing a nose-down pitch action to recover from a stall while the aircraft is already plummeting) can be both stressful and counter-intuitive for pilots (IATA, 2018b; e.g., Air France Flight 446: BEA, 2012; Colgan Air Flight 3407: NTSB, 2010). Moreover, the alerts that warn the pilots of the impending stall or upset only indicate that a parameter is dangerously deviating (e.g., aural "STALL" or "BANK ANGLE"; Ewbank, et al., 2016) and the recovery maneuver is not "*primed*" as with the GPWS alert. We assume that gesture alarms could help pilots take action and perform these recovery maneuvers before the deviation of the flight path becomes unrecoverable.

The implementation of gesture alarm systems may also help improve the reactions of operators in many fields where fast and accurate responses are needed. Gesture alarms may be particularly adapted to nuclear powerplants and transportation (i.e., air, rail, or road transport), to improve emergency shutdown or collision avoidance maneuvers for instance. They could also be used in the medical field. It is long known that the design of many auditory warning signals in hospitals is far from optimal (Alizezaee et al., 2017; Edworthy, 2013). Unlike the field of aviation where a unique constructor is responsible for the design of all the alarm systems that are implemented in the cockpit, the medical devices used in the hospital are designed by different constructors, making the harmonization of the many alarm systems extremely difficult (Bach et al., 2018). It creates a noisy and annoying environment − especially in intensive care units (ICU) and operation rooms (OR) − that is likely to

jeopardize safety (for a review, see Konkani et al., 2012). For instance, this ill-adapted auditory environment was found to trigger alarm fatigue (i.e., the desensitization to the alarms occurring when an individual is exposed to a high number of alarms; Oliveira, et al., 2018) in healthcare professionals. Moreover, so many different auditory alarms sound in OR and the ICU that medical staff is often unable to properly discriminate and/or react to them, putting patients' safety at jeopardy (Bridi et al., 2014; Momtahan et al., 2007). Combining each of the aural alerts played in the OR or the ICU with the corresponding gesture alarm could help healthcare professional deal with the important amount of emergency signals they are exposed to and improve safety in medical environments.

However, before gesture alarms could be implemented, further research is necessary to investigate how the association with the corresponding aural alerts would affect operators. Based on past research, it is probable that adding the corresponding auditory modality to the gesture-based visual modality would improve even more the efficiency of gesture alarms (Alirezaee et al., 2017; Child & Wendt, 1938; Hughes et al., 1994; Liu, 2001). Even though this assumption remains to be confirmed, our study provides a step forward towards the implementation of such gesture-based alarm aiming at triggering faster and more scrupulous reactions in operators.

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.

Bach, T. A., Berglund, L. M., & Turk, E. (2018). Managing alarm systems for quality and safety in the hospital setting. *BMJ open quality*, *7*(3), e000202. http://dx.doi.org/10.1136/bmjoq-2017-000202

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.

Bridi, A. C., Silva, R. C. L. D., Farias, C. C. P. D., Franco, A. S., & Santos, V. D. L. Q. D. (2014). Reaction time of a health care team to monitoring alarms in the intensive care unit: implications for the safety of seriously ill patients. *Revista Brasileira de terapia intensiva*, *26*, 28-35. https://doi.org/10.5935/0103-507X.20140005

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

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

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.

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.

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

Dehais, F., Causse, M., Vachon, F., Régis, N., Menant, E., & Tremblay, S. (2014). Failure to Detect Critical Auditory Alerts in the Cockpit Evidence for Inattentional Deafness. *Human Factors: The Journal of the Human Factors and Ergonomics Society*. https://doi.org/10.1177/0018720813510735.

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.

Doll, T. J., Folds, D., & Leiker, L. A. (1984). *Auditory Information Systems in Military Aircraft: Current Configurations Versus the State of the Art*. Atlanta Systems Engineering Lab: Georgia Institute of Technology.

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

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.

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

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.

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

IATA (2018b). *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 (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.

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.

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.

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.

Konkani, A., Oakley, B., & Bauld, T. J. (2012). Reducing hospital noise: a review of medical device alarm management. *Biomedical Instrumentation & Technology*, *46*(6), 478-487. https://doi.org/10.2345/0899-8205-46.6.478

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

Lin, C. J., Lin, P. H., Chen, H. J., Hsieh, M. C., Yu, H. C., Wang, E. M. Y., & Ho, H. L. (2012). Effects of controller-pilot communication medium, flight phase and the role in the cockpit on pilots' workload and situation awareness. *Safety science*, *50*(9), 1722-1731. https://doi.org/10.1016/j.ssci.2012.04.007

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

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/10.1027/2192-0923/a000092

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/a0000 92.

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

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

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

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

Oliveira, A. E. C. D., Machado, A. B., Santos, E. D. D., & Almeida, É. B. D. (2018). Alarm fatigue and the implications for patient safety. *Revista brasileira de enfermagem*, *71*, 3035-3040. https://doi.org/10.1590/0034-7167-2017-0481.

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

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.

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.

Rouwhorst, W., Verhoeven, R., Suijkerbuijk, M., Bos, T., Maij, A., Vermaat, M., & Arents, R. (2017, September). Use of touch screen display applications for aircraft flight control. In *2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC)* (pp. 1-10). IEEE. https://doi.org/10.1109/DASC.2017.8102060

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. *Focusing Attention on Aviation Safety*.

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

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

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

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.

Wiggins, M., & O'Hare, D. (1995). Expertise in aeronautical weather-related decision making: A cross-sectional analysis of general aviation pilots. *Journal of Experimental Psychology: Applied*, *1*(4), 305. https://doi.org/10.1037/1076-898X.1.4.305

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.

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).

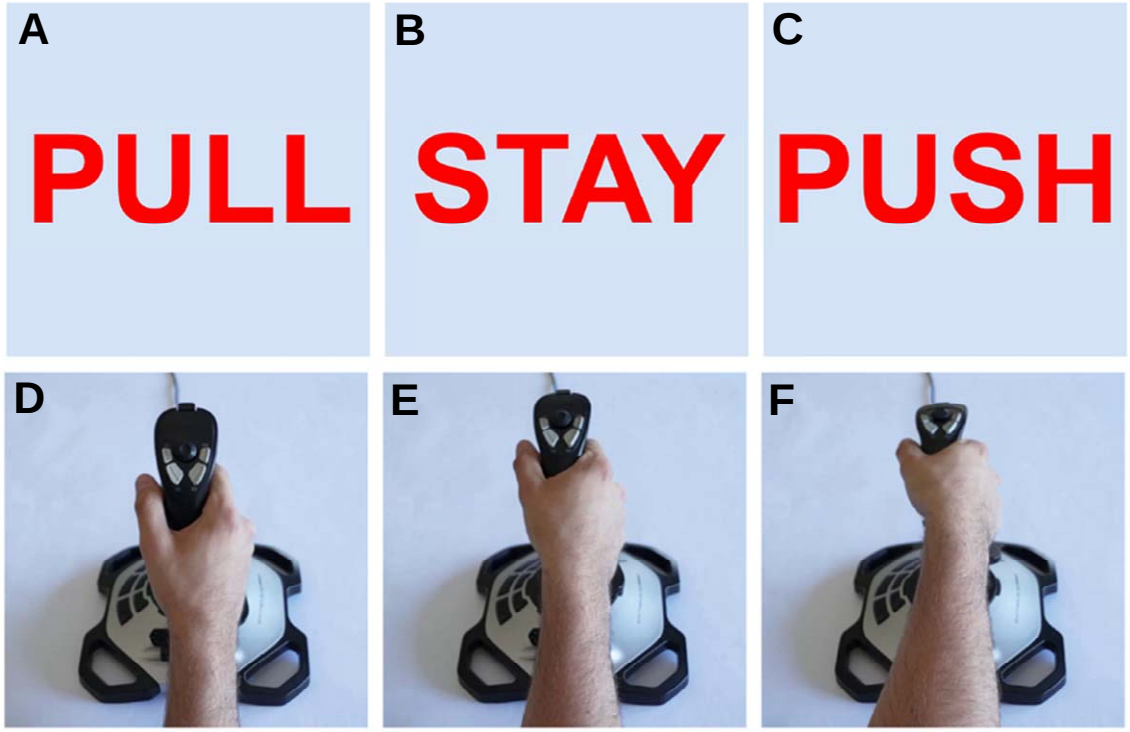


Figure 1. Illustration of the three written alarms A) Pull, B) Stay, C) Push and the three corresponding gesture alarms D) Pull, E) Stay, F) Push.

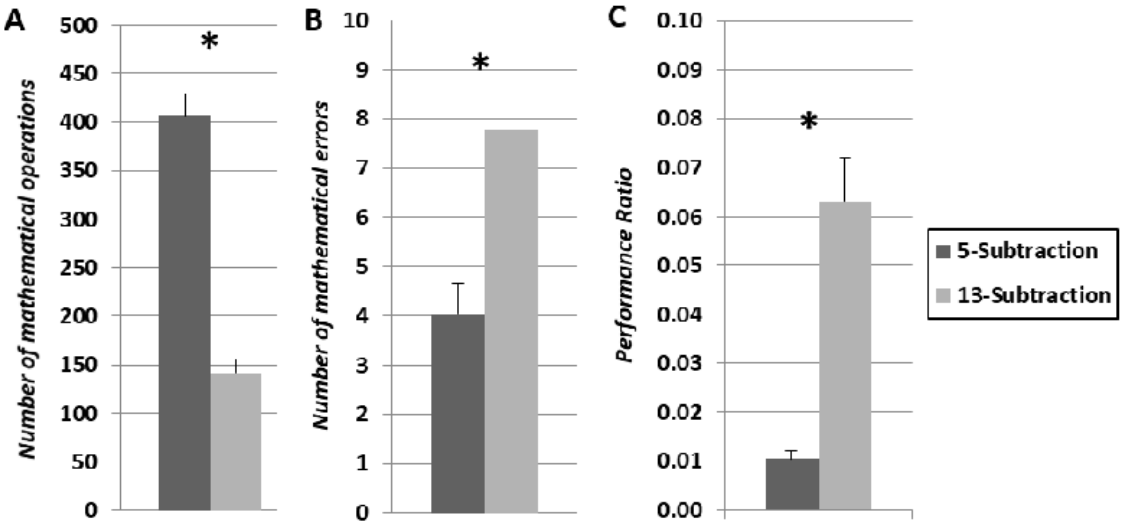
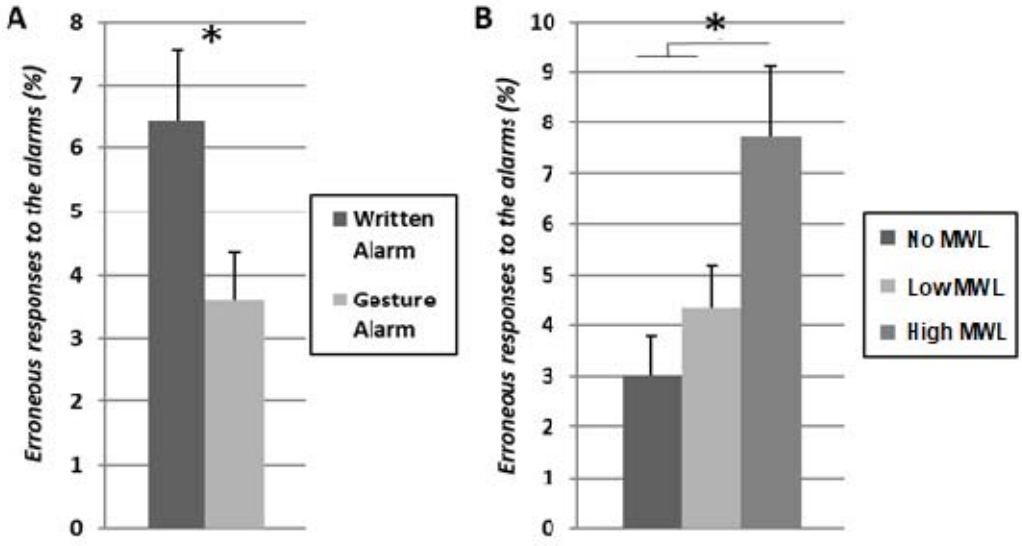


Figure 2. Illustration of (A) the number of mathematical operations, (B) the number of mathematical errors and (C) the error ratio in the low MWL run (dark grey) and in the high MWL run (light grey). * $p < .01$

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). *: $p < .05$



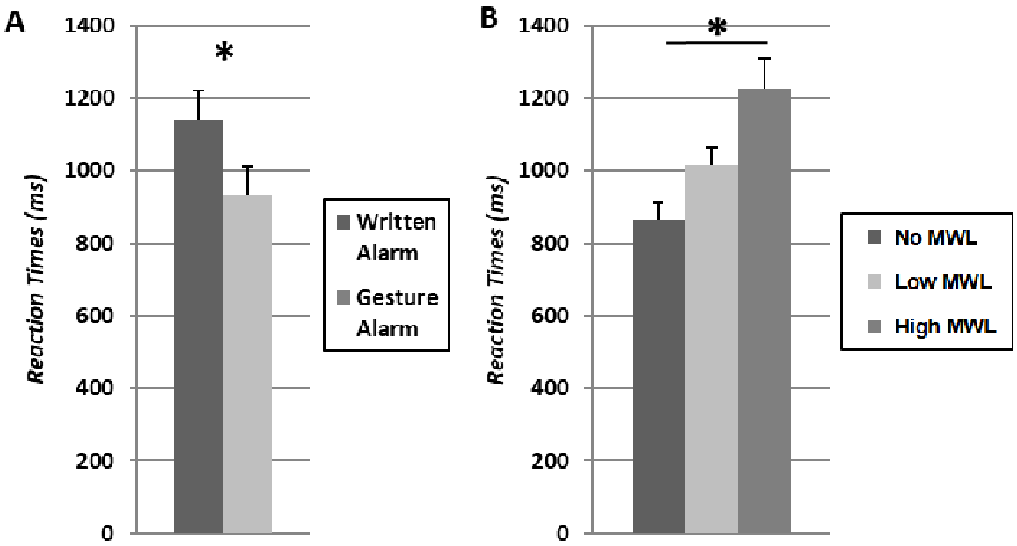


Figure 4. Illustration of the reaction times to the alarm (ms) as a function of (A) the type of alarms with written alarms (dark grey) and gesture alarms (light grey); and (B) the level of MWL with no MWL (dark grey), low MWL (light grey) and high MWL (mid grey). *: $p < .05$

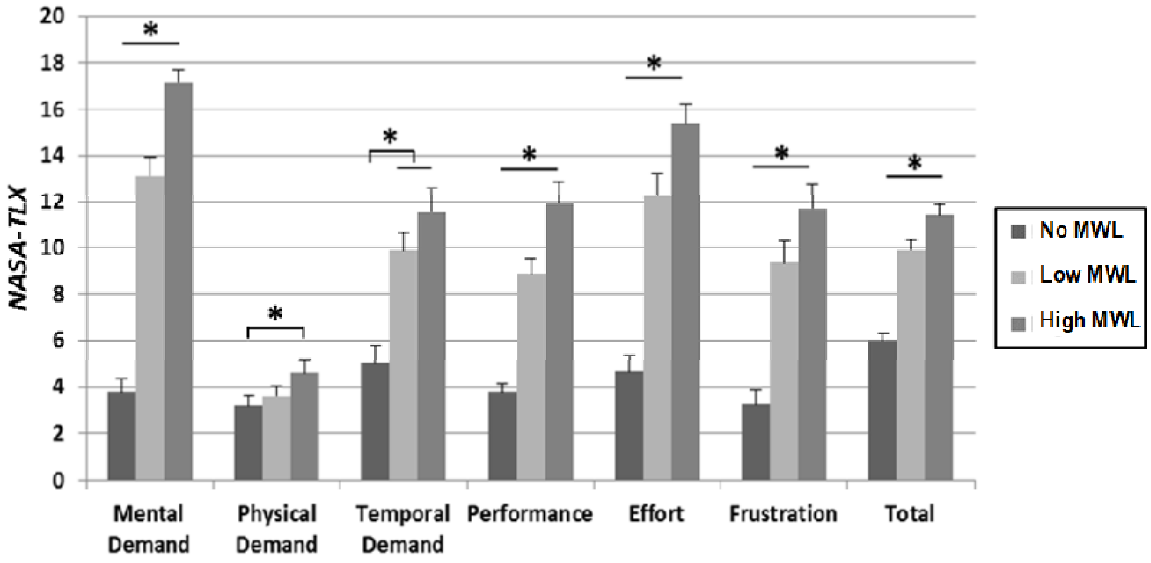


Figure 5. Illustration of the NASA-TLX ratings observed in the no MWL run (dark grey), the low MWL run (light grey) and the high MWL run (middle grey).*: $p < .05$

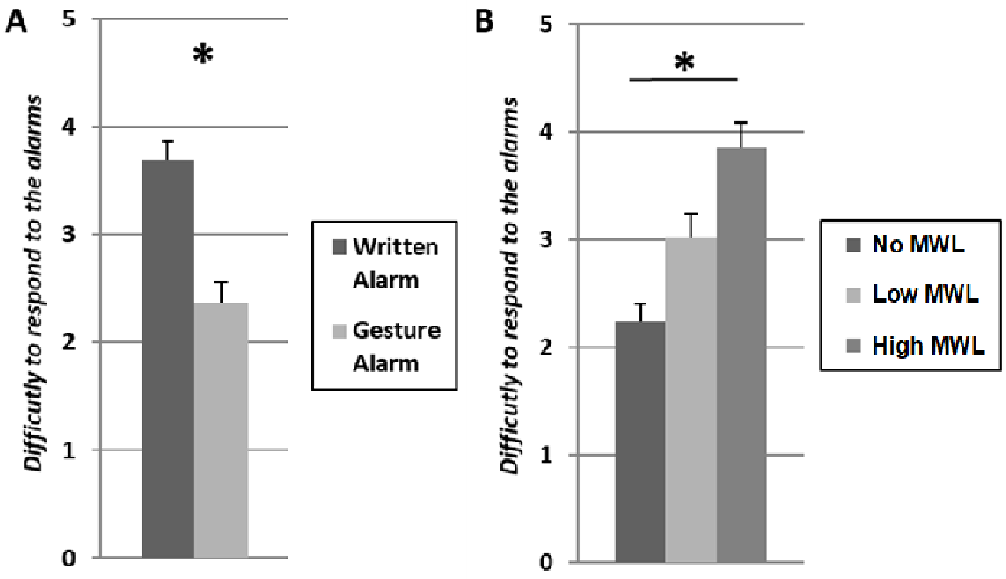


Figure 6. Illustration of the difficulty to respond to the alarms as a function of (A) the type of alarms with written alarms (dark grey) and gesture alarms (light grey); and (B) the level of MWL with no MWL (dark grey), low MWL (light grey) and high MWL (mid grey). *: $p < .001$.

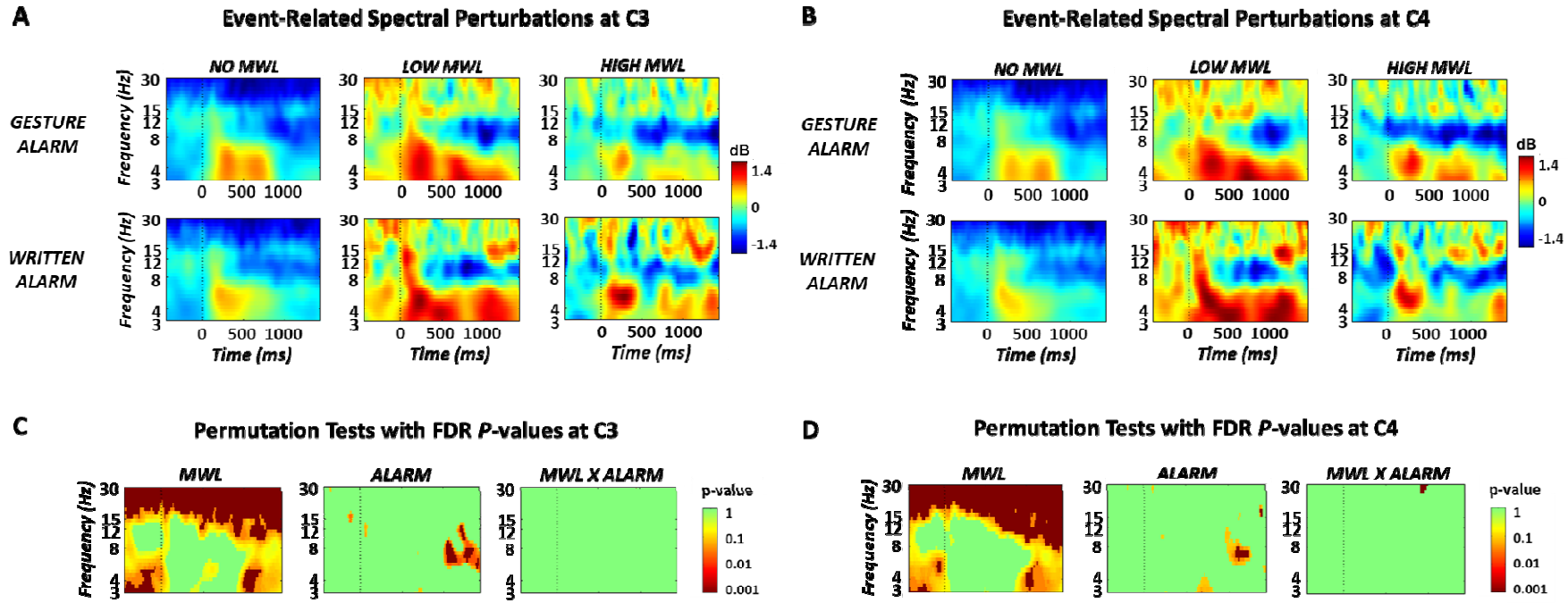


Figure 7. Event Related Spectral Perturbations (ERSPs) showing increases (positive values – red) and decreases (negative values – blue) in dB-corrected power in the three MWL levels (no MWL – left, low MWL – middle, and high MWL – right) for the two types of alarms (Gesture – top and written – bottom) 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).