



HAL
open science

Fast & scrupulous: Gesture-based alarms improve accuracy and reaction times under various mental workload levels. An ERSP study

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

► **To cite this version:**

Eve Floriane Fabre, Bertille Somon, Valeria Baragona, Quentin Uhl, Mickaël Causse. Fast & scrupulous: Gesture-based alarms improve accuracy and reaction times under various mental workload levels. An ERSP study. *Applied Ergonomics*, 2023, 113, pp.104082. 10.1016/j.apergo.2023.104082 . hal-04165468

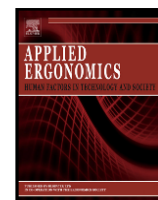
HAL Id: hal-04165468

<https://hal.science/hal-04165468>

Submitted on 19 Jul 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Fast & scrupulous: Gesture-based alarms improve accuracy and reaction times under various mental workload levels. An ERSP study[☆]

Eve Floriane Fabre^{a, *}, Bertille Somon^b, Valeria Baragona^a, Quentin Uhl^a, Mickaël Causse^a

^a ISAE-SUPAERO, Neuroergonomics and Human Factors Research Group, DCAS, Toulouse University, France

^b ONERA, The French Aerospace Lab, Information Processing and Systems Department, Salon Cedex Air, France

ARTICLE INFO

Keywords:

Gesture alarms
Emergency situations
Operators
Safety
EEG
Time-frequency analysis

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 observed at C3 and C4 electrodes. 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, 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 (e.g., Brown et al., 2000; Torralba et al., 2007; Wu and Li, 2018).

Alarm systems are expected to distract operators from the task currently at hand and provide them with crucial information such as the severity of the failure, its cause, its location and even sometimes how to respond to it (Guillaume, 2011). Emergency alarm systems are generally multimodal and combine symbolic (e.g., flashing lights, bibles, or bips) and language-based alarms (Guillaume, 2011). The latter can either be *simple awareness warnings*, which indicate the nature of the emergency but not how to respond to it (e.g., Stall alarm; IATA, 2018a; Whittemore and Woods, 2021), or *pro-active instructions* – such as the Ground Proximity Warning System (GPWS or “Pull Up” alarm) – that

state how to respond to the emergency without explicitly outlining its nature (e.g., EASA, 2007; IATA, 2018b; Moroze and Snow, 1999).

1.2. Why operators sometimes fail to appropriately react to emergency situations

Despite the important efforts 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 and Efthymiou, 2019; Scoot, 1996; Whittemore and Woods, 2021). For instance, pilots' delayed or inadequate reaction to the Stall or the GPSW alarms remains an important contributive factor in most Controlled Flight Into Stall (CFIS; IATA, 2018a; Oliver et al., 2019; Sherry and Mauro, 2014) and Controlled Flight Into Terrain (CFIT; Cooper, 1995; Moroze and Snow, 1999; Shappell and Wiegmann, 2001, 2003) accidents, which account for a large part of commercial aviation fatalities (IATA, 2021; ICAO, 2019).

There are various explanations to this phenomenon. First, due to ergonomics flaws, emergency alarms are likely to blend with concurrent visual and auditory information (Corwin, 1995; Guillaume, 2011; Momtahan et al., 1993), which can impair their detection and/or recognition (Scoot, 1996; Whittemore and Woods, 2021). Second, individu-

[☆] 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).

* Corresponding author.

E-mail address: fabreeve@gmail.com (E.F. Fabre).

<https://doi.org/10.1016/j.apergo.2023.104082>

Received 4 April 2022; Received in revised form 24 June 2023; Accepted 27 June 2023
0003-6870/© 20XX

ally or combined, mental workload (MWL) and perceptual workload can disrupt the allocation of cognitive resources to an incoming emergency alarm (Wickens, 2008, 2020), affecting its detection and process (Causse et al., 2016b; Giraudet et al., 2015; Lavie et al., 2014; Raveh and Lavie, 2015), and eventually impairing the operators' response (e.g., Bliss and Dunn, 2000; Bliss, 2003; Causse et al., 2016a; Causse et al., 2022; Simonson et al., 2022). This phenomenon is referred as *inattention deafness* or *blindness* depending on the nature of the stimulus (e.g., Dehais et al., 2019; White and O'Hare, 2022; Zhu et al., 2022). Finally, even though an alarm has been successfully detected and processed by operators, different factors, such as stress or MWL, can still impair their cognitive ability and psychomotor skills (Martin et al., 2012; Schmidt et al., 2008), and trigger ill-adapted responses like freezing (DeCelles, 1991; TSB, 1998), delayed (Shaw et al., 2018; NTSB, 2000) or startle reactions (Martin et al., 2015; BEA, 2012).

1.3. Improving operators' reactions with gesture alarms?

As 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 and Jones, 2004; Stevens et al., 2009) that can spread out MWL and allow optimal responses (Bliss and Gilson, 1998). Therefore, Whittemore and Woods (2021) recommend using proactive instruction alarms in emergency situations, instead of simple awareness warnings.

Because action videos both strongly attract attention (Abrams and Christ, 2003) and pre-activate the motor cortex (Rizzolatti et al., 1996), they might facilitate the detection of emergencies and the initiation of the expected motor responses (Brucker et al., 2015; Iacoboni et al., 1999; Jeannerod, 1994; Meltzoff and Prinz, 2002). Therefore, action videos represent a potential new and efficient way to provide operators with proactive instructions in emergency situations (Fabre et al., 2021).

To test this assumption, Fabre et al. (2021) conducted a two-experiment study investigating 1) the response performances to two types of alarms, 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. 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 response to gesture alarms might result from a greater motor preparation and pre-activation of the "Pull Up" recovery maneuver (Fadiga et al., 1995). However, no difference in response accuracy was found between the two alarm types, supposedly because participants had to execute the same action (i.e., Pull Up) in response to both alarm types, which might have triggered automatic responses and erased a potential difference in accuracy.

Given the effect of MWL on detection, processing and responding to alarms (Simonson et al., 2022), it is essential to evaluate whether and to what extent high levels of MWL might affect the response to new emergency systems, such as action video-based alarms, 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 different levels of MWL, and the associated event-related spectral perturbations (ERSPs). Participants performed an *alarm response task* drawn from the Stall and GPWS alarms and the expected responses to the associated 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" – and concurrently perform no other competing task (no MWL), a competing mental calculation task generating a low level of MWL and a high level of MWL. We predicted to observe faster reaction times to gesture alarms than to written alarms and no difference in response accuracy between the two alarm types in the no-MWL run (Fabre et al., 2021). We also predicted that increasing participants' MWL would overall negatively affect the accuracy and the speed of the responses to both alarm types; and that participants would be faster and more accurate when responding to gesture alarms than to written alarms as the level of MWL increased.

Additional insights into the deciphering of the mechanisms allowing improved performances to gesture alarms can also be provided through EEG measures. 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 desynchronizations are found at central sites when participants observe and/or imitate an action (Meyer et al., 2020; Quandt et al., 2012), and also when they read action language (Moreno et al., 2013; Niccolai et al., 2014). This desynchronization reflects the preparation and/or the execution of actions (and supposedly the activity of mirror neuron system; Fox et al., 2016). We predicted that greater mu and beta power decreases would arise in response to gesture alarms than to written alarms (Fabre et al., 2021), reflecting the greater facilitation of the response preparation and/or execution triggered by gesture alarms. We also predicted that these differences in mu and beta power between the two types of alarms would vary as a function of the level of MWL (Behmer and Fournier, 2014; Jenson et al., 2019; Muthukumaraswamy and Singh, 2008).

2. Material and methods

2.1. Participants

A review of the literature for sample size estimation – performed with the Pingouin package (v.0.5.3) in Python (v3.9) – revealed the necessity to record at least 14 participants to observe both the effect of the action type on performances and on EEG ($n \geq 8$ for both; Behmer and Fournier 2014; Fabre et al., 2021) and the effect of workload on performances and on EEG ($n \geq 14$ for both; Behmer and Fournier, 2014; Jaquess et al., 2017, 2018; Reiser et al., 2019) with a power $(1 - \beta) = 95\%$ and a significance level $\alpha = 5\%$. In addition, the sample size estimation for the effect of the interaction between action type and workload revealed a minimal sample size of 24 participants (Behmer and Fournier, 2014) required to observe this effect with a power of 95% and a significance of 5%. 28 participants ($M_{age} = 26$, $SD = 8$, age range: 20–48 years old; 12 females) took part in the present study. They were recruited using posters at the University. They were university students, PhD students or postdoctoral fellows with a background in either engineering or neuroscience. All had normal or corrected-to-normal vision and reported no history of prior neurological disorder.

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). All participants were informed of their rights and gave written informed consent.

2.3. Material

2.3.1. Alarms

Fig. 1 illustrates the six different stimuli used in this study. The stimuli of interest were two *written alarms* consisting in a red written in-

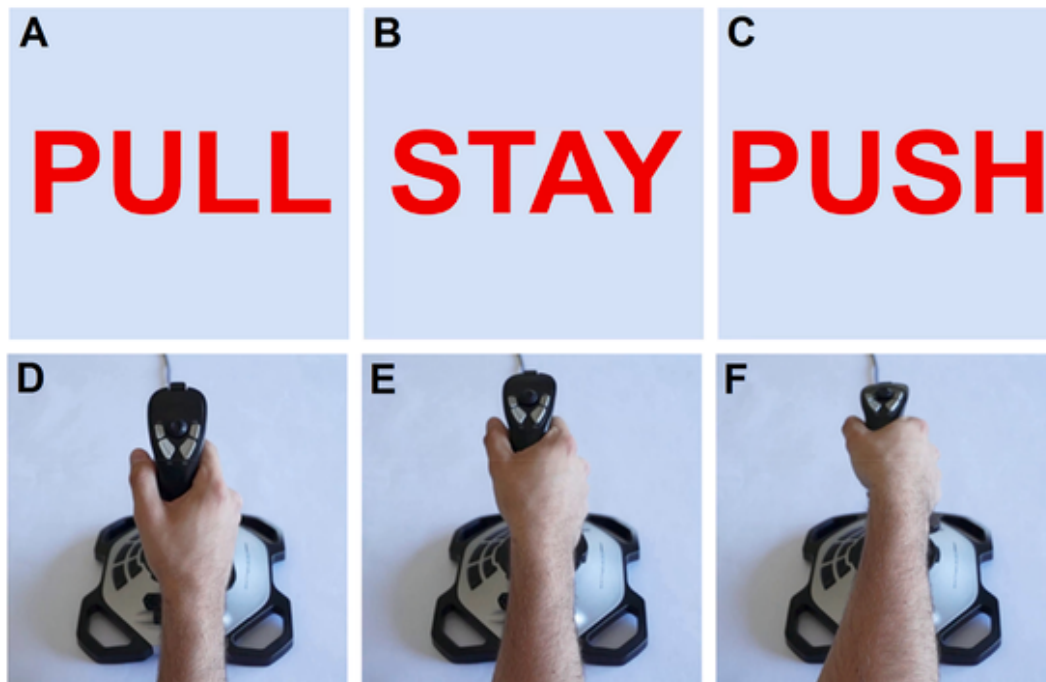


Fig. 1. Illustration of the two (A) PULL and (C) PUSH written alarms and the (B) STAY written filler; and the corresponding (D) PULL and (F) PUSH gesture alarms and (E) gesture filler. For the written alarms and written filler, the verb was displayed twice for 1800ms followed by a grey background for 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.

scription of the verb “PULL” or “PUSH” (respectively Fig. 1A and 1C) and two *gesture alarms* consisting in videos of a hand pulling or pushing a joystick (respectively Fig. 1D & 1F).

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 Fig. 1B & 1E.).

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.

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 (placed below both eyes) and horizontal (placed near both eyes’ outer canthi) 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 s, with a variable inter-trial interval lasting between 4 and 6 s. Participants’ task consisted in responding to the alarms of interest according to the action they described by 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 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 had to repeatedly subtract 5 in the low-MWL run and 13 in the high-MWL run from a fixed starting number (1022). They orally stated their response and continued their subtraction loop until they made a mistake or reached the closest number to zero. In the former case, participants were notified by the experimenter that they made a mistake and had to start all over again from the starting number. In the latter case, they performed the same task a second and sometimes a third time from a different starting number (i.e., 1023 and 1024 respectively) until the end of the run.

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. 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 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 as the number of mathematical operations, number of mathematical errors and error ratio (i.e., 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 they just completed by filling the six subscales of the NASA-TLX questionnaire – which assess the associated perceived mental demand, physical demand, temporal demand, performance, effort and frustration (Hart and Staveland, 1988). Afterwards, they were asked to evaluate the extent to which they found it difficult to respond to the written alarms and 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 ERSPs time-locked to the onset of written alarms and gesture alarms were investigated in all three levels of MWL (i.e., runs). EEG data were analyzed using EEGLAB v.13.6.5b open-source software (Delorme and Makeig, 2004) on Matlab 2017a. First, EEG signals were band-pass filtered using a basic [0.1–30] Hz FIR filter. An independent component analysis was performed to isolate and remove eye-blinks and movement-related artifacts from the signal (Jung et al., 1998). A visual inspection of the data was performed to reject residual artifact intervals. Data were segmented into epochs from -1000 to 2000ms around the stimuli onset. Four participants were removed from the analysis due to the low quality of their EEG signal. For the remaining 24 participants, epoched data were separated for each experimental condition according to the run and the event type. ERSPs were computed with Morlet wavelets, time-locked to the event onset. They were computed with 30 frequencies in a window from 3 to 30 Hz, three wavelet cycles at the lowest frequency, an increasing factor of 0.8 and 350 time-points in the -441 to 1441ms time-window (with 0 corresponding to the onset of the stimuli). Finally, 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 informed 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 pixels 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 (no-MWL run), once simultaneously to a moderately demanding *mental arithmetic task* (low-MWL run), and once simultaneously to a highly demanding *mental arithmetic task* (high-MWL run).

They were expected to respond as accurately and as fast as possible to different alarms in the *alarm response task* and to make as many operations as possible without committing any calculation error in the *mental arithmetic task*. 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 min), participants were asked to fill the NASA-TLX questionnaire (Hart and 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 run started. The experimenter sat next to participants during the three runs (except when they were filling the scales), evaluating the calculations correctness. The experimenter's proximity aimed at simulating the presence of a second operator and increasing participants' MWL (Belletier et al., 2015). At the end of the experiment, participants were debriefed by the experimenter and thanked for participating 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 the number of mathematical operations, the number of mathematical errors and the error ratio.

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 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 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 0.017 (0.05/3 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 per-

formed 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 Fig. 7C & 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 Fig. 7C & D).

3. Results

3.1. Behavioral results

3.1.1. Mental arithmetic task

Fig. 2 illustrates participants' performances to the *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 Fig. 2A.). 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 Fig. 2B.). Finally, their error ratio was significantly greater in the high-MWL run [$t(23) = -6.22, p < .001, CI_{95\%} (-0.07, -0.04; M = .06, SD = .04)$] than in the low-MWL run ($M = .01, SD = .01$; see Fig. 2C.).

3.1.2. Alarm response task

Erroneous responses to the alarms. Fig. 3 illustrates participants' accuracy to the *alarm response task*. The analysis revealed a significant main effect of alarm [$B(SE) = 2.833(1.106), CI_{95\%} = (0.666, 5.001), Wald \chi^2(1) = 6.565, p = .010$; see Fig. 3A.], 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: $B(SE) = -4.679(1.270), CI_{95\%} = (-7.169, -2.189), Wald \chi^2(1) = 13.563, p < .001$; low-MWL: $B(SE) = -3.377(1.362), CI_{95\%} = (-6.048, -0.707), Wald \chi^2(1) = 6.143, p = .013$; with high-MWL as dummy; see Fig. 3B.]. The high-MWL run ($M = 7.60\%, SD = 8.94$) predicted for a lower accuracy than both the low-MWL run ($M = 4.34\%, SD = 5.50, p = .013$)

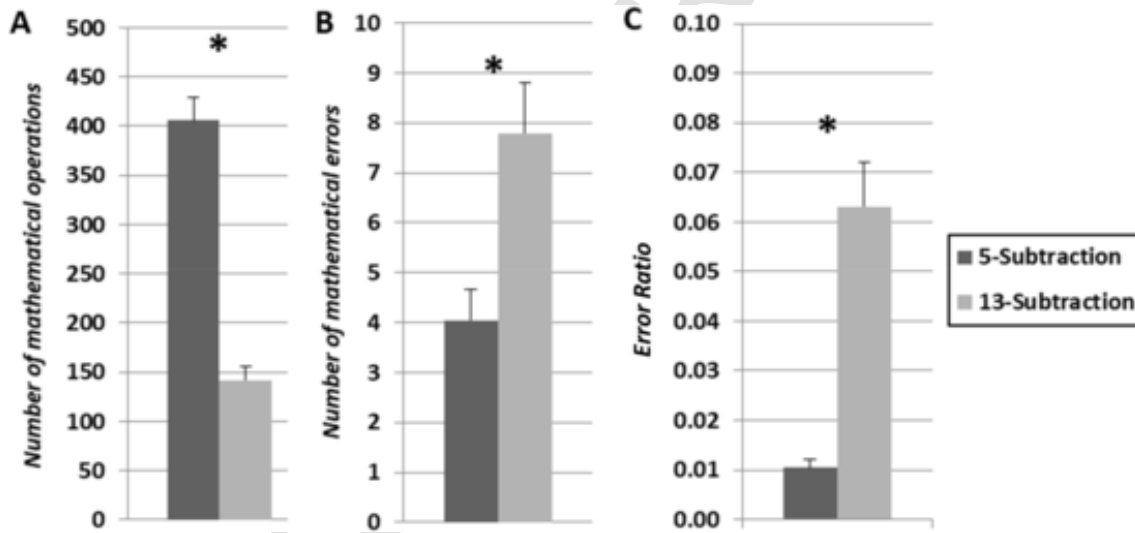


Fig. 2. Illustration of (A) the number of mathematical operations, (B) the number of mathematical errors and (C) the error ratio in the low MWL run (dark grey) and in the high MWL run (light grey). * $p < .01$.

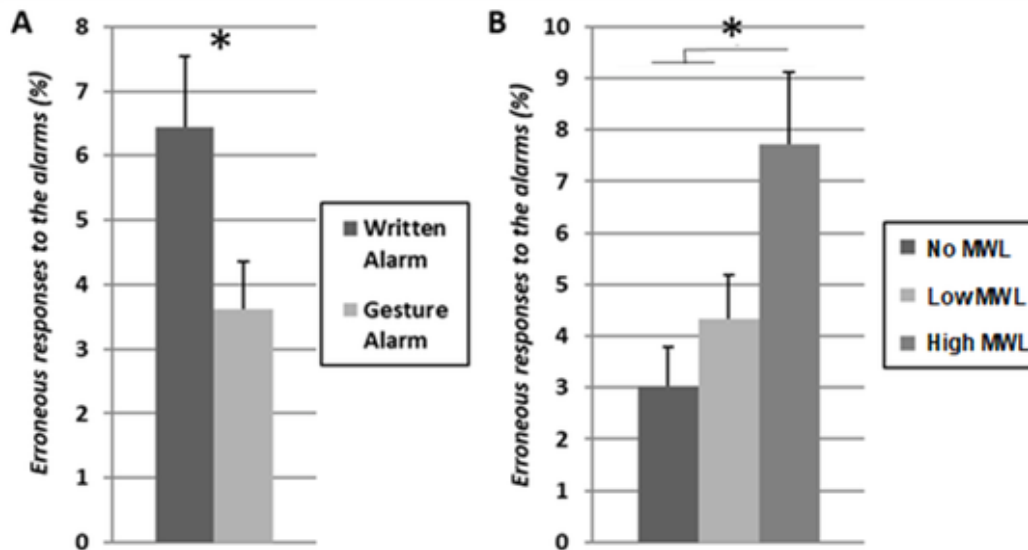


Fig. 3. Illustration of the erroneous responses to the alarm (%) as a function of (A) the type of alarms with written alarms (dark grey) and gesture alarms (light grey); and (B) the level of MWL with no-MWL (dark grey), low MWL (light grey) and high MWL (mid grey). * $p < .05$.

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 ($p = .056$).

Reaction times to the alarms. Fig. 4 illustrates participants' reaction times in the *alarm response task*. The analysis revealed a significant main effect of alarm [$F(1, 23) = 34.56$, $p < .001$, $\eta_p^2 = .60$; see Fig. 4A.] 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 [$F(2, 46) = 46.75$, $p < .001$, $\eta_p^2 = .67$; Fig. 4B.], 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 [$F(2, 46) = 1.06$, $p = .35$, $\eta_p^2 = .04$].

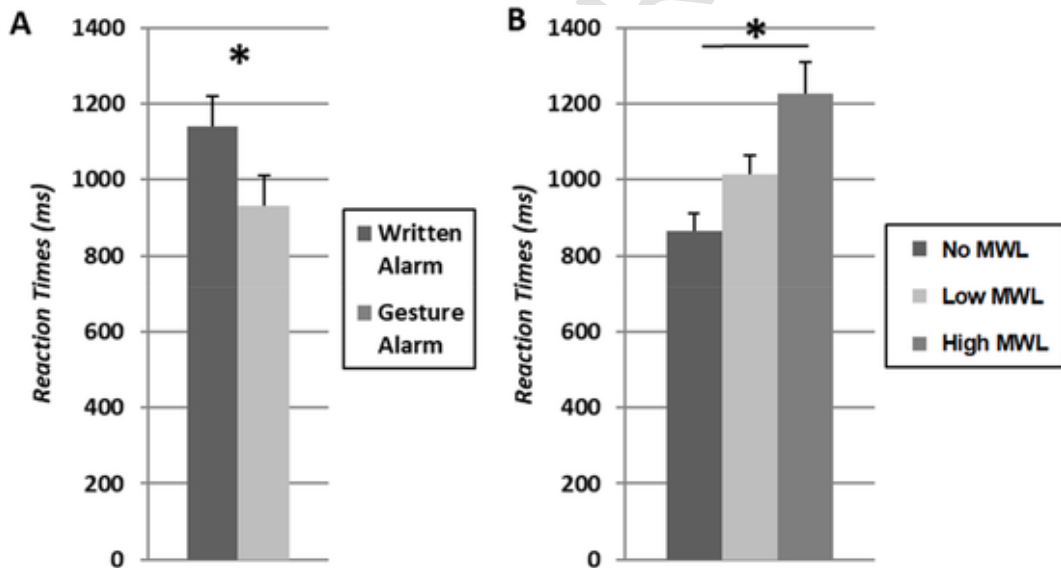


Fig. 4. Illustration of the reaction times to the alarm (ms) as a function of (A) the type of alarms with written alarms (dark grey) and gesture alarms (light grey); and (B) the level of MWL with no MWL (dark grey), low MWL (light grey) and high MWL (mid grey). *: $p < .05$.

3.2. Subjective results

3.2.1. NASA-TLX questionnaire

Fig. 5 illustrates participants' NASA-TLX ratings. The analysis revealed a significant main effect of MWL [$F(2, 46) = 109.89$, $p < .001$, $\eta_p^2 = 1.00$; see Fig. 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 ($p < .001$).

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

3.2.2. Difficulty to respond to the alarm

Fig. 6 illustrates participants' ratings of the difficulty to respond to the alarms. The statistical analysis revealed a significant main effect of MWL [no-MWL: $B(SE) = -1.625(0.244)$, $CI_{(95\%)} = (-2.103, -1.147)$, $Wald \chi^2(1) = 44.371$, $p < .001$; low-MWL: $B(SE) = -0.833(0.244)$, $CI_{(95\%)} = (-1.311, -0.355)$, $Wald \chi^2(1) = 11.67$, $p < .001$; with

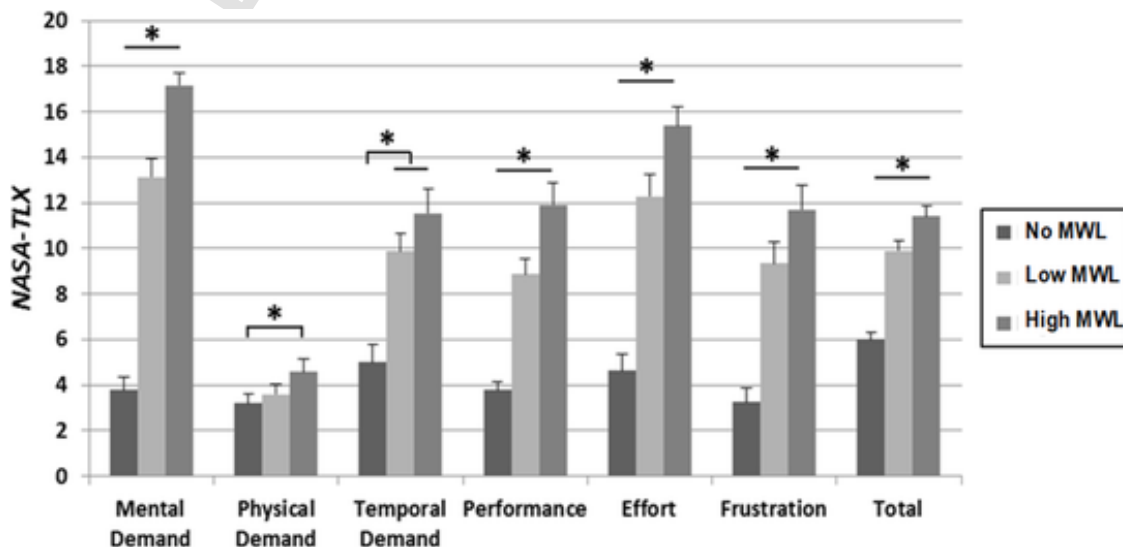


Fig. 5. Illustration of the NASA-TLX ratings observed in the no MWL run (dark grey), the low MWL run (light grey) and the high MWL run (middle grey). *: $p < .05$.

Table 1

Summary of the means, standard deviations and statistical analysis performed on the ratings of NASA-TLX subscales, in the no-MWL, the low-MWL and the high-MWL runs.

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	X ² F (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*

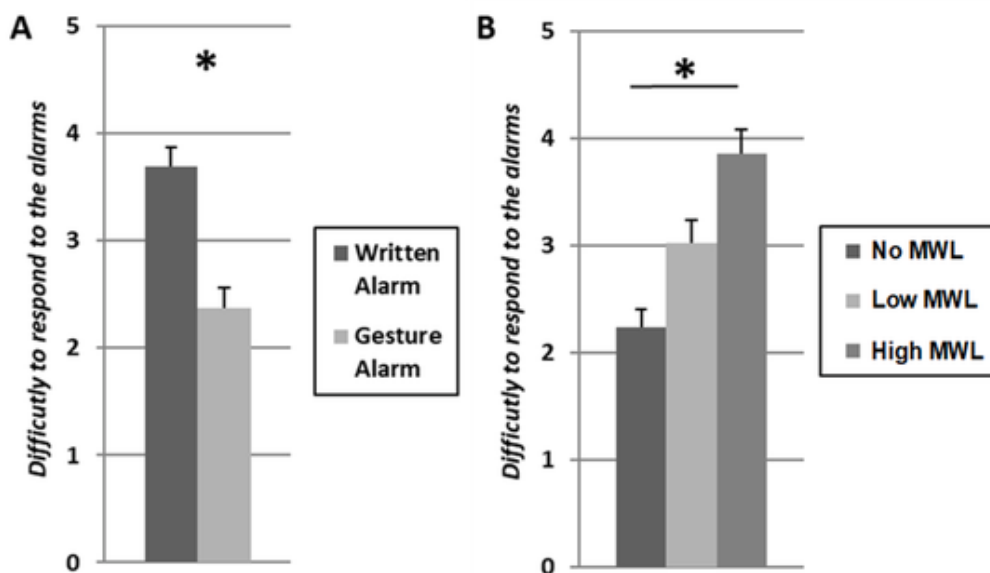


Fig. 6. Illustration of the difficulty to respond to the alarms as a function of (A) the type of alarms with written alarms (dark grey) and gesture alarms (light grey); and (B) the level of MWL with no MWL (dark grey), low MWL (light grey) and high MWL (mid grey). *: $p < .001$.

high-MWL as dummy; Fig. 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 [$B (SE) = 1.325 (0.157)$, $CI (95\%) = (1.016, 1.634)$, $Wald \chi^2 (1) = 70.735$, $p < .001$; Fig. 6B] with gesture alarms ($M = 2.37$, $SD = 1.41$) predicting for a lower difficulty to respond than written alarms ($M = 3.70$, $SD = 1.34$).

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, the low-MWL and the high-MWL runs in the 0–1441 ms time window after the onset of the alarms.

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	η^2_p	No-MWL vs. Low-MWL	No-MWL vs High-MWL	Low-MWL vs. High-MWL	
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

3.3. Brain electrophysiological results

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 Fig. 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 $\eta^2_p > 0.34$; see Fig. 7C & D). Post-

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 [12–15] Hz 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).

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 [12–15] Hz	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 [12–15] Hz	/	/	/	/	/

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.

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).

4. Discussion

4.1. 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* and the *alarm response task* 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 participants' cognitive resources were not sufficient to enable the completion of the *alarm response task* in the high-MWL run with the level of performance observed 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 (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, workload and/or 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 the efficiency of this *mental arithmetic task* to trigger MWL in participants.

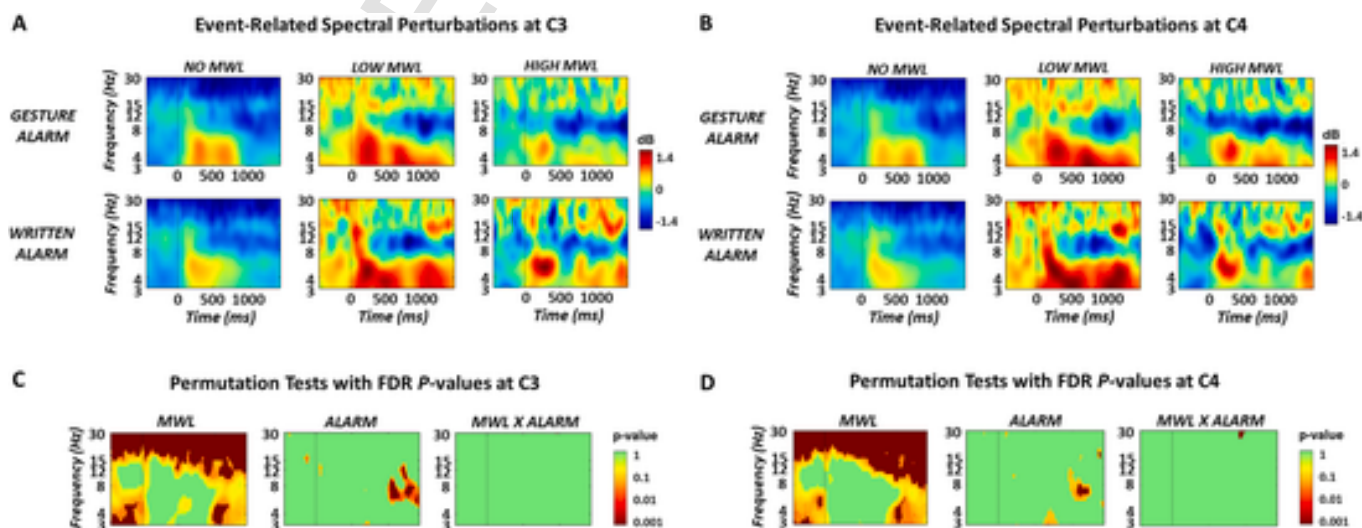


Fig. 7. Event Related Spectral Perturbations (ERSPs) showing increases (positive values – red) and decreases (negative values – blue) in dB-corrected power in the three MWL levels (no MWL – left, low MWL – middle, and high MWL – right) for the two types of alarms (Gesture – top and written – bottom) 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). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

4.2. Gesture alarms versus written alarms

Regardless of the level of MWL, participants were faster, more accurate and found it easier to respond to gesture alarms than to written alarms (Fabre et al., 2021), suggesting a greater efficiency of gesture alarms to trigger fast and accurate responses, even in high MWL situations.

In line with our predictions and previous results (Fabre et al., 2021), greater μ and low beta power decreases were found in response to gesture alarms than to written alarms. μ 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 μ 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), suggesting that the greater decrease in power observed in response to gesture alarms might 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 responses to gesture and written alarms (i.e., no triggers were sent when participants performed the action), we cannot conclude with certainty on this point.

4.3. Applications, limitations and future works

The results of the present study suggest that gesture alarms could be a suitable solution to improve both the visual modality of emergency alarms and operators' performances, even under high levels of MWL (Bliss and Gilson, 1998; Endsley and Jones, 2004).

In the context of aviation, gesture alarms could be used to improve the GPWS alarm, and more importantly the Stall alarm (Whittemore and Woods, 2021). Even though the maneuver to be performed in reaction to the Stall alarm is both stressful and counter-intuitive for pilots (i.e., pulling while the aircraft is stalling; BEA, 2012; IATA, 2018a), it only consists in a simple awareness warning (Ewbank et al., 2016). Replacing the visual modality of this alarm by a gesture alarm might significantly enhance its design and improve pilots' response performance to it.

The implementation of gesture alarm systems might also help improve operators' reactions in many other fields where fast and accurate responses are expected (e.g., transportation in general, nuclear powerplants, medical field; Edworthy, 2013). However, further research is needed before gesture alarms could be implemented in real operational environments. First, we focused exclusively on the visual component of the alarms, and while adding an auditory component might facilitate even more the operators' responses (Alirezaee et al., 2017; Hughes et al., 1994; Liu, 2001), it remains to be demonstrated. Second, gesture alarms should be tested on real-life operators in their usual work environment before they could be safely implemented.

5. 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 level of MWL. Brain electrophysiological results suggest that this greater efficiency might be due to the fact that gesture alarms facilitate the execution (and also possibly the preparation) of the associated action. While further research is necessary to confirm these results 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.

Uncited references

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Abrams, R.A., Christ, S.E., 2003. Motion onset captures attention. *Psychol. Sci.* 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. <https://doi.org/10.21785/icad2017.066>. 175–181.
- 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. *Behav. Brain Res.* 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. *Psychon. Bull. Rev.* 22 (5), 1410–1416. <https://doi.org/10.3758/s13423-015-0804-9>.
- Birkett, M.A., 2011. The Trier Social Stress Test protocol for inducing psychological stress. *JoVE* (56), e3238. <https://doi.org/10.3791/3238>.
- Bliss, J.P., 2003. Investigation of alarm-related accidents and incidents in aviation. *Int. J. Aviat. Psychol.* 13 (3), 249–268. https://doi.org/10.1207/S15327108IJAP13_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., et al., 2016. Independent causal contributions of alpha- and beta-band oscillations during movement selection. *J. Neurosci.* 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). Brookhaven National Lab, Upton, NY.
- 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. *Learn. Instruct.* 36, 27–37. <https://doi.org/10.1016/j.learninstruc.2014.11.003>.
- 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 Res.* 1793, 148035. <https://doi.org/10.1016/j.brainres.2022.148035>.
- Causse, M., Imbert, J.P., Giraudet, L., Jouffrais, C., Tremblay, S., 2016a. The role of cognitive and perceptual loads in inattention deafness. *Front. Hum. Neurosci.* 10, 344. <https://doi.org/10.3389/fnhum.2016.00344>.
- Causse, M., Peysakhovich, V., Fabre, E.F., 2016b. High working memory load impairs language processing during a simulated piloting task: an ERP and pupillometry study. *Front. Hum. Neurosci.* 10, 240. <https://doi.org/10.3389/fnhum.2016.00240>.
- Cooper, J., 1995. *Controlled Flight into Terrain*, pp. 305–831. Aerospace (UK).
- 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. Inattention deafness to auditory alarms: inter-individual differences, electrophysiological signature and single trial classification. *Behav. Brain Res.* 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. *J. Neurosci. 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. *Int. J. Sport Exerc. Psychol.* 15 (4), 423–437. <https://doi.org/10.1080/1612197X.2015.1121508>.
- Edworthy, J., 2013. Medical audible alarms: a review. *J. Am. Med. Inf. Assoc.* 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%2020Decision%25202007-017-R.pdf>.
- Endsley, M.R., Jones, D.G., 2004. *Designing for Situation Awareness: an Approach to User-*

- Centered Design, second ed. CRC Press, Taylor & Francis Group, Boca Raton, FL.
- Ewbank, C.E., Mumaw, R.J., Snow, M.P., 2016. Development of the enhanced bank angle warning. In: 2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC). IEEE, pp. 1–9. <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. *Appl. Psychophysiol. 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. *J. Neurophysiol.* 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. *Psychol. Bull.* 142 (3), 291–313. <https://doi.org/10.1037/bul0000031>.
- Guillaume, A., 2011. Intelligent auditory alarms. In: Hermann, T., Hunt, A., Neuhoff, J.G. (Eds.), *The Sonification Handbook*. Logos Publishing House, Berlin, Germany, pp. 493–508. <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 inattention deafness? *PLoS One* 10, e0118556. <https://doi.org/10.1371/journal.pone.0118556>.
- Hart, S.G., Staveland, L.E., 1988. Development of NASA-TLX (task load index): results of empirical and theoretical research. *Adv. Psychol.* 52, 139–183. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9). North-Holland.
- Hart, S.G., 2006. NASA-task load index (NASA-TLX); 20 years later. No. 9. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50. Sage Publications, Los Angeles, CA, pp. 904–908. Sage CA.
- 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, 2018a. Guidance Material and Best Practices for the Implementation of Upset Prevention and Recovery Training, second ed. International Air Transport Association, Montreal-Geneva. https://www.iata.org/contentassets/b6eb2adc248c484192101edd1ed36015/gmpb_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., et al., 2017. Empirical evidence for the relationship between cognitive workload and attentional reserve. *Int. J. Psychophysiol.* 121, 46–55. <https://doi.org/10.1016/j.jpsycho.2017.09.007>.
- Jaquess, K.J., Lo, L.C., Oh, H., Lu, C., Ginsberg, A., Tan, Y.Y., et al., 2018. Changes in mental workload and motor performance throughout multiple practice sessions under various levels of task difficulty. *Neuroscience* 393, 305–318. <https://doi.org/10.1016/j.neuroscience.2018.09.019>.
- Jeannerod, M., 1994. The representing brain: neural correlates of motor intention and imagery. *Behav. Brain Sci.* 17 (2), 187–202. <https://doi.org/10.1017/S0140525X0003402>.
- Jenson, D., Thornton, D., Harkrider, A.W., Saltukdaroglu, T., 2019. Influences of cognitive load on sensorimotor contributions to working memory: an EEG investigation of mu rhythm activity during speech discrimination. *Neurobiol. Learn. Mem.* 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. *Adv. Neural Inf. Process. Syst.* 894–900.
- Kelly, D., Efthymiou, M., 2019. An analysis of human factors in fifty controlled flight into terrain aviation accidents from 2007 to 2017. *J. Saf. Res.* 69, 155–165. <https://doi.org/10.1016/j.jsr.2019.03.009>.
- Kumar, N., Kumar, J., 2016. Measurement of cognitive load in HCI systems using EEG power spectrum: an experimental study. *Procedia Comput. Sci.* 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. *Phil. Trans. Biol. Sci.* 369 (1641), 20130205. <https://doi.org/10.1098/rstb.2013.0205>.
- 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., 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. 140–145.
- Martin, W.L., Murray, P.S., Bates, P.R., Lee, P.S., 2015. Fear-potentiated startle: a review from an aviation perspective. *Int. J. Aviat. Psychol.* 25 (2), 97–107. <https://doi.org/10.1080/10508414.2015.1128293>.
- Meltzoff, A.N., 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. *Med. Biol. Eng. Comput.* 28 (5), 389–397. <https://doi.org/10.1007/BF02441666>.
- Meyer, M., Endeldijk, 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>.
- 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. *J. Surg. Educ.* 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 Cognit.* 82 (3), 236–242. <https://doi.org/10.1016/j.bandc.2013.04.010>.
- Moroze, M.L., 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>.
- Nicolai, 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, 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>.
- 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. *Appl. Ergon.* 25 (6), 355–365. [https://doi.org/10.1016/0003-6870\(94\)90054-X](https://doi.org/10.1016/0003-6870(94)90054-X).
- Oliver, N., Calvard, T., Potočník, K., 2019. Safe limits, mindful organizing and loss of control in commercial aviation. *Saf. Sci.* 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. *Int. J. Psychophysiol.* 26 (1–3), 121–135. [https://doi.org/10.1016/S0167-8760\(97\)00760-5](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. *Clin. Neurophysiol.* 110 (11), 1842–1857. [https://doi.org/10.1016/S1388-2457\(99\)00141-8](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 Res. Rev.* 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. *Psychol. Bull.* 114 (3), 510. <https://doi.org/10.1037/0033-2909.114.3.510>.
- Raveh, D., Lavie, N., 2015. Load-induced inattention deafness. *Atten. Percept. Psychophys.* 77 (2). <https://doi.org/10.3758/s13414-014-0776-2>. 483–49.
- Reiser, J.E., Wascher, E., Arnau, S., 2019. Recording mobile EEG in an outdoor environment reveals cognitive-motor interference dependent on movement complexity. *Sci. Rep.* 9 (1), 13086. <https://doi.org/10.1038/s41598-019-49503-4>.
- 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. *Biol. Psychol.* 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. *Cognit. Brain Res.* 3 (2), 131–141. [https://doi.org/10.1016/0926-6410\(95\)00038-0](https://doi.org/10.1016/0926-6410(95)00038-0).
- Schmidt, N.B., Richey, J.A., Zvolensky, M.J., Maner, J.K., 2008. Exploring human freeze responses to a threat stressor. *J. Behav. Ther. Exp. Psychiatr.* 39 (3), 292–304. <https://doi.org/10.1016/j.jbtep.2007.08.002>.
- Scout, W.B., 1996. New research identifies causes of CFIT. *Aviat Week Space Technol.* 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, p. 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. *Biol. Psychol.* 134, 39–51. <https://doi.org/10.1016/j.biopsycho.2018.01.009>.
- Sherry, L., Mauro, R., 2014. Controlled Flight into Stall (CFIS): functional complexity failures and automation surprises. In: 2014 Integrated Communications, Navigation and Surveillance Conference (ICNS) Conference Proceedings. IEEE, p. D1. <https://doi.org/10.1109/ICNSurv.2014.6819980>.
- 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. *Appl. Ergon.* 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. *Adv. Cognit. Psychol.* 5, 84. <https://doi.org/10.2478/v10053-008-0064-6>.
- 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. In: *Human Factors and Power Plants and HPRCT 13th Annual Meeting, 2007 IEEE 8th. IEEE, Monterey, CA*, pp. 92–97.
- Wickens, C.D., 2008. Multiple resources and mental workload. *Hum. Factors* 50 (3), 449–455. <https://doi.org/10.1518/001872008X28839>.
- Wickens, C.D., 2020. Processing resources and attention. In: *Multiple-task Performance. CRC Press*, pp. 3–34.
- White, A., O'Hare, D., 2022. In plane sight: inattention blindness affects visual detection of external targets in simulated flight. *Appl. Ergon.* 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 the activation of a Stall alarm. *Journal of Aviation/Aerospace Education & Research* 30 (1). <https://doi.org/10.15394/jaer.2021.1881>.
- Wu, X., Li, Z., 2018. A review of alarm system design for advanced control rooms of nuclear power plants. *Int. J. Hum. Comput. Interact.* 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 inattention deafness during landing decision-making. *Appl. Ergon.* 99, 103629. <https://doi.org/10.1016/j.apergo.2021.103629>.

CORRECTED PROOF