

RESEARCH



Task-independent auditory probes reveal changes in mental workload during simulated quadrotor UAV training

Shaodi Wang[†], Heng Gu[†], Qunli Yao, Chao Yang, Xiaoli Li^{*} and Gaoxiang Ouyang^{*} 

Abstract

Objective: The event-related potential (ERP) methods based on laboratory control scenes have been widely used to measure the level of mental workload during operational tasks. In this study, both task difficulty and test time were considered. Auditory probes (ignored task-irrelevant background sounds) were used to explore the changes in mental workload of unmanned aerial vehicle (UAV) operators during task execution and their ERP representations.

Approach: 51 students participated in a 10-day training and test of simulated quadrotor UAV. During the experiment, background sound was played to induce ERP according to the requirements of oddball paradigm, and the relationship between mental workload and the amplitudes of N200 and P300 in ERP was explored.

Main results: Our study shows that the mental workload during operational task training is multi-dimensional, and its changes are affected by bottom-up perception and top-down cognition. The N200 component of the ERP evoked by the auditory probe corresponds to the bottom-up perceptual part; while the P300 component corresponds to the top-down cognitive part, which is positively correlated with the improvement of skill level.

Significance: This paper describes the relationship between ERP induced by auditory probes and mental workload from the perspective of multi-resource theory and human information processing. This suggests that the auditory probe can be used to reveal the mental workload during the training of operational tasks, which not only provides a possible reference for measuring the mental workload, but also provides a possibility for identifying the development of the operator's skill level and evaluating the training effect.

Keywords: EEG, ERP, Mental workload, Mental schema, Human information processing

Introduction

Unmanned systems have been used widely in now world. Now the unmanned aerial vehicle (UAV) systems are complex to operate, which usually require the operator to undergo specialized training and pass several tests before taking the task. Understanding the variation of the mental workload of operators is very important for evaluating the level of mental workload of operators, improving the training effect and further building a safer and

healthier training system [1–3]. High mental workload levels significantly affect the performance of UAV operators. Accurate assessment of mental workload can help prevent human error and allow intervention before performance declines due to mental overload or underload. Recognition of mental workload is one of the core problems in neuroergonomics [4, 5].

Mental workload is a multi-dimensional structure [6]. Many factors, such as the difficulty of the task, the environment in which it is performed, and the skill level and cognitive state of the operator, might have an impact on it. At present, the interpretation of mental workload by multi-dimensional resource theory [7] is accepted by most researchers. The theory holds that cognitive resources are limited and negatively correlated with

[†]Shaodi Wang, Heng Gu have contributed equally to this work.

*Correspondence: xiaoli@bnu.edu.cn; ouyang@bnu.edu.cn
State Key Laboratory of Cognitive Neuroscience and Learning, Beijing
Normal University, Beijing 100875, People's Republic of China

the level of mental workload. When engaged in mentally demanding tasks, there will be fewer resources left, reducing the brain's ability to recognize visual or auditory events. This means that changes in mental workload levels can be identified by measuring either the amount of resources occupied or remaining.

Neuroimaging can be used to directly measure the state of the central nervous system [8–10], so as to achieve a relatively objective measurement of the mental workload of the operator. The methods to estimate mental workload by extracting EEG features are promising [11]. Among them, the most widely studied features are the power spectrum density (PSD) and event-related potentials (ERP) of persistent EEG. Previous studies have shown that the use of ERP has better classification performance and cross-task consistency [12–14]. Such studies are often associated with Oddball experimental paradigm [15], and most of them focus on P300 and N200 components induced by this paradigm. Current studies have shown that the N200 and P300 components of ERP are closely related to the cognitive processes of perception and selective attention [16–19]. Studies on measuring the mental workload of operators based on ERP can be divided into two categories according to whether the active attention of the operators is required. One is the dual-task paradigm [12, 14, 20, 21]. This kind of paradigm requires the subjects to devote part of resources to the secondary task while completing the operational task (main task), and to measure the cognitive state by measuring the ERP induced by the secondary task (for example, counting or pressing the auditory stimulus during the execution of the main task). The other is the single-task paradigm [18, 22–24], which induces ERP using task-independent probes that do not require a response actively. This is to minimize the competition for cognitive resources and to avoid the influence of probes on operational tasks. Almost all of these studies use auditory probes to induce ERP because visual channels are needed for almost all operational tasks. The ERP induced by the dual-task paradigm are generally higher, but it requires the operator to actively invest extra cognitive resources into the secondary task, which leads to distraction. For systems with high safety requirements such as UAV operations, the requirement that the operator needs to respond to a secondary task during the execution of a task is not applicable because it has been found to disrupt the execution of the primary task [12]. Although the amplitude of ERP induced by auditory probe is relatively small, it has the practical advantage of solving problems where the dual-task paradigm may affect the main task [17, 25].

Some studies have used the single-task paradigm with task-independent auditory stimuli to assess cognitive

workload and report a reduction in ERP amplitude as cognitive workload increases [18, 19, 26, 27]. Most of these studies are based on the reciprocity hypothesis [28, 29], which describes the inverse relationship between task demands and the level of reserved attentional ability. Other studies have designed classifiers based on the amplitude characteristics of ERP to realize the classification of different levels of psychological workload [23, 30, 31].

Most of the previous studies using auditory probes were designed to induce different degrees of mental workload by setting various task difficulties in one experiment. Through such experimental design, the ERP changes of the operator when facing different difficult tasks were studied. However, training in reality often requires several days and multiple times. At present, the experimental results of using task-independent auditory probe to study mental load at both test time and difficulty levels have not been reported.

In order to explore the change rule of mental workload and its representation in ERP under the training scenario which is closer to the reality, the following three aspects were carried out in this study. First, we designed UAV operation experiments that better fit the real training environment and included both task difficulty and test time. Secondly, task-independent auditory probe was used to induce ERP in the experiment, and the changes of mental workload during training and its representation at ERP were analyzed. Thirdly, a theoretical framework is proposed to explain the corresponding relationship and change rule between the perceptual and cognitive links and ERP components in the process of human information processing, which provides a theoretical explanation for the changes of mental workload in the training process of operational tasks (such as UAV) in the real scene. The development of theories may facilitate the development of more sensitive methods for measuring mental workload [32].

The rest of this paper is as follows: In “[Materials and methods](#)” Section, we explain our experimental procedures and data processing methods. In “[Results](#)” Section reports the results of changes in subjects' ERP during the experiment, which are verified by repeated-measures ANOVA. In “[Discussion](#)” Section, a theoretical framework is proposed to explain the possible reasons for the changes of P300 and N200 components in ERP affected by task difficulty and test time during training.

Materials and methods

Participants

The data set used in this study has been reported in our previously published study [33]. 51 right-handed male volunteers (23.25 ± 2.12 years old) were recruited for the

experiment, all of whom had no experience in operating UAV. This study was approved by the Ethics Committee of Beijing Normal University, and all subjects provided informed consent.

Experiment protocol

All subjects participated in a total of 10 days of UAV simulated flight training and testing tasks. The experimental process has previously been described in detail [33]. In brief, each subject was asked to fulfil six experimental tasks over 10 days (days 1–4, 7, and 10), which were made up of four test sessions (Days 1, 4, 7, and 10) and two training sessions (Days 2 and 3). The purpose of the training is to improve the skill levels of the operator. Only the task with difficulty level 1 is used for training, so that the skills of the operator can reach the basic qualified level, that is, they can complete all the four difficult tasks, and the score of the operation task is above 60 points on the premise that the full score is 100 points. No test or training session lasted more than 1 h. Figure 1 shows an overview of the experimental environment, experimental flow arrangement, and task scenarios. Subjects were asked to use a handle to fly the drone on an ∞ trajectory calibrated by a fixed stake. The tasks are set to four levels of difficulty (LEVEL1, no wind; LEVEL2: directional breeze; LEVEL3: strong wind in fixed direction; LEVEL4: strong wind in random direction). In test sessions, subjects need to complete the flight task, and get the flight score after each successful completion of the flight; otherwise, the score is 0. The test session consisted of four difficulty levels and 16 trials (4 levels \times 4 trials/level = 16 trials), sorted in ascending or descending order. After each trials, the subjects were required to complete the Cooper-Harper Scale [34], the score of which reflects the self-evaluation of the operator's load on the flight operation task. In two training sessions on days 2 and 3, the subjects only had to train for 1 h using task with difficulty

level 1. Before and after each experiment, we asked the subjects to fill in the Karolinska Sleepiness Scale [35] to evaluate the level of mental fatigue caused by the experiment (1–10, the larger the number is, the more serious the mental fatigue is), and fill in the self-evaluation scale of the impact of background sound stimulus on the operational task (1–10, The higher the number, the more affected the task is, with 1 being no affected and 10 being very affected). During each test, the subjects' EEG was recorded, along with their behavioral data and a subjective rating scale.

The 1200Hz and 2000Hz tones of sound (100ms duration, 10ms rise/fall time, 75–80 dB SPL, 1000 \pm 100ms intervals) were presented about 80 cm behind the subject's head through Lenovo computer speakers with the probabilities of 80% and 20%, respectively. The amplitudes of N200 and P300 were positively modulated by the stimulus intensity [36]. Exposure for more than 8 h in the scene with a volume above 85 dB will cause ear damage [37]. In order to avoid damage to subjects' ears and obtain high quality signal, we used a volume lower than 85 dB in the experiment. Figure 2 shows the procedure and design principle of the auditory probe experiment.

The operation task is only visually based, and the stimulus sounds are presented in the background. We did not give the subjects any instructions about the sound. At the end of each experiment, subjects filled out the rating scale for the effect of the sound on the operational task.

Data analysis

A polygraph recorded data at a sampling rate of 1 kHz. According to the International 10–20 system [38], 24 electrodes were placed: FP1, F2, F7, F8, F3, F4, FZ, C3, C4, FC5, FC6, CP5, CP6, T3, T4, TP7, TP8, P3, P4, P7, P8, PZ, O1 and O2. All electrodes are grounded to the Fpz channel by referring to the Cz channel, and their impedance is kept below 10k Ω . The original collected

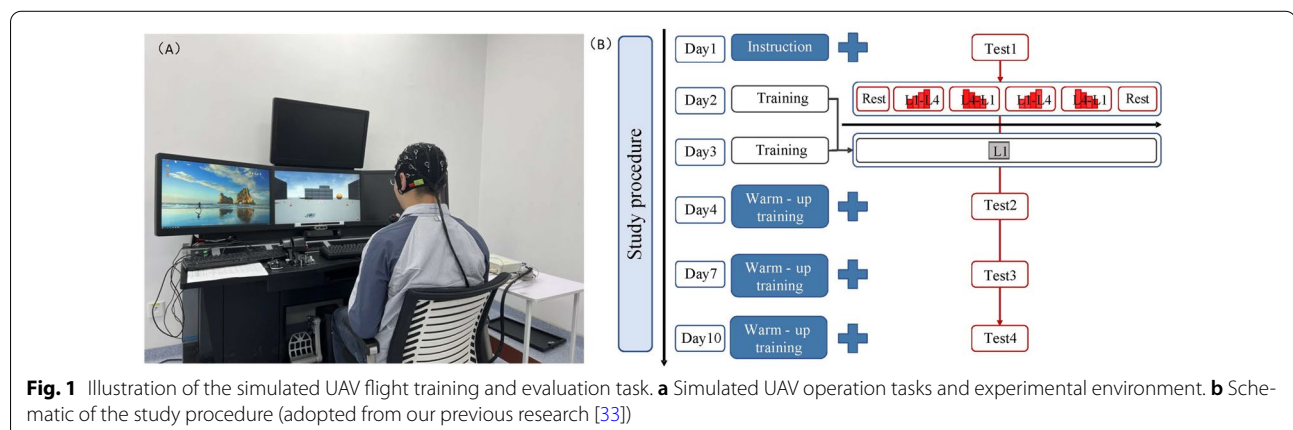
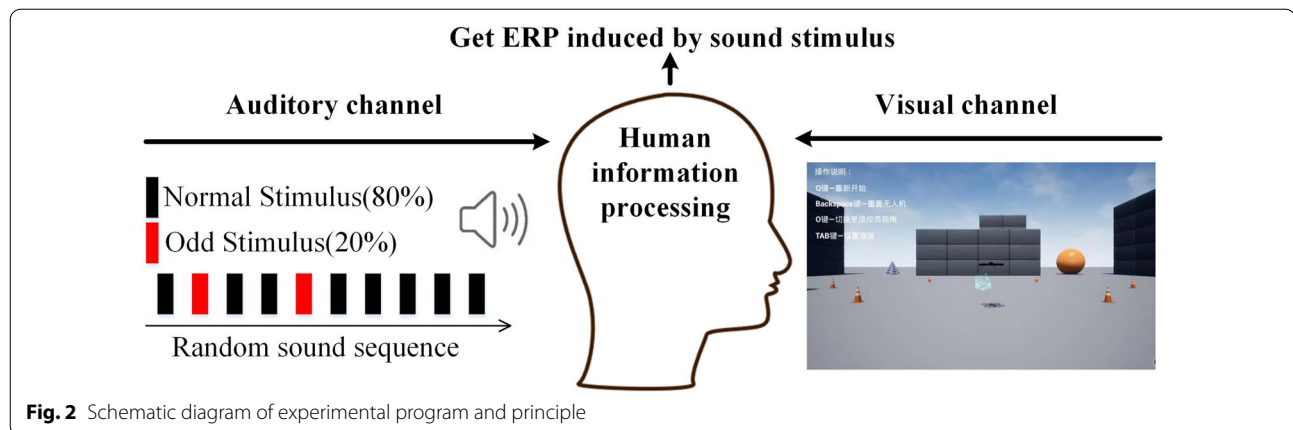


Fig. 1 Illustration of the simulated UAV flight training and evaluation task. **a** Simulated UAV operation tasks and experimental environment. **b** Schematic of the study procedure (adopted from our previous research [33])



EEG signals are first converted into an average reference form. Then, a finite impulse response filter is used to digitally filter the EEG data in the range of 0.5 to 45 Hz. Afterwards, the data are sampled to 256Hz. Finally, artifacts related to blinking, horizontal eye movements, and ECG activity were removed from EEG data using independent component analysis [39].

The ERP data were obtained using a time window from 200ms before the onset of auditory stimulation to 800ms after the stimulation, and then the baseline was corrected with reference to the pre-stimulus interval. The amplitude of each ERP component was calculated using the mean-amplitude method [40] which recommended the use of a delimited time window centred on the peak of the component in the large mean waveform, with good results for large components such as N200 and P300. The time windows used were as follows: N200=180–250ms; P300=280–350ms. The data were analyzed with MATLAB R2020b software using scripts based on the EEGLAB 2021 toolbox [41]. RStudio (Windows 1.4) is used to calculate statistics. The behavioral data, scale data and amplitude data of subjects at four task difficulty levels in four tests were submitted to a repeated-measures ANOVA, with test time (level 4) and task difficulty (level 4) as factors. The Geisser-Greenhouse correction is used when a sphericity violation occurs during the integrated test. A paired T-test with a 0.05 significance criterion was used to determine the locus of the main effect. The Bonferroni method was used for multiple comparison correction. The estimated effect size was given using the η_g^2 statistic.

Results

In this study, data from four tests were selected for analysis, namely, the first day of the experiment (TEST1), the fourth day after two training sessions to improve the level (TEST2), and the seventh day (TEST3) and the tenth day

(TEST4). Each test consists of tasks with four levels of difficulty (LEVEL1 to LEVEL4).

Measurement of behavioral performance

In TEST1, only 24 subjects completed tasks of four levels (LEVEL1 to LEVEL4). The remaining 27 subjects failed to complete all tests because they were unable to achieve the required scores. Therefore, in the statistical test of behavioral performance, only 24 subjects who completed all tasks were used for analysis. Two main effects were significant: task difficulty ($F_{(3,69)} = 66.399$, $p < 0.0001$, $\eta_g^2 = 0.743$); test time ($F_{(2,117,48,693)} = 44.293$, $p < 0.0001$, $\eta_g^2 = 0.658$). The interaction between test time and task difficulty was not significant ($F_{(9,207)} = 43.918$, $p = 0.694$, $\eta_g^2 = 0.03$). As shown in Fig. 3, the test scores of subjects improve significantly with the test time, which also indicates the improvement of the skill level of the operational task. With the increase of difficulty level, the test scores of subjects decrease significantly, indicating that the experiment successfully set different operational task difficulties. However, the distinction between LEVEL3 and LEVEL4 is not obvious. After inquiry, some subjects think that the difficulty of LEVEL3 is higher than that of LEVEL4.

Self-reported ratings

The Cooper-Harper Scale scores of all subjects were included in the analysis, and the interaction between the test time and the task difficulty was not significant, ($F_{(4,370,218,511)} = 1.559$, $p = 0.181$, $\eta_g^2 = 0.030$). Both the test time and the task difficulty had significant effects: task difficulty ($F_{(1,383,69,141)} = 175.021$, $p < 0.0001$, $\eta_g^2 = 0.778$); test time ($F_{(1,488,74,386)} = 50.328$, $p < 0.0001$, $\eta_g^2 = 0.502$). The results showed that with the progress of the test, the mental workload level of the subjects during the execution of the task significantly decreased, and the mental workload level of the subjects

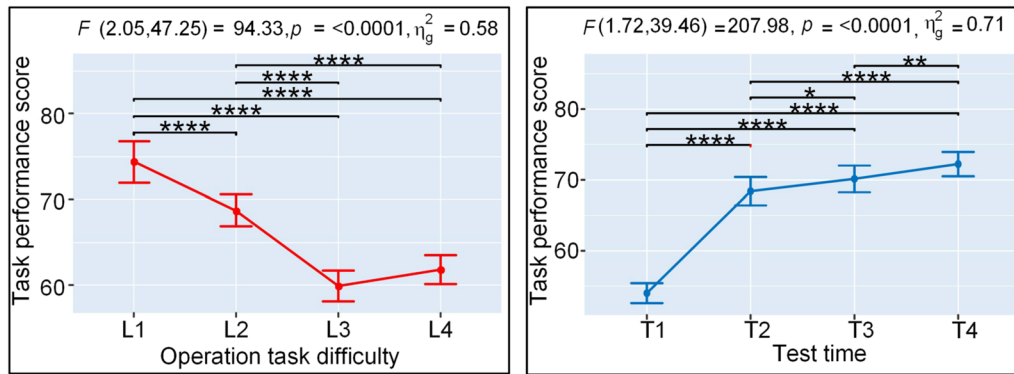


Fig. 3 Total score of UAV task. The results in the left panel indicate that task performance decreases significantly with increasing task difficulty. The results in the right subfigure show that after repeated training and testing, the performance of the subjects driving the UAV has significantly improved (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.0001$)

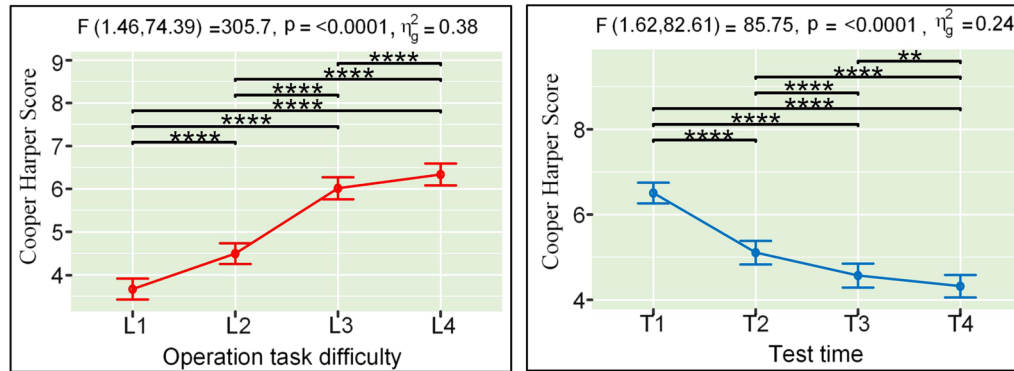


Fig. 4 Self-rated scores of the subjects measured by Cooper-Harper scale. The results on the left show that tasks designed successfully triggered different levels of mental workload. The results on the right show that after repeated training and testing, the subjects' perceived difficulty in the same task situation decreased (** $p < 0.01$, *** $p < 0.0001$)

significantly increased with the increase of the task difficulty (Fig. 4).

We also analyzed the results of the Karolinska Narcolepsy Scale [35] and found no significant differences in scores before and after each test, indicating that our experiment did not cause significant mental fatigue. The scale score of the effect of auditory stimulation on task performance recorded after each experiment was 1.55 ± 1.20 , indicating that the effect of auditory probes on the task was small.

EEG results

We used mean amplitude between two fixed latencies [15] to represent the amplitudes of P300 and N200 on the selected channels. Figure 5 shows the topographic map of the grand average of the N200 and P300 components from four tests \times four difficulty levels after max-abs normalization. As Fig. 5 shows, the amplitudes of N200 and

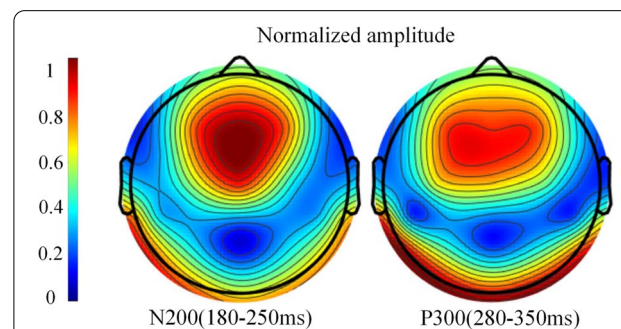


Fig. 5 Grand average topographic maps of the N200 and P300 components for four difficulty levels at four tests

P300 were larger in the parietal, occipital and temporal lobes. We used the data from five channels (Fz, O1, O2, P7 and P8) in the above brain regions for further analysis.

Figures 6 and 7 shows the statistical results, and we found the following results:

The amplitude of P300 on the selected channel increases with the test time

The main effect of the test time is significant, while the interaction between the test time and the task difficulty and the main effect of the task difficulty are not significant. The five figures in the first column of figure 6 show the relationship between the amplitude of P300 on the selected channel and the task difficulty. The Bonferroni post-test results of the test time showed that the amplitude of P300 in the first test was significantly lower than that in the subsequent three tests after training.

The amplitude of N200 on the selected channel decreases as the task difficulty increases

The main effect of the task difficulty is significant, while the interaction between the test time and the task difficulty and the main effect of the test time are not significant. The five figures in the second column of Fig. 6 show the relationship between the amplitude of N200 on the selected channel and task difficulty. The Bonferroni post-test results of task difficulty showed that the amplitude of N200 obtained at task difficulty Level=2,3,4 tended to decrease compared to that obtained at task difficulty Level=1.

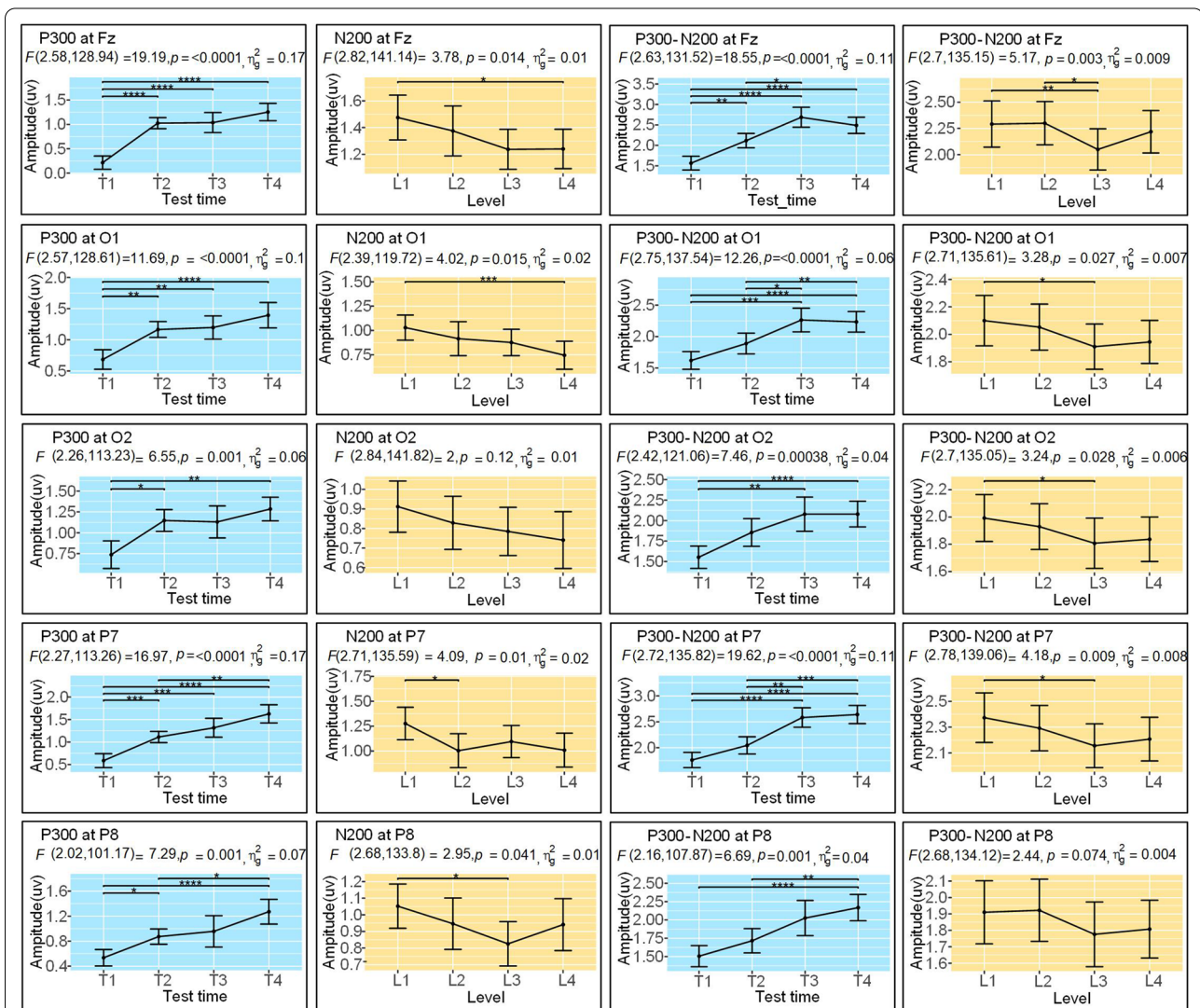


Fig. 6 Results of repeated-measures ANOVA calculation of the amplitude of selected ERP components versus test difficulty or test time. (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$). Because of the use of average reference, the amplitude of channel Fz is inverted for consistency

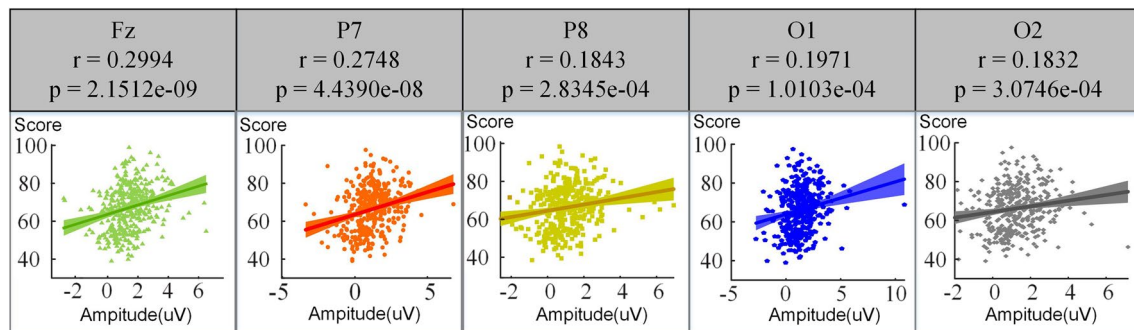


Fig. 7 The amplitude of P300 is correlated with the task performance score (Since the average reference was used in this study, the amplitude of channel Fz is inverted for consistency). The shadow indicates the 95% confidence interval

The calculation results of P300-N200 increase with the test time and decrease with the task difficulty level

The interaction between the test time and the task difficulty is not significant, while the main effect of the test time and the task difficulty are significant. The five figures in the third column of Fig. 6 show the relationship between the calculation results of P300-N200 and the test time. The Bonferroni post-test results of the test time show that, the P300-N200 obtained in the first test was significantly lower than that in the next three tests after training. The five figures in the fourth column show the relationship between P300-N200 and the difficulty of the task. The Bonferroni post-test results of task difficulty showed that the calculation results of P300-N200 decreased with the increase of task difficulty.

The amplitude of P300 and task performance scores were positively correlated

Figure 7 shows the correlation results between the amplitude of P300 and the performance of all four tests calculated by spearman method [42]. The amplitude of P300 on all channels selected are positively correlated with the task performance. The correlation was highest at the Fz channel, located on the frontal region. However, the correlation results between the amplitude of N200 and the performance of all four tests are not significant.

Discussion

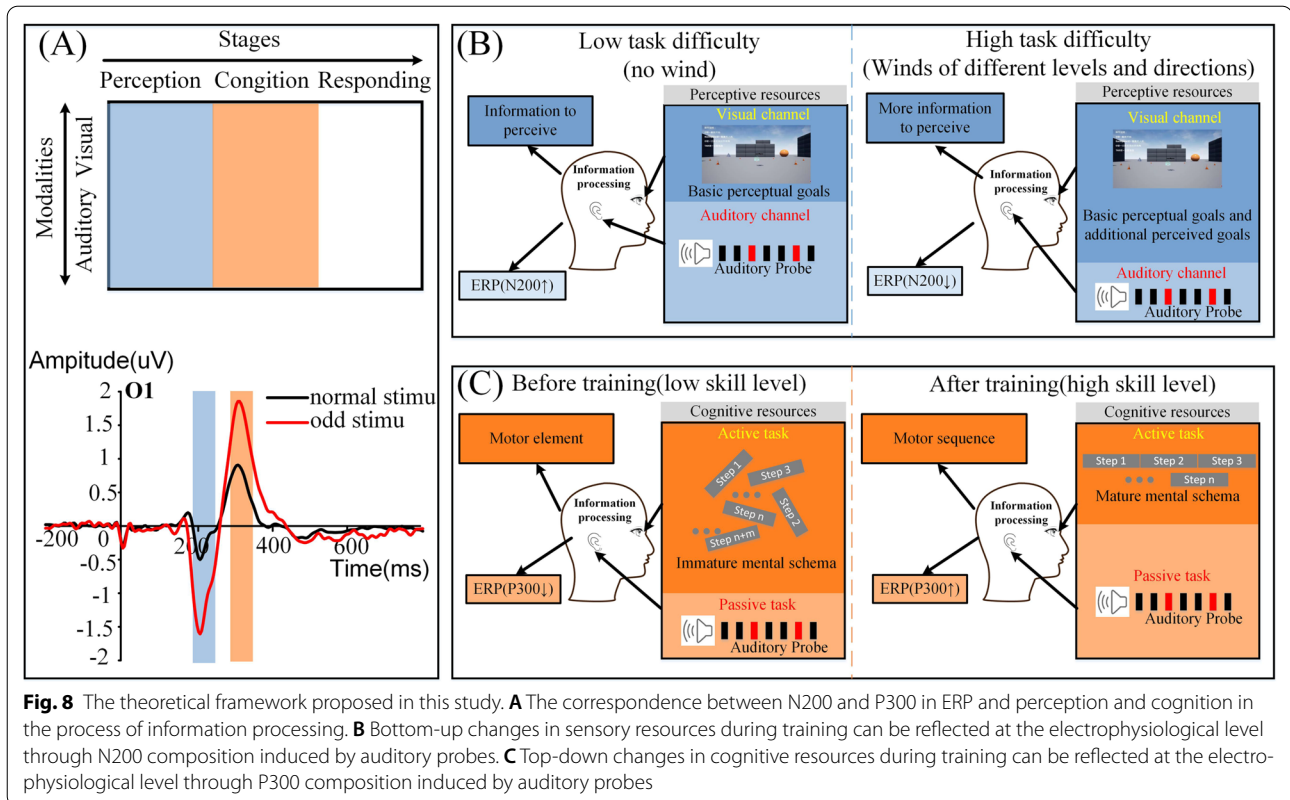
The objective of this study was to explore the changes in mental workload during operational task and its relationship with ERP induced by auditory probes. Different from the previous works in which a simple single task was set with several difficulty levels to induce different levels of mental workload, we designed a more complex experimental scheme that was closer to the realistic training situation. As far as we know, we are the first to introduce a

factor of skill level development in mental workload studies based on auditory probes. As shown in Fig. 8, based on the human information processing model [7] and the cognitive framework of sequential motor behavior [43], we proposed a framework for the corresponding relationship between ERP induced by auditory probes and mental workload changes during training. The framework demonstrates the feasibility of using ERP to assess the varying levels of mental workload due to the varying skill levels and operational task difficulties during UAV operation training.

In brief, human information processing for external stimuli is carried out according to the process of perception, cognition and response. This time sequence is also reflected in the ERP waveform induced by auditory probes. The N200 corresponds to the bottom-up perception part and the P300 corresponds to the top-down cognition part. Our experimental results verify the theoretical framework described above, which is discussed in detail below in combination with the previous statistical results.

- The N200 component induced by auditory probes is related to the external information perception process, which is a bottom-up process, and its amplitude decreases with the increase of task difficulty.

In all tests, our results confirmed the previous results on auditory probes [22], namely, the amplitude of N200 decreases with the increase of difficulty, further proving the reliability of this conclusion. Figure 8B shows how bottom-up perceived resources change with task difficulty during training. The perception process of task-independent auditory probes is expected to consume perceptual resources corresponding to attentional orientation responses. We believe that the skill level of the operator will not fluctuate drastically in the same



test session. Under this premise, when faced with a low-difficulty task in the same test session, the operator only needs to perceive the basic goals. As the difficulty of the task increases, the perceived goals required by the operator increase, such as wind direction, wind intensity and inclination of the UAV, etc., so the operation task requires more perceptual resources. The difficulty distinction in our experiment was mainly achieved in this way. This results in the formation of varying degrees of competition between the perceptual resources required for the auditory probe and those required for the operation task. There is a performance requirement for the UAV task, but for the background auditory probe, there is no requirement for the operator. Therefore, when such resource competition occurs, the perceptual resources available to task-independent auditory probes are compressed. At the electrophysiological level, the amplitude of N200 corresponding to the perceived resources decreases. At the same task difficulty level, the perceived goal required for each test was the same, so there was no significant difference in test time.

- The P300 component induced by auditory probes is related to the cognitive processing of external stimuli, which is a top-down process, and its amplitude will increase with the test time. This indicator can indi-

rectly reflect the development of skills in operational tasks.

Previous studies have shown that P300 has an obvious attentional effect, and its amplitude is small under non-attentional conditions, and its amplitude is related to the amount of cognitive resources actively invested [44, 45]. It is related to the active allocation of information processing resources. We suggest that the amplitude of P300 induced by auditory probes can indirectly reflect the development of skill level.

Figure 8C shows how top-down cognitive resources during training vary with test time. When the skill level is low, the operator has no concept of the operation task and does not form the mental diagram. A mental schema is a structure of knowledge about a subject or event that allows the brain to work more efficiently. It is based on experience and is used to guide current understanding or action [46]. Mental schema is dynamic and developed with experience, which also echoes the concept of plasticity in brain development. According to Bartlett's schema theory [47], procedural memory, as a part of long-term memory formed by practice, is responsible for controlling the limbs to perform tasks with certain strategies, also known as motor skills. When mature mental schema are not formed, the

operator executes the task in the form of a large number of motor elements with simple content and weak correlation [43]. Such a decentralized, step-by-step process leads to a higher occupancy of cognitive resources and a lower efficiency [48]. After a period of training, the operator becomes more familiar with the operation task, and the execution of the operation becomes more automatic, forming a mature mental schema. At this time, the action is based on the internal model. The action at this time is controlled through the motor sequences or action block [43, 49]. There is increasing continuity between the steps of an operational task and fewer cognitive resources are required to perform it.

The results in the first column of Fig. 6 show that the amplitude of P300 increases with the sequence of tests. This is because the skill level of the operator increases with the sequence of tests. Figure 8C shows that with the improvement of skill level, the operators form an increasingly mature mental schema [33, 50], and thus they can deal with difficult situations more easily, and multiple operation steps form automatic and continuous motion blocks. The cognitive resources occupied by the manipulative task are reduced so the remaining available cognitive resources are increased, and operators are more flexible in allocating resources. For these reasons, we believe that the amplitude of P300 can indirectly indicate the development of skill level during operational tasks. This conclusion is also supported by the results shown in Fig. 7, namely, the positive correlation between the amplitude of P300 and the score of the operation task.

- N200-P300 can be used as an effective supplementary indicator for more reliable mental workload detection methods.

The use of auditory probes for mental workload research is of wide significance and highly feasible because it requires only a small number of task-independent stimuli (in our experience, a relatively reliable ERP waveform can be obtained with more than 50 trials at the stimulus volume we adopted). Since it will not interfere with the main task and loudness of the used sound stimulus will not cause any damage to the operator, it can easily be compatible with the traditional load assessment methods. At present, there have been some studies using ERP components as indexes to design classifiers and realize the classification of different levels of mental workload [51, 52]. Therefore, it is a useful supplement to state assessment in the training of various cognitive skills (such as maneuvering aircraft and playing chess), and can be used in a long term in a variety of learning environments and human-machine interfaces.

N200 and P300 are often studied together, not only because they can be induced by the Oddball paradigm, but also because they are major components sensitive to mental workload [53]. The third and fourth columns in Fig. 6 prove this point. The N200 and P300 can reflect the mental workload state during the execution of operational tasks from the bottom-up perceptual level and the top-down cognitive level respectively. Therefore, the index P300-N200, which is formed by combining them, reflects the unified theory of multi-dimensional cognitive resources and can explain the mental workload in the changing environment. In other words, N200 reflects the consumption of earlier perceptual resources. However, P300 reflects the consumption of processing resources actively invested in the later stage [54]. This index takes into account the influence of task difficulty and skill level development, which may be a better and more comprehensive indicator of mental workload. Therefore, this index can provide supplementary information for other traditional mental load measurement. Its applications in future system evaluation and design should be investigated.

Limitations and further research

There are some limitations in the study. First of all, we only used the components of ERP with the largest amplitude, namely N200 and P300, which have been widely concerned in mental workload studies. Some studies suggest that other components may also contain information related to mental workload, but we did not conduct a unified analysis. Secondly, this study analyzes at the group level and demonstrates the law of change at the group level, which reduces the influence of individual differences. Thirdly, since most UAV operators are male at present, we only conducted experiments on male subjects in this experiment, which caused a critical gender unbalance problem in the data collection. We will recruit female subjects to further verify our conclusions. Despite the above shortcomings, based on the multi-resource model, we suggest that the use of auditory probes can reflect the perception-cognitive mental workload. Future work will revolve around: In the aspect of experimental design, designing a more complex oddball paradigm, i.e., one involving three acoustic stimuli, to compare with the results of this study. In the aspect of data analysis, operators were further grouped according to performance and other conditions to analyze individual differences, and more ERP components were analyzed. In the aspect of theoretical to practical application, we plan to explore the calibration method of mental workload and skill level development based on ERP in the next step, and consider

the combination with brain rhythm and other factors to establish a more comprehensive mental workload measurement method.

Conclusion

Training in real scenarios often can't be completed in one day. Therefore, we designed a UAV operation experiment which is closer to the real training environment and contains both task difficulty (LEVEL1 to LEVEL4) and test time (TEST1 to TEST4) as factors. In this design scheme, we not only examine the effects of task difficulty from an electrophysiological perspective, but also analyze the changes in skill level development that accompany the test time. Our study demonstrates the value of auditory probes in assessing the mental workload of the operator during training tasks, especially for skill-based tasks such as UAV. Our proposed framework can describe the electrophysiological changes of the operator's perceptual and cognitive resources in UAV control tasks from the perspective of information processing, namely, N200 is related to task difficulty and P300 is related to skill level development. The statistical results of ERP demonstrate the validity of our model to a certain extent. Our results lay a foundation for the application of auditory probes in more realistic training scenarios. The P300 and N200 in ERP components induced by auditory probes can reflect the changes in the mental workload of the manipulator from the perspective of cognition and perception, respectively, so these two components should be considered comprehensively. P300-N200 takes into account both internal (skill level) and external (task difficulty) factors that affect the level of mental workload, and can be used as a reference index to identify the mental load of operators in the execution of operational tasks. The introduction of this index into future mental workload detection systems may provide a more accurate measurement of mental load and may also provide insight into the development of the operator's skill level.

Author contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by SW, HG, QY, CY, GO and XL; The first draft of the manuscript was written by SW and all authors commented on previous versions of the manuscript. All authors read and approved the all manuscript.

Funding

This research was supported by grants from the National Defence Basic Scientific Research Program of China.

Data availability

Not applicable.

Code availability

Not applicable.

Declarations

Conflict of interest

The authors have no relevant financial or non-financial interests to disclose.

Research involving human and animals rights

Not applicable.

Consent for publication

All authors have checked the manuscript and have agreed to the submission.

Received: 30 November 2022 Accepted: 22 January 2023

Published: 7 March 2023

References

- Hancock P, Parasuraman R. Human factors and safety in the design of intelligent vehicle-highway systems (ivhs). *J Safety Res.* 1992;23(4):181–98.
- Parasuraman R, Rizzo M. *Neuroergonomics: the brain at work*, vol. 3. New York: Oxford University Press; 2006.
- Parasuraman R, Wilson GF. Putting the brain to work: neuroergonomics past, present, and future. *Hum Factors.* 2008;50(3):468–74.
- Ayaz H, Dehais F. *Neuroergonomics: the Brain at work and in everyday life*. Cambridge: Academic Press; 2018. p. 111–4.
- Mehta RK, Parasuraman R. Neuroergonomics: a review of applications to physical and cognitive work. *Front Hum Neurosci.* 2013;7:889.
- Wickens CD, Helton WS, Hollands JG, Banbury S. *Engineering psychology and human performance*. London: Routledge; 2021. p. 3–4.
- Wickens CD. Multiple resources and mental workload. *Hum Factors.* 2008;50(3):449–55.
- Ayaz H, Shewokis PA, Bunce S, Izzetoglu K, Willems B, Onaral B. Optical brain monitoring for operator training and mental workload assessment. *Neuroimage.* 2012;59(1):36–47.
- Haynes JD, Rees G. Decoding mental states from brain activity in humans. *Nat Rev Neurosci.* 2006;7(7):523–34.
- Tharawadeepimuk K, Wongsawat Y. Quantitative eeg in sports: performance level estimation of professional female soccer players. *Health Inf Sci Syst.* 2021;9(1):1–15.
- Tan W, Xu Y, Liu P, Liu C, Li Y, Du Y, Chen C, Wang Y, Zhang Y. A method of vr-eeg scene cognitive rehabilitation training. *Health Info Sci Syst.* 2021;9(1):1–9.
- Kramer AF, Trejo LJ, Humphrey D. Assessment of mental workload with task-irrelevant auditory probes. *Biol Psychol.* 1995;40(1–2):83–100.
- Ke Y, Jiang T, Liu S, Cao Y, Jiao X, Jiang J, Ming D. Cross-task consistency of electroencephalography-based mental workload indicators: comparisons between power spectral density and task-irrelevant auditory event-related potentials. *Front Neurosci.* 2021;15: 703139.
- Ullsperger P, Freude G, Erdmann U. Auditory probe sensitivity to mental workload changes-an event-related potential study. *Int J Psychophysiol.* 2001;40(3):201–9.
- Luck SJ. *An introduction to the event-related potential technique*. Cambridge: MIT press; 2014. p. 144–6.
- Patel SH, Azzam PN. Characterization of n200 and p300: selected studies of the event-related potential. *Int J Med Sci.* 2005;2(4):147.
- Ghani U, Signal N, Niazi IK, Taylor D. A novel approach to validate the efficacy of single task erp paradigms to measure cognitive workload. *Int J Psychophysiol.* 2020;158:9–15.
- Miller MW, Rietschel JC, McDonald CG, Hatfield BD. A novel approach to the physiological measurement of mental workload. *Int J Psychophysiol.* 2011;80(1):75–8.
- Allison BZ, Polich J. Workload assessment of computer gaming using a single-stimulus event-related potential paradigm. *Biol Psychol.* 2008;77(3):277–83.
- Causse M, Fabre E, Giraudet L, Gonzalez P, Maysakhovich V. Eeg/erp as a measure of mental workload in a simple piloting task. *Procedia Manufacturing.* 2015;3:5230–6.

21. Sirevaag EJ, Kramer AF, Coles MG, Donchin E. Resource reciprocity: an event-related brain potentials analysis. *Acta Physiol (Oxf)*. 1989;70(1):77–97.
22. Ghani U, Signal N, Niazi IK, Taylor D. Efficacy of a single-task erp measure to evaluate cognitive workload during a novel exergame. *Front Hum Neurosci*. 2021. <https://doi.org/10.3389/fnhum.2021.742384>.
23. Tang S, Liu C, Zhang Q, Gu H, Li X, Li Z. Mental workload classification based on ignored auditory probes and spatial covariance. *J Neural Eng*. 2021;18(4):0460–9.
24. Xu J, Ke Y, Liu S, Song X, Xu C, Zhou G, Ming D. (2020) Task-irrelevant auditory event-related potentials as mental workload indicators: a between-task comparison study. In: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE, pp 3216–3219
25. Papanicolaou AC, Johnstone J. Probe evoked potentials: theory, method and applications. *Int J Neurosci*. 1984;24(2):107–31.
26. Dyke FB, Leiker AM, Grand KF, Godwin MM, Thompson AG, Rietschel JC, McDonald CG, Miller MW. The efficacy of auditory probes in indexing cognitive workload is dependent on stimulus complexity. *Int J Psychophysiol*. 2015;95(1):56–62.
27. Takeda Y, Inoue K, Kimura M, Sato T, Nagai C. Electrophysiological assessment of driving pleasure and difficulty using a task-irrelevant probe technique. *Biol Psychol*. 2016;120:137–41.
28. Rösler F, Heil M, Röder B. Slow negative brain potentials as reflections of specific modular resources of cognition. *Biol Psychol*. 1997;45(1–3):109–41.
29. Wickens C, Kramer A, Vanasse L, Donchin E. Performance of concurrent tasks: a psychophysiological analysis of the reciprocity of information-processing resources. *Science*. 1983;221(4615):1080–2.
30. Horat SK, Herrmann FR, Favre G, Terzis J, Debatisse D, Merlo MC, Missonnier P. Assessment of mental workload: a new electrophysiological method based on intra-block averaging of erp amplitudes. *Neuropsychologia*. 2016;82:11–7.
31. Roy RN, Charbonnier S, Campagne A, Bonnet S. Efficient mental workload estimation using task-independent eeg features. *J Neural Eng*. 2016;13(2):026019.
32. Kramer AF, Parasuraman R. Neuroergonomics: applications of neuroscience to human factors. In: Cacioppo JT, Tassinari LG, Berntson GG, editors. *Handbook of psychophysiology*. Cambridge: Cambridge University Press; 2007. p. 704–22.
33. Gu H, Yao Q, Chen H, Ding Z, Zhao X, Liu H, Feng Y, Li C, Li X. The effect of mental schema evolution on mental workload measurement: an eeg study with simulated quadrotor uav operation. *J Neural Eng*. 2022;19(2):026058.
34. Cooper GE, Harper RP. The use of pilot rating in the evaluation of aircraft handling qualities. Washington: National Aeronautics and Space Administration; 1969.
35. Shahid A, Wilkinson K, Marcu S, Shapiro CM. Karolinska sleepiness scale (kss). In: Shahid A, Wilkinson K, Marcu S, Shapiro C, editors. *STOP, THAT and one hundred other sleep scales*. New York: Springer; 2011. p. 209–10.
36. O'Reilly JA. Can intensity modulation of the auditory response explain intensity-decrement mismatch negativity? *Neurosci Lett*. 2021;764:136199.
37. Daniel E. Noise and hearing loss: a review. *J Sch Health*. 2007;77(5):225–31.
38. Jasper HH. The ten-twenty electrode system of the international federation. *Electroencephalogr Clin Neurophysiol*. 1958;10:370–5.
39. Iriarte J, Urrestarazu E, Valencia M, Alegre M, Malanda A, Viteri C, Artieda J. Independent component analysis as a tool to eliminate artifacts in eeg: a quantitative study. *J Clin Neurophysiol*. 2003;20(4):249–57.
40. Handy TC. *Event-related potentials: a methods handbook*. Cambridge: MIT press; 2005.
41. Delorme A, Makeig S. Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *J Neurosci Methods*. 2004;134(1):9–21.
42. De Winter JC, Gosling SD, Potter J. Comparing the pearson and spearman correlation coefficients across distributions and sample sizes: a tutorial using simulations and empirical data. *Psychol Methods*. 2016;21(3):273.
43. Verwey WB, Shea CH, Wright DL. A cognitive framework for explaining serial processing and sequence execution strategies. *Psychon Bull Rev*. 2015;22(1):54–77.
44. Isreal JB, Chesney GL, Wickens CD, Donchin E. P300 and tracking difficulty: evidence for multiple resources in dual-task performance. *Psychophysiology*. 1980;17(3):259–73.
45. Debener S, Kranczioch C, Herrmann CS, Engel AK. Auditory novelty oddball allows reliable distinction of top-down and bottom-up processes of attention. *Int J Psychophysiol*. 2002;46(1):77–84.
46. Boutyline A, Soter LK. Cultural schemas: what they are, how to find them, and what to do once you've caught one. *Am Sociol Rev*. 2021;86(4):728–58.
47. Bartlett FC, Bartlett FC. *Remembering: a study in experimental and social psychology*. New York: Cambridge University Press; 1995.
48. Abrahamse EL, Jiménez L, Verwey WB, Clegg BA. Representing serial action and perception. *Psychon Bull Rev*. 2010;17(5):603–23.
49. Verwey WB, Dronkers WJ. Skill in discrete keying sequences is execution rate specific. *Psychol Res*. 2019;83(2):235–46.
50. Van Merriënboer JJ, Schuurman JG, De Croock M, Paas F. Redirecting learners' attention during training: effects on cognitive load, transfer test performance and training efficiency. *Learn Instr*. 2002;12(1):11–37.
51. Dehais F, Duprès A, Blum S, Drougard N, Scannella S, Roy RN, Lotte F. Monitoring pilot's mental workload using erps and spectral power with a six-dry-electrode eeg system in real flight conditions. *Sensors*. 2019;19(6):1324.
52. Roy RN, Bonnet S, Charbonnier S, Campagne A. (2015) Enhancing single-trial mental workload estimation through xdown spatial filtering. In: 2015 7th International IEEE/EMBS Conference on Neural Engineering (NER), IEEE. pp 360–363
53. Solis-Marcos I, Kircher K. Event-related potentials as indices of mental workload while using an in-vehicle information system. *Cogn Technol Work*. 2019;21(1):55–67.
54. Sokhadze EM, Casanova MF, Casanova EL, Lamina E, Kelly DP, Khachidze I. Event-related potentials (erp) in cognitive neuroscience research and applications. *NeuroRegulation*. 2017;4(1):14.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.