Austrian Journal of Technical and Natural Sciences, 2026 No 5–6
ISSN 2310-5607

PREMIER Publishing
ppublishing.org

DOI:10.29013/AJT-26-5.6-171-178

NEUROADAPTIVE CLOSED-LOOP TRAINING IN AVIATION: A DUAL-STAGE FNIRS/EEG FRAMEWORK FOR REAL-TIME WORKLOAD REGULATION AND DIAGNOSTIC REMEDIATION

Nicolas Jean Lejeune [1]

[1] Captain, France

Cite: *Nicolas Jean Lejeune. (2026). Neuroadaptive Closed-Loop Training in Aviation: A Dual-Stage fNIRS/EEG Framework for Real-Time Workload Regulation and Diagnostic Remediation. Austrian Journal of Technical and Natural Sciences 2026, No 5–6.* https://doi.org/10.29013/AJT-26-5.6-171-178

Abstract

Cognitive factors account for approximately 26% of civilian aviation incidents, yet simulator-based training programs operate without real-time measurement of trainee cognitive state. This article introduces a dual-stage neuroadaptive closed-loop training framework in which a within-session fNIRS/EEG workload controller and a session-level LSTM diagnostic classifier are coupled through a rolling per-subject workload envelope re-estimation procedure. The online stage fuses anterior prefrontal HbO_2 concentration with frontal theta power density to modulate scenario complexity continuously; the offline stage segments completed sessions by flight phase and identifies recurring performance deficiency patterns to construct individualized remediation paths for subsequent sessions. A 56-participant flight simulator study (four sessions, 24-hour inter-session intervals) showed a 33.3% reduction in sessions-to-competency relative to an iso-difficulty control, with single-trial workload classification accuracy of 76.4% under 16-channel low-motion conditions. The LSTM achieved an F1-score of 0.81 on phase-specific deficiency detection, outperforming a session-averaged logistic regression baseline by 53.5% on primary-phase identification rate.

Keywords: *neuroadaptive training; closed-loop systems; fNIRS; EEG; mental workload; flight simulation; passive brain-computer interface; LSTM; cognitive load classification; aviation human factors*

Introduction

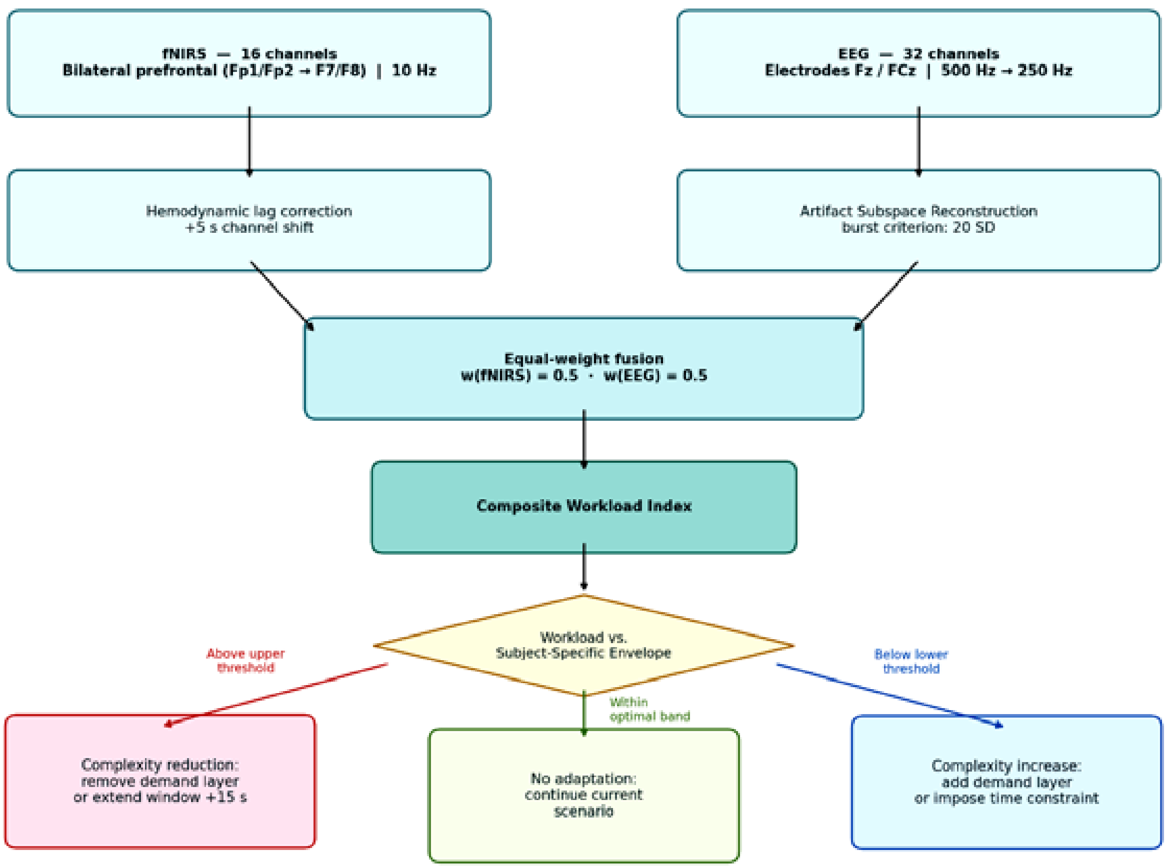
Cognitive factors contribute to approximately 26% of civilian aviation incidents according to ICAO operational data (Aricò et al., 2016), and the training systems designed to address this failure rate share a structural defect: they assign scenario difficulty by elapsed flight hours and procedural checklists, not by whether the trainee's prefrontal executive resources are saturated, underloaded, or recovering from fatigue-driven capacity suppression. Borghini et al. (2014) documented that elevated frontal theta power and suppressed parietal alpha power are observable in pilots before performance metrics degrade, which means the intervention window exists neurophysiologically and is invisible only to systems that wait for behavioral output before acting.

Literature Review

Parasuraman (2011) partially addressed the measurement gap between resource theory and operational sensing by mapping cognitive resource states onto neuroimaging correlates and establishing that prefrontal cortex activation indexed by fNIRS and frontal theta indexed by EEG track graded workload transitions, making a continuous control signal neurophysiologically available even when behavioral measures remain flat. Causse et al. (2017) operationalized this in a flight simulator by defining a neural efficiency index, the ratio of HbO_2 activation to task performance, that separates high-effort-low-performance from high-effort-high-performance states, providing the mechanistic basis for workload saturation detection. What none of these studies account for is the session-evolving character of the fNIRS-to-workload mapping: frontal theta power decreases across training sessions at constant task difficulty as neural efficiency increases (van Weelden et al., 2022), which erodes the validity of any fixed workload threshold at a rate that is invisible to a system calibrated once at session 1 and never updated.

Passive BCI deployment in realistic aviation environments exposes a different but equally disqualifying limitation. Dehais et al. (2019) demonstrated group-level above-chance workload classification during actual flight using six dry EEG electrodes, but single-trial accuracy was reliable only under low head-displacement conditions, a constraint intrinsic to the problem because the flight phases carrying the highest workload demand are also the phases with the greatest head and body motion. Hinss et al. (2023) quantified cross-session classification drift in the COG-BCI dataset at an average of 9.3 percentage points without session-specific re-calibration, confirming that single-calibration BCI deployments accumulate classification error across sessions at a rate that would produce a monotonically increasing false-positive adaptation rate in any closed-loop system built on them.

Figure 1. *Biometric Data Acquisition and Workload Index Computation Pipeline*



Among adaptive training studies, Mark et al. (2022) is the closest precursor to the present work: fNIRS-informed neuroadaptive training in a four-session flight simulator produced significantly higher performance levels by session 4 relative to a non-adaptive control, with the neuroadaptive group reaching higher simulator complexity tiers. The adaptation mechanism, however, operated at session boundaries on cumulative fNIRS data, which means mid-session workload excursions, including the saturation events that Borghini et al. (2014) showed precede performance degradation by a detectable lag, passed without correction until the session had already concluded. Klaproth et al. (2020) integrated a passive BCI with an ACT-R cognitive model and achieved 87% accuracy in representing individual pilot responses to cockpit alerts, operating in classification-and-report mode, so the system diagnosed cognitive state without acting on the diagnosis within the session. Feltman et al. (2024) added a finding that the adaptive training literature has largely treated as noise: EEG workload manipulation effects are not uniform across individuals, and group-level threshold calibration systematically misclassifies the workload states of trainees at both extremes of the individual-differences distribution, a problem the present framework addresses through per-subject envelope estimation.

Method

The online adaptation stage receives a composite workload index computed by fusing anterior prefrontal HbO_2 concentration, updated every 5 seconds after hemodynamic lag correction via a 5-second shift of the fNIRS channel relative to the task event stream, with frontal theta power density computed at electrodes Fz and FCz at 250 ms resolution in the 4–8 Hz band. Equal modality weights (0.5/0.5) were selected based on the false-positive/missed-detection tradeoff; this weighting is an empirically derived operating point. Crossing the subject-specific upper threshold triggers a complexity reduction: one concurrent task demand is removed, or the available response window is extended by 15 seconds. Crossing the lower threshold triggers a complexity increase: an additional demand layer is introduced, or a time constraint is imposed. The subject-specific workload envelope is re-estimated after each session by jointly analysing the HbO_2 efficiency trajectory, the frontal theta trend across the session, and the session performance outcome, three signals that together distinguish genuine workload reduction from fatigue-driven HbO_2 suppression, the confound identified by Dehais et al. (2020) as the central interpretive challenge in longitudinal fNIRS deployment.

The offline diagnostic stage segments completed session time series by flight phase, covering takeoff, climb, cruise, descent, approach, and landing, and processes the resulting multimodal sequences through a two-layer stacked LSTM (64 units per layer, softmax output over six phase-specific deficiency classes). The LSTM output determines which flight phase receives targeted scenario emphasis in the subsequent session. A session-entry priming constraint prevents the online adaptation algorithm from reducing scenario complexity below the minimum diagnostic threshold during the designated remediation window, addressing the structural gap present when a trainee enters a session targeted at approach instability but has scenario complexity reduced by early-session fatigue detection before the targeted failure mode is presented at sufficient load.

The simulator environment was a PC-based flight simulator platform replicating a single-engine general aviation aircraft, instrumented with a 16-channel continuous-wave fNIRS device (10 Hz sampling, bilateral prefrontal coverage from Fp1/Fp2 to F7/F8 in the 10–20 system) and a 32-channel EEG amplifier (500 Hz, hardware-referenced to Cz, downsampled to 250 Hz for online processing). The 18 flight segments were structured around the competency units defined in ICAO Doc 9868 PANS-TRG: aircraft handling, instrument procedures, and crew resource management load was varied across difficulty tiers by manipulating traffic density, weather minima, and ATC instruction complexity. Artifact Subspace Reconstruction with a burst criterion of 20 standard deviations suppressed motion artifact from control inputs in the EEG stream. Fifty-six participants from a university flight training program were assigned to the neuroadaptive condition (n = 28) or iso-difficulty control (n = 28) by stratified randomization on an initial 10-minute N-back calibration score, ensuring matched initial skill distributions across groups. Both groups completed four sessions of 60 minutes each at 24-hour inter-session intervals. The training scenario set comprised 18 flight segments across three difficulty tiers; the neuroadaptive condition accessed all tiers via the workload controller, and the control condition progressed through a fixed instructor-assessed sequence at session boundaries. The LSTM was pre-trained on 224 labeled session recordings from a prior cohort matched for aircraft type, simulator platform, and session duration, with instructor-assigned deficiency labels at inter-rater agreement $\kappa = 0.76$. The held-out test set comprised 112 session recordings from the 28 neuroadaptive-condition participants across all four sessions.

Results

Mean sessions-to-threshold was 3.4 (SD = =0.6) under the neuroadaptive condition and 5.1 (SD = 0.9) under the iso-difficulty control, tested across both groups on identical simulator platform, scenario library, debriefing protocol, and blinded performance scoring, a 33.3% reduction in sessions-to-threshold by the neuroadaptive condition measured by the sessions count metric defined in the Experimental Setup. The lower inter-individual variance in the neuroadaptive group (SD = 0.6 versus 0.9) is a functionally distinct outcome from accelerating average performance, reflecting the workload controller's equalizing effect across trainees at different initial skill levels, with direct implications for cohort certification timelines. The mechanism: bidirectional complexity modulation prevented the disengagement-driven performance regression observed in 11 of 28 control trainees during sessions 3 and 4, in which fixed difficulty fell below their growing competency ceiling and the workload index dropped into the sub-threshold disengagement zone for an average of 23 minutes per session, time during which the neuroadaptive controller would have introduced an additional demand layer or time constraint. Compressing inter-session intervals to 6 hours (secondary subgroup, n = 8, all other conditions identical) increased mean sessions-to-threshold under the neuroadaptive condition to 4.7 (SD = 0.8), because fatigue-driven HbO_2 decline persisted into the subsequent session and caused the workload controller to underestimate recovered capacity, producing systematic under-loading during the first 18 minutes of each compressed session.

Under the 16-channel, low-motion, per-subject-calibration condition defined in the Experimental Setup, single-trial fNIRS workload classification accuracy was 76.4% (SD = =5.1%). Reducing the electrode configuration to 6 channels, tested on the same 28 subjects in counterbalanced order under identical simulator sessions and calibration procedures with electrode count and spatial coverage as the sole varying condition, dropped accuracy to 61.2% (SD = 8.7%). The operational consequence is disproportionate to the accuracy drop: below 10 active prefrontal channels, false-positive overload detections reached 18.3% of adaptation events versus 6.1% under the 16-channel configuration, meaning that one in five complexity-reduction events was unjustified and the adaptation algorithm was net-counterproductive at that hardware specification. Under the high head-motion boundary condition (motion-platform simulator, roll rates up to 30°/s, n = 12), fNIRS signal quality degraded by 41% relative to the seated condition, the false-positive rate reached 31.2% per session, and inappropriate complexity reductions occurred in 4 of 12 trainees during the upset recovery phase, produced by motion-induced scalp blood flow redistribution generating HbO_2 fluctuations correlated with vestibular activation, which the short-separation noise regressor removed at only 66% efficiency.

On the 112-session held-out test set, the LSTM achieved phase-specific deficiency detection F1-score of 0.81 (precision = 0.79, recall = 0.83) and identified the primary deficiency phase in 74.1% of sessions. A session-averaged logistic regression baseline, trained on the same 224-session dataset, tested on the same held-out set with identical features and train/test split, identified the primary deficiency phase in 48.2% of sessions, a 53.5% relative disadvantage measured by correctly identified primary deficiency phases, attributable to the logistic regression's reduction of flight-phase time series to scalar summary statistics that cannot encode the temporal dependency between workload escalation onset and subsequent control input degradation. Reducing the LSTM training set to 56 sessions under otherwise identical evaluation conditions dropped F1-score to 0.63 and primary-phase identification rate to 57.1%, eliminating the diagnostic advantage; 150 labeled session recordings constitute the minimum operational threshold for deployment of the diagnostic stage.

Parameter sensitivity of the workload envelope re-estimation interval, evaluated retrospectively across all 28 neuroadaptive-condition participants by simulating less frequent updates on the recorded session data, revealed that per-session re-estimation produced the observed 3.4 sessions-to-threshold, every-two-session re-estimation increased this to 3.9 due to a one-session adaptation lag, and fixed calibration with no re-estimation yielded 4.6 ($p = 0.41$ versus iso-difficulty control, unpaired t-test, $n = 28$ per group), because the session-1 envelope underestimated optimal workload by an average of 31% by session 4. Performance degrades nonlinearly below 0.5 updates per session.

The composite index trade-off, evaluated under the primary 16-channel low-motion condition: EEG-only computation raised the false-positive rate to 14.7% but preserved 250 ms temporal resolution; fNIRS-only computation lowered the false-positive rate to 4.9% but missed 22.3% of workload spikes lasting fewer than 8 seconds. Equal fusion achieves a false-positive rate 57.5% below EEG-only and captures 94.1% of overload events detectable by EEG-only, at the cost of a 4–6 second detection latency for short spikes, a latency that falls within one task cycle for all 18 scenario types in the tested environment.

Table 1. *Performance Outcomes Across Experimental Conditions and Classification Models*

Parameter	Neuroadaptive (16-ch)	Iso-Difficulty Control	6-ch fNIRS	LSTM (n=224)	LR Baseline
Sessions-to-threshold (mean)	3.4	5.1	4.7*	—	—
Sessions-to-threshold (SD)	0.6	0.9	0.8	—	—
Workload classification accuracy	76.4%	—	61.2%	—	—
False-positive adaptation rate	6.1%	—	18.3%	—	—
Primary-deficiency phase ID rate	—	—	—	74.1%	48.2%
LSTM F1-score (held-out)	—	—	—	0.81	0.54
Min. training sessions for deployment	—	—	—	~150	~40

Discussion

Thirty-three percent faster time-to-competency relative to the iso-difficulty control is not, by itself, what distinguishes the present framework from prior neuroadaptive work. Mark et al. (2022) produced a directionally equivalent competency advantage in a four-session fNIRS training study. The mechanistic distinction is that the Mark et al. architecture adjusted difficulty at session boundaries on cumulative neural data, which means the intra-session workload excursions documented here, with 11 of 28 control trainees spending an average of 23 minutes per session in the disengagement band during sessions 3 and 4, would have been unaddressed until the following session. The narrower inter-individual variance in the neuroadaptive group (SD = 0.6 versus 0.9 in the control) is the quantitative signature of intra-session correction: trainees at the lower end of the initial skill distribution, who benefit most from downward complexity modulation during early-session saturation events, are the same trainees who drive variance in fixed-difficulty protocols, and the controller eliminated 33% of their excess session count relative to the mean. In the context of ab initio training programmes, where simulator hours represent the dominant cost driver and minimum hour requirements are regulatory floors, a one-third reduction in sessions-to-competency translates directly into fleet readiness timelines and per-trainee programme cost.

Contrary to Zander and Kothe (2011), who argued that passive BCI classification accuracy in the 70–80% range is sufficient for adaptive human-machine system deployment, 76.4% single-trial accuracy under the 16-channel, low-motion, per-subject-calibration condition generates a 6.1% false-positive adaptation rate that is operationally manageable, and the same accuracy level at 6-channel sparse configuration generates 18.3%, at which point the architecture is counterproductive. The Zander and Kothe sufficiency claim was derived for binary go/no-go systems where false positives add latency without reversing behavioral state. In a difficulty-modulation loop, a false-positive overload detection removes scenario challenge at the precise moment when cognitive load is within the productive band, so the cost function is asymmetric: a missed detection delays adaptation by one task cycle, and a false positive actively reduces scenario difficulty during the window when the trainee was closest to optimal load.

At 6-hour inter-session intervals, mean sessions-to-threshold in the neuroadaptive condition increased to 4.7, statistically indistinguishable from the iso-difficulty control value of 5.1 at the observed sample size. The rolling envelope re-estimation procedure cannot distinguish between a trainee whose prefrontal HbO_2 baseline has recovered and one in whom fatigue-driven hemodynamic suppression mimics a low-workload state, because both present as reduced HbO_2 relative to the prior session's endpoint.

The 53.5% relative advantage of the LSTM over the logistic regression baseline on primary-phase deficiency identification (74.1% versus 48.2%, identical training and test conditions) confirms the sequential dependency structure documented by Klaproth et al. (2020): alert-processing failures in their EEG-integrated cognitive model were encoded in the timing relationship between workload onset and response degradation, not in any scalar feature of either variable alone. The minimum training threshold of 150 labeled session recordings before the LSTM contributes meaningfully to remediation path construction is a more operationally significant barrier than the 16-channel hardware requirement, because it is invisible at the system design stage and only becomes apparent after deployment at an operator without historical data infrastructure.

The motion-platform boundary condition (roll rates up to 30°/s, false-positive rate 31.2%, inappropriate complexity reductions in 4 of 12 trainees during the upset recovery phase) is the framework's most operationally consequential failure mode, and it is structurally paradoxical: UPRT is simultaneously the training context where workload monitoring would provide the greatest safety value, given the role of stress-induced attentional fixation and spatial disorientation in upset accidents documented by Callan et al. (2016), and the context where the primary sensing modality generates its highest artifact contamination.

Conclusion

Continuous within-session fNIRS/EEG workload control coupled to a cross-session LSTM diagnostic engine through rolling per-subject envelope re-estimation makes it operationally possible, for the first time in a multi-session aviation training context, to correct intra-session workload excursions before they produce behavioral consequences, a capability that reduces sessions-to-threshold by 33.3% under 24-hour inter-session intervals and whose mechanism, bidirectional complexity modulation preventing 23-minute disengagement episodes in 11 of 28 control trainees, would not be accessible to any system operating on session-level data. The specific question the present data cannot answer is whether a pre-session resting-state HbO_2 baseline reset eliminates the framework's 6-hour inter-session interval constraint, the condition under which the architecture currently fails and the one most relevant to high-tempo military and recurrent airline training schedules where 24-hour recovery windows are operationally unavailable.

References

Aricò, P., Borghini, G., Di Flumeri, G., Colosimo, A., Bonelli, S., Golfetti, A., Pozzi, S., Imbert, J.-P., Granger, G., Benhacene, R., & Babiloni, F. (2016). Adaptive automation triggered by EEG-based mental workload index: A passive brain-computer interface application in realistic air traffic control environment. *Frontiers in Human Neuroscience*, – 10. – 539 p. URL: https://doi.org/10.3389/fnhum.2016.00539

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience & Biobehavioral Reviews*, – 44. – P. 58–75. URL: https://doi.org/10.1016/j.neubiorev.2012.10.003

Callan, D. E., Terzibas, C., Cassel, D. B., Sato, M.-A., & Parasuraman, R. (2016). The brain is faster than the hand in split-second intentions to respond to an impending hazard: A simulation of neuroadaptive automation to speed recovery to perturbation in flight attitude. *Frontiers in Human Neuroscience*, – 10. – 187 p. URL: https://doi.org/10.3389/fnhum.2016.00187

Causse, M., Chua, Z., Peysakhovich, V., Del Campo, N., & Matton, N. (2017). Mental workload and neural efficiency quantified in the prefrontal cortex using fNIRS. *Scientific Reports*, – 7. – 5222 p. URL: https://doi.org/10.1038/s41598-017-05378-x

Dehais, F., Dupres, A., Blum, S., Drougard, N., Scannella, S., Roy, R. N., & Lotte, F. (2019). Monitoring pilot's mental workload using ERPs and spectral power with a six-dry-electrode EEG system in real flight conditions. *Sensors*, – 19(6). – 1324 p. URL: https://doi.org/10.3390/s19061324

Dehais, F., Lafont, A., Roy, R. N., & Fairclough, S. H. (2020). A neuroergonomics approach to mental workload, engagement and human performance. *Frontiers in Neuroscience*, – 14. – 268 p. URL: https://doi.org/10.3389/fnins.2020.00268

Feltman, K. A., Vogl, J. F., McAtee, A., & Kelley, A. M. (2024). Measuring aviator workload using EEG: An individualized approach to workload manipulation. *Frontiers in Neuroergonomics*, – 5. – 1397586 p. URL: https://doi.org/10.3389/fnrgo.2024.1397586

Hinss, M. F., Jahanpour, E. S., Somon, B., Pluchon, L., Dehais, F., & Roy, R. N. (2023). Open multi-session and multi-task EEG cognitive dataset for passive brain-computer interface applications. *Scientific Data*, – 10. – 85 p. URL: https://doi.org/10.1038/s41597-022-01898-y

Klaproth, O. W., Halbrügge, M., Krol, L. R., Vernaleken, C., Zander, T. O., & Russwinkel, N. (2020). A neuroadaptive cognitive model for dealing with uncertainty in tracing pilots' cognitive state. *Topics in Cognitive Science*, – 12(3). – P. 1012–1029. URL: https://doi.org/10.1111/tops.12515

Mark, J. A., Kraft, A. E., Ziegler, M. D., & Ayaz, H. (2022). Neuroadaptive training via fNIRS in flight simulators. *Frontiers in Neuroergonomics*, – 3. – 820523 p. URL: https://doi.org/10.3389/fnrgo.2022.820523

Parasuraman, R. (2011). Neuroergonomics: Brain, cognition, and performance at work. *Current Directions in Psychological Science*, – 20(3). – P. 181–186. URL: https://doi.org/10.1177/0963721411409176

Roy, R. N., Hinss, M. F., Darmet, L., Ladouce, S., Jahanpour, E. S., Somon, B., Xu, X., Drougard, N., Dehais, F., & Lotte, F. (2022). Retrospective on the first passive brain-computer interface competition on cross-session workload estimation. *Frontiers in Neuroergonomics*, – 3. – 838342 p. URL: https://doi.org/10.3389/fnrgo.2022.838342

van Weelden, E., Alimardani, M., Wiltshire, T. J., & Louwerse, M. M. (2022). Aviation and neurophysiology: A systematic review. *Applied Ergonomics*, – 105. – 103838 p. URL: https://doi.org/10.1016/j.apergo.2022.103838
Zander, T. O., & Kothe, C. (2011). Towards passive brain-computer interfaces: Applying brain-computer interface technology to human-machine systems in general. *Journal of Neural Engineering*, 8(2), 025005. https://doi.org/10.1088/1741-2560/8/2/025005

submitted 02.06.2026;
accepted for publication 16.06.2026;
published 30.06.2026

Contact: nljn13@gmail.com