Human Factors in Design, Engineering, and Computing, Vol. 238, 2026, 39–48
https://doi.org/10.54941/ahfe1008239

EEG-Based Classification of Reward, Penalty, and Movement Events Across Four Video Games of Varying Complexity

Laura Martinez-Tejada[1], Dechathron Dasri[2], Yuta Yamamura[1], and Natsue Yoshimura[1,3,4]

[1]School of Computing, Department of Computer Science, Institute of Science Tokyo, Yokohama, Japan
[2]School of Information, Computer, and Communication Technology, Thammasat University, Bangkok, Thailand
[3]Institute of Biomedical Engineering, Institute of Science Tokyo, Yokohama, Japan
[4]Neural Information Analysis Laboratories, ATR, Kyoto, Japan

ABSTRACT

Longitudinal studies have tested video games as tools to alleviate mental health disorders such as depression and anxiety, with promising results. However, to fully assess the efficacy of video games for treating and diagnose mental health disorders, it is necessary to understand how they affect cognitive events in real time during self-paced video game interaction, and how reproducible these events are across different types of video game environments. This study characterises ERPs to three interactive mechanics: reward, penalty, and movement, across four games of increasing complexity, and tests whether the mechanics can be decoded from single-events. Seven healthy male participants played four video games of different complexity for approximately ten minutes in randomised order while 64-channel EEG was recorded. ERPs were examined over the FCz electrode and single-events were then classified with an EEGNet convolutional network across three binary tasks in cross-subject and within-subject settings. Reward events elicited a fronto-central positivity in the 200–500 ms range in every game, consistent with the reward positivity. Penalty events elicited fronto-central deflections in a comparable window but were less consistent across games. Movement events elicited a negative shift around key-press onset over fronto-central sites. All three binary classification tasks exceeded chance. Penalty versus movement was the most separable (73.3% pooled cross-subject; 81.0% best single game) and reward versus penalty the least (64.3% pooled). Within-subject models with subject-specific analysis windows performed better than cross-subject models. Both the ERP and classification results indicate that game context and individual differences shape these responses.

Keywords: Event-related potentials, Video games, Reward processing, EEGNet, Single-trial classification, Ecological validity

INTRODUCTION

Longitudinal studies have examined video games as a means of alleviating mental-health disorders such as depression and anxiety (Bergmann et al., 2023; Butler et al., 2020), and have reported a significant decrease in symptoms

Received August 28, 2026; Revised October 2, 2026; Accepted October 10, 2026; Available online December 1, 2026

© 2026 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

after exposure to video games like Super Mario Odyssey and Tetris for six weeks. These results are reported using different techniques such as self-report questionnaires and structural MRI. Mostly, research related to the use of video games to treat depression or anxiety uses questionnaires before or after frequent exposure to the video game, raising the question of how video games can affect cognitive or neurological processes in real time during the exposure. One possible way to analyze real-time neurological dynamics is event-related potentials (ERPs), which are changes in electroencephalography (EEG) signals that are time-locked to sensory, motor, or cognitive events (Sur & Sinha, 2009).

Depression and anxiety studies have found that ERPs are sensitive to negativity bias (the tendency to preferentially attend to negative cues), captured by components such as P1, P2, N2, and the late positive complex (e.g., P3, late positive potential [LPP]), which are amplified in response to negative stimuli (relative to positive or neutral stimuli) such as pictures of sad faces or negative words (Klumpp & Shankman, 2018). In anxiety research, components including P1, P2, P3, N2, N1, error related negativity (ERN), correction related negativity (CRN), feedback-related negativity (FRN), mismatch negativity (MMN), error positivity (Pe), and LPP have been studied in oddball and emotional-stimulus paradigms comparing patients with healthy controls (Xie et al., 2025). For depression, ERP components such as P1, N170, P3, and LPP, which are related to auditory, reward, and emotional stimuli, have been studied compared with healthy controls (Li et al., 2023; Proudfit et al., 2015; Zhang et al., 2016; Zhou et al., 2019).

Some studies have researched how video game mechanics can instantiate the kinds of events used in laboratory cognitive tasks to study attention, cognitive control, movement preparation, reward processing and performance monitoring. A scoping review by Fleischmann et al., identified 23 studies addressing stimulus (sensory input), response (behavioral) and feedback (performance monitoring) related ERPs across video games differing in genre (arcade, action, simulation, puzzle and sports), visual presentation (2D, 3D and VR), type of interface (abstract or naturalistic visual elements), and player perspective such as side view, top-down, third-person, or first-person (Fleischmann et al., 2026). Studies of feedback-related ERPs report components such as P3, ERN, Pe, FRN, and LPP in response to unexpected game events or valence-related feedback (Cavanagh & Castellanos, 2016; Martínez-Tejada et al., 2021; Spüler & Niethammer, 2015; Wimmer et al., 2024).

Most studies, however, rely on controlled events designed to elicit specific ERPs: pseudorandom perturbations of avatar control, pseudorandom mechanics such as sudden or novel events, and go/no-go paradigms with visual or auditory feedback. This raises two questions. First, to what extent can response- and feedback-related ERP components be identified in a self-paced video game containing no events inserted specifically to elicit ERP components? Second, do these ERP components replicate across video game environments? In other words, if a participant exhibits a specific ERP in response to a feedback mechanic (e.g., reward), will the same or a similar response appear in a different game environment when an equivalent mechanic occurs?

Purpose of This Study

To study the potential of video games and EEG, especially ERPs, for the diagnosis and treatment of depression and anxiety, it is important to know to what extent video games can naturally elicit ERP components related to response (keyboard manipulation) and feedback (reward, penalty) mechanics, and whether they do so consistently across different game environments.

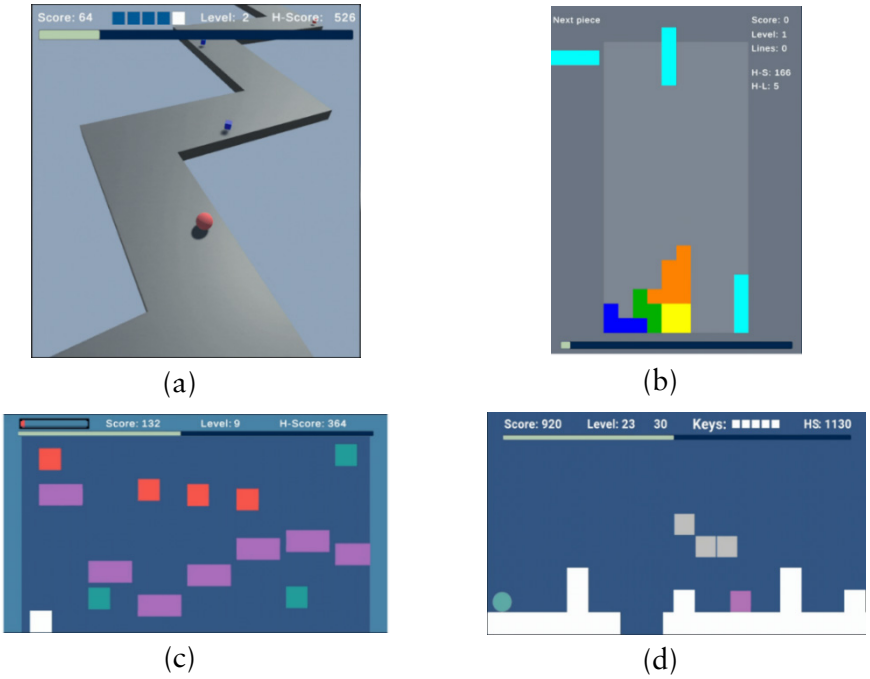
For this study, we analysed ERPs from healthy participants playing four video games of increasing complexity: Zig-Zag, Up-Down Scroller, Platformer, and Tetris. We focused on three mechanics: two feedback-related (reward and penalty) and one response-related (movement through keyboard presses).

The guiding research question is: To what extent is it possible to characterize and classify similar interactive mechanics across games of varying complexity using EEG? The study proceeds in two stages: an ERP characterization of the response to each mechanic, and artificial intelligence classification of the mechanics from single-event EEG.

METHODOLOGY

Games and Event Mechanics

Three interactive mechanics were studied across four games of increasing complexity: reward (gaining points or lives), penalty (losing points or lives), and movement (keyboard presses). All games used simple geometric shapes (spheres, cubes, circles, squares) as visual elements. Games are as follows:



(a) (b) (c) (d)

Figure 1: Interfaces of each game: (a) Zig-zag, (b) Tetris, (c) platformer, (d) up-down scroller.

- Zig-Zag (Figure 1a). The participant keeps a ball rolling along a narrow endless platform, changing direction by 90° with each [down] key press, and earning one point per second while remaining on the platform. Collecting blue cubes yields points and restores lives back (reward); colliding with a red cube or missing a blue cube reduces points or costs a life (penalty). Difficulty increases through additional concurrent tokens and changes in platform shape. The game ends when the participant falls off the platform or loses five lives.
- Up-Down Scroller (Figure 1c). The participant navigates ten levels, moving a white square with [up]/[down] keys to collect or avoid tokens scrolling from right to left while a life meter depletes over time. Green squares give points and replenish life (reward); red and purple squares cost points and accelerate depletion (penalty). The game ends when life runs out.
- Platformer (Figure 1d). The participant controls a green ball and must reach the next level within 30 seconds, navigating obstacles while avoiding enemies and hazard blocks (penalty) and collecting a set number of keys (reward) to open a gate. Movement uses [up]/[left]/[right]. Collision with an enemy or hazard causes injury, recoverable by finding a health kit. The game ends on timeout, on falling into a gap, or after two consecutive injuries without recovery.
- Tetris (Figure 1b). The participant clears horizontal lines by arranging falling pieces, moving and rotating them with the arrow keys [up]/[down]/[left]/[right]. Clearing a line yields points and reduces stack height (reward). Garbage lines are added at fixed intervals; creating holes, deepening them, or raising the stack is treated as a penalty. The game ends when the stack reaches the top of the board.

Participants

Seven healthy male participants with ages ranging from 21–31 years (mean 25.6 ± 3.5) participated in the study. All reported normal or corrected-to-normal vision and no history of neurological or psychiatric disorder. The study was approved by ethics committee of the Institute of Science Tokyo (Approval No. 2025157) and conducted in accordance with the Declaration of Helsinki. The procedures were explained to each participant prior to the experiment, and all participants gave written informed consent.

EEG Acquisition Methodology and Preprocessing

Participants played each game for approximately 10–15 minutes (M = 12.5 min) in randomized order. One minute of eyes-open resting-state EEG was recorded before each game, and participants completed questionnaires and rested between games. EEG was recorded from 64 channels at 2048 Hz using a BioSemi ActiveTwo amplifier system with active electrodes and two earlobe references (Biosemi, Amsterdam, Netherlands).

Lab Streaming Layer (LSL; Swartz Center for Computational Neuroscience, University of California, San Diego, CA, USA) recorded two synchronized streams: the EEG signal and event markers generated by the video game, each sample timestamped using the computer's local high-resolution clock. Data were processed using MATLAB R2023b (The MathWorks, Inc., Natick, MA, USA). Signals were band-pass filtered with an FIR filter (1–45 Hz) and resampled to 512 Hz. Corrupted segments were rejected by visual inspection; files from the four games were then concatenated so that bad channels could be interpolated, a common average reference applied, and independent component analysis used to remove ocular and muscular components, after which the data were split back into individual games.

ERP Analysis

Epochs were extracted from −500 to 800 ms around each event and baseline-corrected to −0.25 to 0 s relative to event onset for reward and penalty events, and −0.5 to −0.35 s relative to onset for movement events. Epochs were rejected using adaptive threshold (median + 3 × IQR of per-epoch peak-to-peak amplitudes, from 60–250 μV depending on the participant); events with fewer than 15 clean epochs were excluded.

Grand-average ERPs were computed at electrode FCz for each mechanic and game. Grand averages were computed as the mean of participant averages, so that participants contributing more epochs do not dominate. Mean amplitudes were extracted in the 200–350 ms window for reward and penalty events, and −100 to 100 ms for movement events, and tested against zero and against each other using t-tests across participants; per-event tests were additionally corrected with Benjamini–Hochberg FDR across all 21 events.

Classification Method

For every labelled event, an epoch was extracted around event onset from the preprocessed (1–45 Hz filtered, ICA-cleaned) data, yielding trials of 64 channels × time samples at 512 Hz. No baseline correction or feature standardization was applied to the classification epochs. Three binary tasks were evaluated (reward vs. movement, reward vs. penalty, penalty vs. movement). For each task, classes were balanced by random undersampling of the majority class, so chance performance is 50%. Trials were split into training (65%), validation (15%), and test (20%) sets, stratified by class label; trials from all participants were pooled; accuracies reflect subject-dependent performance rather than generalization to unseen participants. Models were evaluated in four settings: pooled participants with all games, pooled participants per game, single participants with all games, and single participants per game.

The classifier was an adapted EEGNet spatio-temporal convolutional network: a temporal convolution (kernel ≈ 125 ms), a depthwise spatial

convolution (depth multiplier = 2), and a separable convolution, each followed by batch normalization, ELU activation, average pooling, and dropout (p = 0.3), with a final softmax layer. Training used Adam (learning rate 1×10^{-3}) with class-weighted cross-entropy, up to 80 epochs with early stopping (patience = 15) on validation macro-F1, and a ReduceLROnPlateau scheduler. Because the latency of the informative response differs between individuals and tasks, the epoch window was grid-searched per model over window start times of 0.2 and 0.5 s before event onset and end times of 0.2, 0.4, 0.5, 0.6, and 0.8 s after onset. The window with the highest validation accuracy was selected, and the corresponding test accuracy was reported.

RESULTS

ERP Analysis

Grand-averaged ERPs were calculated and compared against zero (Figure 3a). Reward events elicited a fronto-central positivity in the 200–500 ms range in all four games, consistent with the reward positivity (Figure 2a). Mean amplitude at FCz was 1.53uV ± 0.67μV, $t(6) = 6.01$, $p < 0.001$, $d = 2.27$, with the effect in the same direction in 7 of 7 participants. Penalty events elicited fronto-central deflections in a comparable window (Figure 2b). Mean amplitude at FCz was 2.36uV ± 1.13μV, $t(6) = 5.51$, $p = 0.002$, $d = 2.08$, with effect in 7 of 7 participants. Movement events elicited a pre-movement negative shift around key-press onset over fronto-central sites (Figure 2c). Mean amplitude at FCz was −0.29uV ± 0.32μV, $t(6) = -2.38$, $p = 0.055$, $d = -0.90$, present in 5 of 7 participants.

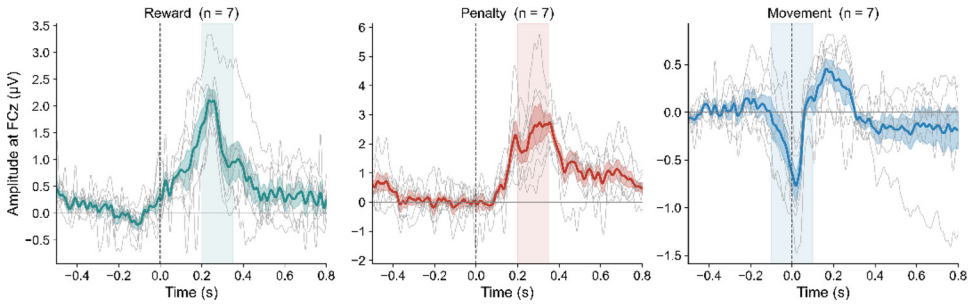


Figure 2: Grand-averaged ERPs from FCz electrode with ±1 SEM across participants, and across games (a) reward, (b) penalty, (c) movement. Shaded region: 200–350 ms for feedback events; −100 to +100 ms for response event.

Results of paired t-test (Figure3b) show that reward and penalty do not differ in amplitude ($\Delta = -0.84$μV, $t(6) = -1.83$, $p = .116$), while both differed from movement (reward: $\Delta = 1.81$μV, $t(6) = 5.58$, $p = .001$, $dz = 2.11$; penalty: $\Delta = 2.65$μV, $t(6) = 6.70$, $p < .001$, $dz = 2.53$).

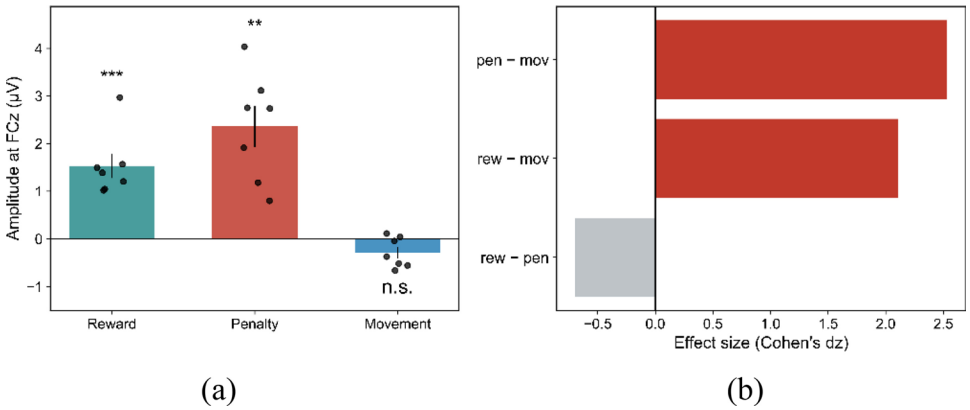


Figure 3: Significant differences from grand-average amplitude (a) mean and SEM across participants for each event category. (b) Pairwise contrast between event categories, reward and penalty (in red) showed significant different against movement, but not against each other.

Classification Results

All three binary tasks exceeded the 50% chance level in every setting (Table 1, Fig. 4). Penalty vs. movement was consistently the most separable (81.0% on the Up-Down scroller; 73.3% pooled), while reward vs. penalty was the least (55.9% on Zig-Zag; 64.3% pooled). This ordering mirrors the ERP analysis results: penalty and movement engage different processes with distinct time courses, whereas reward and penalty produce overlapping fronto-central components. Per-game models outperformed the pooled model, indicating that game context shapes the response.

Table 1: Cross-subject test accuracy (%) by game and task.

Task	Platformer	Tetris	Up-Down Scroller	Zig-Zag	All
Reward vs Movement	78.5	70.0	65.1	69.0	70.1
Reward vs Penalty	72.5	64.0	68.7	55.9	64.3
Penalty vs Movement	78.8	73.2	81.0	73.0	73.3

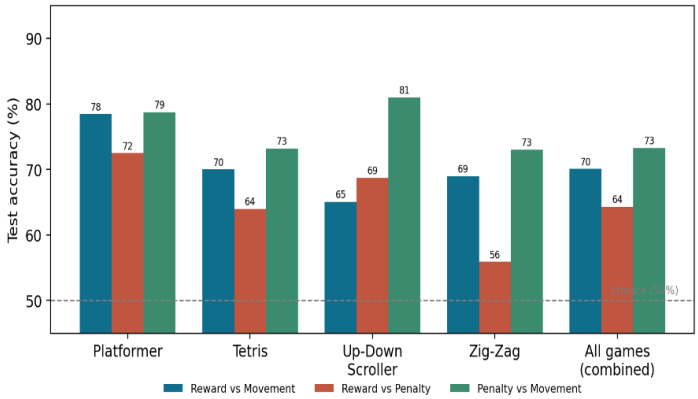


Figure 4: Cross-subject accuracy by game and task.

Within-subject models performed better and varied considerably between participants (Fig. 5). Averaged across participants, all-games accuracies were 70.2% (reward vs. movement), 69.1% (reward vs. penalty), and 74.3% (penalty vs. movement). The best single-participant, single-game models reached up to 90.9% (participant 6, platformer) and 95.0% (participant 7, zig-zag). The window selected by the grid search varied across participants and tasks, and tailoring it per participant improved accuracy, consistent with the individual differences seen in the ERPs. Penalty remained the limiting factor, as penalty events are rare by design.

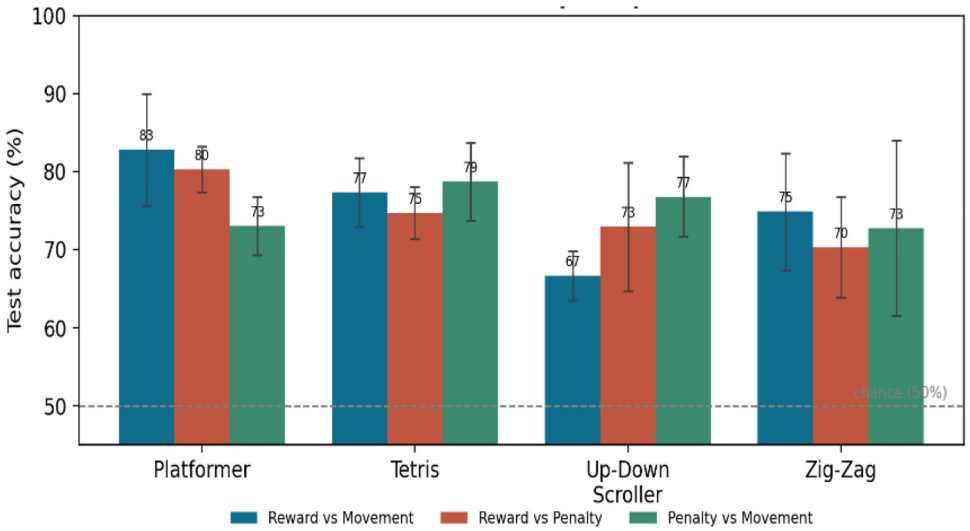


Figure 5: Within-subject accuracy by participant and task.

DISCUSSION

The two stages tell a consistent story. The ERP analysis shows that reward and penalty engage feedback-related cognitive components that overlap in time (approximately 200–500 ms) and space (fronto-central midline), while movement engages a distinct pre-movement response. The classifier reproduces exactly this structure: penalty vs. movement is the easiest task to classify and reward vs. penalty the hardest. Game complexity also matters; models specialized to a single game outperform the pooled model, so a fully game-independent representation of each event is only partially recovered. Both stages point to individual differences: ERP latency and morphology differ between participants, and adapting the analysis window per participant produced the largest gain in classification accuracy.

The clearest opportunity is to move from the current time × channel input to an explicit spatio-temporal representation of the EEG. Each time point can be projected onto a two-dimensional scalp map to form a sequence of topographic frames, or encoded as a functional-connectivity graph. Architectures that exploit this structure include 3D or CNN-LSTM networks over topographic frames, graph neural networks, and spatio-temporal transformers. These models can capture how components

propagate across the scalp—information that one-dimensional temporal models discard—and are particularly promising for the difficult reward-versus-penalty contrast.

Complementary directions include making the per-participant window an explicit calibration step or a learned temporal-attention mechanism; using transfer learning to combine cross-subject data with participant-specific fine-tuning; increasing the number of penalty events, for example by introducing additional penalty mechanics or raising game difficulty; fusing EEG with the recorded biosignals; and addressing onset time-shifts to enable real-time classification and eventual clinical translation.

LIMITATIONS

This study has several limitations. The sample is small (n = 7) and entirely male, which limits generalisation. Penalty events were rare relative to reward and movement events, so penalty estimates are the least reliable. Penalty events were not equivalent across games, which may explain their inconsistency. Events were self-paced and therefore overlapped in time, so responses to adjacent events may contaminate one another. Only events related with reward, penalty and movement were included in the study, other mechanics related with events from each game must be analysed to understand their contribution to the ERPs. Reward and penalty events were pooled without regard to the underlying game mechanic (e.g., colliding with objects, creating holes, or raising the stack height), which may mask mechanic-specific differences.

CONCLUSION

Reward, penalty, and movement events can be characterized by ERPs and decoded from single-event EEG across video games of varying complexity. Reward events elicited a fronto-central positivity consistently across games; penalty events elicited components in a comparable window; and movement events elicited a pre-movement negativity. All three binary classification tasks exceeded chance, with penalty versus movement the most separable and reward versus penalty the least. Accounting for individual differences through participant-specific analysis windows produced the largest improvement in accuracy, and the scarcity and heterogeneity of penalty events are the principal practical limitations.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number JP25K18952.

REFERENCES

Bergmann, M., Wollbrandt, I., Gittel, L., Halbe, E., Mackert, S., Philipsen, A., & Lux, S. (2023). Effects of a video game intervention on symptoms, training motivation, and visuo-spatial memory in depression. *Frontiers in Psychiatry*, *14*, 1173652. https://doi.org/10.3389/FPSYT.2023.1173652

Butler, O., Herr, K., Willmund, G., Gallinat, J., Kuhn, S., & Zimmermann, P. (2020). Trauma, treatment and Tetris: video gaming increases hippocampal volume in male patients with combat-related posttraumatic stress disorder. *Journal of Psychiatry & Neuroscience, 45*(4), 279–287. https://doi.org/10.1503/JPN.190027

Cavanagh, J. F., & Castellanos, J. (2016). Identification of canonical neural events during continuous gameplay of an 8-bit style video game. *NeuroImage*, *133*, 1–13. https://doi.org/10.1016/j.neuroimage.2016.02.075

Fleischmann, B., Mulready, S., & Ostwald, D. (2026). Video games in cognitive event-related EEG research: A scoping review [Preprint]. *OSF Preprints*. https://doi.org/10.31234/OSF.IO/5KTWB_V3

Klumpp, H., & Shankman, S. A. (2018). Using Event-Related Potentials and Startle to Evaluate Time Course in Anxiety and Depression. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *3*(1), 10–18. https://doi.org/10.1016/J.BPSC.2017.09.004

Li, M., Zhang, J., Jiang, C., Wang, J., Sun, R., Jin, S., Zhang, N., & Zhou, Z. (2023). The Neural Correlates of the Recognition of Emotional Intensity Deficits in Major Depression: An ERP Study. *Neuropsychiatric Disease and Treatment*, *19*, 117–131. https://doi.org/10.2147/NDT.S393264

Martínez-Tejada, L. A., Puertas-González, A., Yoshimura, N., & Koike, Y. (2021). Exploring EEG Characteristics to Identify Emotional Reactions under Videogame Scenarios. *Brain Sciences 2021, Vol. 11, Page 378*, *11*(3), 378. https://doi.org/10.3390/BRAINSCI11030378

Proudfit, G. H., Bress, J. N., Foti, D., Kujawa, A., & Klein, D. N. (2015). Depression and event-related potentials: emotional disengagement and reward insensitivity. *Current Opinion in Psychology*, *4*, 110–113. https://doi.org/10.1016/J.COPSYC.2014.12.018

Spüler, M., & Niethammer, C. (2015). Error-related potentials during continuous feedback: Using EEG to detect errors of different type and severity. *Frontiers in Human Neuroscience*, *9*(MAR), 123266. https://doi.org/10.3389/FNHUM.2015.00155

Sur, S., & Sinha, V. (2009). Event-related potential: An overview. *Industrial Psychiatry Journal*, *18*(1), 70. https://doi.org/10.4103/0972-6748.57865

Wimmer, M., Weidinger, N., Veas, E., & Müller-Putz, G. R. (2024). Multimodal decoding of error processing in a virtual reality flight simulation. *Scientific Reports 2024 14:1*, *14*(1), 9221-. https://doi.org/10.1038/s41598-024-59278-y

Xie, C., Xue, C., Li, Y., Liu, X., Zhong, D., Jin, Q., Li, J., & Jin, R. (2025). The characteristics of event-related potentials in generalized anxiety disorder: A systematic review and meta-analysis. *Journal of Psychiatric Research*, *181*, 470–483. https://doi.org/10.1016/J.JPSYCHIRES.2024.12.016

Zhang, D., He, Z., Chen, Y., & Wei, Z. (2016). Deficits of unconscious emotional processing in patients with major depression: An ERP study. *Journal of Affective Disorders*, *199*, 13–20. https://doi.org/10.1016/J.JAD.2016.03.056

Zhou, L., Wang, G., Nan, C., Wang, H., Liu, Z., & Bai, H. (2019). Abnormalities in P300 components in depression: an ERP-sLORETA study. *Nordic Journal of Psychiatry*, *73*(1), 1–8. https://doi.org/10.1080/08039488.2018.1478991