Exploring Movie Recall Prediction Using Functional Descriptors of the EEG Signal

Iván Martín-Fernández[1,*], Mario Lobo-Alonso[1], Sergio Esteban-Romero[1], Manuel Gil-Martín[1] and Fernando Fernández-Martínez[1]

[1]*Grupo de Tecnología del Habla y Aprendizaje Automático (THAU), Information Processing and Telecommunications Center, E.T.S.I. de Telecomunicación, Universidad Politécnica de Madrid (UPM), Madrid, 28040, Spain*

Abstract

We present the contribution of THAU Group to the MediaEval Predicting Movie and Commercial Memorability Challenge, focusing on Challenge 1.2: predicting subject-wise movie clip recall from EEG data. The goal is to identify brain processes linked to long-term recall in naturalistic settings and to test whether established neural markers can support predictive models. Focusing on five functional domains—visual, auditory, semantic, attentional, and emotional—we extract oscillatory features and train ElasticNet classifiers with Leave-One-Subject-Out validation. Semantic and visual descriptors prove most predictive (mean AUC = 0.556), though performance varies across individuals. Clustering by per-area AUCs reveals four subject profiles, and training only on data belonging to the subject being evaluated yields the best test result on the testing set (AUC = 0.656). These findings underscore strong subject dependency in EEG-based recall prediction and point toward adaptive, domain-specific models for long-term audiovisual memory.

1. Introduction and Related Work

Despite decades of research, our understanding of how the brain encodes and retrieves complex audiovisual experiences remains limited. Data-driven approaches and machine learning offer new opportunities, but models must remain interpretable to ensure that results connect back to cognitive and neural processes. EEG oscillations provide a natural entry point, as they reflect sensory, cognitive, and emotional functions that can be directly related to recall mechanisms.

EEG research has linked oscillatory dynamics to visual, auditory, semantic, emotional, and attentional processes [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]. Within MediaEval, Hamelink [11] examined the P300, a positive EEG deflection around 300 ms, relating Event Related Potential (ERP) amplitudes over occipital and temporal sites with memorability. Kleinlein *et al.* [12] used statistical summaries and coherence maps for classification. More recently, Nguyen [13] combined time- and time–frequency features with subject-specific encoding.

In this paper, we report the contribution of THAU Group to the MediaEval 2025 Predicting Movie and Commercial Memorability Challenge [14], focusing on Challenge 1.2: predicting subject-wise movie clip recall from EEG. We ask whether established EEG markers of brain function can inform this challenge of long-term recall in naturalistic conditions. To this end, we construct domain-specific feature sets for visual, auditory, semantic, emotional, and attentional processes, and evaluate them with ElasticNet logistic regression under Leave-One-Subject-Out Cross Validation (LOSO-CV). We also analyze subject variability through clustering, aiming to uncover cognitive profiles that account for differences in recall prediction. We refer to the overview paper for a detailed description of the task, data and evaluation process [14].

MediaEval'25: Multimedia Evaluation Workshop, October 25–26, 2025, Dublin, Ireland and Online
*Corresponding author.
ivan.martinf@upm.es (I. Martín-Fernández)
0009-0004-2769-9752 (I. Martín-Fernández); 0009-0009-8567-8526 (M. Lobo-Alonso); 0009-0008-6336-7877 (S. Esteban-Romero); 0000-0002-4285-6224 (M. Gil-Martín); 0000-0003-3877-0089 (F. Fernández-Martínez)

Table 1
Feature inventory by brain area. Each set corresponds to established EEG markers of perception, attention, or memory processes. ERD: Event Related Desynchronisation, P: Parietal, PO: Parieto-Occipital, O: Occipital, F: Frontal, T: Temporal., FP: Fronto-Parietal.

Area (# feat.)	Feature	Description
Visual ERD (1)	α + β power over P/PO/O [1]	Mean ERSP
Auditory (3)	θ power over T7/T8 [2] F3-T7 θ coherence [2]	Mean ERSP Mean weighted-Phase Lag Index (wPLI) coherence
Semantic (5)	α + θ power over CP [3]	Full ERSP power + diff. between early (0-2s) and late (3-5s) + percentage below baseline
Emotion (7)	Frontal Alpha Asymmetry [4] Frontal-Midline Theta [4] Temporal β [5] F3-T7, Fz-T7, Fz-T8, F4-T8 θ coherence [6]	Diff. between mean alpha ERSP power at F3 and F4 Mean ERSP power over θ band at Fz Mean ERSP power over β band at T7 and T8 Mean wPLI coherence
Attention (11)	Frontal Engagement [7] Fz-Pz, F3-P3, F4-P4 α + β coherence [8] FP Spectral Entropy [9] Individual Alpha Frequency [10]	Ratio $\frac{\beta}{\alpha+\theta}$ for Fz, F3, F4 Mean wPLI coherence Avg. $\alpha + \beta$ spectral entropy in the Fx+Px channels Argmax ERSP frequency over P3/P4/Pz channels
ERP baseline (4)	P300 ERP [11]	Peak amplitude (0.3-0.6s) in Oz,O1,O2 and P8.

2. Experiment Design

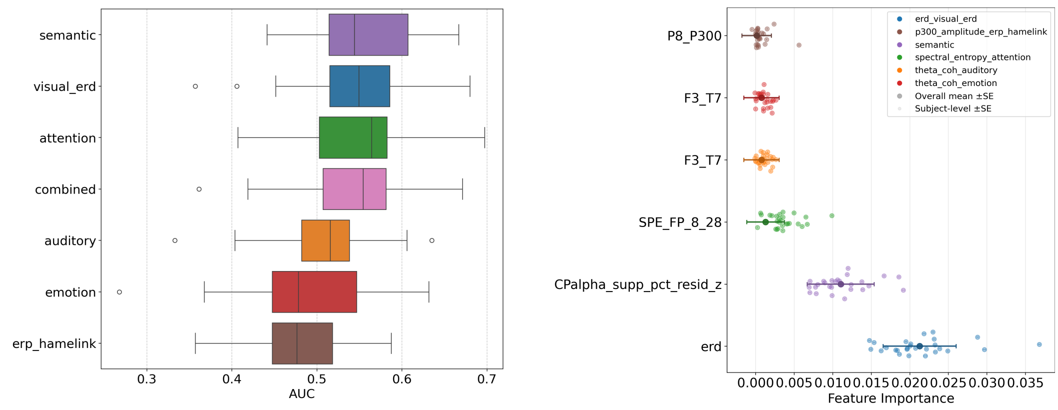
Our study investigates whether oscillations in specific brain functions correlate with movie recall performance. To this end, we curated feature sets corresponding to five functional domains: visual processing, auditory processing, semantic encoding, emotion, and attention. Each feature set was designed based on established EEG markers from the literature and was computed either directly from the provided epochs or from the pre-computed ERPs and Event-Related Spectral Potentials (ERSPs) of the dataset. These descriptors have previously been used to model or predict the corresponding brain functions, and here we examine whether they also relate to long-term memory for movie clips. For every epoch, a per-area feature vector is built from the statistics described in Table 1. These vectors are normalized across subjects and used for classification. As a baseline, we replicated the P300 effects reported by Hamelink [11], extracting peak amplitudes in occipital and parietal electrodes for direct comparison with prior work.

We employ a simple ElasticNet logistic regressor for classification and evaluate generalization through LOSO-CV, ensuring robustness across individuals. To interpret model behavior, we estimated permutation-based feature importance. Finally, we design a combined classifier by selecting the most relevant descriptor from each domain based on importance scores, constructing a compact multi-area representation. This strategy enables us to contrast the predictive capacity of domain-specific features with that of integrated models.

3. Results and Discussion

3.1. Area Performance Comparison

Figure 1a shows the per-subject distribution of AUC scores for each functional area. No single feature set emerges as a strong predictor of clip recall, although semantic features perform best on average (mean AUC = 0.556). This suggests that brain activations related to interpreting the



(a) Per-subject AUC distributions for each functional area.

(b) Permutation-based feature importance in the early fusion model.

Figure 1: Performance analysis by functional area and fused feature importance.

symbolic meaning of stimuli are most correlated with subsequent recall. However, performance varies widely across subjects: for instance, visual features reach an AUC of 0.680 for subject *cgrxsi*, but drop to 0.357 for subject *cnojgj*, underscoring the strong subject-dependent nature of the task and the difficulty of building models that generalize well.

The combined classifier yields a mean AUC of 0.548, slightly below the semantic, visual and attentive sets. Because the adequacy of each descriptor is highly individual, combining them without adaptation can degrade performance. Feature importance analysis of the fused model (Figure 1b) shows that visual cues dominate overall, followed by centro-parietal alpha percentage, consistent with the stronger performance of the semantic and visual domains.

3.2. Performance-Based Subject Clustering

To examine inter-individual variability across the 27 subjects, we applied hierarchical clustering with Ward linkage to per-subject AUCs across functional areas, selecting four clusters. Although their heterogeneity is limited (silhouette score ≈ 0.2), they can be interpreted qualitatively to extract prospective subject profiles. The **Attention+Semantic reliant group** (*adqftq, cnojgj, gqthjh, llavjb, okrfkk, rfexms*) showed relatively greater dependence on attentional and semantic cues, although overall performance was low. The **Visual specialists** cluster (*arqkyq, cgrxsi, cxvkvn, kqsutv, twbmgh, vvobdn, xigjja, ynaysa*) achieved the highest predictability from visual features (mean AUC = 0.60), suggesting recall anchored in perceptual cues. The **Balanced multimodal** group (*asljaa, ayabuh, gzomle, iyapba, lzkbut, moaxib, pnomiz, sgzais, snbuvk, umvakd*) achieves overall decent results, with no specific area highlight apart from ERP underperforming. Finally, the **Semantic specialists** (*oqkubg, tzeatc, unbzlb*) achieved the highest predictability (mean AUC = 0.63) when classifiers relied exclusively on semantic features. As shown in Figure 2, predictive power varies strongly across these groups, highlighting the subject-specific nature of the task and the challenge of generalizing across individuals.

3.3. Experiments on the test set

We prepare our test runs by applying the lessons learned in the previous sections, selecting the most relevant area features for each individual based on their assigned cluster (attention and semantic, visual, semantic, or all but ERP) and reporting late-fusion prediction values when

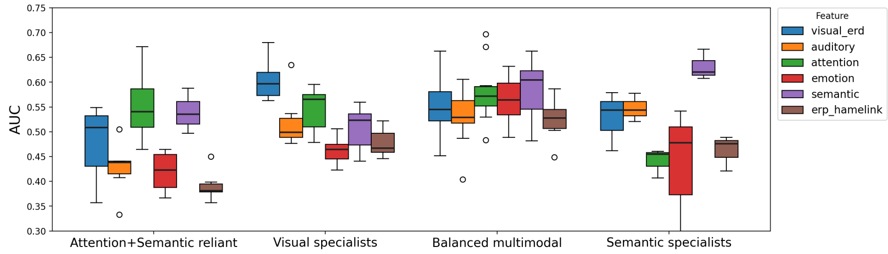


Figure 2: AUC scores across subject clusters. Distinct reliance profiles (visual, semantic, attention+semantic, and balanced multimodal) illustrate the subject-dependent nature of recall prediction.

Table 2
Official AUC results in the test portion of the challenge dataset.

	All Participants	Excluding Evaluated Subject	Only Evaluated Subject
AUC	0.570	0.530	0.656

needed. To assess subject dependency in the test set, we submit three model versions trained on different data: all available epochs, only the epochs belonging to the current subject, and all epochs except those belonging to the current subject. In the results shown in Table 2 it can be observed that training only with the data corresponding to the evaluated subject yields better results than training with all available data. This points out that EEG signals from other subjects may interfere with actual relevant information from the current individual under study, which poses the challenge of generalisable predictive systems.

4. Conclusions and Future Work

This paper has reported on the participation of THAU Group in Challenge 1.2 of the MediaEval Predicting Movie and Commercial Memorability Challenge. Results show that predicting long-term recall of movie clips from EEG is feasible but highly subject-dependent. Semantic and visual descriptors were the most consistently predictive, yet performance varied widely across individuals. Clustering revealed distinct reliance profiles—some subjects benefited from meaning-based cues, others from perceptual, attentional, or emotional markers—underscoring the need for adaptive approaches. Future work should exploit these differences through subject-aware or expert models and refine feature selection and model architectures. Overall, progress in this area depends on building systems that are both predictive and interpretable.

Acknowledgements

The research of Iván Martín-Fernández was supported by the Universidad Politécnica de Madrid (Programa Propio I+D+i). This work was funded by the Spanish Ministry of Science and Innovation through the project TRUSTBOOST (PID2023-150584OB-C21), funded by MCIN/AEI/10.13039/501100011033 and by the European Union "NextGenerationEU/PRTR".

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT in order to: Improve writing style, Grammar and spelling check. After using these tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] J. Sattelberger, H. Haque, J. J. Juvonen, F. Siebenhühner, J. M. Palva, S. Palva, Local and interareal alpha and low-beta band oscillation dynamics underlie the bilateral field advantage in visual working memory, Cerebral Cortex 34 (2024) bhae448. doi:10.1093/cercor/bhae448.

[2] Y. Xie, Y. Li, H. Duan, X. Xu, W. Zhang, P. Fang, Theta oscillations and source connectivity during complex audiovisual object encoding in working memory, Frontiers in Human Neuroscience Volume 15 - 2021 (2021). URL: https://www.frontiersin.org/journals/human-neuroscience/articles/10.3389/fnhum.2021.614950.

[3] P. A. Packard, T. K. Steiger, L. Fuentemilla, N. Bunzeck, Neural oscillations and event-related potentials reveal how semantic congruence drives long-term memory in both young and older humans, Scientific Reports 10 (2020) 9116.

[4] G. Zhao, Y. Zhang, Y. Ge, Frontal eeg asymmetry and middle line power difference in discrete emotions, Frontiers in Behavioral Neuroscience Volume 12 - 2018 (2018). URL: https://www.frontiersin.org/journals/behavioral-neuroscience/articles/10.3389/fnbeh.2018.00225. doi:10.3389/fnbeh.2018.00225.

[5] P. Mesa-Gresa, J.-A. Gil-Gómez, J. A. Lozano-Quilis, K. Schoeps, I. Montoya-Castilla, Electrophysiological correlates of the emotional response on brain activity in adolescents, Biomedical Signal Processing and Control 89 (2024) 105754. URL: https://www.sciencedirect.com/science/article/pii/S1746809423011874.

[6] B. Güntekin, E. Başar, A review of brain oscillations in perception of faces and emotional pictures, Neuropsychologia 58 (2014) 33–51. URL: https://www.sciencedirect.com/science/article/pii/S0028393214000992.

[7] B. Raufi, L. Longo, An evaluation of the eeg alpha-to-theta and theta-to-alpha band ratios as indexes of mental workload, Frontiers in Neuroinformatics Volume 16 - 2022 (2022). URL: https://www.frontiersin.org/journals/neuroinformatics/articles/10.3389/fninf.2022.861967.

[8] N. Liegel, D. Schneider, E. Wascher, S. Arnau, Task prioritization modulates alpha, theta and beta eeg dynamics reflecting proactive cognitive control, Scientific reports 12 (2022) 15072.

[9] Y. Tian, H. Zhang, W. Xu, H. Zhang, L. Yang, S. Zheng, Y. Shi, Spectral entropy can predict changes of working memory performance reduced by short-time training in the delayed-match-to-sample task, Frontiers in Human Neuroscience Volume 11 - 2017 (2017). URL: https://www.frontiersin.org/journals/human-neuroscience/articles/10.3389/fnhum.2017.00437.

[10] Y. Zhang, Y. Lu, D. Wang, C. Zhou, C. Xu, Relationship between individual alpha peak frequency and attentional performance in a multiple object tracking task among ice-hockey players, Plos one 16 (2021) e0251443.

[11] R. Hamelink, Visual recognition of the memento10k dataset: A p300 component erp study, in: Working Notes Proceedings of the MediaEval 2022 Workshop, volume 3583 of *CEUR Workshop Proceedings*, CEUR-WS.org, Bergen, Norway, 2023. URL: https://ceur-ws.org/Vol-3583/paper20.pdf.

[12] R. Kleinlein, E. Rodríguez-Sebastián, F. Fernández-Martínez, Understanding media memorability from event-related potential records and visual semantics, in: Working Notes Proceedings of the MediaEval 2022 Workshop, volume 3583 of *CEUR Workshop Proceedings*, CEUR-WS.org, Bergen, Norway, 2023. URL: https://ceur-ws.org/Vol-3583/paper35.pdf.

[13] M.-Q. Nguyen, M.-H. Trinh, H.-G. Bui, K.-T. Vo, M.-T. Tran, T.-P. Tran, H.-D. Nguyen, SELAB-HCMUS at MediaEval 2023: A cross-domain and subject-centric approach towards the memorability prediction task., in: Working Notes Proceedings of the MediaEval 2023 Workshop, volume 3658 of *CEUR Workshop Proceedings*, CEUR-WS.org, Amsterdam, The Netherlands, 2024. URL: https://ceur-ws.org/Vol-3658/paper29.pdf.

[14] I. Martín-Fernández, M. G. Constantin, C.-H. Demarty, M. Gil-Martín, S. Halder, B. Ionescu, A. Matran-Fernandez, R. Savran Kiziltepe, A. García Seco de Herrera, Overview of the mediaeval 2025 predicting movie and commercial memorability task, in: Proc. of the MediaEval 2025 Workshop, Dublin, Ireland and Online, 2025.