

# Cinematic Memories: EEG-Based Decoding of Recall Using Naturalistic Stimuli

Ana Matran-Fernandez and Sebastian Halder

**Abstract**—Identifying neural correlates of long-term memory retrieval has been an area of considerable research. Previous studies used custom-built, static stimuli to elicit and study memory processes, limiting the insights into brain activity in a natural environment. Instead, we used movie clips to identify neural correlates of memory retrieval. We extracted features from the time-frequency domain from electroencephalography (EEG) signals to classify whether a 10-second-long video clip had been remembered, obtaining greater-than-chance memory decoding (median area under the receiver operating characteristic curve = 0.64) in a naturalistic movie-watching task with greater real-world validity than word-list or image recall paradigms. Feature analysis revealed that remembered clips, compared to not recognised clips, showed increased event-related desynchronisation (ERD) in the theta and low alpha bands early after stimulus onset, followed by stronger ERD in the alpha bands. While beta-band differences were observed at the group level, high within-class variability precluded their use for reliable single-trial classification, with theta and alpha activity emerging as the primary discriminative features. Our ERD/S analysis generalises observations previously demonstrated in the literature using simpler stimuli and more controlled experimental designs. Finally, we observed a significant correlation between classifier performance and retrospective memory ratings of participants ( $\rho = 0.49$ ) demonstrating that individual differences in memory ability are reflected in neural activity patterns. This work demonstrates the potential of naturalistic stimuli to advance memory research using EEG, revealing neural dynamics that may not be detectable with simpler stimuli, with possible implications for the development of clinical tools in memory assessment.

**Index Terms**—EEG, recall, memory retrieval, long-term memory, neural decoding, episodic memory, machine learning, FBCSP, feature importance, interpretability.

## I. INTRODUCTION

**M**EMORY retrieval is a fundamental cognitive process that plays a critical role in human experience, influencing learning, decision-making, and emotional regulation [1]. Understanding the neural correlates of memory recall has significant implications for numerous fields, including education, health care, and generally through the creation of affective technologies that are more aware of human experience. Advances in neuroimaging techniques and brain-computer interface (BCI) technologies have facilitated the exploration of neural data, particularly electroencephalography (EEG), in decoding cognitive states associated with memory tasks [2].

Previous studies have used EEG to reveal neural signatures associated with memory encoding and retrieval [3], [4], [5].

The authors are with the Essex Brain-Computer Interfaces and Neural Engineering Lab, University of Essex, Colchester, CO4 3SQ, UK (email: amatra@essex.ac.uk; s.halder@essex.ac.uk).

Manuscript received August 12, 2025; revised AA.

EEG correlates of memory include oscillatory activity (in particular in the theta and alpha bands), event-related potentials, and connectivity [6], [7], [8]. Most of these investigations have focused on simple stimuli, often relying on static images or word lists [9], [10], [11], [12], [13] that participants are asked to memorise. However, most memories are multimodal in nature and often evoke strong emotional responses and can be vividly recalled even after significant time has passed. Consequently, there is increasing evidence that naturalistic and dynamic stimuli such as videos may be more ecologically valid [14], [15], [16].

Machine learning can be used to advance our understanding of cognitive processes, such as memory mechanisms [17], [18], [19]. While advanced machine learning models have shown promise in decoding neural signals, their “black box” nature often limits interpretability [20], [21].

Our research departs from traditional work on memory by exploiting effective BCI pipelines to analyse EEG data collected during the viewing of previously watched movie fragments to decode memory retrieval of complex, time-varying stimuli. By employing dynamic, multimodal content, we seek to capture the richness and complexity of real-world memory retrieval in a more naturalistic setting. This approach not only enhances the ecological validity of our research but also allows us to explore how the brain processes and remembers continuous, narrative-driven information [19].

Moreover, departing traditional memory paradigms which begin with an encoding phase in which participants get familiar with the stimuli they will later be tested on, we focus only on the recall of memories that had been formed organically during the participants’ lives. Our research uses a public dataset of EEG signals from participants watching clips from movies they reported having previously watched (together with clips from movies they had not watched) to study the neural correlates of memory recall [22].

A key aspect of our approach is our focus on explainable methods. By emphasizing explainable techniques in our analyses, we aim to provide insights into the specific EEG features and patterns associated with memory recall when using dynamic stimuli, enhancing our understanding of the underlying neural mechanisms and comparing them with those from traditional memory research.

In the following sections, we will provide an overview of prior efforts to decode memory processes using EEG (Section II), detail our methodology (Section III), present our findings (Section IV), and discuss the implications of our research, offering suggestions for possible improvements

and avenues for future work in Sections V and VI. By combining the engaging nature of cinematic experiences with EEG analysis techniques and explainable AI, we aim to shed new light on the processes underpinning memory retrieval when audiovisual stimuli are used.

## II. RELATED WORK

In neurophysiological terms, memory is described as a physical, integrated system that performs encoding, storage, and retrieval of information [23], [24], [25], [26]. The relationship between EEG oscillations and memory has been extensively studied, with oscillations in the theta and alpha bands being primarily linked to memory processes [27], [28], [8], [29], [30], together with activity in the beta and gamma bands [31], [32], [33].

### A. Neural correlates of memory

Event-related synchronisation (ERS) in the theta band (4–8 Hz) has been associated to memory performance during retrieval [34], particularly in episodic memory [35], [28], [36], [37]. This ERS is believed to reflect sustained attention to the processing of new information and its maintenance in short-term memory [29], [30]. Theta is also believed to hold a critical role in memory retrieval, with power changes in frontal areas being related to the access of stored information in the brain [38], [34].

Alpha-band oscillations during memory tasks comprise at least two distinct sub-bands with different characteristics. The lower alpha band (8–10 Hz) might be interpreted as an index of attention to the stimulus being presented, while activity in the upper alpha band (10–12 Hz) represents long-term memory (LTM) retrieval [36], [39], [30]. Brief event-related desynchronisation (ERD) in lower alpha following stimulus presentation is a sign of allocation of processing resources [40]. Significant lower and upper alpha ERD in parietal and occipital regions has been observed during LTM retrieval [35], particularly on the left brain hemisphere [41].

Memory performance has been linked to variations in neural activity across participants. For instance, Hanouneh *et al.* found that power in the alpha and beta frequency bands were positively correlated with performance in frontal brain areas, but negatively correlated in parietal and temporal regions. In contrast, absolute power in the frontal delta and theta bands showed a negative correlation with performance [31]. Additionally, better memory performance has been associated with increased theta activity in the left parietal lobe [42]. Furthermore, the individual alpha frequency has been proposed as a marker for distinguishing between “good” and “poor” memory performers [40].

### B. Static vs. naturalistic stimuli

Most research on memory has relied on stationary stimuli, such as lists of words or images [9], [10], [11], [12], [13]. However, dynamic stimuli are more likely to be recognised in a recall test than static images [43], particularly in the long-term [12].

Although memory works as an integrated system, activation of its subcomponents give varying patterns, reflecting the brain regions that were engaged during encoding [44], [45], [46], which depend on the modality of the stimulus used. Since most natural memories are multimodal, it remains to be seen whether the knowledge acquired with such simple stimuli generalises to a more naturalistic scenario.

In a significant departure from traditional memory paradigms, Park and Donaldson used mobile EEG to allow participants to follow a predefined route while memorising locations with objects [47], and found that the elicited EEG patterns were similar to those observed in a laboratory setting. Desai *et al.* also found that models of EEG encoding trained with naturalistic audiovisual stimuli generalise to traditional, highly-controlled audio stimuli, but that the opposite is not true [48].

Naturalistic videos were also used in a learning experiment in which students saw educational content and were then tested on the material through multiple choice questions [49], [31], finding correlations between power and functional connectivity (at the time of encoding) with memory recall.

These works show that it is possible to study memory using complex, time-varying stimuli. The use of naturalistic stimuli, such as videos, stories, or real-life scenarios, to study memory processes in EEG research offers several advantages over simpler stimuli, as these engage memory processes in a more ecologically valid manner, closely resembling the types of stimuli we encounter in daily life [14], [15], [16]. In turn, this can lead to a more comprehensive understanding of how the brain processes and retains complex information. Additionally, naturalistic stimuli tend to elicit richer and more dynamic brain responses, capturing a broader range of cognitive and emotional processes that are crucial for memory. In contrast, simple stimuli often fail to replicate the complexity and variability of everyday experiences, potentially limiting the external validity of the findings.

### C. Memory classification from EEG

Recent memory research has shifted from analysing averaged neural activity across stimulus categories to classifying individual trials. This new approach aims to predict, for each stimulus presentation, whether an item will be remembered.

The following works typically rely on traditional BCI pipelines [50] which extract features prior to feeding them to a machine learning classifier, but there have also been some attempts using deep learning for this task.

1) *Encoding phase:* A key question in memory classification is whether it is possible to predict, at the time of encoding, whether something will be later remembered. When the stimuli used are static, information from the event-related potentials (ERPs) evoked at the time of presentation can be used to predict subsequently remembered vs. forgotten trials. This was the approach used to predict short-term memory of written words by Chakravarty *et al.* [51]. More advanced techniques additionally rely on features extracted from the frequency domain. Sun *et al.* employed a convolutional neural network to discriminate between trials using theta band

power from pre- and post-stimulus periods of spoken word presentations [19]. By using ERPs combined with features extracted from filter-bank common spatial patterns (FBCSP), Noh *et al.* were able to distinguish between recollection and familiarity using information extracted from the alpha band [52]. In an attempt to create a systematic approach to memory encoding classification, Mirajalili *et al.* compared different combinations of features (statistical, entropy-based, phase-based, and CSP), feature selection methods (wrappers and filters), and classifiers (e.g., Lasso, support vector machines, logistic regression) across several datasets, concluding that a combination of features is better than just features from one type [53]. Rudoler *et al.* tried to detect optimal states for learning in a cross-session study in order to optimise the presentation of information to participants for optimising memory, using power spectral features from frequencies between 6–180 Hz [5]. Shin *et al.* extracted power features from a short-term recall task to identify, using deep learning, which of the items remembered in short-term memory would make it into long-term memory (24 hours after the first recall test) [54]. Finally, Matran-Fernandez and Halder used explainable deep learning to predict whether a participant would remember a short clip 1–3 days after initial encoding [55]. However, their manuscript lacks a description of the features that were learnt by their neural network.

Performance across all these studies is fairly uniform, with models not being able to perform above random chance level for a number of participants in each case (e.g., 5/18 in [52]) and with modest performance across the board. Due to the class imbalance, most studies report the area under the receiver operating characteristic curve (AUC) when evaluating classifier performance. Chakravarty *et al.* reported 95% confidence intervals for the AUC in the range of 0.51–0.55 [51]; while Rudoler and colleagues achieved an average AUC of 0.53 across participants [5]; Matran-Fernandez and Halder showed an average AUC of  $0.51 \pm 0.12$  for subject-dependent prediction, and 0.67 for a multi-participant approach [55]. Shin *et al.*'s highest reported Cohen's kappa value was 0.24 [54].

2) *Retrieval phase:* There have also been attempts to determine, at recall time, whether something that is being presented is remembered or not. To the best of our knowledge, this has only been done using ERPs elicited during the test phase [56], [9]. In an attempt to find generalisable ERP features, Liao *et al.* tested their methods in a leave-one-subject-out fashion, reaching their highest performance (AUC of 0.66) for the classification of trials equivalent to the remember vs. know paradigm [56], [57].

As in the case of classification using the EEG data from the encoding phase, classifier performance in these works is relatively low (e.g., Noh *et al.* reported an average accuracy of 61% in [9], but the level of class imbalance in this work was not reported), highlighting the difficulty of this task.

All the studies reviewed in Section II-C, with the exception of [55], used static stimuli either as study materials or at test time (and most of the time at both), which offers researchers the advantage of having a clear time to study ERPs time-locked to the appearance of the stimulus. Still, in many cases, features were extracted from the frequency domain.

When audiovisual stimuli are used to study memory, as is the case in our work, there is no clear trigger or time point at which memory encoding or retrieval can be studied. A solution to this problem could be to edit the video to trigger ERPs (e.g., by introducing luminance changes) to analyse differences in ERPs for recalled vs. not recalled frames from the videos [58], but this would not allow for a fully naturalistic watching experience. On the other hand, scene changes are known to induce ERPs, and these could be compared across different conditions [59], [60], [61] when a recall test is administered after the initial watching/encoding while keeping the experience intact, although this may pose limits on the material that can be studied. Finally, a third option would be to emulate an asynchronous system, similarly to self-paced BCIs [62], which tries to detect memory encoding and/or recall in real-time using sliding windows over time.

### III. MATERIALS AND METHODS

#### A. Dataset

The data analysed here are from the Essex EEG Movie Memory Dataset (EEMMD), a public dataset which is fully described in [22].

Briefly, the dataset contains raw EEG signals from 27 participants while they watched 10-second long clips from movies. The clips had previously been independently annotated in terms of their probability of being remembered by someone who had watched the movie [63]. Participants filled a survey to report which movies they had watched so that a subject-specific list of clips could be generated to balance clips that were likely to be reported as “remembered” and those which were not—either because the participant had not watched the movie or because they had low memorability scores.

The experimental protocol was as follows: chosen clips were shown in randomised order. After the clip, the participant was asked if they recognised the movie it belonged to. If they did not, the clip was labelled as “not recognised”. If they recognised the clip, they were asked whether they remembered watching it previously. If the answer was positive, the clip was labelled as “remembered”. Clips which were recognised but not remembered were labelled as “recognised”.

Clips that were either recognised or remembered were shown a second time to participants, who were then asked to mark the time points that caused them to recognise/remember the clip.

#### B. Data preprocessing

EEG data from the EEMMD were collected using a 64-channel BioSemi ActiveTwo recording system at a sampling rate of 2048 Hz. We first downsampled the EEG recordings to 512 Hz before running the PREP pipeline [64] to reference the data. The data were then band-pass filtered between 1–30 Hz with a zero-phase Butterworth filter using MNE [65].

We performed artifact correction on epochs extracted starting one second before the beginning of each clip and lasting 5 seconds. We used Autoreject [66] on these epochs prior to fitting an independent component analysis estimator using the epochs marked as bad only. ICLabel [67] was used to remove

eye-related components from the continuous EEG recordings. Results from ICLabel were manually confirmed for a random selection of participants. All data were visually checked before and after processing.

We then extracted epochs time-locked to the start of each clip, beginning 1 second before the clip started and ending 10 seconds after the start of the clip (so each epoch includes the neural signals corresponding to the viewing of that clip). Autoreject was used again to remove bad epochs and interpolate bad channels. On average, 4.5% of epochs were rejected across participants.

Finally, epochs were downsampled to 128 Hz and labelled as “not recognised”, “recognised” or “remembered” based on the answers provided by the participants during the experiment.

### C. Memory questionnaire

Participants of the EEMMD answered a subset of items from the Prospective and Retrospective Memory Questionnaire (PRMQ) [68], a standard memory scale with several subscales, where each question (which asks how often a person makes specific types of memory errors) is answered using a 5-point scale (1 = never; 5 = very often). Since we are interested in retrospective memory, we averaged the answers to items 2, 4, 6, 8, 9, and 15 to calculate a retrospective memory score [68] for each participant (Cronbach’s  $\alpha = 0.73$ ).

### D. Time-Frequency Analysis

Time-frequency representations (TFRs) were computed for the training set epochs (80% of data per participant) using the multitaper method implemented in MNE-Python [65]. TFRs were calculated for frequencies between 3 and 26 Hz. ERD/S was quantified as the percentage change in power relative to a pre-stimulus baseline (−1.0 to 0 s). ERD/S maps were computed and averaged for each participant individually and then averaged over the group for visualisation.

### E. Statistical analysis

To evaluate differences between “remembered” and “not recognised” conditions, we performed a non-parametric cluster-based permutation test on the ERD/S differences of the individual participants. Spatio-temporal-spectral clusters were identified using a one-sample t-test (clustering across sensors, time, and frequency) with a cluster-forming threshold corresponding to  $p < 0.001$ . Statistical significance was determined using 1,000 permutations to control the family-wise error (FWE) rate at  $\alpha = 0.05$ . Significant clusters are indicated by black contours on the time-frequency plots. For each ROI, clusters that were significant for any individual channel of that ROI were marked as significant.

### F. FBCSP classification

We followed a standard pipeline for filter-bank CSP [69] followed by a logistic regression classifier with *ElasticNet* penalty (i.e., including both  $l1$  and  $l2$  penalty terms) [70] to classify epochs into two classes (“not recognised” vs.

“remembered”). Epochs were baselined to the 250 ms interval before the start of the clip.

The data for each participant were filtered in the theta (4–8 Hz), low alpha (8–10 Hz), high alpha (10–12 Hz), and beta (12–25 Hz) frequency bands before extracting 2-second-long windows centred at 1, 2, 3, 4, 5, 6, and 7 s from the start of the clip. We trained 2 CSP filters for each combination of time window and frequency band. CSP filters are subject-specific spatial filters which maximise the difference in variance between the classes in the data. This reduces the number of features (in our case, we used two filters, which reduces the temporal dimension from 64 channels to 2) and improves the signal-to-noise ratio by emphasising the oscillatory components which differ between the two classes [71]. Furthermore, the inverse of the filters (the so-called patterns) can be interpreted as a projection of the brain sources to the scalp. It is important to note that CSP patterns represent discriminative spatial filters that maximize variance differences between classes, rather than direct measures of absolute neural activity amplitude. Regions with high pattern weights indicate where activity patterns most reliably differ between conditions, which may not correspond to regions with highest absolute power [72].

The features were concatenated (4 frequency bands  $\times$  7 time windows  $\times$  2 filters = 56 features) and used to train the regularised logistic regression models.

For each participant, we divided the epochs available into training and test sets using a stratified 80:20 split. Table I shows the number of epochs available for each participant after preprocessing and epoch rejection. For visualisation, throughout this paper only data from the participants’ training sets were used. We performed stratified nested 5-fold cross-validation on the training set to identify the best value of regularisation ratio (values between 0–1, with 0 =  $l2$  penalty only and 1 =  $l1$  penalty only; values in the middle represent a combination of both). We then re-trained the model using the best value found from the cross-validation loop using the full training set. We report the results from stratified nested 5-fold cross-validation and on the final test set using the AUC.

All hyperparameters and analysis choices were determined independently of the test data. Frequency bands and time windows were selected a priori based on established literature on memory processes and not based on our data. The number of CSP filters ( $n=2$ ) was chosen to minimize feature dimensionality. Regularization hyperparameters ( $l1/l2$  ratios for *ElasticNet*) were optimized exclusively on the training set using stratified nested 5-fold cross-validation, with no access to test data.

### G. Feature analysis

Since our classifiers incorporate  $l1$  regularization, feature selection is inherently embedded within the models. However, the amplitudes of the coefficients can not be interpreted as coefficient importance, even when the features are normalised [73]. Instead, we assessed feature importance by analysing which features are more likely to be excluded from the classification task vs. which ones are selected more

TABLE I

DATASET DESCRIPTION. NUMBERS OF MOVIES WATCHED BY EACH PARTICIPANT ACCORDING TO THEIR ANSWERS TO THE SURVEY, AS WELL AS NUMBERS OF NOT RECOGNISED (NR), RECOGNISED (REC), AND REMEMBERED (REM) EPOCHS FOR EACH PARTICIPANT, AND THE TOTAL EPOCHS. ALL EPOCH-RELATED VALUES REPRESENT NUMBERS AFTER PREPROCESSING AND EPOCH REJECTION.

| Participant | Movies watched | NR / Rec / Rem epochs | Total epochs (NR + Rem) |
|-------------|----------------|-----------------------|-------------------------|
| 1           | 28             | 84 / 20 / 48          | 152                     |
| 2           | 15             | 85 / 7 / 19           | 111                     |
| 3           | 33             | 100 / 19 / 12         | 131                     |
| 4           | 35             | 66 / 25 / 39          | 130                     |
| 5           | 33             | 88 / 25 / 31          | 144                     |
| 6           | 32             | 74 / 22 / 40          | 136                     |
| 7           | 61             | 70 / 4 / 101          | 175                     |
| 8           | 59             | 87 / 2 / 80           | 169                     |
| 9           | 54             | 76 / 19 / 75          | 170                     |
| 10          | 65             | 106 / 37 / 37         | 180                     |
| 11          | 32             | 84 / 27 / 31          | 142                     |
| 12          | 70             | 45 / 0 / 81           | 126                     |
| 13          | 51             | 61 / 12 / 78          | 151                     |
| 14          | 38             | 97 / 13 / 26          | 136                     |
| 15          | 55             | 75 / 55 / 28          | 158                     |
| 16          | 37             | 65 / 3 / 66           | 134                     |
| 17          | 47             | 93 / 21 / 34          | 148                     |
| 18          | 54             | 86 / 26 / 42          | 154                     |
| 19          | 34             | 51 / 0 / 62           | 113                     |
| 20          | 66             | 81 / 3 / 78           | 162                     |
| 21          | 34             | 75 / 19 / 32          | 126                     |
| 22          | 68             | 79 / 24 / 56          | 159                     |
| 23          | 73             | 68 / 37 / 58          | 163                     |
| 24          | 45             | 81 / 33 / 35          | 149                     |
| 25          | 72             | 71 / 35 / 58          | 164                     |
| 26          | 40             | 78 / 17 / 39          | 134                     |
| 27          | 55             | 96 / 3 / 76           | 175                     |
| Average     | 47.6           | 78.6 / 18.8 / 50.4    | 147.9                   |

frequently [74], [75]. Moreover, the reliability of the learned patterns depends on the quality of the weights, which tend to be suboptimal when performance is low [73], [18], so including data from poorly performing participants would not provide meaningful insights into feature selection. Therefore, we only include in this analysis the data from participants whose test AUC scores exceeded random-chance performance.

It is important to note that these feature importances are conditional on our analytical choices: the FBCSP + *ElasticNet* architecture, pre-defined frequency bands (theta, alpha subbands, beta), and 2-second time windows. Different feature representations or extraction methods might reveal different patterns of importance.

#### IV. RESULTS

##### A. Time-frequency analysis

For illustration purposes, we have chosen to summarise results by dividing the scalp into several regions of interest (ROIs). These are: left frontal (Fp1, AF7, AF3, F7, F3, F5), midline frontal (Fpz, AFz, F1, Fz, F2), right frontal (Fp2, AF8, AF4, F4, F6, F8), left central (FC3, FC5, C3, C5, CP3, CP5), midline central (FC1, FCz, FC2, C1, Cz, C2, CP1, CPz, CP2), right central (FC4, FC6, C4, C6, CP4, CP6), left parietal (P3, P5, P7, PO3, PO7, O1), midline parietal (Pz, P1, P2, POz, Oz), and right parietal (P4, P6, P8, PO4, PO8, O2).

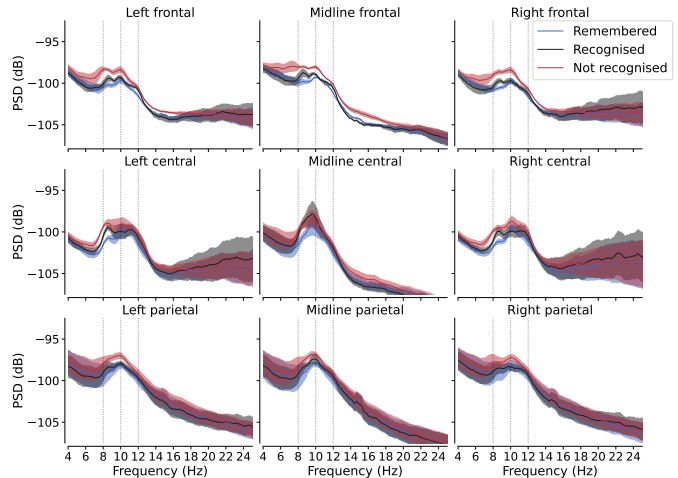


Fig. 1. Power spectral density for different brain regions (see Section IV-A for a list of electrodes included in each region) and types of trials, focusing on the theta and alpha bands. Vertical dashed lines highlight the boundaries between the theta (4–8 Hz), low alpha (8–10 Hz), and high alpha (10–12 Hz) bands. Only trials from the participants’ training sets were used.

We first computed the power spectral density (PSD) on our nine ROIs using Welch’s method [76] on epochs starting 500 ms after the start of the clip and ending at 9.5 s to avoid border effects at  $t = 0$  and  $t = 10$  s (start and end of clip, respectively). PSDs were computed for each electrode and then averaged across the electrodes in each ROI. The PSDs for different ROIs are shown in Fig. 1 using only the participants’ training sets. We can observe differences across the three conditions, which are most prominent in the frontal regions and comprise the theta and alpha bands. Power in these bands is highest for non-recognised clips and lowest for remembered clips, with values for recognised clips lying in between. There are no differences in the beta band across the different conditions in any of the regions, except for the low beta band in the midline frontal ROI (up to 18 Hz). The beta band also showed notably larger within-class standard deviations compared to theta and alpha bands, indicating high trial-to-trial variability (see confidence intervals in Fig. 1).

Fig. 2 shows the average of the ERD/S differences of the individual participants between remembered minus not recognised clips across the ROIs for epochs from the participants’ training sets. Significant clusters are indicated by black contours. Focusing first on the lower bands, significant ERD was observed in the early theta band primarily in parietal and right central ROIs. Further we observed significant alpha ERD in right and left central and parietal ROIs and a consistent ERD in the low beta band particularly on frontal ROIs.

With respect to beta, we find bursts of alternating signs in the left and right central ROIs. The low beta frequencies predominantly indicate higher ERD for remembered than for not recognised trials throughout the full length of the trial, and this difference is most prominent in parietal and frontal ROIs but was not significant.

Based on these results and prior literature, we chose these bands (theta, the alpha subbands, and beta) to train our FBCSP classifiers for each participant. Due to the low number of

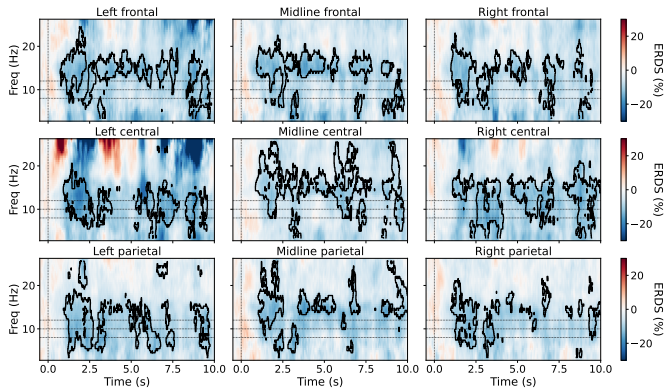


Fig. 2. Time-frequency representations of the average individual differences in ERD/S between remembered and not recognised epochs across several ROIs covering the full scalp. Only epochs from the training set were used. Each subplot represents the difference of average remembered minus average not recognised ERD/S for a region of interest. Positive—red—values indicate higher ERS for remembered than for not recognised trials; negative—blue—values indicate higher ERD for remembered than for not recognised trials. Black contours indicate areas of significant differences between conditions (FWE  $\alpha = 0.05$ ). The x-axis represents time and the y-axis maps frequency. The vertical dashed line represents the start of the trial at  $t = 0$ . Horizontal dashed lines show the limits between theta, the alpha sub-bands, and the beginning of the beta band, at 8, 10, and 12 Hz, respectively.

epochs labelled as “recognised” (18.8 per participant, on average), we trained our classifiers to discriminate between not recognised and remembered trials only.

### B. FBCSP classification

The classification results are shown in Fig. 3, which includes the AUCs from 5-fold cross-validation (diamonds) and on the test set (crosses) for each participant. The median AUC on the test set was 0.64 (95% CI = [0.55, 0.69]). A Wilcoxon test confirmed that performance on the test set is significantly above chance ( $W = 30.5$ ,  $p = 4 \times 10^{-4}$ ), with 70% of our participants reaching better-than-chance performance. Bayesian analysis with a range of priors (see Appendix A for details) further showed posterior mean estimates for the proportion of participants performing above chance to range from 62 to 70% (i.e., 16–19 participants for our sample size).

### C. Feature analysis from FBCSP classification

As mentioned in Section III-G, we only include in our feature analysis the models from participants whose test AUC scores exceeded random-chance performance (19/27 participants). Across these, our classifiers discarded an average of  $13.2 \pm 8$  features through  $l_1$  regularisation. Fig. 4 shows the probability that a specific CSP filter was selected. Features in the theta band were the most likely to be selected (average probability = 0.86), followed by the low and upper alpha sub-bands (average probabilities = 0.81 and 0.80, respectively). On the other hand, the features that were most likely to be discarded corresponded to CSP filters in the beta band (average probability of a feature in the beta band being selected = 0.58)—matching our PSD observations from Fig. 1, where no differences were observed between classes for this frequency band.

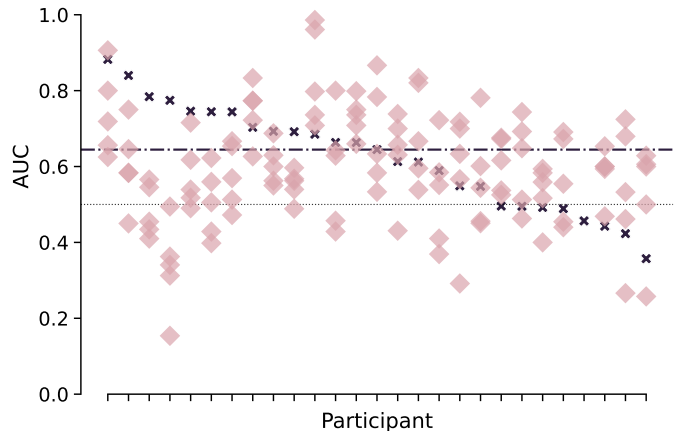


Fig. 3. AUC scores from FBCSP for each participant obtained through 5-fold cross-validation (diamonds) and for the test set (dark crosses). The dashed line represents the median test AUC across all participants (AUC = 0.64). The dotted line represents chance-level performance (AUC = 0.5). Participants are sorted in decreasing value of their test AUC score.

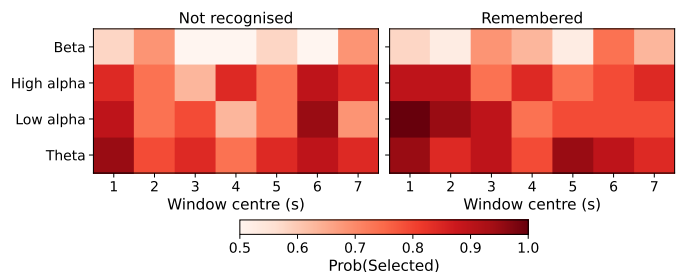


Fig. 4. Analysis of feature importances across all participants with better than random performance in the test set (19 out of 27 participants) in the form of probability that a given coefficient will be important for classification and therefore selected (due to  $l_1$  regularisation, features with small coefficients will be excluded). The plot on the left represents the probabilities for coefficients from the not recognised CSP patterns, and the one on the right corresponds to the CSP patterns for the remembered class. Darker colours indicate higher probability of a feature being kept.

With respect to time windows, no clear pattern was distinguishable. Features from the beginning and end of the trial were most likely to be selected. In particular, the first window (0–2 s) had the highest probability of being selected (average probability = 0.84), consistent with the findings by the EEMMD authors, who reported median recognition times of 1.5 and 1.7 s for different types of clips [22]. This window was followed by windows centred at 6, 2, and 7 s, respectively. The window less likely to be selected was 3–5 s, but the probability was still relatively high at 0.7.

To explore the influence of beta-band activity on classification we ran an ablation study in which we replicated our pipeline removing all beta-related features. The cross-validation AUC did not change significantly with respect to the version that included beta features according to a two-sided Wilcoxon test (median  $AUC_{with\_beta} = 0.58$ , median  $AUC_{without\_beta} = 0.59$ ;  $W = 187$ ,  $p = 0.97$ ).

Finally, in Fig. 5 we represent the patterns of the FBCSP classification model for a single participant, chosen due to

their test AUC being close to the median test AUC across all participants (cross-validation AUC=0.58, test AUC=0.61). These CSP patterns identify the spatial configurations of neural activity that are most discriminative for classification and should not be directly compared to the PSD results in Fig. 1, which show absolute power averaged across participants. CSP patterns reflect where differences between conditions contribute most to classification rather than where absolute neural activity is highest.

Focusing on the top 5 features according to Fig. 4, we will look at the patterns in brain regions that show the strongest differences between both conditions in low alpha (0–2 s, 1–3 s, and 5–7 s) and theta (0–2 s and 4–6 s). The first time window (0–2 s) shows frontal activation in both low alpha and theta for the remembered class (incidentally, this pattern also appears in the high alpha band). In contrast, in theta we see parieto-occipital activation for the not recognised class; in low alpha activity appears in the periphery. In later time windows, in low alpha we can see that the patterns become lateralised at 1–3 s, before returning to a frontal vs. parieto-occipital pattern at 5–7 s. In the theta band we see that the patterns between both classes alternate between frontal and occipital activation without lateralisation. In contrast, the CSP patterns for the beta band appear very “fragmented”, suggesting that they are not good candidates for classification. Indeed, the 5 features that are least likely to be selected comprise filters from the beta band between 1–7 s.

#### D. Memory score

As an exploratory analysis, we calculated the Pearson correlation between a participant’s test AUC and their corresponding retrospective memory score. The correlation coefficient was moderate, but significant  $\rho = 0.49$  (95% CI = [0.14, 0.73];  $p = 9 \times 10^{-3}$ ; see Fig. 6).

### V. DISCUSSION

This study investigated the neural correlates of LTM retrieval in an audiovisual recall task and classified EEG signals from participants watching video clips according to whether or not they remembered watching them previously.

Specifically-constructed stimuli are often used to study memory [9], [10], [11], [12], [13]. While these stimuli are beneficial, as they are easier to control by the scientist [77], it remains unclear whether the findings from such experiments reflect the way memory (and, more generally, other cognitive processes) works with real-world stimuli. In this study, we made an additional effort to incorporate an organic element: we ensured an ecologically valid encoding phase by choosing the EEMMD [59], a public dataset in which participants were presented with clips from movies that they had watched prior to the experiment. This causes our work to deviate from others in that the timing of the encoding was not specified. In contrast, most LTM research paradigms focus on retrieval following hours or, at most, weeks from encoding. Instead, we looked at real-life LTM retrieval [78]. Crucially, in the EEMMD participants had freely decided to watch the movies as part of their lives, not trying to encode

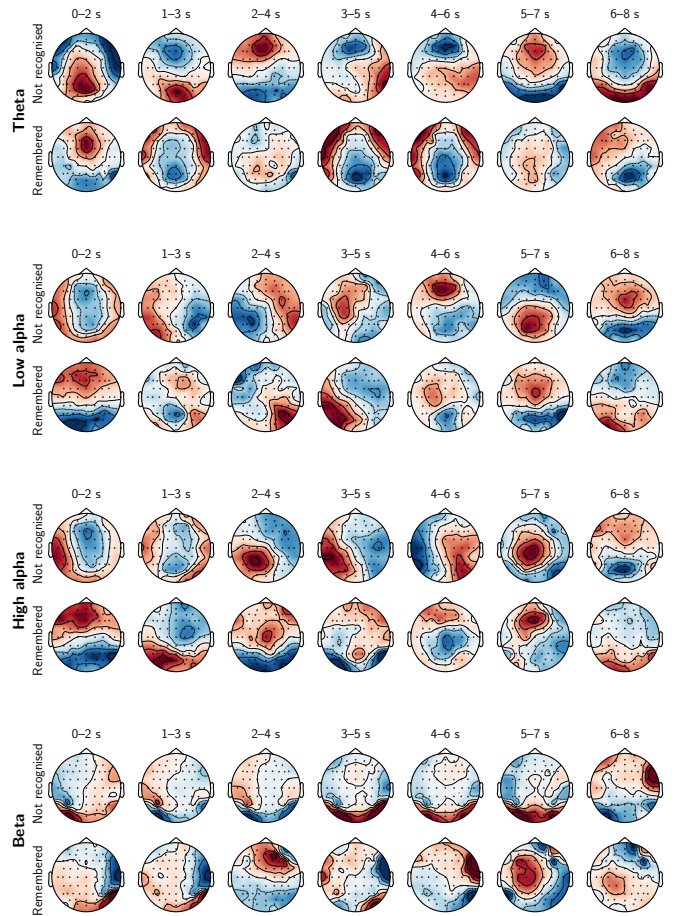


Fig. 5. Scalp maps of the filter-bank common spatial patterns for each not recognised and remembered classes, for each combination of time window and frequency band, for a participant with test AUC = 0.61. The FBCSPs were trained on the full training+validation set for this participant (80% of their epochs). We chose this participant because they lie near the median AUC performance. The colours represent the contribution of each channel to the pattern, with red (resp. blue) colours representing positive (resp. negative) weights, and darker intensities correspond to greater contributions.

them for a subsequent memory test. Given the number of movies that participants were surveyed about in the onboarding questionnaire and the limited amount of time between filling the survey and EEG data collection, it is extremely unlikely that they could have watched the movies again in preparation for the experiment. This was a limitation of other studies that used real-life information from participants, as they require the participants to provide the materials, which necessitates re-exposure prior to the experiment [78]. On the downside, the participant annotations relied on self-report of movie-watching, which can be subjective and prone to misremembering (e.g., when asked “when was the last time you watched this movie?”).

Overall, despite the methodological departure from traditional experimental designs that use controlled, abstract stimuli, our findings align with previous research on the neural correlates of LTM retrieval. This consistency highlights the robustness of memory processes across different experimental

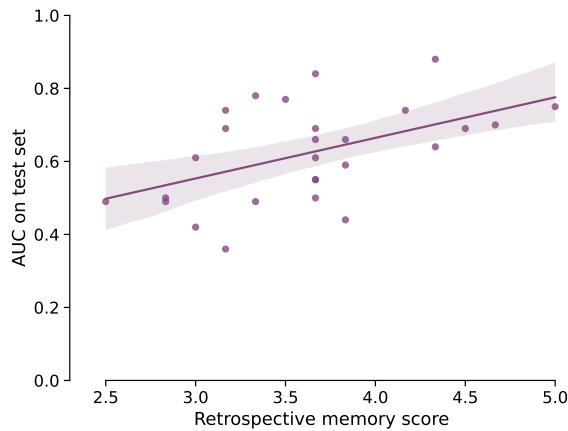


Fig. 6. Scatterplot showing the correlation between retrospective memory score and classifier performance (test AUC) for our participant sample ( $\rho = 0.49$ ; 95% CI = [0.14, 0.73];  $p = 9 \times 10^{-3}$ ). We also show the linear regression fit between the two variables. The shaded area represents the 95% confidence interval, calculated through 1000 bootstraps.

contexts and underscores the value of incorporating naturalistic stimuli in research on memory and other cognitive processes [79]. Unlike laboratory-based stimuli, naturalistic stimuli offer a more ecologically valid representation of how we encode and recall information in daily life. This approach can yield insights into the more complex, context-dependent aspects of memory. However, there are limitations, including the inherent variability of naturalistic stimuli, which can introduce noise into the data and complicate the interpretation of underlying neural mechanisms. Furthermore, the higher complexity of naturalistic stimuli may pose challenges in isolating specific cognitive processes, making it harder to draw direct comparisons to the more controlled conditions often used in traditional memory research. Nevertheless, the integration of naturalistic stimuli provides a richer, more holistic understanding of memory processes that aligns with real-world experiences, making it a valuable complement to more traditional experimental approaches.

Another consideration for future work involves the influence of stimulus-level properties on neural correlates of memory. While the EEMMD dataset captures ecologically valid, naturalistic viewing, it was not designed to systematically vary or control for clip-level properties such as genre, narrative complexity, emotional valence, or low-level visual statistics. Investigating whether these properties modulate the strength or reliability of memory-related EEG signatures would require a specifically designed stimulus set with balanced representation across such dimensions [63]. Understanding how different types of audiovisual content influence neural patterns during recall could inform both theoretical models of memory and practical applications in multimedia design and educational technology.

In the frequency domain, we found differences in the PSD across frontal, central, and left parietal ROIs of the brain depending on whether a clip had been categorised as not recognised, recognised, or remembered (see Fig. 1). These differences spanned high theta and both alpha sub-bands. In frontal ROIs, differences were also found in the low beta

band. In particular, we found lower power in these bands for remembered clips than for the other two classes, and there appears to be an order between the three conditions (not recognised > recognised > remembered), suggesting there may be a dependency of the certainty of the memory on the degree of desynchronisation [44], [11]. These differences are in agreement with the findings from Martín-Buro *et al.*, who also found differences in power in the alpha band between different types of recall (and with respect to non-remembered stimuli) [11].

Unfortunately, due to the limited number of trials from the recognised class (average: 18.8 epochs per participant, with several participants having  $\leq 4$  trials), we were not able to explore this through classification, particularly for single-trial classification. Future work with studies specifically designed to collect balanced samples across memory states (not recognised, recognised/familiar, and remembered/recalled) would enable proper investigation of familiarity vs. recollection using classification approaches, providing insights into the neural distinctions between these memory processes. We also did not explore the influence of time-since-last-watch (which was reported by participants of the EEMMD during the onboarding survey) on the strength of neural correlates, but plan to do this in the future, as old vs. recent memories show activation in different parts of the brain [78], [44] which might affect classification performance.

To localise these changes in time and frequency, we compared ERD/S between the not recognised and remembered epochs across the whole clip duration. We found that theta is consistently more synchronised at the beginning of the clip for remembered than for non-recognised clips (see Fig. 2) across most of the scalp, as previously reported in [80]. This was followed by a consistent pattern of alpha desynchronisation, most prominent in the posterior region of the head, and low beta desynchronisation across the scalp. These identified ERD/S patterns agreed with the well-established roles of theta, alpha, and beta in cognitive and memory performance [4], [31], [32], [38], [30], [35], [81].

We classified between the remembered and not recognised trials using FBCSP and a logistic regressor. We implemented a typical BCI pipeline and employed robust machine learning methods to find the best classification hyperparameters on a subject-by-subject basis. Our classifiers showed, on average, higher level of performance than earlier attempts (see Section II-C) [51], [19], [52], [5], [9]. Although average performance remained modest, we showed that it was significantly above random chance. Other classifiers might offer better performance than the logistic regression used here, but potentially at the cost of losing interpretability. Consistently with prior research, there is relatively high variability across participants, although the majority of them (70%) reached performance levels significantly above chance. This high variability is not only a feature of memory-related EEG correlates, but it is a well-known problem in EEG classification and BCIs [82], [83], [84].

In our classification feature analysis, we found that the models from good performers consistently selected features in the theta and alpha sub-bands, as we expected from the ERD/S

analysis and prior literature [31], [32], [34], [30]. Particularly, activity in the theta band was the most likely to be used to discriminate between the two classes, followed by alpha-band activity. Other studies also emphasised the role of theta in classification, with some using only activity in this band for the task [19] or together with alpha [85].

Alpha band activity, particularly in early time windows following stimulus onset, emerged as highly discriminative for memory classification, consistent with the higher ERD observed in this band at clip onset (Fig. 2). This aligns with established literature showing that alpha ERD in parietal regions, including the left parietal area where we observed higher power for not recognised trials (Fig. 1), is associated with successful long-term memory retrieval. The differential alpha power between conditions—with lower power (greater ERD) for remembered stimuli—reflects the well-documented role of parietal alpha desynchronization in accessing stored memory traces [35], [41].

Interestingly, despite the well-known role of beta in memory recall processes (e.g., [31], [32], [81]), this does not seem to be reflected in the features chosen by our classifiers. Although we did find differences in this band in the ERD/S between remembered and not recognised epochs (Fig. 2), in particular in the low beta range, features extracted from this band were substantially less likely to be chosen by our classifiers (average selection probability: 0.58 for beta vs. 0.86 for theta and 0.81–0.80 for alpha sub-bands), and completely removing beta-band features did not affect classification performance (median AUC with beta: 0.58 vs. without beta: 0.59;  $W = 187$ ,  $p = 0.97$ ). The high beta band exhibited high within-class standard deviations, as evidenced by the large confidence intervals in Fig. 1 and the alternating patterns of synchronization and desynchronization observed in central ROIs in Fig. 2. A potential reason for this might be beta’s high trial-to-trial variability, making it an unreliable biomarker for single-trial memory recall classification despite its established role in memory processes at the group level. It should be noted that this limited contribution may reflect both the high within-class variance observed in beta activity and our particular feature representation (variance-based CSP filters), rather than indicating an absence of beta-related mechanisms in memory processes. Alternative feature extraction approaches, such as connectivity or phase-based measures, might reveal different roles for beta activity. Finally, as we defined beta over 12–25 Hz and the lower half of the beta band did exhibit significant changes in the ERD/S analysis we will explore low beta as a separate feature in future work.

The PSD from Fig. 1 showed large within-class standard deviations in the beta band, and similarly the time-frequency ERD/S differences shown in Fig. 2 seemed smaller for beta than for lower frequency bands (which were selected). Thus, our results match earlier work in memory classification, where others predominantly relied on theta and alpha band activity to discriminate between classes at the time of *encoding* [52], [19], [54]—to the best of our knowledge, no prior work exists for memory *recall* classification using ERD/S or power features.

Finally, a particularly noteworthy finding is the observed correlation between individual classifier performance and ret-

spective memory scores ( $\rho = 0.49$ , 95% CI = [0.14, 0.73];  $p = 9 \times 10^{-3}$ ; see Fig. 6). This demonstrates that behavioural measures of memory ability are reflected in the neural signatures captured during recall [31], [39], [42]. This finding supports the hypothesis that the strength and stability of memory representations play a critical role in the effectiveness of EEG-based memory classification. Importantly, this relationship suggests that individual differences in memory ability modulate the consistency and reliability of neural correlates, which has significant implications for developing personalized memory assessment tools and understanding the neural basis of memory variability across individuals. This would match with earlier results regarding different brain activation patterns based on the age of the memory we are trying to recall [78], [44]. Future work should validate this relationship in larger samples and investigate whether similar correlations exist with other standardized memory assessments, potentially enabling EEG-based biomarkers for memory function in clinical populations.

In addition to looking at the strength of neural correlates of LTM with respect to when the memory was encoded, future work should aim to uncover the neural correlates that underpin the influence of emotion on memory. This is one of the areas where the limitations of traditional memory research stimuli become apparent, but where naturalistic stimuli excel. Emotional experiences significantly impact memory retention [86], [87], [88], so having both emotion and memorability annotations would allow us to investigate the relationship between emotional responses and how well clips are remembered. These additional labels would therefore be very valuable to study why some fragments of a movie are remembered and others are not, and whether and how the neural signatures of recall vary depending on the emotional content of the stimuli, paving the way for better affective interfaces.

From the point of view of potential applications of this technology, future work should also explore the development of person- or session-independent classifiers which can be used in clinical or educational settings, respectively [89]. Finally, transfer learning techniques remain largely unexplored in memory classification and should be further investigated.

## VI. CONCLUSION

We demonstrate above-chance decoding of long-term memory retrieval from naturalistic movie stimuli using EEG, achieving a median AUC of 0.64 in single-trial classification. Our analysis revealed that theta and alpha band activity, particularly in early post-stimulus windows (0–2s) and parietal regions, were reliable discriminative features for detecting memory recall, while low beta frequencies predominantly showed higher ERD for remembered than for not recognised trials whereas higher beta frequencies showed high within-class variability limiting their utility for single-trial classification. Notably, neural signatures of memory retrieval correlated significantly with participants’ self-reported memory ability ( $\rho = 0.49$ ), demonstrating that behavioural ratings are reflected in objective neural activity patterns.

Despite using naturalistic audiovisual stimuli and examining memories encoded organically months or years prior—a

significant departure from traditional controlled paradigms— we demonstrated that established neural correlates of memory generalize to ecologically valid contexts.

These findings have implications across multiple domains. In cognitive neuroscience, our work advances understanding of memory dynamics with complex, time-varying stimuli. Clinically, the correlation between neural decoding and memory ability suggests potential for developing EEG-based biomarkers of memory function in patients with cognitive impairments [53], [5], [19], [89]. For BCI research, our methods could inform applications in educational technology, mental health, and human-computer interaction, where understanding how audiovisual stimuli trigger memories could enable personalized learning environments [53] that adapt to individual memory processes.

## REFERENCES

- [1] E. Tulving, *Elements of episodic memory*. Oxford University Press, 1983.
- [2] J. Żygierewicz, R. A. Janik, I. T. Podolak *et al.*, “Decoding working memory-related information from repeated psychophysiological EEG experiments using convolutional and contrastive neural networks,” *Journal of Neural Engineering*, vol. 19, no. 4, p. 046053, 2022.
- [3] W. Klimesch, H. Schimke, and J. Schwaiger, “Episodic and semantic memory: an analysis in the EEG theta and alpha band,” *Electroencephalography and Clinical Neurophysiology*, vol. 91, no. 6, pp. 428–441, 1994.
- [4] W. Klimesch, “EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis,” *Brain research reviews*, vol. 29, no. 2-3, pp. 169–195, 1999.
- [5] J. H. Rudoler, J. P. Bruska, W. Chang *et al.*, “Decoding EEG for optimizing naturalistic memory,” *Journal of Neuroscience Methods*, vol. 410, p. 110220, 2024.
- [6] M. Werkle-Bergner, V. Müller, S.-C. Li *et al.*, “Cortical EEG correlates of successful memory encoding: implications for lifespan comparisons,” *Neuroscience & Biobehavioral Reviews*, vol. 30, no. 6, pp. 839–854, 2006.
- [7] J. H. Rudoler, N. A. Herweg, and M. J. Kahana, “Hippocampal theta and episodic memory,” *Journal of Neuroscience*, vol. 43, no. 4, pp. 613–620, 2023.
- [8] J. Fell and N. Axmacher, “The role of phase synchronization in memory processes,” *Nature Reviews Neuroscience*, vol. 12, no. 2, pp. 105–118, 2011.
- [9] E. Noh, K. Liao, M. V. Mollison *et al.*, “Single-trial EEG analysis predicts memory retrieval and reveals source-dependent differences,” *Frontiers in Human Neuroscience*, vol. 12, p. 258, 2018.
- [10] M. Gärtner and M. Bajbouj, “Encoding-related EEG oscillations during memory formation are modulated by mood state,” *Social Cognitive and Affective Neuroscience*, vol. 9, no. 12, pp. 1934–1941, 2014.
- [11] M. C. Martín-Buro, M. Wimber, R. N. Henson *et al.*, “Alpha rhythms reveal when and where item and associative memories are retrieved,” *Journal of Neuroscience*, vol. 40, no. 12, pp. 2510–2518, 2020.
- [12] W. Matthews, C. Benjamin, and C. Osborne, “Memory for moving and static images,” *Psychonomic Bulletin & Review*, vol. 14, no. 5, pp. 989–993, 2007.
- [13] S. L. Schneider and M. Rose, “Intention to encode boosts memory-related pre-stimulus EEG beta power,” *NeuroImage*, vol. 125, pp. 978–987, 2016.
- [14] H. Saarimäki, “Naturalistic stimuli in affective neuroimaging: A review,” *Frontiers in human neuroscience*, vol. 15, p. 675068, 2021.
- [15] I. P. Jääskeläinen, M. Sams, E. Glerean *et al.*, “Movies and narratives as naturalistic stimuli in neuroimaging,” *NeuroImage*, vol. 224, p. 117445, 2021.
- [16] S. Sonkusare, M. Breakspear, and C. Guo, “Naturalistic stimuli in neuroscience: critically acclaimed,” *Trends in Cognitive Sciences*, vol. 23, no. 8, pp. 699–714, 2019.
- [17] A. Jafarpour, A. Horner, L. Fuentemilla *et al.*, “Decoding oscillatory representations and mechanisms in memory,” *Neuropsychologia*, vol. 51, no. 4, pp. 772–780, 2013.
- [18] T. Grootswagers, S. G. Wardle, and T. A. Carlson, “Decoding dynamic brain patterns from evoked responses: A tutorial on multivariate pattern analysis applied to time series neuroimaging data,” *Journal of Cognitive Neuroscience*, vol. 29, no. 4, pp. 677–697, 2017.
- [19] X. Sun, C. Qian, Z. Chen *et al.*, “Remembered or forgotten?—An EEG-based computational prediction approach,” *PLoS one*, vol. 11, no. 12, p. e0167497, 2016.
- [20] G. Shmueli, “To Explain or to Predict?” *Statistical Science*, vol. 25, no. 3, pp. 289 – 310, 2010.
- [21] A. Diwali, K. Saeedi, K. Dashtipour *et al.*, “Sentiment analysis meets explainable artificial intelligence: A survey on explainable sentiment analysis,” *IEEE Transactions on Affective Computing*, 2023.
- [22] A. Matran-Fernandez and S. Halder, “An EEG dataset to study neural correlates of audiovisual long-term memory retrieval,” *Scientific Data*, vol. 12, no. 1, p. 1933, 2025.
- [23] E. Tulving, “How many memory systems are there?” *American Psychologist*, vol. 40, no. 4, p. 385, 1985.
- [24] K. A. Paller and A. D. Wagner, “Observing the transformation of experience into memory,” *Trends in Cognitive Sciences*, vol. 6, no. 2, pp. 93–102, 2002.
- [25] R. A. Poldrack, J. Clark, E. a. Paré-Blagoev *et al.*, “Interactive memory systems in the human brain,” *Nature*, vol. 414, no. 6863, pp. 546–550, 2001.
- [26] A. M. Schedlbauer, M. S. Copara, A. J. Watrous *et al.*, “Multiple interacting brain areas underlie successful spatiotemporal memory retrieval in humans,” *Scientific Reports*, vol. 4, no. 1, p. 6431, 2014.
- [27] Y. Agam and R. Sekuler, “Interactions between working memory and visual perception: An ERP/EEG study,” *NeuroImage*, vol. 36, no. 3, pp. 933–942, 2007.
- [28] W. Klimesch, “Memory processes, brain oscillations and EEG synchronization,” *International Journal of Psychophysiology*, vol. 24, no. 1-2, pp. 61–100, 1996.
- [29] W. Klimesch, R. Freunberger, P. Sauseng *et al.*, “A short review of slow phase synchronization and memory: evidence for control processes in different memory systems?” *Brain Research*, vol. 1235, pp. 31–44, 2008.
- [30] W. Klimesch, “Alpha-band oscillations, attention, and controlled access to stored information,” *Trends in Cognitive Sciences*, vol. 16, no. 12, pp. 606–617, 2012.
- [31] S. Hanouneh, H. U. Amin, N. M. Saad *et al.*, “EEG Power and Functional Connectivity Correlates with Semantic Long-Term Memory Retrieval,” *IEEE Access*, vol. 6, pp. 8695–8703, 2018.
- [32] S. Hanslmayr, T. Staudigl, and M.-C. Fellner, “Oscillatory power decreases and long-term memory: the information via desynchronization hypothesis,” *Frontiers in Human Neuroscience*, vol. 6, p. 74, 2012.
- [33] E. Nyhus and T. Curran, “Functional role of gamma and theta oscillations in episodic memory,” *Neuroscience & Biobehavioral Reviews*, vol. 34, no. 7, pp. 1023–1035, 2010.
- [34] R. J. Addante, A. J. Watrous, A. P. Yonelinas *et al.*, “Prestimulus theta activity predicts correct source memory retrieval,” *Proceedings of the National Academy of Sciences*, vol. 108, no. 26, pp. 10702–10707, 2011.
- [35] P. H. Khader and F. Rösler, “EEG power changes reflect distinct mechanisms during long-term memory retrieval,” *Psychophysiology*, vol. 48, no. 3, pp. 362–369, 2011.
- [36] W. Klimesch, M. Doppelmayr, H. Schimke *et al.*, “Theta synchronization and alpha desynchronization in a memory task,” *Psychophysiology*, vol. 34, no. 2, pp. 169–176, 1997.
- [37] W. Klimesch, M. Doppelmayr, A. Yonelinas *et al.*, “Theta synchronization during episodic retrieval: neural correlates of conscious awareness,” *Cognitive Brain Research*, vol. 12, no. 1, pp. 33–38, 2001.
- [38] W. Klimesch, M. Doppelmayr, W. Stadler *et al.*, “Episodic retrieval is reflected by a process specific increase in human electroencephalographic theta activity,” *Neuroscience letters*, vol. 302, no. 1, pp. 49–52, 2001.
- [39] W. Klimesch, M. Doppelmayr, T. Pachinger *et al.*, “Brain oscillations and human memory: EEG correlates in the upper alpha and theta band,” *Neuroscience Letters*, vol. 238, no. 1-2, pp. 9–12, 1997.
- [40] W. Klimesch, “EEG-alpha rhythms and memory processes,” *International Journal of Psychophysiology*, vol. 26, no. 1-3, pp. 319–340, 1997.
- [41] W. Klimesch, M. Doppelmayr, T. Pachinger *et al.*, “Event-related desynchronization in the alpha band and the processing of semantic information,” *Cognitive Brain Research*, vol. 6, no. 2, pp. 83–94, 1997.
- [42] J. Jacobs, G. Hwang, T. Curran *et al.*, “EEG oscillations and recognition memory: theta correlates of memory retrieval and decision making,” *NeuroImage*, vol. 32, no. 2, pp. 978–987, 2006.
- [43] A. G. Goldstein, J. E. Chance, M. Hoisington *et al.*, “Recognition memory for pictures: Dynamic vs. static stimuli,” *Bulletin of the Psychonomic Society*, vol. 20, no. 1, pp. 37–40, 1982.

- [44] J. F. Danker and J. R. Anderson, "The ghosts of brain states past: remembering reactivates the brain regions engaged during encoding," *Psychological Bulletin*, vol. 136, no. 1, p. 87, 2010.
- [45] M. E. Wheeler, S. E. Petersen, and R. L. Buckner, "Memory's echo: vivid remembering reactivates sensory-specific cortex," *Proceedings of the National Academy of Sciences*, vol. 97, no. 20, pp. 11 125–11 129, 2000.
- [46] J. Chen, Y. C. Leong, C. J. Honey *et al.*, "Shared memories reveal shared structure in neural activity across individuals," *Nature Neuroscience*, vol. 20, no. 1, pp. 115–125, 2017.
- [47] J. L. Park and D. I. Donaldson, "Detecting the neural correlates of episodic memory with mobile EEG: Recollecting objects in the real world," *NeuroImage*, vol. 193, pp. 1–9, 2019.
- [48] M. Desai, J. Holder, C. Villarreal *et al.*, "Generalizable EEG encoding models with naturalistic audiovisual stimuli," *Journal of Neuroscience*, vol. 41, no. 43, pp. 8946–8962, 2021.
- [49] H. U. Amin, A. S. Malik, N. Kamel *et al.*, "P300 correlates with learning & memory abilities and fluid intelligence," *Journal of Neuroengineering and Rehabilitation*, vol. 12, pp. 1–14, 2015.
- [50] F. Lotte, L. Bougrain, A. Cichocki *et al.*, "A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update," *Journal of Neural Engineering*, vol. 15, no. 3, p. 031005, 2018.
- [51] S. Chakravarty, Y. Y. Chen, and J. B. Caplan, "Predicting memory from study-related brain activity," *Journal of Neurophysiology*, vol. 124, no. 6, pp. 2060–2075, 2020.
- [52] E. Noh, G. Herzmann, T. Curran *et al.*, "Using single-trial EEG to predict and analyze subsequent memory," *NeuroImage*, vol. 84, pp. 712–723, 2014.
- [53] S. Mirjalili, P. Powell, J. Strunk *et al.*, "Evaluation of classification approaches for distinguishing brain states predictive of episodic memory performance from electroencephalography," *NeuroImage*, vol. 247, p. 118851, 2022.
- [54] G.-H. Shin, Y.-S. Kweon, and M. Lee, "Predicting the transition from short-term to long-term memory based on deep neural network," in *2021 9th International Winter Conference on Brain-Computer Interface (BCI)*. IEEE, 2021, pp. 1–5.
- [55] A. Matran-Fernandez and S. Halder, "Data-driven detection of memory encoding from EEG in an audiovisual task," in *5th International NeuroErgonomics Conference*. Frontiers, July 2024.
- [56] K. Liao, M. V. Mollison, T. Curran *et al.*, "Single-trial EEG predicts memory retrieval using leave-one-subject-out classification," in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2018, pp. 2613–2620.
- [57] J. M. Gardiner and A. Richardson-Klavehn, *Remembering and knowing*. Oxford University Press, 2000.
- [58] E. Erdoğdu, E. Kurt, A. D. Duru *et al.*, "Measurement of cognitive dynamics during video watching through event-related potentials (ERPs) and oscillations (EROs)," *Cognitive Neurodynamics*, vol. 13, pp. 503–512, 2019.
- [59] A. Matran-Fernandez and R. Poli, "Event-Related Potentials induced by cuts in feature movies and their exploitation for understanding cut efficacy," in *7th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 2015, pp. 74–77.
- [60] P. Francuz and E. Zabielska-Mendyk, "Does the Brain Differentiate Between Related and Unrelated Cuts When Processing Audiovisual Messages? An ERP Study," *Media Psychology*, vol. 16, no. 4, pp. 461–475, 2013.
- [61] V. M. Reid and T. Striano, "N400 involvement in the processing of action sequences," *Neuroscience Letters*, vol. 433, no. 2, pp. 93–97, 2008.
- [62] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," *Sensors*, vol. 12, no. 2, pp. 1211–1279, 2012.
- [63] R. Cohendet, K. Yadati, N. Q. Duong *et al.*, "Annotating, understanding, and predicting long-term video memorability," in *Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval*, 2018, pp. 178–186.
- [64] N. Bigdely-Shamlo, T. Mullen, C. Kothe *et al.*, "The PREP pipeline: standardized preprocessing for large-scale EEG analysis," *Frontiers in Neuroinformatics*, vol. 9, p. 16, 2015.
- [65] A. Gramfort, M. Luessi, E. Larson *et al.*, "MEG and EEG data analysis with MNE-Python," *Frontiers in Neuroscience*, vol. 7, no. 267, pp. 1–13, 2013.
- [66] M. Jas, D. A. Engemann, Y. Bekhti *et al.*, "Autoreject: Automated artifact rejection for MEG and EEG data," *NeuroImage*, vol. 159, pp. 417–429, 2017.
- [67] A. Li, J. Feitelberg, A. P. Saini *et al.*, "MNE-ICLabel: Automatically annotating ICA components with ICLabel in Python," *Journal of Open Source Software*, vol. 7, no. 76, p. 4484, 2022.
- [68] G. Smith, S. Del Sala, R. H. Logie *et al.*, "Prospective and retrospective memory in normal ageing and dementia: A questionnaire study," *Memory*, vol. 8, no. 5, pp. 311–321, 2000.
- [69] Z. J. Koles, "The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG," *Electroencephalography and Clinical Neurophysiology*, vol. 79, no. 6, pp. 440–447, 1991.
- [70] H. Zou and T. Hastie, "Regularization and variable selection via the Elastic Net," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 67, no. 2, pp. 301–320, 2005.
- [71] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 4, pp. 441–446, 2000.
- [72] B. Blankertz, R. Tomioka, S. Lemm *et al.*, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal processing magazine*, vol. 25, no. 1, pp. 41–56, 2007.
- [73] S. Haufe, F. Meinecke, K. Görgen *et al.*, "On the interpretation of weight vectors of linear models in multivariate neuroimaging," *NeuroImage*, vol. 87, pp. 96–110, 2014.
- [74] S. Jauhiainen, J.-P. Kauppi, M. Leppänen *et al.*, "New machine learning approach for detection of injury risk factors in young team sport athletes," *International Journal of Sports Medicine*, vol. 42, no. 02, pp. 175–182, 2021.
- [75] M. Saarela and S. Jauhiainen, "Comparison of feature importance measures as explanations for classification models," *SN Applied Sciences*, vol. 3, no. 2, p. 272, 2021.
- [76] P. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70–73, 1967.
- [77] J. Bures, O. Burešová, and J. P. Huston, *Techniques and basic experiments for the study of brain and behavior*. Elsevier, 2016.
- [78] W. A. Bainbridge and C. I. Baker, "Multidimensional memory topography in the medial parietal cortex identified from neuroimaging of thousands of daily memory videos," *Nature Communications*, vol. 13, no. 1, p. 6508, 2022.
- [79] A. Bhatti, P. Angkan, B. Behinaein *et al.*, "CLARE: Cognitive Load Assessment in Real-Time With Multimodal Data," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 17, no. 6, pp. 1337–1349, 2025.
- [80] H. K. Gomar, M. Althaus, A. A. Wijers *et al.*, "The effects of memory load and stimulus relevance on the EEG during a visual selective memory search task: an ERP and ERD/ERS study," *Clinical Neurophysiology*, vol. 117, no. 4, pp. 871–884, 2006.
- [81] G. T. Waldhauser, M. Johansson, and S. Hanslmayr, "Alpha/beta oscillations indicate inhibition of interfering visual memories," *Journal of Neuroscience*, vol. 32, no. 6, pp. 1953–1961, 2012.
- [82] Z. Lan, O. Sourina, L. Wang *et al.*, "Domain Adaptation Techniques for EEG-Based Emotion Recognition: A Comparative Study on Two Public Datasets," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 11, no. 1, pp. 85–94, 2019.
- [83] D. Wu, Y. Xu, and B.-L. Lu, "Transfer Learning for EEG-Based Brain–Computer Interfaces: A Review of Progress Made Since 2016," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 14, no. 1, pp. 4–19, 2022.
- [84] L. Zhang, Y. Zhou, P. Gong *et al.*, "Speech imagery decoding using eeg signals and deep learning: A survey," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 17, no. 1, pp. 22–39, 2025.
- [85] M. Mirjalili, R. Zomorodi, Z. J. Daskalakis *et al.*, "Individualized real-time prediction of working memory performance by classifying electroencephalography signals," *International Journal of Imaging Systems and Technology*, vol. 32, no. 2, pp. 575–589, 2022.
- [86] C. M. Tyng, H. U. Amin, M. N. Saad *et al.*, "The influences of emotion on learning and memory," *Frontiers in Psychology*, vol. 8, p. 235933, 2017.
- [87] Y. Baveye, R. Cohendet, M. Perreira Da Silva *et al.*, "Deep learning for image memorability prediction: The emotional bias," in *Proceedings of the 24th ACM international conference on Multimedia*, 2016, pp. 491–495.
- [88] L.-W. Mai and G. Schoeller, "Emotions, attitudes and memorability associated with TV commercials," *Journal of Targeting, Measurement and Analysis for Marketing*, vol. 17, pp. 55–63, 2009.
- [89] Z. Unnisa, A. Tariq, I. Din *et al.*, "On Harnessing EEG Signals for the Comprehensive Assessment of Neurological Disorders: A Review," *IEEE Transactions on Cognitive and Developmental Systems*, pp. 1–22, 2026.

- [90] A. Gelman, J. B. Carlin, H. S. Stern *et al.*, *Bayesian Data Analysis*. CRC Press, 2013.
- [91] R. Van de Schoot, S. Depaoli, R. King *et al.*, “Bayesian statistics and modelling,” *Nature Reviews Methods Primers*, vol. 1, no. 1, p. 1, 2021.

## VII. BIOGRAPHY SECTION



**Ana Matran-Fernandez** received her MSc in Biomedical Engineering in 2012 (University of Surrey) and her PhD degree in 2012 for her work on collaborative BCIs. After a 3-year secondment at the Institute of Analytics and Data Science, in 2021 she became a Lecturer in Neural Engineering and Artificial Intelligence in the University of Essex. Her research interests focus on passive BCIs and the use of real-life stimuli to study the brain, particularly memory.



**Sebastian Halder** received his MSc in Bioinformatics in 2006 and PhD in Computer Science in 2011, both from the University of Tübingen, Germany. From 2006 to 2012, Dr. Halder worked in Prof. Niels Birbaumer’s group, focusing on BCIs for communication with severely motor-impaired individuals. His postdoctoral training included positions at the University of Würzburg, Germany; the National Rehabilitation Center for Persons with Disabilities in Japan; and the University of Oslo, Norway. In 2019, Dr. Halder joined the University of Essex as a Lecturer in Brain-Computer Interfacing, later becoming a Senior Lecturer in October 2022. His research interests encompass auditory BCIs for communication, neural mechanisms of BCI learning, neural markers of pain, and disorders of consciousness.

TABLE A.1

POSTERIOR DISTRIBUTIONS, TOGETHER WITH THEIR 95% CREDIBLE INTERVALS (CI) AND EXPECTED NUMBER OF PARTICIPANTS (OUT OF 27) WHO WOULD REACH PERFORMANCE ABOVE RANDOM CHANCE FOR DIFFERENT PRIORS.

| Prior              | Posterior Mean | 95% CI         | Expected Count |
|--------------------|----------------|----------------|----------------|
| Uniform            | 0.690          | [0.513, 0.841] | 18.6           |
| Weakly skeptical   | 0.677          | [0.506, 0.827] | 18.3           |
| Strongly skeptical | 0.617          | [0.475, 0.749] | 16.7           |
| Optimistic         | 0.700          | [0.528, 0.847] | 18.9           |
| Pessimistic        | 0.667          | [0.492, 0.821] | 18.0           |

that classification performance in our system was reliably above chance across participants.

## APPENDIX

### A. Statistical Evidence for Above-Chance Classification Performance

To assess whether classification performance exceeded random-chance levels across participants, we analysed the test AUCs. We defined above-chance performance as  $AUC > 0.5$  and evaluated whether the true proportion of participants with test AUCs  $> 0.5$  was greater than what would be expected by chance.

We used a Bayesian approach to estimate the posterior distribution over this proportion, comparing posterior distributions and 95% credible intervals under several plausible prior beliefs. Let  $k = 19$  be the number of participants with above-chance performance in our sample and  $n = 27$  the total number of participants. Assuming a binomial likelihood, and placing a Beta prior  $\text{Beta}(\alpha_0, \beta_0)$  on the proportion  $\theta$ , the posterior distribution is:

$$\theta \mid \text{data} \sim \text{Beta}(\alpha_0 + k, \beta_0 + n - k)$$

We chose the following commonly used prior beliefs to test our approach [90], [91]:

- **Uniform prior** ( $\text{Beta}(1, 1)$ ): a non-informative prior treating all proportions as equally likely.
- **Weakly skeptical prior** ( $\text{Beta}(2, 2)$ ): lightly centres belief around chance level (0.5), expressing mild regularization.
- **Strongly skeptical prior** ( $\text{Beta}(10, 10)$ ): reflects prior skepticism by placing most prior mass near 0.5.
- **Optimistic prior** ( $\text{Beta}(2, 1)$ ): expresses a belief that above-chance performance is more likely.
- **Pessimistic prior** ( $\text{Beta}(1, 2)$ ): expresses a belief that below- or chance-level performance is more likely.

Posterior probabilities  $P(\theta > 0.5)$  were computed as:

$$P(\theta > 0.5) = 1 - F_{\text{Beta}(\alpha_0+k, \beta_0+n-k)}(0.5)$$

where  $F_{\text{Beta}}$  is the cumulative distribution function of the Beta distribution. The results are summarized in Table A.1.

Across all priors, posterior probabilities that  $\theta > 0.5$  exceed 90%, suggesting robust evidence for above-chance performance (i.e., there is a very high probability that more than half of participants had AUC values above chance). Under strongly skeptical and optimistic priors, the 95% credible intervals include 0.5, but the bulk of posterior mass remains above chance level, indicating that the conclusion is not overly sensitive to prior assumptions and providing strong evidence