

UC San Diego

UC San Diego Previously Published Works

Title

Using single-trial EEG to predict and analyze subsequent memory

Permalink

<https://escholarship.org/uc/item/2zq184wt>

Authors

Noh, Eunho

Herzmann, Grit

Curran, Tim

et al.

Publication Date

2014

DOI

10.1016/j.neuroimage.2013.09.028

Peer reviewed



Published in final edited form as:

Neuroimage. 2014 January ; 84: . doi:10.1016/j.neuroimage.2013.09.028.

Using Single-trial EEG to Predict and Analyze Subsequent Memory

Eunho Noh^{a,*}, Grit Herzmann^b, Tim Curran^b, and Virginia R. de Sa^c

^aDepartment of Electrical and Computer Engineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, California, 92093

^bDepartment of Psychology and Neuroscience, University of Colorado Boulder

^cDepartment of Cognitive Science, University of California, San Diego

Abstract

We show that it is possible to successfully predict subsequent memory performance based on single-trial EEG activity before and during item presentation in the study phase. Two-class classification was conducted to predict subsequently remembered vs. forgotten trials based on subjects' responses in the recognition phase. The overall accuracy across 18 subjects was 59.6 % by combining pre- and during-stimulus information. The single-trial classification analysis provides a dimensionality reduction method to project the high-dimensional EEG data onto a discriminative space. These projections revealed novel findings in the pre- and during-stimulus period related to levels of encoding. It was observed that the pre-stimulus information (specifically oscillatory activity between 25–35Hz) –300 to 0 ms before stimulus presentation and during-stimulus alpha (7–12 Hz) information between 1000–1400 ms after stimulus onset distinguished between recollection and familiarity while the during-stimulus alpha information and temporal information between 400–800 ms after stimulus onset mapped these two states to similar values.

Keywords

EEG; Memory; SME; Prediction; Recollection; Familiarity

1. Introduction

Many studies have shown evidence of differences in the electroencephalography (EEG) signals during learning of pictures or words that will later be remembered compared to items that will be forgotten (Sanquist et al., 1980; Paller and Wagner, 2002). In addition to brain activity during learning, many studies have found evidence that anticipatory activity preceding the onset of a stimulus can contribute to subsequent episodic memory encoding (Otten et al., 2006, 2010; Park and Rugg, 2010; Guderian et al., 2009; Fell et al., 2011). These differences in brain activity between the subsequently remembered and forgotten trials before or during stimulus presentation are often referred to as *subsequent memory effects* or SMEs.

© 2013 Elsevier Inc. All rights reserved.

*Corresponding author. Voice: +1 858 534 6775. Fax: +1 858 534 1128., EUNOH@UCSD.EDU (Eunho Noh).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

The difference in event-related potential (ERP) to presentation of the subsequently remembered and forgotten trials is known as difference due to memory (Dm) (Paller et al., 1987). It is typically measured as a posterior positivity between 400 and 800 ms in the study phase of a memory task (Paller and Wagner, 2002). However, the size and timing of the effect varies depending on the paradigm of the experiment (Johnson, 1995).

Several studies have successfully demonstrated that brain oscillations in multiple EEG frequency bands during encoding can distinguish between remembered and forgotten trials (see Hanslmayr and Staudigl (2013) for a review). It was found that power increases for the remembered items (positive spectral SMEs) typically occurred in the theta and high gamma bands (Klimesch et al., 1996a; Sederberg et al., 2003; Staudigl and Hanslmayr, 2013) and power decreases for the remembered items (negative spectral SMEs) typically occurred in the alpha and low beta bands (Klimesch et al., 1996b; Hanslmayr et al., 2009, 2012) of the EEG signal.

It has been recently shown that successful encoding also depends on anticipatory brain activity before encoding elicited by presenting cues before each study item. Using an incidental memory paradigm, Otten et al. (2006, 2010) showed that there is a significant difference in the ERPs to cue presentation during the *pre-stimulus* period of the study phase between the subsequently remembered and forgotten words. In a functional magnetic resonance imaging (fMRI) study, Park and Rugg (2010) found significant differences in the level of hippocampal BOLD activity during the cue-item interval between words with subsequent memory contrasts. It has also been reported that anticipatory brain activity is not only related to memory formation but reward anticipation, where differences in ERP and theta power were only observed for words following high reward cues (Gruber and Otten, 2010; Gruber et al., 2013).

A number of studies have shown that subsequent memory can be predicted from pre-stimulus spectral (oscillatory) activity without informative cues. This was identified by analysing power in different frequency bands of the pre-stimulus brain activity (Guderian et al., 2009; Fell et al., 2011). For instance, Guderian et al. (2009) used MEG to show that later recalled words, as compared to later forgotten items, are associated with stronger pre-stimulus increases in theta power (3–8 Hz) starting 200 ms before study item presentation (a fixation-cross was presented 500 ms before each stimulus). In an intracranial EEG study, Fell et al. (2011) found that the rhinal cortex and hippocampus show enhancement of pre-stimulus theta power during the jittered inter-stimulus interval (ISI) for successful memory formation. It was also found that this pre-stimulus effect extends from theta all the way up to the beta range (up to 34 Hz) within the rhinal cortex.

Studies discussed above averaged over multiple trials to reveal the underlying SMEs. However, pattern classification approaches on fMRI data have been successful in predicting subsequent memory in single trials. Single trial prediction of subsequent recognition performance has been demonstrated using multivoxel pattern analysis (MVPA) of fMRI data during encoding of phonogram stimuli (Watanabe et al., 2011). Watanabe et al. (2011) found that activity in the MTL (medial temporal lobe) acquired during encoding is predictive of subsequent recognition performance. In a very recent fMRI study, Yoo et al. (2012) monitored the activation in parahippocampal cortex (PHC) in real-time and presented study items when subjects entered *good* or *bad* brain states for learning of novel scenes. The brain states were determined by computing the pre-stimulus difference between the BOLD signal activations in the parahippocampal place area (PPA) and reference ROI (region of interest). They found that subsequent recognition memory was more accurate for items presented when PPA activation was lower than the reference ROI activation by a subject-specific threshold. The good/bad brain states defined by Yoo et al. (2012) are unlikely to

reflect a general encoding-related state but rather a context specific encoding-related state (good/bad brain state for encoding scenes in this case).

While single-trial classification results using fMRI are encouraging, there has not been any research on single-trial analysis of SME using a more mobile and affordable recording procedure such as EEG. Our study aims to identify the characteristics of the various SMEs in pre- and during-stimulus EEG on a single trial basis. This can potentially be developed as a practical system to predict preparedness for, and success of, memory encoding which could be used to improve memory performance. By presenting stimuli at predicted optimal memory encoding times (and repeating presentations when the during-stimulus classifier deems them not likely to be well encoded) users may be able to learn material with fewer presentations. With prolonged use of the system, users may become more aware of when they are in, and how to get into, better states for remembering from the implicit feedback provided by the timing and repetition of the presented items. This may eventually improve the memory performance of the users even without the system.

Classification was conducted on remembered vs. forgotten trials by combining the pre- and during-stimulus information in the EEG signal. Three separate classifiers were trained to learn the spectral features of the pre-stimulus SME, temporal features of the during-stimulus SME, and spectral features of the during-stimulus SME. The results from the individual classifiers were then combined to predict subsequent memory in single trials. The single-trial classification analysis can be considered as a non-linear dimensionality reduction method to effectively project the high-dimensional EEG data onto a discriminative space. These projections further revealed novel findings in the pre- and during-stimulus period related to levels of encoding which would have been difficult to find by simply averaging over the high-dimensional EEG data. The classifier scores (i.e. projections of the EEG signals onto the discriminative space defined by the classifier) were grouped by the different response options given in the recognition phase to examine the relationship between the classifier scores and levels of encoding represented by subjects' recognition confidence. In order to better understand the brain activity underlying SMEs utilized by the classifiers, temporal and spectral analyses were conducted on the EEG signals.

2. Materials and methods

EEG for the present study was previously recorded in 61 healthy right-handed males (consisting of car experts and novices) during a visual memory task (Herzmann and Curran, 2011). In the study phases, subjects memorized pictures of birds and cars (in separate blocks). In the recognition phases, participants had to discriminate these study items from random distractors using a rating scale with 5 options (*recollect*, *definitely familiar*, *maybe familiar*, *maybe unfamiliar*, and *definitely unfamiliar*). Timings of trials in the study and recognition phases are given in Figure 1.

2.1. Participants

Subjects were right-handed males (age 18–29) who volunteered for paid participation in the experiment. Out of the 61 subjects, 30 were self-reported car experts while none were bird experts based on a self-report questionnaire. For the classification study, 18 subjects were pre-selected from the group based on the criteria given below. Inclusion criteria were set up to acquire a dataset with 1) a sufficient number of remembered/forgotten trials for classifier training; 2) subjects who were attentive during the experiment based on their performance in the memory task. Subjects who did not meet these criteria were excluded in a stepwise manner. As a result, 18 subjects were pre-selected for analysis (10 subjects were car experts).

1. Subject's behavioral performance

10 subjects who were not effectively participating in the given memory task were discarded from further analysis. These subjects had behavioral performance lower than 56.3% (50 % chance performance) were excluded. A response was considered correct if they responded with old (*recollect, definitely familiar, maybe familiar*) to a target item or new (*maybe unfamiliar, definitely unfamiliar*) to a distractor. Note that the threshold 56.3% was calculated by subtracting the standard deviation from the average of the behavioral accuracies of all 61 subjects.

2. Number of trials after rejection of trials with artifacts

33 subjects were excluded due to insufficient number of trials to train a reliable classifier. Subjects that had less than 64 trials within each of the two classes after trial rejection were excluded from further analysis to ensure the number of trials available was equal to the number of electrodes in the worst case.

2.2. Stimulus presentation and EEG recording

The experiment was divided into 8 blocks consisting of a study and recognition phase. Stimuli consisted of color photographs of cars and birds where cars were given in the odd blocks and birds in the even blocks. The pictures were presented on a 17-inch flat-panel LCD monitor (Apple Studio Display SP110, refresh rate 59 Hz) at a viewing distance of one meter.

During the study phase, subjects were instructed to memorize forty target pictures. A fixation cross appeared for 200 ms then a study item was shown for 2 seconds. The ISI between the items in the study phase was 800 ms. After approximately 10 minutes, the subjects were given a recognition test. In the recognition phase, targets learned in the study phase had to be discriminated from forty new, unfamiliar distractors. A fixation cross appeared for 200 ms then a study or distractor item was shown for 1.5 seconds. All items were presented in random order. The participants had to decide without time limit if they had seen the picture in the study phase or not using a rating scale with 5 options (*recollect, definitely familiar, maybe familiar, maybe unfamiliar, and definitely unfamiliar*). Subjects were asked to select *recollect* if they had a conscious recollection of learning the picture in the study phase. If they did not recollect the stimulus, they were asked to give familiarity ratings for it by pressing one of the keys that corresponded to one of the four options from the rating scale. The order of stimuli and assignment of response buttons were kept constant for all participants to ensure comparability of task demands.

EEG was recorded with a 128-channel Geodesic Sensor NetTM (HydroCel GSN 128 1.0, Tucker, 1993) using an AC-coupled 128-channel, high-input impedance amplifier (200 M Ω , Net AmpsTM, Electrical Geodesics Inc., Eugene, OR). Amplified analog voltages (0.1–100 Hz bandpass) were digitized at 250 Hz. Initial common reference was the vertex channel (Cz). Individual sensor impedances were adjusted until the levels were lower than 50 k Ω .

2.3. Pre-processing

EEG epochs from the study phase of the experiment were extracted and recalculated to average reference. Trials that included high noise were automatically discarded using the rejkurt function in EEGLAB (Delorme and Makeig, 2004) which rejects trials based on the kurtosis of each trial. Then each trial was manually inspected to exclude trials which showed eye movement or muscle artifacts. An average of 40 trials were rejected for each subject. To further remove eye movement artifacts, independent component analysis (infomax ICA) (Hyvärinen et al., 2001; Makeig et al., 1996) was performed to identify and remove them. The degrees of freedom of the EEG signal is reduced after removing eye movement

components. A subset of 73 electrodes which is an approximate equivalent of the 10–20 system was selected for further analysis in order to reduce the dimensionality of the data set and ensure a full rank covariance matrix for eigenvalue decomposition (for common spatial patterns) even after removing independent components. The locations of the selected electrodes are given in Figure 2.

2.4. Classification problem

The classification problem was set up as follows. First, trials that were presented in the study phase were labelled according to the results of the recognition phase. There were two labels: remembered (class 1) and forgotten (class 2). The remembered class consisted of trials where the subjects pressed the button *recollect* and the forgotten class consisted of trials where the subjects pressed buttons *maybe unfamiliar* and *definitely unfamiliar*. Trials with *definitely familiar* or *maybe familiar* responses were not included in the remembered class to maximize the difference in encoding strength between the classes (trials with *maybe unfamiliar* were considered forgotten trials due to the limited number of trials with *definitely unfamiliar* responses), but they were used to compare the classifier scores and the subjects responses in the recognition phase (see Section 3.3). Sets of labelled examples were acquired from the shaded areas (A) (–300 to 0 ms before stimulus presentation) and (B) (400–800 and 1000–1400 ms after stimulus onset) of each trial in Figure 1. Note that separate classification analysis on item type (car/bird) was omitted since the number of car/bird items was insufficient to build a reliable classifier for most of the subjects.

Classifier performance was evaluated based on the number of trials considered for classification. Chance level in a simple 2-class classification problem is not exactly 50%, but 50% with a confidence interval for a given p value depending on the number of trials. These intervals were calculated using Wald intervals with adjustments for a small sample size (Müller-Putz et al., 2008; Agresti and Caffo, 2000). This gives a much more accurate interval for small samples compared to the ordinary Wald interval. The Wald interval is the normal approximation of the binomial confidence interval.

2.5. Classification

Based on previous findings on pre-stimulus spectral SME that found power differences between the remembered and forgotten items ranging from theta to the beta band (Fell et al., 2011), linear classifiers were designed to learn the power differences (i.e. amplitude differences) between the two classes in multiple subbands ranging from theta to low gamma of the pre-stimulus EEG data. Common spatial patterns (CSPs) were used to learn spatial filters which maximize the power difference between the two classes (Blankertz et al., 2008). The CSP algorithm is designed to increase the discriminability by finding spatial filters that maximize the power of the filtered signal while minimizing for the other class. The 300 ms subsequence preceding the to-be-learned stimulus (portion noted as (A) in Figure 1) was extracted from each trial before any pre-processing was performed to prevent any temporal smearing from the signal during actual encoding. We used a total of 9 bandpass filters with pre-selected subbands to account for the wide range of frequency bands associated with pre-stimulus SME. The subbands were selected based on well known rhythmic activities of EEG signals between 4–40 Hz and overlapping frequencies in between. The passband for each filter was 4–7 Hz (theta band), 6–10 Hz, 7–12 Hz (alpha band), 10–15 Hz, 12–19 Hz (low beta band), 15–25 Hz, 19–30 Hz (high beta band), 25–35 Hz, 30–40 Hz (low gamma band). The overlapping frequencies were used to compensate for individual differences in the EEG subbands (Doppelmayr et al., 1998) and timing of the pre-stimulus SME. Subbands with informative patterns for subsequent memory prediction were identified from the training set and only the classifiers corresponding to those *informative* subbands were used to classify the validation set. The output of the pre-stimulus classifier

(denoted as $0 \leq p_A \leq 1$) can be interpreted as the *pre-stimulus classifier score* of how good the classifier deems the brain state for remembering pictures.

Two separate classifiers were designed to extract the temporal and spectral characteristics of the during-stimulus period of the remembered/forgotten trials. Temporal features were learned by exploiting the ERP differences (namely the Dm effect) between the two classes in the spatio-temporal domain. The during-stimulus temporal classifier was trained to learn these features of the EEG data between 400–800 ms after stimulus presentation from four channel groups (*CM* centro medial, *LPS* left posterior superior, *RPS* right posterior superior, and *PM* posterior-medial as given in Figure 2) where the Dm effect is known to be prominent (Paller and Wagner, 2002). Significant spectral SME in the alpha band (7–12 Hz) has been robustly observed in various memory experiments (Klimesch et al., 1996b; Hanslmayr et al., 2009, 2012), hence spectral features were extracted (using the CSP algorithm) by learning the spatial patterns that best distinguish the alpha power difference between the two classes. The data suggested that the early and late alpha SME showed considerably different patterns. Hence the during-stimulus spectral classifier learned the power difference between the remembered and forgotten trials by combining the information from the two separate time windows (400–800 ms and 1000–1400 ms after stimulus presentation). The during-stimulus temporal and spectral classifier results were averaged to determine the final output of the during-stimulus classifier (denoted as $0 \leq p_B \leq 1$) for a given test trial. This value can be interpreted as the *during-stimulus classifier score* on the success of the encoding process.

The scores p_A and p_B from the pre- and during-stimulus classifiers were averaged and compared to the average score of the training set to determine the final label for a given test trial. A given trial was classified as remembered if $(p_A + p_B)/2 \geq (m_A + m_B)/2$ and forgotten if $(p_A + p_B)/2 < (m_A + m_B)/2$ where m_A and m_B are the mean pre- and during-stimulus classifier scores of the training set respectively. The classification accuracies for the pre- and during- classifiers were evaluated by comparing p_A to m_A and p_B to m_B respectively. More details on the classifier design can be found in Appendix A.1.

2.6. Temporal and spectral analyses

Temporal and spectral analyses were conducted in order to better understand the brain activity differences that are available for use by the three classifiers. Even though some channels were excluded from classification, all channels were considered here to reveal any significant SME across subjects. Significant SMEs were identified by conducting a non-parametric randomization test using cluster-based correction for multiple comparisons (Maris and Oostenveld, 2007). First, the test statistic between the remembered and forgotten trials was calculated for each sample (each time point for temporal analysis, each electrode position for spatial analysis). Clusters were then identified by finding adjacent samples with significant difference between the two conditions ($p < 0.05$). The cluster-level statistic was calculated by summing up these differences for each cluster and selecting the cluster with the maximum value. This result was compared to the cluster-based statistic of the permutation distribution generated from 10,000 random within subject permutations of trial labels (Maris and Oostenveld, 2007). In order to adjust for multiple tests across frequency bands in the pre-stimulus period, significant cluster-level statistics in adjacent frequency bands were summed and compared to the corresponding permutation distribution.

3. Results

3.1. Classification accuracy

Table 1 gives the classification accuracies for all 18 subjects. By combining the pre- and during-stimulus classifiers, the overall classification accuracy (calculated for all trials from the 18 subjects) achieved 59.64% which is approximately a 2% increase from the individual pre- and during-stimulus classifier results. The pre-stimulus and during-stimulus classifiers each gave individual classification results significantly over chance (significantly over 50% with $p < .05$) for 9 subjects with none going significantly below 50%. By combining the two time periods, we were able to achieve significantly over chance results for 13 subjects out of the 18 subjects. Significance level was calculated based on the total number of trials in the cross-validation and left-out sets for each subject (Müller-Putz et al., 2008; Agresti and Caffo, 2000) as described in Section 2.4.

Out of the 13 subjects with significantly over chance results, 8 subjects were self-reported car experts. However, there were no significant differences in accuracy for any of the classifiers between the two groups based on the Kruskal-Wallis test (pre-: $p = 0.33$, during-: $p = 0.79$, combined: $p = 0.92$), which should not be surprising since memory for both birds and cars was included in all analyses.

3.2. Temporal and spectral SME

Subsequent memory effects in the pre- and during-stimulus period were identified using methods given in Section 2.6. Oscillatory power in the pre-stimulus period was examined separately on 5 non-overlapping subbands (theta, alpha, low beta, high beta, and low gamma). For a given subband, within-subject averages of the power difference between the remembered and forgotten trials were calculated on all electrode positions. Afterwards, electrode positions with significantly large power difference for a given subband were identified by conducting a paired-sample t-test. This effect was adjusted for multiple comparison using the cluster-based correction explained in Section 2.6. The pre-stimulus period showed consistent positive spectral SME across subjects in the high beta (19–30 Hz) and low gamma (30–40 Hz) bands in the parietal electrodes as given in Figure 3.

The temporal during-stimulus classifier performance depends on the size of the Dm in channel groups CM, LPS, RPS, and PM within 400–800 ms. Time segments with significant Dm effect across subjects were identified based on the cluster-based analysis. Subject-specific ERPs were calculated for the two classes on all channel groups. Time points with significantly large Dm were identified by conducting a paired-sample t-test on the ERPs ($p < 0.05$). Cluster-based correction was used to adjust for multiple comparison. Channel groups LPS, RPS, and PM had significant Dm effects within this time segment as given in Figure 4.

Differences in alpha power between the remembered and forgotten trials were analyzed separately in the two time windows used for the during-stimulus spectral classifier (400–800 and 1000–1400 ms after stimulus onset). For each time window, the alpha event related desynchronization (ERD) (Pfurtscheller and Lopes da Silva, 1999) measurements for the remembered and forgotten trials were calculated using EEG power relative to the average power during the baseline period. Alpha power difference between the remembered and forgotten trials was defined as the difference of the ERD measurements between the two classes. For each subject, the average alpha power difference between the remembered and forgotten trials was calculated on all electrode positions. These values were used in the same manner as the pre-stimulus analysis to reveal clusters of channels that showed significant difference between the two classes. The two time windows gave significantly different scalp patterns as given in Figure 5. There was significantly stronger alpha desynchronization for

the forgotten trials compared to the remembered trials (positive spectral SME) in the left central area during the 400–800 ms window ($p < 0.05$); while there was significantly stronger alpha desynchronization for the remembered trials (negative spectral SME) in the posterior area during the 1000–1400 ms window ($p < 0.05$).

3.3. Classifier scores for all rating scale responses

We also examined the relationship between subjects' responses and classifier scores. Even though trials with *maybe familiar* and *definitely familiar* responses were excluded from the previous analysis due to a desire to maximize difference in encoding strength, we can acquire the classifier scores for these trials using the same classification procedure (see Section Appendix A.1 for details). The classifier score is a projection of the high-dimensional EEG data onto a 1-dimensional hyperplane which best discriminates between the remembered and forgotten classes. These hyperplanes (or projections) are defined by the features used by the different classifiers. Hence, it is possible to efficiently reveal underlying factors related to subsequent memory from the EEG data by examining the scores given from the different classifiers. This analysis was conducted on the combined classifier scores as well as the three individual classifier (pre-, during-temporal, and during-spectral) scores. Both analysis of variance (ANOVA) and the Kruskal-Wallis test were used to compare the classifier scores from the recollect trial to the 4 other responses. Since both tests gave similar results, we only report results based on the repeated measure ANOVA with Bonferroni adjustment for multiple comparisons on different responses and classifiers. The results are illustrated in Figure 6e.

For the combined classifier, *recollect* trials had mean score significantly different from all other responses ($p < 2 \times 10^{-4}$). For the pre-stimulus classifier, trials with *recollect* responses also had mean score significantly different from all other responses ($p < 9 \times 10^{-4}$). For the during-stimulus temporal classifier, trials with *recollect* responses had mean score significantly different from *maybe familiar* and *all unfamiliar* trials ($p < 2 \times 10^{-8}$). For the during-stimulus spectral classifier, trials with *recollect* responses also had mean score significantly different from *maybe familiar* and *all unfamiliar* trials ($p < 4 \times 10^{-5}$). These results indicate that the pre-stimulus classifier gives significantly smaller scores to the *definitely familiar* trials compared to the *recollect* group while the two during-stimulus classifiers map the *definitely familiar* trials closer to the *recollect* trials.

Since the pre-stimulus classifier combines information from multiple bands, each subband had to be isolated to examine how the different frequencies contributed to the difference in classifier scores between the different responses. It was revealed that the recollect trials had significantly larger mean score than the familiar trials between 25–35 Hz. This implies that the pre-stimulus classifier's ability to distinguish between recollect and definitely familiar trials is carried mostly by information in the high beta and low gamma bands. All mean scores and significant results from the ANOVA test are given in Table 2. Here, we only adjusted for multiple comparisons across the 4 response options and not across the multiple frequencies since the goal was to reveal underlying activities that may account for the effect found in the pre-stimulus scores.

The during-stimulus spectral classifier combines information from two distinct time windows (400–800 and 1000–1400 ms after stimulus onset). Hence, classifier scores were recomputed using classifiers trained on individual windows. The classifier scores for the early window (400–800 ms) showed similar values for the *recollect* and *definitely familiar* trials. However, the classifier scores for the later window (1000–1400 ms) were significantly different between the two responses ($p = 3 \times 10^{-4}$). All mean scores and significant results from the ANOVA test are given in Table 3.

4. Discussion

These results show that it is possible to successfully predict subsequent episodic memory performance based on single-trial scalp EEG activity recorded before and during item presentation. The prediction rate improved by 2%, by combining information from the pre- and during-stimulus periods. However, many factors can influence whether a subject will remember a stimulus, not all of which could be controlled in our study including how intrinsically memorable the stimulus is and the subject's brain state during the recognition phase. These factors add noise to the trial labels which may lower classifier accuracy.

- a. Combined: d-unfam ($P < 5 \times 10^{-20}$); m-unfam ($P < 9 \times 10^{-26}$); m-famil ($P < 7 \times 10^{-11}$); d-famil ($P < 2 \times 10^{-4}$).
- b. Pre-stimulus: d-unfam ($P < 8 \times 10^{-11}$); m-unfam ($P < 2 \times 10^{-12}$); m-famil ($P < 0.002$); d-famil ($P < 9 \times 10^{-4}$).
- c. During-stimulus temporal: d-unfam ($P < 2 \times 10^{-8}$); m-unfam ($P < 5 \times 10^{-12}$); m-famil ($P < 6 \times 10^{-11}$).
- d. During-stimulus spectral: d-unfam ($P < 2 \times 10^{-7}$); m-unfam ($P < 2 \times 10^{-10}$); m-famil ($P < 4 \times 10^{-5}$).
- e. The estimated means and the approximate 95 % confidence intervals of the classifier scores (Hochberg and Tamhane, 1987) for all 5 response options (*d-unfam: definitely unfamiliar, m-unfam: maybe unfamiliar, m-famil: maybe familiar, d-famil: definitely familiar, recollect*). Responses with significantly different means from the recollect trials are given with a star and the corresponding *P*-values are given below the figure. All results are based on the ANOVA test with Bonferroni adjustment for multiple comparisons.

There has not been any study that combines information from the pre- and during-stimulus periods of the data to predict subsequent memory, but the two time periods have been used to predict subsequent memory separately in two different fMRI studies. Watanabe et al. (2011) showed that it is possible to predict subsequent memory with approximately 66 % accuracy using fMRI data while subjects attend to the stimuli. Since EEG has lower spatial resolution compared to fMRI a lower prediction rate might be expected (56.8 % accuracy for the during-stimulus classifier). Also, it is difficult to separate out the brain signal prior to and during encoding using fMRI due to the slowness of the vascular response. Hence, the classifier may have incorporated information from the pre-stimulus as well as the during-stimulus period. The proportion of subjects with significantly over chance results in our study are comparable to that found by Watanabe et al. (2011) (6 out of 13 subjects¹ for Watanabe et al. (2011) and 13 out of 18 subjects for the current study).

Yoo et al. (2012) used the pre-stimulus period of the fMRI data to predict good/bad brain states for learning novel scenes. Their predictions gave 48.8 % hit rate (percentage of remembered items) during good brain states and 41.9 % hit rate (percentage of forgotten items) during bad brain states. Though it is difficult to directly compare the results due to the differences in the experiment paradigm and other settings such as recording technique, online/offline² setting etc., the results from the present study are numerically higher than the results from Yoo et al. (2012). The average hit rate during the good brain states (trials with p_A over 0.5) of the pre-stimulus classifier was 56.5 % while the average hit rate during the

¹This was computed by averaging over the main and confirmatory results given in Watanabe et al. (2011) with threshold for chance performance at 66.1 % which was calculated using methods given in Agresti and Caffo (2000).

²We refer to a system as *online* when it interprets the data and predicts the receptiveness of a subject to stimuli in real-time. An *offline* analysis uses data recorded from past experiments where subjects had no knowledge of the system's predictions.

bad brain states (trials with p_A below 0.5) was 42.0 %. The hit rate of a random selection of trials was 53.5 % across all subjects.

Table 5 shows how often each band was chosen for the pre-stimulus classifier. For example, the first value 0.82 in the table indicates that for subject S03, frequency band 4–7 Hz gave better than chance training error (and identified as *informative*) 82 % of the time over all cross-validation folds. There are individual differences in the frequency bands utilized by the pre-stimulus classifiers (Table 5). Subjects S26, S40 and S62 have no certain *informative* band that has better than chance training error. This suggests that these subjects' EEG data could be too noisy for the pre-stimulus classifier to work properly or the pre-stimulus EEG does not contain any useful information (Nijholt et al., 2008). Subjects S16, S20, S24, and S26 have at least one subband that is selected 60 % of the time, but the pre-stimulus accuracies are not significantly over chance. This suggests that the training set does not well represent the entire data set for these subjects. This may be due to non-stationarity in the data which may result in non-optimal CSP filters. A consistent cross-subject pre-stimulus spectral SME was only observed in the high beta and low gamma bands (Figure 3).

Our data did not show the significant theta power difference observed in Guderian et al. (2009). This may be due to the difference in timing of the pre-stimulus theta SME. Theta difference may occur earlier in the current study due to difference in experiment set-up. Fell et al. (2011) observed that power difference in the theta band occurred earlier in time than the higher frequencies. Also, Fellner et al. (2013) demonstrated that pre-stimulus theta SME occurred from –900 to –300 ms, but not immediately before stimulus onset. Hence if a majority of the subjects showed theta enhancement in the remembered trials prior to –300 ms before stimuli presentation, the data would not show significant SME in the theta band and only in the higher bands. The pre-stimulus SME observed in the higher frequencies supports this hypothesis. One other possibility is that, due to the small number of theta cycles possible in the 300 ms pre-stimulus window, the phase shifts may be confusable with power differences making the power differences related to subsequent memory difficult to detect.

Extra post-hoc spectral analysis in the during-stimulus window was conducted on additional frequencies to verify whether spectral SME found in previous studies could be identified in the current dataset. Analysis on the theta (4–7 Hz), low beta (12–19 Hz), and high gamma (55–70 Hz) bands revealed that 1) the positive theta SME within the posterior area in the 200–600 ms window and 2) the negative low beta SME within the posterior area within the 800–1200 ms window were significant ($p < 0.05$) as given in Figure 6. These results agree with findings in Hanslmayr et al. (2009, 2012). Single-trial analysis was conducted on the theta (4–7 Hz), low beta (12–19 Hz) band features to confirm whether information in those bands were classifiable. The overall classification results were 49.3 % for the theta band and 53.0 % for the low beta band. The during-stimulus theta classifier gave significantly lower results than the two during-stimulus alpha classifiers based on the rank sum test ($p = 0.001$) suggesting that the theta band features were not appropriate for single-trial classification. The during-stimulus low beta classifier gave slightly lower accuracy than the two during-stimulus alpha classifiers but the results were not significantly different ($p = 0.87$). However, adding the low beta features to the classifier gave an overall accuracy of 59.03 % which did not improve the overall classification results. The reason the theta SME did not give useful features for single-trial analysis may be due to the early timing of the effect (200–600 ms). The subjects' responses to the stimulus itself may act as artifacts on a single-trial basis, whereas this aspect of the brain activity is diminished when the SME is computed on all available trials. Also the single-trial phase shifts may add noise to the power estimation in the 400 ms window. The low beta band features may partially be present in the late alpha band features (1000–1400 ms) due to the spectral/temporal proximity and spatial

similarity (negative spectral SME in the posterior area) of the two features. This may explain why the overall classification does not improve by including the beta band features in the during-stimulus spectral classifier.

The alpha SME during 400–800 ms gave considerably different patterns from the alpha SME during 1000–1400 ms (given in Figure 5). The negative SME in the posterior area found between 1000–1400 ms is consistent with previous studies (Klimesch et al., 1996b; Hanslmayr et al., 2009, 2012). The early positive alpha SME may be related to previous findings which showed that high alpha power over task-irrelevant regions is important for the participants to perform optimally in covert attention tasks (Händel et al., 2011; Haegens et al., 2012). Thus, the early during-stimulus spectral classifier may be utilizing information reflecting attention. The asymmetric alpha power difference between the remembered and forgotten trials may be due to increased activity associated with the left hemisphere such as subvocal speech (or internal thoughts) during the forgotten trials (Ehrlichman and Wiener, 1980) which could interfere with the visual encoding task.

The classifiers were originally trained to give high scores for the *recollected* trials and low scores for the *unfamiliar* trials. However, the different classifiers showed interesting trends on their classification of the untrained *definitely familiar* trials. The during-stimulus temporal scores (Figure 6e (c)) and spectral scores from the 400–800 ms window (1st row in Table 3) did not distinguish between the *recollected* and *definitely familiar* trials while the pre-stimulus spectral scores between 25–35 Hz (8th row in Table 2) and the during-stimulus spectral scores from the 1000–1400 ms window (2nd row in Table 3) gave significantly lower scores to the *definitely familiar* trials than the *recollected* trials. Subsequent analyses showed that the *definitely familiar* scores were significantly higher than the *unfamiliar* trials for the first group of classifiers while there were no significant differences for the second group as given in Table 4. Moreover, it was found that the *definitely familiar* scores given by the first group were significantly higher than the second group ($p < 10^{-7}$) (values in column 3 of Table 4). Thus, the familiarity judgments revealed that the different classifiers are utilizing distinct neural processes for their classification of subsequent memory.

Recent research has raised doubts about the extent to which remember/familiar judgments can be used to estimate separate recollection and familiarity processes rather than merely reflecting confidence differences attributable to a single continuously varying memory signal (Dunn, 2004; Rotello et al., 2005; Wixted and Stretch, 2004). The scores from the first group of classifiers seem consistent with the continuous confidence perspective because both of the high confidence “old” responses (*definitely familiar* and *recollect*) gave significantly higher scores than the *unfamiliar* trials, but there were no significant differences between *definitely familiar* and *recollect* trials. On the other hand, the second group of classifiers showed a pattern that seem to differentiate only *recollect* responses from all other responses (without being sensitive to gradations in confidence between the *familiar* and *unfamiliar* trials). Thus, EEG differences in the –300 to 0 ms window (specifically oscillatory activity between 25–35Hz) and alpha activity between 1000–1400 ms appear to be differentiating subsequent familiarity from recollection in a manner that is not synonymous with confidence, so may reflect aspects of encoding preparation and processes that would differentiate these responses. For example, although contextual influences on familiarity have been demonstrated (Addante et al., 2012; Elfman et al., 2008; Mollison and Curran, 2012; Speer and Curran, 2007), contextual influences are widely regarded to be stronger on recollection than familiarity (Davachi et al., 2001; Cansino et al., 2002; Ranganath et al., 2004; Duarte et al., 2004; Summerfield and Mangels, 2005). Perhaps pre-stimulus activity between 25–35 Hz is important for encoding contextual information, which may include contextual information taken from the pre-stimulus period itself (e.g., whatever the subject was thinking about prior to encoding). Also, during stimulus presentation, the

brain activity may shift from encoding the stimulus early in the trial to also encoding the contextual information in that period.

We cannot completely rule out the possibility that the pre-stimulus classifier may be using the brain activity of the evoked response to the fixation-cross rather than the ongoing pre-stimulus neural activities for classification. However the pre-stimulus ERP did not show any significant difference between the remembered and forgotten trials. This decreases the possibility that the evoked response from the fixation-cross holds any information that discriminates between the two classes. In a follow-up study, the effects of these different signals on classification results will be further investigated using an appropriate experiment paradigm.

In summary, this study shows that pre- and during-stimulus EEG can be used to predict subsequent memory performance. We discovered that the pre-stimulus classifier (especially in frequencies around 25–35 Hz) using the –300–0 ms window and during-stimulus alpha band classifier using the 1000–1400 ms window distinguished recollection from familiarity, whereas the during-stimulus temporal and alpha band classifiers using the 400–800 ms did not. These results suggest that 1) the brain activity before item presentation contributes to how well context gets encoded with the upcoming item and 2) the brain activity during item presentation initially focuses on item encoding then shifts to also encoding the contextual information. Finally, these findings could provide an inexpensive and non-invasive way to monitor learning preparedness to optimally determine the time to present a stimulus and present the stimulus again at a later time point if the encoding process is unsuccessful.

Acknowledgments

This research was funded by NSF grants # CBET-0756828 and # IIS-1219200, NIH Grant MH64812, NSF grants # SBE-0542013 and # SMA-1041755 to the Temporal Dynamics of Learning Center (an NSF Science of Learning Center), and a James S. McDonnell Foundation grant to the Perceptual Expertise Network, and the KIBM (Kavli Institute of Brain and Mind) Innovative Research Grant. We would like to thank Marta Kutas and Tom Urbach for helpful comments on the manuscript.

References

- Addante RJ, Ranganath C, Yonelinas A. Examining ERP correlates of recognition memory: Evidence of accurate source recognition without recollection. *Neuroimage*. 2012; 62:439–450. [PubMed: 22548808]
- Agresti A, Caffo B. Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. *The American Statistician*. 2000; 54:280–288.
- Blankertz B, Lemm S, Treder MS, Haufe S, Müller KR. Single-trial analysis and classification of ERP components – a tutorial. *NeuroImage*. 2011; 56:814–825. [PubMed: 20600976]
- Blankertz B, Tomioka R, Lemm S, Kawanabe M, Müller KR. Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal Processing Magazine*. 2008; 25:41–56.
- Cansino S, Maquet P, Dolan RJ, Rugg MD. Brain activity underlying encoding and retrieval of source memory. *Cerebral Cortex*. 2002; 12:1048–1056. [PubMed: 12217968]
- Chang CC, Lin CJ. Training ν -support vector classifiers: Theory and algorithms. *Neural Computation*. 2001; 13:2119–2147. [PubMed: 11516360]
- Chang CC, Lin CJ. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2. 2011; 27:1–27. 27.
- Davachi L, Maril A, Wagner AD. When keeping in mind supports later bringing to mind: Neural markers of phonological rehearsal predict subsequent remembering. *J Cognitive Neuroscience*. 2001; 13:1059–1070.

- Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods*. 2004; 134:9–21. [PubMed: 15102499]
- Doppelmayr M, Klimesch W, Pachinger T, Ripper B. Individual differences in brain dynamics: important implications for the calculation of event-related band power. *Biological Cybernetics*. 1998; 79:49–57. [PubMed: 9742677]
- Dornhege G, Blankertz B, Curio G, Müller KR. Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms. *IEEE Trans Biomed Eng*. 2004; 51:993–1002. [PubMed: 15188870]
- Duarte A, Ranganath C, Winward L, Hayward D, Knight RT. Dissociable neural correlates for familiarity and recollection during the encoding and retrieval of pictures. *Cognitive Brain Research*. 2004; 18:255–272. [PubMed: 14741312]
- Dümbgen L, Igl BW, Munk A. P-values for classification. *Electronic Journal of Statistics*. 2008; 2:468–493.
- Dunn JC. Remember-know: A matter of confidence. *Psychological Review*. 2004; 111:524–542. [PubMed: 15065921]
- Ehrlichman H, Wiener MS. EEG asymmetry during covert mental activity. *Psychophysiology*. 1980; 17:228–235. [PubMed: 7384372]
- Elfman KW, Parks CM, Yonelinas AP. Testing a neurocomputational model of recollection, familiarity, and source recognition. *Journal of Experimental Psychology Learning, Memory, and Cognition*. 2008; 34:752–768.
- Fell J, Ludowig E, Staresina BP, Wagner T, Kranz T, Elger CE, Axmacher N. Medial temporal theta/alpha power enhancement precedes successful memory encoding: evidence based on intracranial EEG. *Journal of Neuroscience*. 2011; 31:5392–5397. [PubMed: 21471374]
- Fellner MC, Bäuml KHT, Hanslmayr S. Brain oscillatory subsequent memory effects differ in power and long-range synchronization between semantic and survival processing. *NeuroImage*. 2013; 79:361–370. [PubMed: 23664950]
- Gruber MJ, Otten LJ. Voluntary control over prestimulus activity related to encoding. *Journal of Neuroscience*. 2010; 30:9793–9800.10.1523/JNEUROSCI.0915-10.2010 [PubMed: 20660262]
- Gruber MJ, Watrous AJ, Ekstrom AD, Ranganath C, Otten LJ. Expected reward modulates encoding-related theta activity before an event. *NeuroImage*. 2013; 0:68–74. [PubMed: 22917987]
- Guderian S, Schott BH, Richardson-Klavehn A, Duzel E. Medial temporal theta state before an event predicts episodic encoding success in humans. *Proceedings of the National Academy of Sciences*. 2009; 106:5365–5370.
- Haegens S, Luther L, Jensen O. Somatosensory anticipatory alpha activity increases to suppress distracting input. *J Cognitive Neuroscience*. 2012; 24:677–685.
- Hammon PS, de Sa VR. Pre-processing and meta-classification for brain-computer interfaces. *IEEE Trans Biomed Eng*. 2007; 54:518–525. [PubMed: 17355065]
- Händel BF, Haarmeier T, Jensen O. Alpha oscillations correlate with the successful inhibition of unattended stimuli. *J Cognitive Neuroscience*. 2011; 23:2494–502.
- Hanslmayr S, Spitzer B, Bäuml KH. Brain oscillations dissociate between semantic and nonsemantic encoding of episodic memories. *Cerebral Cortex*. 2009; 19:1631–1640. [PubMed: 19001457]
- Hanslmayr S, Staudigl T. How brain oscillations form memories - a processing based perspective on oscillatory subsequent memory effects. *NeuroImage*. 2013
- Hanslmayr S, Staudigl T, Fellner MC. Oscillatory power decreases and long-term memory: The information via desynchronization hypothesis. *Frontiers in Human Neuroscience*. 2012;6. [PubMed: 22375109]
- Herzmann G, Curran T. Experts' memory: an ERP study of perceptual expertise effects on encoding and recognition. *Memory & Cognition*. 2011; 39:412–32. [PubMed: 21264603]
- Hochberg, Y.; Tamhane, AC. Multiple comparison procedures. John Wiley & Sons, Inc; New York, NY, USA: 1987.
- Hyvärinen, A.; Karhunen, J.; Oja, E. Independent Component Analysis. Wiley; New York: 2001.

- Johnson, RJ. Event-related potential insights into the neurobiology of memory systems. In: Boller, F.; Grafman, J., editors. *The Handbook of Neuropsychology*. Vol. 10. Amsterdam: Elsevier Science Publishers; 1995. p. 134-164.
- Klimesch W, Doppelmayr M, Russegger H, Pachinger T. Theta band power in the human scalp eeg and the encoding of new information. *Neuroreport*. 1996a; 17:1235–1240. [PubMed: 8817539]
- Klimesch W, Schimke H, Doppelmayr M, Ripper B, Schwaiger J, Pfurtscheller G. Event-related desynchronization (ERD) and the Dm effect: Does alpha desynchronization during encoding predict later recall performance? *International Journal of Psychophysiology*. 1996b; 24:47–60. [PubMed: 8978435]
- Makeig S, Bell AJ, Jung TP, Sejnowski TJ. Independent component analysis of electroencephalographic data. *Advances in Neural Information Processing Systems*. 1996; 8:145–151.
- Maris E, Oostenveld R. Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*. 2007; 164:177–190. [PubMed: 17517438]
- Mollison MV, Curran T. Familiarity in source memory. *Neuropsychologia*. 2012; 50:2546–2565. [PubMed: 22789677]
- Müller-Putz G, Scherer R, Brunner C, Leeb R, Pfurtscheller G. Better than random? a closer look on BCI results. *International Journal of Bioelectromagnetism*. 2008; 10:52–55.
- Nijholt A, Tan D, Pfurtscheller G, Brunner C, del R, Millán J, Allison B, Graimann B, Popescu F, Blankertz B, Müller KR. Brain-computer interfacing for intelligent systems. *IEEE Intelligent Systems*. 2008; 23:72–79. <http://doi.ieeecomputersociety.org/10.1109/MIS.2008.41>.
- Otten LJ, Quayle AH, Akram S, Ditlew TA, Rugg MD. Brain activity before an event predicts later recollection. *Nature Neuroscience*. 2006; 9:489–491.
- Otten LJ, Quayle AH, Puvaneswaran B. Prestimulus subsequent memory effects for auditory and visual events. *Journal of Cognitive Neuroscience*. 2010; 22:1212–1223. [PubMed: 19583467]
- Paller KA, Kutas M, Mayes AR. Neural correlates of encoding in an incidental learning paradigm. *Electroencephalography and Clinical Neurophysiology*. 1987; 67:360–371.10.1016/0013-4694(87)90124-6 [PubMed: 2441971]
- Paller KA, Wagner AD. Observing the transformation of experience into memory. *Trends in Cognitive Sciences*. 2002; 6:93–102. [PubMed: 15866193]
- Park H, Rugg MD. Neural correlates of encoding within- and across-domain inter-item associations. *Journal of Cognitive Neuroscience*. 2010; 9:2533–2543.
- Pfurtscheller G, Lopes da Silva FH. Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical neurophysiology*. 1999; 110:1842–1857. [PubMed: 10576479]
- Ranganath C, Yonelinas AP, Cohen MX, Dy CJ, Tom SM, D'Esposito M. Dissociable correlates of recollection and familiarity within the medial temporal lobes. *Neuropsychologia*. 2004; 42:2–13. [PubMed: 14615072]
- Rotello CM, Macmillan NA, Reeder JA, Wong M. The remember response: Subject to bias, graded, and not a process-pure indicator of recollection. *Psychonomic Bulletin & Review*. 2005; 12:865–873. [PubMed: 16524003]
- Sanquist TF, Rohrbaugh JW, Syndulko K, Lindsley DB. Electrocortical signs of levels of processing: perceptual analysis and recognition memory. *Psychophysiology*. 1980; 17:568–576. [PubMed: 7443924]
- Sederberg PB, Kahana MJ, Howard MW, Donner EJ, Madsen JR. Theta and gamma oscillations during encoding predict subsequent recall. *Journal of Neuroscience*. 2003; 23:10809–10814. [PubMed: 14645473]
- Speer NK, Curran T. ERP correlates of familiarity and recollection processes in visual associative recognition. *Brain Research*. 2007; 1174:97–109. [PubMed: 17825273]
- Staudigl T, Hanslmayr S. Theta oscillations at encoding mediate the context-dependent nature of human episodic memory. *Current Biology*. 2013; 23:1101–1106. [PubMed: 23746635]
- Summerfield C, Mangels JA. Coherent theta-band EEG activity predicts item-context binding during encoding. *NeuroImage*. 2005; 24:692–703. [PubMed: 15652304]

- Watanabe T, Hirose S, Wada H, Katsura M, Chikazoe J, Jimura K, Imai Y, Machida T, Shirouzu I, Miyashita Y, Konishi S. Prediction of subsequent recognition performance using brain activity in the medial temporal lobe. *NeuroImage*. 2011; 54:3085–3092. [PubMed: 21035553]
- Wixted JT, Stretch V. In defense of the signal-detection interpretation of remember/know judgments. *Psychonomic Bulletin & Review*. 2004; 11:616–641. [PubMed: 15581116]
- Yoo JJ, Hinds O, Ofen N, Thompson TW, Whitfield-Gabrieli S, Tri-antafyllou C, Gabrieli JD. When the brain is prepared to learn: Enhancing human learning using real-time fMRI. *NeuroImage*. 2012; 59:846–852.10.1016/j.neuroimage.2011.07.063 [PubMed: 21821136]

Appendix A. Appendix

Appendix A.1. Classifier training procedure

Depending on the performance (recollection rate) of each subject, the difference between the number of trials for the remembered class and the forgotten class ranged from 1 to 82. Rather than discarding subjects with unbalanced classes (Watanabe et al., 2011), enough trials from the larger class were set aside from training as the *left-out* set to balance the number of trials per class in the *cross-validation* set. Trials in the left-out set were evenly distributed over time (epochs and blocks) to minimize the effect of drift or bias in the cross-validation set. The cross-validation set was evaluated based on a balanced leave-two-out cross-validation procedure where one example from each class is randomly selected and left out of *any* training procedure as the validation set (to ensure they were not used in any manner to train the classifier) while the remaining trials are used as the training set for each fold. The left-out set was evaluated using the classifier trained from all trials in the cross-validation set. This procedure allowed us to eliminate any effect from unbalanced classes during classifier training while conducting classification on all available trials. The classifiers to compute the classifier scores for trials with *definitely familiar* and *maybe familiar* responses were also trained for each subject using all trials in the cross-validation set.

Appendix A.1.1. Pre-stimulus classifier

Zero-phase filtering was used to extract desired subband signals while preserving the timing of the features from the pre-stimulus period. Since a non-causal filter was used, the 300 ms subsequence preceding the to-be-learned stimulus was extracted before filtering to prevent any temporal smearing from the signal during actual encoding. 25 extra samples in the 100 ms period before the fixation cross were included to estimate a better covariance matrix for CSP analysis. 20 tap zero-phase FIR filters were used to design the 9 bandpass filters (4–7 Hz, 6–10 Hz, 7–12 Hz, 10–15 Hz, 12–19 Hz, 15–25 Hz, 19–30 Hz, 25–35 Hz, 30–40 Hz). Nine separate passband signals were generated for each trial through this procedure.

Separate classifiers were constructed using the training sets of the 9 subbands. For each subband group, CSP filters were learned to extract features that maximally discriminate between the remembered (class 1) and forgotten (class 2) trials. CSP is a supervised dimensionality reduction algorithm commonly used for EEG classification. CSP utilizes the covariance matrices of the two classes (estimated from the bandpass filtered EEG data) to find spatial filters that maximize the variance of spatially filtered signals under one condition while minimizing it for the other condition. The 73 channels of EEG data were used to estimate the spatial filters. Three spatial filters were selected from each class resulting in 6 filtered signals as in Blankertz et al. (2008). The log power was calculated by

$$P_i = \frac{1}{T} \log \sum_{t=1}^T s_{i,t}^2 \quad (\text{A.1})$$

where $s_{i,t}$ is the sample for time t from filtered signal i ($i = 1, \dots, 6$ and $t = 1, \dots, T$ where T is the number of samples within an example). This resulted in a 6 dimensional vector $P = [P_1, \dots, P_6]$ for each trial.

The soft margin³ support vector classifier machine (ν -SVM) (Chang and Lin, 2001) with a linear kernel was used to classify the 6 dimensional vectors. LIBSVM (Chang and Lin, 2011) was utilized for this part of the simulation. The parameter $0 < \nu < 1$ can be interpreted as an upper bound on the proportion of margin errors and the lower bound on proportion of support vectors. ν was selected based on a 4-fold cross-validation on the set $\{P\}$ acquired from the training set.

The training error for each subband group was calculated by conducting a balanced cross-validation on the *training set*. Subband groups that gave better than chance (with $p < 0.10$) training error were identified as *informative*. If none of the subbands gave better than chance training error, all 9 subbands were selected. The decision of the pre-stimulus classifier for a given trial in the validation or left-out set (p_A) was determined by averaging over the scores given by SVM classifiers from all *informative* subbands. This meta-classification approach was used based on previous studies which found that meta-classification strategies generally outperform single classifiers (Dornhege et al., 2004; Hammon and de Sa, 2007).

Appendix A.2. During-stimulus classifier

Different bandpass filters and spatial filters were used to extract features for the during-stimulus temporal and spectral classifiers.

In order to learn the ERP patterns of the Dm effect, the baselined signal (baseline offset corrected using -200 to 0 ms of each trial) was bandpass filtered between 0.1 – 5 Hz using a 40 tap zero-phase FIR filter. Based on previous research on the Dm, the 400 – 800 ms time window and four channel groups were selected for evaluation (*CM* centro medial, *LPS* left posterior superior, *RPS* right posterior superior, and *PM* posterior-medial as given in Figure 2). Mean amplitudes for each channel group were calculated by averaging over the channels within each group. For each channel group, a 5-dimensional template for remembered/forgotten trials were calculated. First, the ERP of the training set was calculated for each class. The dimensionality of the ERP was reduced to 5 by averaging over 80 ms length non-overlapping windows between 400 – 800 ms. Finally, templates from all channel groups were concatenated to create a 20-dimensional template for remembered/forgotten trials. A soft margin⁴ linear classifier using LDA (linear discriminant analysis) was trained based on these templates and the dispersion of the training examples. LDA is a simple classifier which is commonly used to classify ERP components (Blankertz et al., 2011).

In order to isolate the alpha band of the EEG signal, the baselined signal (baseline offset corrected using -200 to 0 ms of each trial) was bandpass filtered between 7 – 12 Hz with a 40

³The soft margin SVM classifier for a two class classification problem gives a pair of scores (P_1 and P_2) corresponding to the probability of potential class membership where $P_1 + P_2 = 1$. Here, we consider the output of the classifier to be $P = P_1$ which represents the probability an example is a remembered trial (classified as remembered if $P > 0.5$ and forgotten if $P < 0.5$).

⁴The probability output for the soft margin LDA classifier was calibrated based on a permutation test with plug-in estimator of Bayesian likelihood ratios for the standard homoscedastic Gaussian model (Dümbgen et al., 2008). As in the soft margin SVM classifier, the classifier gives a pair of numbers (P_1 and P_2) corresponding to the probability of class membership. We consider the output of the classifier to be $P = P_1$ which represents the probability an example is a remembered trial.

tap zero-phase FIR filter. The data were divided into two time windows (400–800 and 1000–1400 ms after the cue). For each time window, 6 CSP filters (3 for each class) were learned using the 73 channel EEG data and the log powers of the spatially filtered signals were computed. The log power values were combined to acquire a 12 dimensional feature vector for each trial. The soft margin ν -SVM with a linear kernel was used for classification. The CSP procedure, log power calculation, and ν parameter selection followed the procedures given in Appendix A.1.1.

The decision of the during-stimulus classifier (p_B) was determined by averaging over the scores given by the temporal and spectral classifiers.

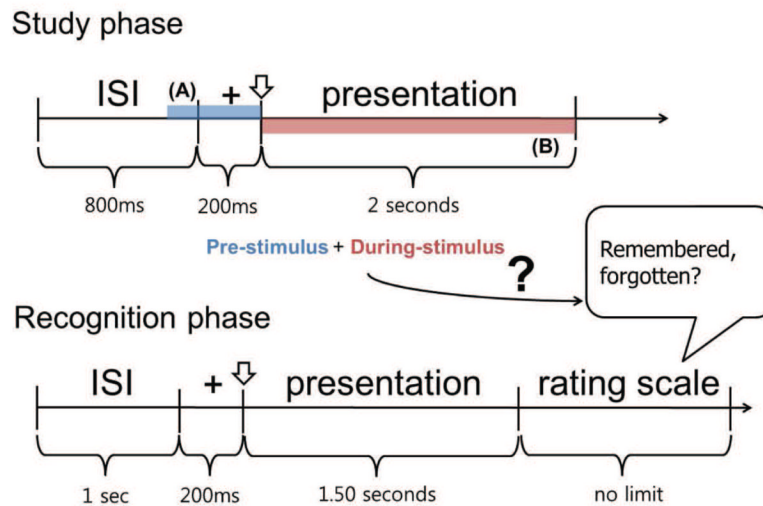


Figure 1. Timing of the visual memory task. The two shaded areas of the study phase noted as (A) and (B) are the pre- and during-stimulus periods considered in our analysis (colored in blue and red respectively). The goal of the classifier is to predict whether the subject remembers a given stimulus using the pre- and during-stimulus EEG of each presentation in the study phase.

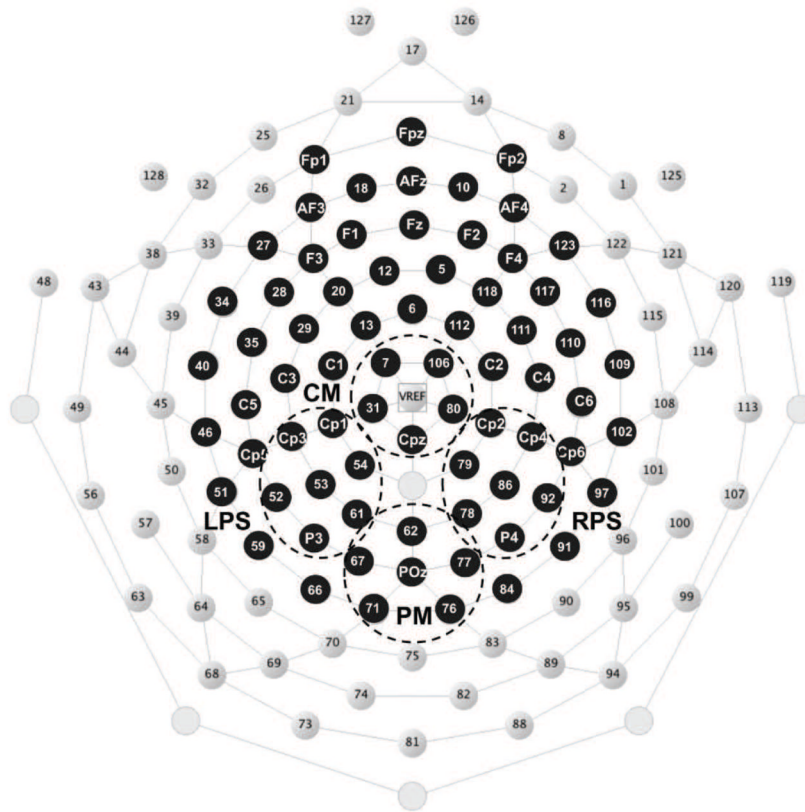


Figure 2. The GSN electrode locations with the 73 electrodes used for analysis are highlighted in black. These electrode locations are an approximate equivalent of the 10–20 system. The four channel groups are regions of interest used by the temporal during-stimulus classifier. *CM* centro medial, *LPS* left posterior superior, *RPS* right posterior superior, *PM* posterior-medial.

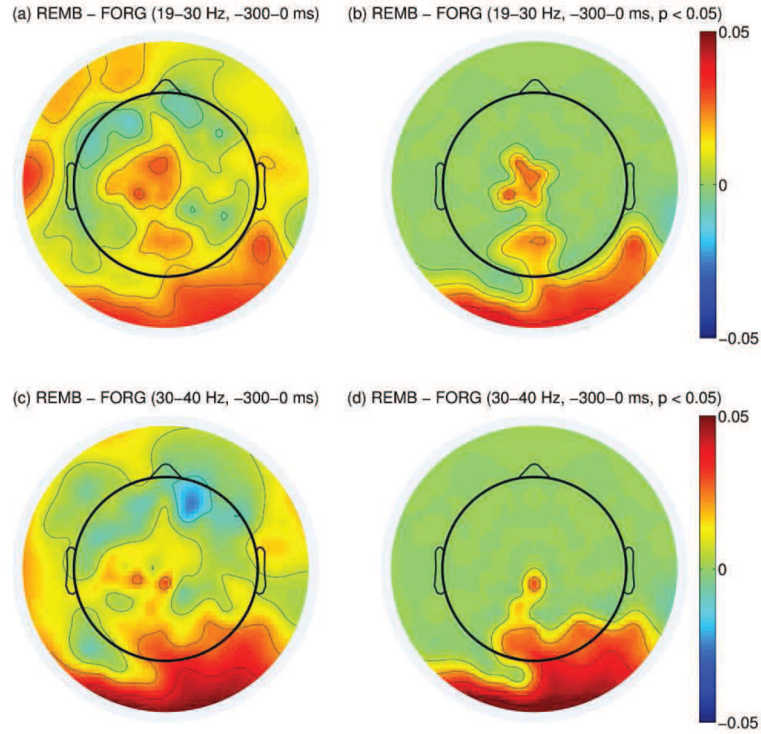


Figure 3. (a): Difference in high beta power between the remembered and forgotten trials between $-300-0$ ms before stimulus presentation ($\log(\mu V^2)$). (b): Same topography as in (a) but masked by the spatial pattern of the most significant cluster resulting from cluster-based analysis across all subjects ($p < 0.05$). (c): Difference in low gamma power between the remembered and forgotten trials between $-300-0$ ms before stimulus presentation. (d): Same topography as in (c) but masked by the spatial pattern of the most significant cluster resulting from cluster-based analysis across all subjects ($p < 0.05$).

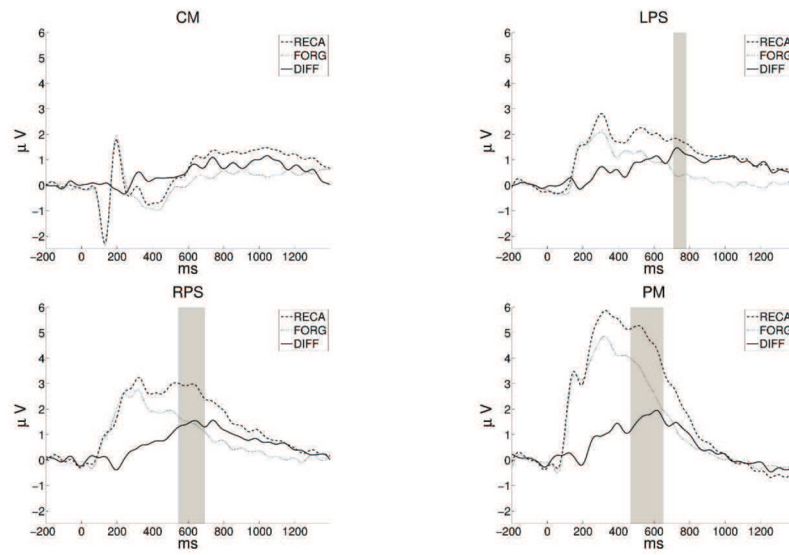


Figure 4. Mean amplitudes for remembered/forgotten trials across channels groups CM, LPS, RPS, and PM. Portions with significant effects resulting from cluster-based analysis are shaded in gray ($p < 0.01$).

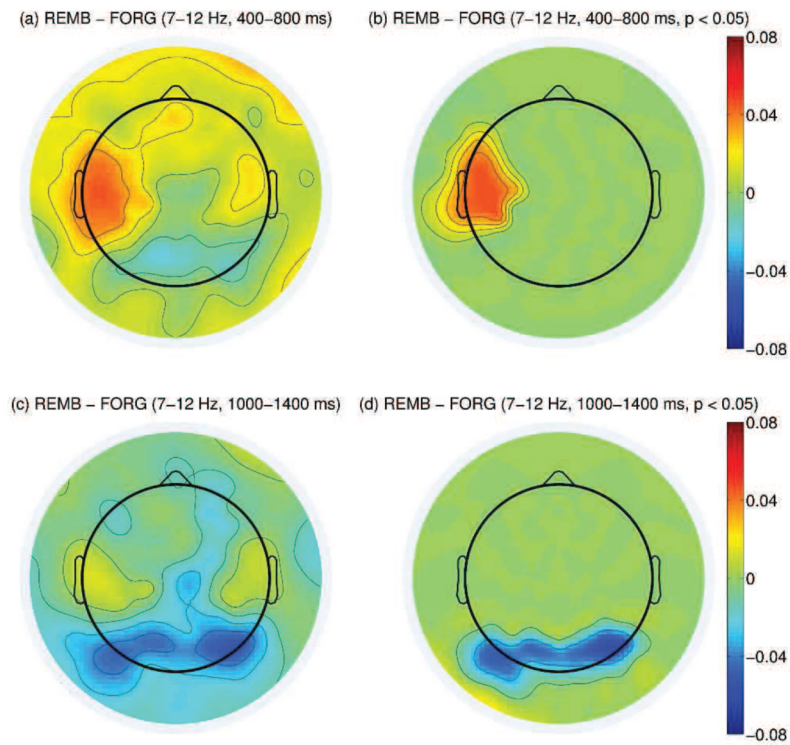
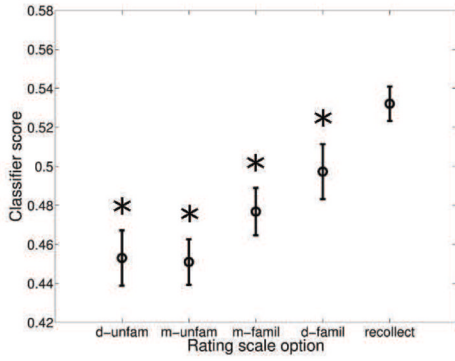
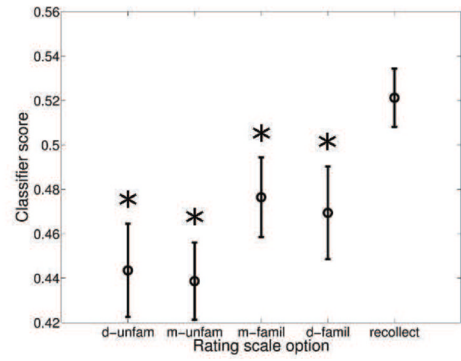


Figure 5.

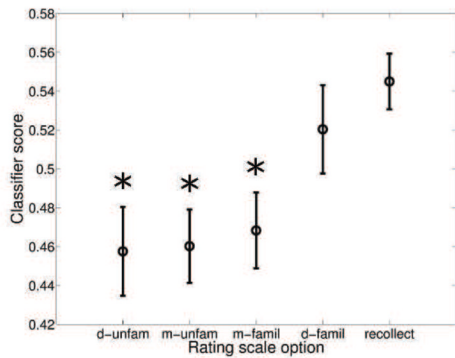
(a): Difference in alpha power between the remembered and forgotten trials between 400–800 ms after stimulus onset ($\log(\mu V^2)$). (b): Same topography as in (a) but masked by the spatial pattern of the most significant cluster resulting from cluster-based analysis across all subjects ($p < 0.05$). (c): Difference in alpha power between the remembered and forgotten trials between 1000–1400 ms after stimulus onset. (d): Same topography as in (c) but masked by the spatial pattern of the most significant cluster resulting from cluster-based analysis across all subjects ($p < 0.05$).



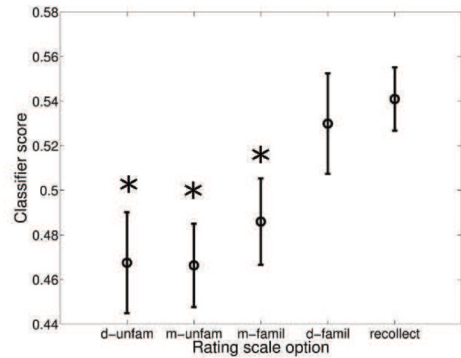
(a) Combined: d-unfam ($p < 5 \times 10^{-20}$); m-unfam ($p < 9 \times 10^{-26}$); m-famil ($p < 7 \times 10^{-11}$); d-famil ($p < 2 \times 10^{-4}$).



(b) Pre-stimulus: d-unfam ($p < 8 \times 10^{-11}$); m-unfam ($p < 2 \times 10^{-12}$); m-famil ($p < 0.002$); d-famil ($p < 9 \times 10^{-4}$).



(c) During-stimulus temporal: d-unfam ($p < 2 \times 10^{-8}$); m-unfam ($p < 5 \times 10^{-12}$); m-famil ($p < 6 \times 10^{-11}$).



(d) During-stimulus spectral: d-unfam ($p < 2 \times 10^{-7}$); m-unfam ($p < 2 \times 10^{-10}$); m-famil ($p < 4 \times 10^{-5}$).

Figure 6.

The estimated means and the approximate 95 % confidence intervals of the classifier scores (Hochberg and Tamhane, 1987) for all 5 response options (*d-unfam*: definitely unfamiliar, *m-unfam*: maybe unfamiliar, *m-famil*: maybe familiar, *d-famil*: definitely familiar, *recollect*). Responses with significantly different means from the recollect trials are given with a star and the corresponding p -values are given below the figure. All results are based on the ANOVA test with Bonferroni adjustment for multiple comparisons.

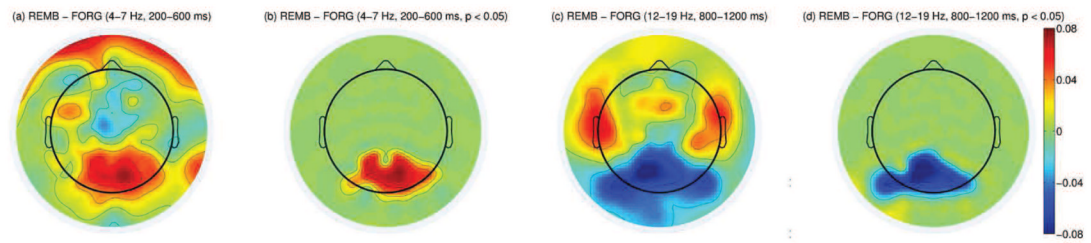


Figure 7.

(a): Difference in theta power between the remembered and forgotten trials between 200–600 ms after stimulus onset ($\log(\mu V^2)$). (b): Same topography as in (a) but masked by the spatial pattern of the most significant cluster resulting from cluster-based analysis across all subjects ($p < 0.05$). (c): Difference in low beta power between the remembered and forgotten trials between 800–1200 ms after stimulus onset. (d): Same topography as in (c) but masked by the spatial pattern of the most significant cluster resulting from cluster-based analysis across all subjects ($p < 0.05$).

Table 1

Average classification accuracy from the pre-stimulus, during-stimulus, and pre-during combined classifiers. Results significantly over chance (based on the number of trials used for classification) are given with their corresponding p -values. The last column gives the number of trials from each class before dividing into cross-validation and left-out sets (R: remembered/F: forgotten). Car experts and novices are noted as (E) and (N), respectively. Overall accuracies given in the last row are the accuracies over all trials considered for classification.

Subject	Pre- (%)	During- (%)	Combined (%)	# trials (R/F)
S03 (E)	58.85($p = 0.010$)	59.81($p = 0.005$)	61.72($p = 7 \times 10^{-4}$)	144/65
S06 (E)	58.06($p = 0.011$)	56.05	58.87($p = 0.005$)	117/131
S10 (E)	55.82	52.21	59.04($p = 0.004$)	104/145
S15 (N)	58.29($p = 0.022$)	53.48	57.75($p = 0.033$)	125/62
S16 (N)	52.00	60.00($p = 0.005$)	56.00	112/88
S17 (E)	58.86($p = 0.018$)	55.43	58.86($p = 0.018$)	84/91
S20 (E)	57.25	57.97	60.87($p = 0.010$)	71/67
S22 (N)	57.05	63.46($p = 7 \times 10^{-4}$)	56.41	94/62
S24 (N)	55.80	60.14($p = 0.016$)	62.32($p = 0.004$)	68/70
S26 (E)	51.88	54.89	55.64	68/65
S40 (N)	52.66	51.21	54.11	75/132
S51 (E)	62.14($p = 2 \times 10^{-4}$)	63.79($p = 2 \times 10^{-5}$)	66.26($p = 4 \times 10^{-7}$)	122/121
S52 (N)	57.80($p = 0.038$)	63.58($p = 4 \times 10^{-4}$)	71.10($p = 2 \times 10^{-8}$)	90/83
S56 (E)	59.11($p = 0.009$)	65.02($p = 2 \times 10^{-5}$)	61.08($p = 0.002$)	121/82
S57 (N)	61.96($p = 0.002$)	55.83	64.42($p = 2 \times 10^{-4}$)	94/69
S59 (E)	62.24($p = 2 \times 10^{-4}$)	57.68($p = 0.016$)	59.75($p = 0.003$)	123/118
S61 (N)	56.47	53.53	58.82($p = 0.020$)	85/85
S62 (E)	50.44	58.41($p = 0.011$)	51.77	154/72
Overall	57.16	57.88	59.64	

Table 2

The mean scores given by the pre-stimulus classifiers trained on the 9 separate bandpass filtered data. Repeated measure ANOVA was conducted between recollect trials and the 4 other response options. Significant *p*-values after Bonferroni adjustment for multiple comparisons are given with *superscripts (*: $p < 0.012$, **: $p < 0.005$, ***: $p < 0.001$, ****: $p < 0.0001$, *****: $p < 0.00001$).

	<i>Recollect</i>	<i>Def fam</i>	<i>Maybe fam</i>	<i>Maybe unfam</i>	<i>Def unfam</i>
4–7 Hz	0.506	0.512	0.506	0.467**	0.459**
6–10 Hz	0.506	0.500	0.493	0.466**	0.454****
7–12 Hz	0.505	0.498	0.492	0.468**	0.459**
10–15 Hz	0.511	0.488	0.487	0.472*	0.474**
12–19 Hz	0.511	0.498	0.496	0.461***	0.482
15–25 Hz	0.499	0.500	0.478	0.462**	0.471*
19–30 Hz	0.492	0.464	0.463	0.457*	0.481
25–35 Hz	0.511	0.449*****	0.460****	0.463*****	0.466*****
30–40 Hz	0.496	0.456	0.461	0.464	0.478

Table 3

The mean scores given by the during-stimulus spectral classifiers trained on the individual time windows. Repeated measure ANOVA was conducted between recollect trials and the 4 other response options. Significant p -values after Bonferroni adjustment for multiple comparisons are given with * superscripts (*: $p < 10^{-3}$, **: $p < 10^{-4}$, ***: $p < 10^{-5}$).

	Recollect	Def fam	Maybe fam	Maybe unfam	Def unfam
400–800 ms	0.543	0.527	0.492*	0.480***	0.475***
1000–1400 ms	0.524	0.473*	0.449***	0.475**	0.472*

Table 4

The mean *definitely familiar* scores given by the 4 different classifiers were compared to the *maybe familiar* and *unfamiliar* scores using repeated measure ANOVA. Significant *p*-values after Bonferroni adjustment for multiple comparisons are given with * superscripts (*: $p < 0.003$, **: $p < 10^{-3}$).

	Classifier	Def fam	Maybe fam	Maybe unfam	Def unfam
Group 1	during-temporal (400–800 ms)	0.520	0.468*	0.460**	0.458**
	during-alpha (400–800 ms)	0.527	0.492	0.480**	0.475**
Group 2	pre-[25–35 Hz] (-300-0 ms)	0.449	0.460	0.463	0.466
	during-alpha (1000–1400 ms)	0.473	0.449	0.475	0.472

Table 5

Proportion of the selected subbands from nested cross-validation. Results over 0.7 are highlighted in increasing shades of gray.

Subject	S03	S06	S10	S15	S16	S17	S20	S22	S24	S26	S40	S51	S52	S56	S57	S59	S61	S62
4-7 Hz	0.82	0.99	0.00	0.76	0.81	0.00	0.00	0.00	0.13	0.11	0.03	0.76	0.66	0.01	0.01	1.00	0.92	0.00
6-10 Hz	0.95	0.99	0.00	0.35	0.66	0.00	0.00	0.00	0.00	0.03	0.12	0.57	0.89	0.00	0.01	1.00	0.88	0.00
7-12 Hz	0.92	0.98	0.01	0.92	0.24	0.00	0.00	0.02	0.00	0.03	0.04	0.72	0.92	0.00	0.01	1.00	0.48	0.00
10-15 Hz	0.06	0.97	0.00	1.00	0.28	0.00	0.00	0.35	0.00	0.02	0.11	0.96	0.99	0.02	0.01	0.69	0.08	0.00
12-19 Hz	0.05	0.97	0.00	0.00	0.00	0.00	0.00	0.53	0.00	0.00	0.19	1.00	0.86	0.00	0.94	0.00	0.14	0.06
15-25 Hz	0.12	0.86	0.96	0.00	0.00	0.06	0.00	0.06	0.00	0.66	0.00	0.99	0.37	0.72	0.51	0.00	0.00	0.18
19-30 Hz	0.11	0.97	0.02	0.00	0.40	0.15	0.00	0.92	0.00	0.02	0.00	1.00	0.29	1.00	0.45	0.00	0.46	0.01
25-35 Hz	0.06	0.31	0.02	0.02	0.06	0.96	0.64	0.87	0.13	0.03	0.00	1.00	0.24	0.39	0.97	0.00	0.02	0.00
30-40 Hz	0.00	0.11	0.01	0.06	0.00	0.01	0.00	0.21	0.85	0.00	0.00	1.00	0.80	0.13	1.00	0.00	0.00	0.01