Dynamics of visual category learning with magnetoencephalography

Yang Xu*

December 2011

Machine Learning Department
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15232, USA
Email: *yx1@cs.cmu.edu*

*Joint work with Christopher J. D'Lauro, John A. Pyles, Michael J. Tarr and Robert E. Kass.

Abstract

Humans have the remarkable ability to distinguish visual categories rapidly, even when the target categories bear a high degree of similarity. It remains unclear how the brain encodes such categories, and more so, how it obtains adequacy through experience. Most fMRI and ERP studies focus on activities of specific brain regions and do not address the rich dynamics of neural encoding of visual categories. In this work, we pursue the neural underpinnings of visual categorization via magnetoencephalography when participants are trained to categorize two novel and resemblant classes of blob-like objects via feedback. Our analysis unfolds the temporal and spatial dynamics of encoding and finds that significant information about the visual categories is embedded at varying sites over multiple temporal stages termed as $M100$, $M170$, $M250$ and $M350$. Furthermore, we find support for our hypotheses that the brain becomes more efficient and informative about the categories through learning, with categorical information inferred from the $M170$ component improving reliably as individuals learn the task with greater accuracy.

1 Introduction

How does the brain encode visual categories, especially those that are highly visually similar? This essential function has important consequences in the natural environment. Ripe berries or poisonous ones, Retriever or Rottweiler, wet road or icy road, wife or sister-in-law. Each of these highly similar visual distinctions requires rapid and accurate categorization to guide future behavior, but the topic of visual categorization in the past has often focused on the end points of visual category discrimination rather than assessing the dynamic process of visual category learning. In this experiment, we use highly similar visual stimuli, constructed to be confusable yet learnable, to assess the behavioral and brain changes associated with visual category learning. Magnetoencephalography (MEG) is used to measure the neural changes associated with visual learning in time and space—both on a micro-scale (timing under a second) and over the course of a learning session (hours). Further, this study investigates the neural markers of interest using a bottom-up, data-driven approach that allows us to assess the most critical spatiotemporal windows for predicting learning in the brain with no a priori assumptions.

The standard model of visual object recognition has long emphasized the importance of ventral visual cortex (VVC), the "what pathway", in the recognition of objects (Mishkin, Ungerleider and Macko, 1983; Ungerleider and Haxby, 1994; Goodale et al., 1994) [1, 2, 3]. This hierarchical process of recognition begins by taking the elementary features from primary visual cortex and assembling them into complex and nuanced objects like a human face. While many investigators have studied different categories of visual stimuli and how they are processed within the ventral stream, fewer studies have addressed how the ventral stream interacts with other parts of the brain (like the prefrontal cortex or PFC) throughout the dynamic process of learning.

The prefrontal cortex is thought to make important contributions to visual object recognition, especially when the recognition process is effortful and involves the application of top-down information, categorization or ongoing visual learning. One prominent previous fMRI study (Jiang et al., 2007) [4] has suggested greater sensitivity to learned object category boundaries in prefrontal cortex and object identity sensitivity in ventral cortex. This forwards the idea that ventral cortex is a category-agnostic area that performs object perception tasks, while categorical information resides in pre-frontal cortex; however, some aspects of the study may actually enhance this pre-frontal-ventral distinction. First, the groups of objects identified were morphed versions of previously known objects. However, this study asked subjects to make distinctions between different groups of already familiar objects (i.e. distinguishing between morphed car models). Learning where to

place a category boundary between groups of different known objects does not capture the full course of learning that we see when the categories are novel. In addition, the temporal resolution of fMRI does not allow careful examination of the temporal processes associated with the encoding of visual stimuli.

Better temporal resolution can be gained using the Event-Related Potential (ERP) method. Category learning research using ERP shows a high degree of visual category sensitivity in components sourced to posterior inferior temporal cortex (Tanaka and Curran, 2001; Rossion, Curran and Gauthier, 2002; Wong et al., 2005; Scott et al., 2008) [5, 6, 7, 8] and also give an idea of the latency of this category sensitivity - approximately $170ms$ post-stimulus. Similarly, recent works on macaques (Freedman et al, 2005; Meyers et al., 2008) [9, 10] have shown that inferior temporal, prefrontal and parietal lobes are implicated in encoding differential information related to categorization at varied temporal stages. Complementary to these findings, our current study aims to unravel the encoding patterns of novel visual categories in the human brain. We conduct a computational analysis to investigate the spatiotemporal dynamics of visual categorical encoding based on whole-brain MEG recordings. Furthermore, we track the development of encoding during the course of learning.

Our experiment involves learning from scratch two categories of novel visual stimuli, termed as *blobs*, that are perceptually similar (see Figure 1a). Participants are trained to distinguish these categories in a continuous, trial-by-trial and feedback-driven session. Meanwhile, we use MEG to assess the neural changes concomitant with the entire course of learning. Our setup is similar to that in Krigolson et al. (2009) [11] where they found, using EEG recordings, that visual category learning is associated with increased negativity at $N250$, a component around $250ms$ after the stimulus onset. Our study, however, differs in several important respects. First, we examine directly how the encoding of visual categories evolves during learning. Although the change in signal negativity can be correlated with learning, it does not explain how the brain becomes adept at categorization—ultimately we need to understand how the categories are encoded, and how such encodings improve through experience. Second, we focus our analysis on a short period after the onset of stimuli—various works have suggested that important visual processing takes place within the first $400ms$ after onset (Halgren et al., 2000; Curran, Tanaka & Weiskopf, 2002; Liu, Harris & Kanwisher, 2002; Sugase-Miyamoto, Matsumoto & Kawano, 2011) [12, 13, 14, 15]. It is unclear from the work of Krigolson et al. whether early temporal components outside $N250$, e.g. those in the vicinity of $100ms$ and $170ms$, play any role in visual category learning—we investigate all of these signature components. Finally, different from the traditional region-of-interest analysis,

we take a more bottom-up approach that studies encoding patterns at the whole-brain scale. The advantage is that any effects potentially missing from the excluded regions (or sensors) can be captured when we consider the brain as a whole.

Figure 2 provides some initial basis for our analysis. Each sub-figure shows the stimulus-evoked waveforms averaged over all trials from two visual blob categories (A and B). These waveforms are taken from two separate MEG sensors from a single participant. Interestingly, the shaded regions highlight the fact that the signals are separable from the two categories, suggesting the recorded brain signals may indeed encode these categories. Moreover, the separations appear at different times for different sensors, suggesting that such encoding changes dynamically in time and space. Our first analysis picks up on these observations and addresses the following questions 1) Does the brain encode statistically significant information about the visual categories? 2) What is the temporal dynamics associated with categorical encoding? 3) What are the corresponding spatial patterns of encoding? To preview our results, we find that significant visual encoding occurs at multiple temporal stages (we term these as $M100$, $M170$, $M250$ and $M350$) in the MEG signals. Moreover, we find that the spatial patterns of encoding vary from early to late temporal stages, migrating from occipital pole to temporal and parietal regions.

Our second analysis relates neural activities to behavior and examines bases of the brain that reflect learning of visual categories. We postulate two hypotheses. First, we predict that the brain should become more efficient through learning. We operationalize this proposal by computing the signal energy at sensors of interest and evaluating their changes over the course of learning. Second, we predict that behavioral improvement in the categorization task should find accompanying improvement in neural encoding of the visual categories. We formalize this idea in terms of a simple probabilistic model that computes the *informativeness* of neural signals throughout the course of learning. We find support for both of our hypotheses from the data.

2 Materials and Methods

2.1 Participants

The experimental procedures were approved by the Institutional Review Board at Carnegie Mellon University and the University of Pittsburgh. Ten able-bodied participants took part in the experiments. Their ages ranged from 17 to 35. All participants gave informed consent before the experiments.

2.2 Stimulus design

The stimuli used in the experiments were two novel categories of "blob" images, abbreviated as A-blobs and B-blobs (Figure 1a). Each category consisted of 300 blobs that were samples slightly jittered from the prototype of the category. The prototypes were randomly generated two-dimensional polygons with 20 edges following Krigolson et al. (2009) [11]. The edges were defined as proportions $(30 - 70\%)$ of the distance between the origin and vertices of an original unit circle. The blob samples were generated by randomly drawing from a multivariate Gaussian distribution for each respective category, where the mean was the 20-dimensional vector (edge-to-origin distances) defined in the prototype, and the covariance was a diagonal matrix with uniform variance of 20% in each independent dimension. Figure 1a shows a handful of samples of the blob stimuli from each category—note that the two categories appear perceptually similar, and the subtle difference lies on the edge contours.

2.3 Experimental procedure

Participants were instructed to distinguish between the two blob categories in a supervised, trial-based and continuous experiment. The experiments were conducted in an electromagnetically shielded room with participants seated comfortably and head-fixed. A non-magnetic back-projection screen was placed about $50cm$ in front of participants to display the visual stimuli and signals. A non-magnetic ear-plug was used to channel the audio signal. A standard glove pad was used for participants to respond.

To reduce fatigue, each experimental session consisting of 600 trials was divided into five blocks of 120 trials, with brief self-paced breaks between the blocks. The trials included 300 A-blobs and 300 B-blobs, where the sequence of their presentations was randomized for each experiment and the frequency of category occurrences was balanced within each block. Each blob stimulus presentation was preceded by a random category indicator (the frequency of indicator categories was also balanced between "A" and "B" within each block). Note that the indicators did not necessarily correspond to the ground-truth categories of the blob, and the participants were asked to judge whether the indicator matches the true category of a given blob image at each trial. The goal was to allow participants to learn the correct blob categories via feedback.

Figure 1b illustrates the timeline of a single trial. First, a machine-generated $630ms$ audio label of "A" or "B" is played at the background to a central fixation cross on the screen—this label is a random indicator for the category. After a $120ms$ continued fixation, either an

A-blob or B-blob image is displayed at the screen center for $750ms$ (this is a very brief period to keep participants engaged in the task). During this period, the participant responds with a click on the glove pad of "yes" or "no" to indicate whether the random indicator label matches the category of the blob stimulus (the "yes" and "no" signs appear near left and right bottoms of the screen with their positions counterbalanced for each session). Finally, a feedback signal of "correct", "wrong" or "two slow" is given at the screen center for $750ms$, preceded by a jittered interval of blank screen. The trial ends with a $500ms$ break before the beginning of the next trial. This paradigm is similar to that in Krigolson et al. (2009) except that in our design, the label and the blob are decoupled both in time and in terms of associated brain processes, which aims to avoid additional visual processing of letter label "A" or "B" during the presentation of a blob stimulus. Moreover, the morphed blob category in the original design to increase the task difficulty was removed to ensure smooth learning during the session.

2.4 Data acquisition and processing

The behavioral data, including the decisional responses on the blob category membership and the response times, were recorded during each experimental session. The brain responses were acquired by a 306-channel whole-head MEG system produced by Elekta Neuromag, Helsinki, Finland. The system has 102 channels where each is a triplet of a magnetometer and two perpendicular gradiometers. The MEG signals were sampled at $1000Hz$. Four head position indicator coils were placed on the participant's scalp to record relative head positions to the MEG machine at each session. Electrooculography and electrocardiography were recorded by additional electrodes placed above, below and lateral to the eyes and at the left chest respectively. The coil and electrode signals were used to correct for movement and artifacts throughout the experiments.

The MEG signals were bandpass filtered between 0 to $50Hz$ for all subsequent analysis. Signal projection methods were used to remove any artifacts. The delay of visual stimulus onset on the screen was measured by special photodiodes and was accounted for in all results reported. Two experimental sessions had trigger failures where the timing of individual trials could not be retrieved, hence our analysis was based on the remaining eight participants, discarding the two that had incomplete data. For all of our analysis, the baseline defined as 1 to $120ms$ prior to the onset of visual stimulus at each trial was removed.

2.5 Data analysis

We investigate the temporal and spatial encoding of visual categorical information in the brain and how such encoding evolves from trial to trial during learning. The following sections explain each of these analyses in turn.

2.5.1 Temporal encoding of visual categories

To examine the precise timing of encoding of visual categories, we conducted a pattern analysis with *millisecond* precision at the whole-brain scale. We focused our analysis on the first $400ms$ after the onset of blob stimuli in each trial. We also parsed the 600 trials into two conditions based on their ground-truth category membership (each condition has 300 trials). Using multivariate analysis of variance or MANOVA (Johnson & Wichern, 1992) [16] and treating trials as independent samples, we tested whether there is a difference in conditions with signals across the visual categories at each instance, with the null hypothesis that there is no difference between category A and B blobs. Each trial, or sample, is a vector of 102 dimensions that represents the signals from all magnetometers at a single time point. Just as ANOVA where each sample is a uni-dimensional scalar, MANOVA extends the input space to multiple dimensions and tests whether there is a difference in conditions on multiple variables, in this case the whole-brain signals while participants viewing the A-blobs or B-blobs. MANOVA test at each time point yields a p-value from the test statistic Wilk's λ, which indicates the significance of conditional difference at that instance. In other words, with repeated MANOVA measures at each instance, we could track the trend of encoding of visual categories along time and locate the temporal processes that embed significant information.

2.5.2 Spatial encoding of visual categories

Following the temporal analysis in Section 2.5.1, we examine the spatial encoding patterns by locating sensors of interest (SOI) within the time periods that encode visual categories with high significance. Section 3.2 discusses in detail the discovered temporal markers termed as $M100$,$M170$,$M250$ and $M350$. Within each of these times of interest, we performed marginal decoding analysis on each of the 102 magnetometers for each individual participant. We then identified sensors that decode the visual categories significantly above chance as SOI. We used a simple Bayesian decoding algorithm that assumed a Gaussian likelihood function for each visual category and a uniform prior over the two categories. By training the decoder on all 600 but one held-out trials, we made a prediction on the blob category in the left-out trial based on the training sensor signals from the other 599 trials. Here

each trial was the mean sensor signal within a short window around the time of interest. We repeated this procedure 600 times holding out a different trial at each turn to obtain the leave-one-out classification accuracy for each sensor. Our baseline decoder was a random guesser with chance-level accuracy of 50%. We claimed sensors significant if their decoding accuracies exceed one deviation from random guesses, namely 55%. This way we located sensors of interest at each of the four temporal markers.

2.5.3 Visual category learning

The temporal and spatial analyses help discover spatial and temporal regions of interest that encode significant information about the visual categories. Our subsequent analysis leverages these findings and examines changes in the neural activities while participants learn the categorization task. In particular, we focus on two metrics of learning, *efficiency* and *informativeness*.

Efficiency We define *efficiency* as the overall energy of the magnetometer signals, which is the sum of squared signal values. To relate efficiency to encoding of visual categories, we restricted to sensors of interest within the four temporal markers from our previous analyses in Section 2.5.1 and 2.5.2. Formally, let $X = \{x_1^r, x_2^r, ..., x_n^r\}$ represent the signals from sensors and time of interest at trial r, the energy of that trial is simply $E = (x_1^r)^2 + (x_2^r)^2 + ...(x_n^r)^2$. An improvement in efficiency should then predict a general decrease in the energy during the course of learning.

Informativeness The efficiency measure does not directly address how the encoding of categories evolve during learning. We define a second metric, *informativeness*, as an indicator of the "confidence" of neural encoding. In other words, it specifies how well we can infer the blob categories in the external stimuli from the neural data.

Let C denote the blob category and X represent the neural signals from sensors of interest within a certain time period. Then the inference over the blob categories given the neural activities can be formalized as the Bayesian posterior odds (see Kass & Raftery, 1995) [17] between the two categories. More specifically, the posterior of a category is proportional to its prior probability $P(C)$ and likelihood $P(X|C)$

$$P(C|X) = \frac{P(X|C)P(C)}{\sum_c P(X|C)P(C)} \propto P(X|C)P(C). \tag{1}$$

Assuming equal prior between the two categories, i.e. $P(C_A) = P(C_B) = 0.5$, the posterior odds is equivalent to the likelihood ratio

$$\frac{P(C_A|X)}{P(C_B|X)} = \frac{P(X|C_A)P(C_A)}{P(X|C_B)P(C_B)} = \frac{P(X|C_A)}{P(X|C_B)}. \tag{2}$$

Since the blob stimuli were generated from the prototype within each category, we approximated the neural signal for each prototype (or the prototypical response) as the trial average of that category

$$\mu_C = \frac{1}{|C|} \sum_{r \in C} X_r \tag{3}$$

where $C = \{A, B\}$. Thus the likelihood can be interpreted as a function that describes how similar the neural signals from a single trial are to the prototypical response of a category. Here we chose the likelihood as a simple radial basis function with a Gaussian kernel

$$P(X|C) = f(X, \mu_C) = e^{-(||X-\mu_C||)^2}. \tag{4}$$

Hence the posterior odds in its logarithmic form (monotonous transform of the original odds) for a particular trial is

$$log\frac{P(C_A|X)}{P(C_B|X)} = log\frac{P(X|C_A)}{P(X|C_B)} = (||X - \mu_B||)^2 - (||X - \mu_A||)^2. \tag{5}$$

Intuitively, this means that if the neural signals are more distant to the prototypical response of A than to that of B, then we would predict, based on the neural signals of that single trial, the category to be B. Moreover, a higher posterior odds (with respect to the true categorization) indicates higher confidence in the choice of the correct category. Thus improved informativeness should predict an increase in the log posterior odds during the course of learning. Note that it is also trivial to convert the log posterior odds to probabilities of success, a measure we use to compare with behavioral odds of successful categorization in Section 3.4.

3 Results

3.1 Behavioral data

We used the categorical responses made by the participants at each trial (these are binary variables) to compute their categorization accuracies and learning curves. We treated all trials where participants failed to respond within the designated deadline ($750ms$ after the onset of a blob) as incorrect. Table 1 summarizes the categorization accuracies and response times for block 1 (pre-learning) and 5 (post-learning). Note participants spent significantly less time in their responses during the post-learning stage than at the beginning. Figure 3 shows the learning accuracies over each block. All except the last participant ($s8$) learned the task with their post-learning accuracies exceeding 70% and a steady improvement in the session.

Given these behavioral results, we next examine the neural dynamics of visual categorization and learning, particularly focusing on how subtle visual categories like these blob stimuli are encoded in the brain, and what neural patterns reflect the experience of learning.

3.2 Spatiotemporal encoding of visual categories

Figure 4a and 4b summarize respectively when and where significant visual categorical information is encoded in the brain. Applying whole-brain (102 magnetometers) MANOVA test at each millisecond after the onset of blob stimuli and treating trials with A or B blobs as separate conditions, we obtained a p-value curve that dynamically records the significance of categorical information along time. Figure 4a shows not only that significant information about the visual categories is embedded in the brain (e.g. instances with $p < 0.005$), but also that there are multiple temporal signatures that constitute the peaks (or troughs on the p-value curve) in the time course. Notably near $100ms$, $170ms$, $250ms$ and $350ms$ (shaded in gray in Figure 4a), we observed strong signals of encoding, suggesting that these temporal processes may be important stages in visual categorization. For our following analysis, we denote these temporal markers as $M100$,$M170$,$M250$ and $M350$. All of these components find correspondence with previous proposals about stages of visual processing (Downing, Liu & Kanwisher, 2001; Curran, Tanaka & Weiskopf, 2002; Halgren et al., 2003; Rieger et al., 2005; Mereen et al., 2008) [18, 13, 19, 20, 21]. However here our method automatically identified these temporal processes in the context of encoding of visual categories.

To further find out where categorical information is encoded, we zoomed into the discovered temporal markers and conducted marginal decoding of visual categories for all sensors—the idea is to locate sensors of interest (SOI) on the scalp. We first obtained a MANOVA p-value curve for each participant. We then found local peaks at each of the temporal markers (to retain locality of these peaks, we bounded the ranges of search respectively for $M100$, $M170$, $M250$ and $M350$ at $70-130ms$, $150-200ms$, $240-290ms$ and $340-390ms$ after stimulus onset). Table 2 summarizes the variability of these peaks identified for each participant—note all of these peaks are statistically significant with $p < 0.05$, suggesting significant encoding of categories exists at the individual level. We next took the mean signal around these peaks at each trial and used a leave-one-trial-out cross-validation to compute the decoding accuracy of each sensor. We used a $30ms$-window for averaging within $M100$ and $M170$ and a $40ms$-window for $M250$ and $M350$ since the peaks in the p-value curves are less sharp in the later processes. We defined SOI by thresholding at decoding accuracy of 55% (50% is chance). Figure 4b displays the SOI on the scalp by

juxtaposing SOI found in each of eight participants—the heat-mapped histogram tallies the spatial locations that qualify as SOI, so more reddish regions correspond to higher tallies. Note that the spread of SOI is relatively sparse and starts off at the occipital pole earlier in time, whereas they migrate more laterally and centrally at later stages. These spatial encoding patterns roughly agree with those in Halgren et al., 2000 [12] where they found significant preference for face stimuli transitions from midline occipital (around $110ms$ after onset) to occipitotemporal (around $165ms$ and $256ms$) regions.

Our results so far suggest that significant information about visual categories is embedded in the brain and the encoding is dynamic in time and space. In the following analysis, we restrict to the sensors of interest found at $M100$, $M170$, $M250$ and $M350$ and examine how their signals and encodings evolve over the course of learning (if at all), meanwhile seeking to explain the differential role of the four temporal markers.

3.3 Improved efficiency in learning

We computed the energy or sum of squared signals from sensors of interest at $M100$, $M170$, $M250$ and $M350$ using the same window lengths for averaging as in Section 3.2. Table 3 (second column) compares the mean energy in block 1 and 5. All four temporal markers showed significant less energy at the post-learning stage than at pre-learning, suggesting there may be a general improvement of efficiency. To test whether such improvement is global in the entire time course, we also computed energies between $0 - 50ms$ and during $120ms$ before stimulus onset (baseline). We found no significant improvement from pre-learning to post-learning stages at these times ($p > 0.5$), suggesting improved efficiency is specific to the temporal instances that encode visual categories. Figure 4c shows the mean energy over all five blocks of 600 trials (with non-learner $s8$ excluded), standardized by the mean and deviation within each participant. Note that the energy patterns across all temporal markers are similar, indicating an overall improvement in efficiency. On the other hand, it is difficult to remark on the differences among the temporal markers based on this metric of learning. More importantly, signal energy does not directly address the encoding of visual categories. The following section explicitly analyzes category encoding in the context of learning and relates the change in encoding to behavioral categorization accuracies.

3.4 Improved informativeness in learning

We computed the log posterior ratio with respect to the true categorization (namely designating the numerator in the ratio as the posterior

probability of the true category) at each trial based on sensors of interest at $M100$, $M170$, $M250$ and $M350$ using the same window lengths for averaging as in Section 3.2. Figure 4d shows the mean ratio over the five blocks (with non-learner $s8$ excluded), standardized by the mean and deviation within each participant. Note there is a clear increase in the posterior ratios computed at $M170$, a mild but insignificant improvement at $M250$ and $M350$, yet no improvement at $M100$. Table 3 (third column) compares the informativeness between pre-learning and post-learning stages and shows that only the improvement at $M170$ is statistically significant, suggesting $M170$ may be crucial in learning to encode the visual categories.

To directly compare the posterior ratios with the behavioral probabilities of successful categorization (or learning curves), we convert these ratios to probabilities of inferring the true categories. We then visualize both the "neural" odds and the behavioral odds of success trial by trial in Figure 5. Both of these are smoothed via Gaussian density estimators with width 30. Note that the neural odds have a high degree of correlation with the behavioral accuracies for most of the participants, particularly in those with steady learning curves. In other words, the neural odds inferred at $M170$ have a fair chance of predicting the participant's performance, suggesting some causal relations between the neural signals and the actual behaviors.

4 Discussion

We used computational methods and modeling to investigate the neural dynamics of visual category learning. We found that it is possible to distinguish perceptually close categories from signals recorded via MEG at the whole-brain level, while participants learn the categories behaviorally. Using multivariate statistics, we discovered four distinct temporal processes, $M100$, $M170$, $M250$ and $M350$, that embed significant information about the visual categories. By attending to each of these temporal components, we located sensors of interest that decode the categories above chance on an individual basis. We found that the spatial encoding locations migrate from occipital poles to temporal and parietal regions during the time course.

To examine the neural dynamics of learning, we introduced two metrics, signal energy and Bayesian posterior ratio, to test our hypotheses that the brain becomes more efficient and informative about the categories through learning. We found a general improvement in efficiency across sensors of interest at times that encode significant categorical information but not elsewhere in the time course. Furthermore, we found support from our data that the encoding at $M170$ improves steadily during the session, suggesting its crucial role in learning of

visual categories.

Spatiotemporal encoding The signature temporal components we found in our study confirm with previous proposals about visual processing. The main difference is that our findings were initially based on global patterns of encoding (finding these components at once) and then converged onto more local patterns of spatial encoding, as opposed to a top-down driven approach that is more amenable to finding discrete component. Many studies have found similar temporal processing stages in correspondence with $M100$, $M170$, $M250$ and $M350$. The general view is that early processing (e.g. prior to $130ms$ after stimulus onset) emphasizes on low-level visual features, whereas later stages process more complex and global patterns (e.g. $150 - 330ms$) and semantics (e.g. beyond $350ms$). Here we found that significant discrimination between visual categories occurs at as early as $70ms$ after onset. Although it is prone to attribute such phenomena to low-level features, our design of stimuli yielded perceptually very similar objects which makes it difficult to specify the nature of these features. However, our later analysis showed that $M100$ component does not improve with learning, suggesting the early encoding is unlikely due to complex processing.

In a similar experiment, Krigolson et al. (2009) [11] found increased negativity at ERP component $N250$. Here we replicate their result in finding an MEG counterpart $M250$ that shows significant decrease in signal energy over the course of learning. This suggests that $M250$ and $N250$ may be qualitatively similar, although further investigations are necessary to establish their equivalence. In addition, however, we found that the decrease in energy is general across $M100$, $M170$, $M250$ and $M350$, hence it is difficult to distinguish their roles in learning from this single metric. Our measure of informativeness directly relates categorical encoding with learning component and indicates that $M170$ encodes the categories more precisely as learning progresses, whereas $M250$ and $M350$ show only mild improvement.

Alternative learning effects Visual categorization is a global process that involves activities at multiple sites in the brain (Freedman & Miller, 2008; Seger & Miller, 2010) [22, 23], and our models of visual category learning only scratch the surface of such complex interactions. The study by Jiang et al., 2007 [4] have shown that learning sharpens the tuning in frontal and lateral occipital areas, and the study by Bar et al., 2005 [24] have suggested top-down influence in object recognition tasks. Furthermore, recent works on changes in network activities during learning (Hipp, Engel & Siegel, 2011) [25] open up new possibilities to study learning from the perspectives of synchrony and oscillation. Future work can build on these results and address the mechanistic development of regional interactions in the context of visual category learning.

Feedback and decision making A separate yet important aspect of our current paradigm is the feedback and decision-making component. In the study of Krigolson et al. (2009) [11], they found that the error-related negativity at feedback decreases during learning while the negativity near response increases, suggesting that participants become more self-evaluative about the decisional errors than reliant on feedbacks as learning progresses. Other works (Gehring et al., 1993; Holroyd and Coles, 2002; Seymour et al., 2004; Holroyd et al., 2005) [26, 27, 28, 29] have suggested that basal ganglia, anterior cingulate and frontal cortices are crucially important in trial-and-error learning and decision making. One challenging factor is whether it would be possible to locate such "deep" structures and their interactions in MEG, which picks up signals mainly from the surface of cortex.

Improving spatiotemporal resolutions Our current study provides a sensor-space analysis with MEG, thus it does not directly address the neural drivers of visual categorization. Previous EEG studies have suggested that ERP components $N170$ and $N250$ may be generated from a single source in the brain (Scott et al., 2008) [8], yet the precise locations of these remains debatable. An important question is how to localize these signatured temporal components with fine precision. Although source localization with MEG is an ill-posed inverse problem, it is possible to improve the precision of localization with anatomical constraints such as fMRI (Dale et al., 2000; Corrigan et al., 2009; Sadeh et al., 2010; Henson et al., 2011) [30, 31, 32, 33]. Leveraging the spatial resolution of fMRI and the temporal resolution of MEG to investigate the spatiotemporal dynamics of visual categorization is an exciting direction for future research.

Acknowledgements

We thank David Munoz and Deborah Johnson for conducting the experiments. We also thank Anna Haridis and Gustavo Sudre for their assistance in the acquisition of MEG data. Finally, we thank Avniel Ghuman for helpful discussions on the project.

References

[1] M. Mishkin, L. G. Ungerleider, and K. A. Macko. Object vision and spatial vision: two cortical pathways. *Trends in Neuroscience*, 6:414–417, 1983.

[2] L. G. Ungerleider and J. V. Haxby. 'What' and 'where' in the human brain. *Current Opinion in Neurobiology*, 4(2):157–165, 1994.

[3] M. A. Goodale, J. P. Meenan, H. H. Bulthoff, D. A. Nicolle, K. J. Murphy, and C. I. Racicot. Separate neural pathways for the visual analysis of object shape in perception and prehension. *Current Biology*, 4(7):604–610, 1994.

[4] X. Jiang, E. Bradley, R. A. Rini, T. Zeffiro, J. Van Meter, and M. Riesenhuber. Categorization training results in shape and category-selective human neural plasticity. *Neuron*, 53(6):891–903, 2007.

[5] J. W. Tanaka and T. Curran. A neural basis for expert object recognition. *Psychological Science*, 12(1):43–47, 2001.

[6] B. Rossion, T. Curran, and I. Gauthier. A defense of the subordinate-level expertise account for the n170 component. *Cognition*, 85(2):189–196, 2002.

[7] A. C. N. Wong, I. Gauthier, B. Woroch, C. Debuse, and T. Curran. An early electrophysiological response associated with expertise in letter perception. *Cognitive, Affective, and Behavioral Neuroscience*, 5(3):306–318, 2005.

[8] L. S. Scott, J. W. Tanaka, D. L. Sheinberg, and T. Curran. The role of category learning in the acquisition and retention of perceptual expertise: A behavioral and neurophysiological study. *Brain Research*, 1210:204–215, 2008.

[9] D. J. Freedman, M. Riesenhuber, T. Poggio, and E. K. Miller. Experience-dependent sharpening of visual shape selectivity in inferior temporal cortex. *Cerebral Cortex*, 16(11):1631–1644, 2005.

[10] E. M. Meyers, D. J. Freedman, G. Kreiman, E. K. Miller, and T. Poggio. namic population coding of category information in inferior temporal and prefrontal cortex. *Journal of Neurophysiology*, 100(3):1407–1419, 2008.

[11] O. E. Krigolson, L. J. Pierce, C. B. Holroyd, and J. W. Tanaka. Learning to become an expert: Reinforcement learning and the acquisition of perceptual expertise. *Journal of Cognitive Neuroscience*, 21(9):1833–1840, 2009.

[12] E. Halgren, T. Raij, K. Marinkovic, V. Jousmaki, and R. Hari. Cognitive response profile of the human fusiform face area as determined by MEG. *Cerebral Cortex*, 10:69–81, 2000.

[13] T. Curran, J. W. Tanaka, and D. M. Weiskopf. An electrophysiological comparison of visual categorization and recognition memory. *Cognitive, Affective and Behavioral Neuroscience*, 2(1):1–18, 2002.

[14] J. Liu, A. Harris, and N. Kanwisher. Stages of processing in face perception: an MEG study. *Nature Neuroscience*, 5(9):910–916, 2002.

[15] Y. Sugase-Miyamoto, N. Matsumoto, and K. Kawano. Role of temporal processing stages by inferior temporal neurons in facial recognition. *Frontiers in Psychology*, 2:141, 2011.

[16] R. A. Johnson and D. W. Wichern. *Applied multivariate statistical analysis*. Prentice Hall, Upper Saddle River, New Jersey, 1992.

[17] R. E. Kass and A. Raftery. Bayes factors. *Journal of the American Statistical Association*, 90:773–795, 1995.

[18] P. Downing, J. Liu, and N. Kanwisher. Testing cognitive models of visual attention with fMRI and MEG. *Neuropsychologia*, 39:1329–1342, 2001.

[19] E. Halgren, J. Mendola, C. D. R. Chong, and A. M. Dale. Cortical activation to illusory shapes as measured with magnetoencephalography. *NeuroImage*, 18:1001–1009, 2003.

[20] J. W. Rieger, C. Braun, H. H. Bulthoff, and K. R. Gegenfurtner. The dynamics of visual pattern masking in natural scene processing: a magnetoencephalography study. *Journal of Vision*, 5:275–286, 2005.

[21] H. K. M. Meeren, N. Hadjikhani, S. P. Ahlfors, M. S. Hamalainen, and B. de Gelder. Early category-specific cortical activation revealed by visual stimulus inversion. *PLoS ONE*, 3(10), 2008.

[22] D. J. Freedman and E. K. Miller. Neural mechanisms of visual categorization: insights from neurophysiology. *Neuroscience and Biobehavioral Reviews*, 32:311–329, 2008.

[23] C. A. Seger and E. K. Miller. Category learning in the brain. *Annual Review of Neuroscience*, 33:203–219, 2010.

[24] M. Bar, K. S. Kassam, A. S. Ghuman, J. Boshyan, A. M. Schmid, A. M. Dale, M. S. Hamalainen, K. Marinkovic, D. L. Schacter, B. R. Rosen, and E. Halgren. Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences*, 103(2):449–454, 2005.

[25] J. F. Hipp, A. K. Engel, and M. Siegel. Oscillatory synchronization in large-scale cortical networks predicts perception. *Neuron*, 69:387–396, 2011.

[26] W. J. Gehring, B. Goss, M. G. H. Coles, D. E. Meyer, and E. Donchin. A neural system for error detection and compensation. *Psychological Science*, 4(6):385–390, 1993.

[27] C. B. Holroyd and M. G. H. Coles. The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, 109(4):679–709, 2002.

[28] B. Seymour, J. P. O'Doherty, P. Dayan, M. Koltzenburg, A. K. Jones, R. J. Dolan, K. J. Friston, and R. S. Frackowiak. Temporal difference models describe high-order learning in humans. *Nature*, 429:664–667, 2004.

[29] C. B. Holroyd, N. Yeung, M. G. H. Coles, and J. D. Cohen. A mechanism for error detection in speeded response time tasks. *Journal of Experimental Psychology: General*, 134:163–191, 2005.

[30] A. M. Dale, A. K. Liu, B. R. Fischl, R. L. Buckner, J. W. Belliveau, J. D. Lewine, and E. Halgren. Dynamic statistical parametric mapping: combining fmri and meg for high-resolution imaging of cortical activity. *Neuron*, 26(1):55–67, 2000.

[31] N. M. Corrigan, T. Richards, S. J. Webb, M. Murias, K. Merkle, N. M. Kleinhans, L. C. Johnson, A. Poliakov, E. Alyward, and G. Dawson. An investigation of the relationship between fMRI and ERP source localized measurements of brain activity during face processing. *Brain Topography*, 22(2):83–96, 2009.

[32] B. Sadeh, I. Podlipsky, A. Zhdanov, and G. Yovel. Event-related potential and functional MRI measures of face-selectivity are highly correlated: a simultaneous ERP-fMRI investigation. *Human Brain Mapping*, 31(10):1490–1451, 2010.

[33] R. N. Henson, D. G. Wakeman, V. Litvak, and K. J. Friston. A parametric empirical bayesian framework for the EEG/MEG inverse problem: generative models for multi-subject and multi-modal integration. *Frontiers in Human Neuroscience*, 5(76), 2011.

Tables and figures

Table 1: Behavioral categorization accuracies (Acc) and response times (Resp Time) during the first (B1) and the last (B5) blocks for all participants. Each block has 120 trials.

Sub.	B1 Acc. (%)	B5 Acc. (%)	B1 Resp. Time (*ms*)	B5 Resp. Time (*ms*)
s1	$50 \pm .89$	$84 \pm .65$	587 ± 18	517 ± 15
s2	$46 \pm .89$	$94 \pm .42$	613 ± 21	541 ± 16
s3	$55 \pm .89$	$91 \pm .52$	581 ± 19	575 ± 17
s4	$62 \pm .87$	$82 \pm .69$	632 ± 16	584 ± 17
s5	$60 \pm .88$	$81 \pm .70$	616 ± 20	575 ± 18
s6	$72 \pm .80$	$78 \pm .75$	600 ± 19	579 ± 17
s7	$51 \pm .89$	$72 \pm .81$	640 ± 18	607 ± 19
s8	$59 \pm .88$	$66 \pm .85$	581 ± 18	535 ± 19
group	57 ± 2.9	81 ± 3.3	606 ± 8.0	564 ± 10.6

Table 2: Individual variabilities in identified peaks from MANOVA p-value curves that reflect encoding of blob categories. The peaks are searched locally around $M100$, $M170$, $M250$ and $M350$ with bounds indicated in brackets.

Sub.	$M100$(70-130ms)	$M170$(140-200ms)	$M250$(210-290ms)	$M350$(310-390ms)
s1	83	176	287	354
s2	100	155	280	340
s3	76	178	261	376
s4	107	172	234	384
s5	113	193	216	323
s6	95	199	251	328
s7	74	168	289	330
group	94 ± 5.3	173 ± 6.8	262 ± 9.4	347 ± 8.0

Table 3: Comparison of *energy* and *informativeness* (see Section 2.5.3 for details) at the first (B1) and the last (B5) blocks in the experiment with statistics gathered from the seven learners (excluding $s8$).

	Energy (B1>B5)	Posterior ratio (B1<B5)
M100	$p < .02$*	$p = .90$
M170	$p < 0.008$*	$p < 0.007$*
M250	$p < 0.0001$*	$p = .13$
M350	$p < 0.0001$*	$p = .32$

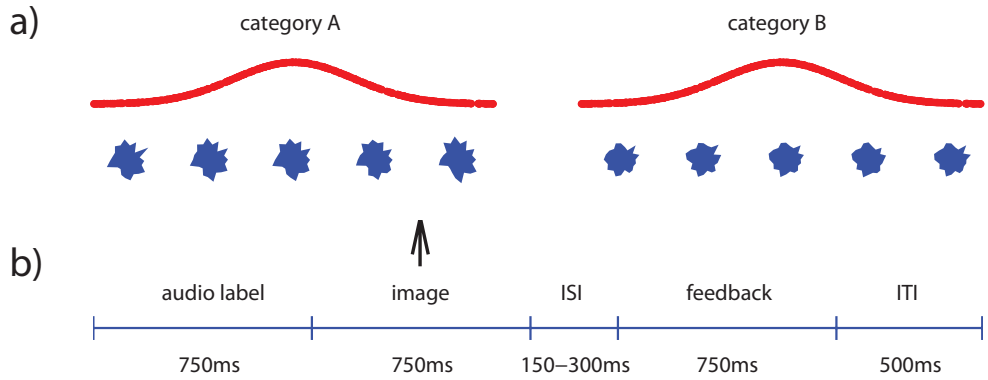

Figure 1: a) Two categories of blob-like stimuli used in the experiment generated from two separate multivariate Gaussian distributions with the prototype of each category as the mean and a diagonal covariance matrix with fixed variability at each dimension. b) Time course of a single trial in the experiment. A trial begins with an audio label randomly chosen as an utterance of "A" or "B" while the participant is instructed to fixate over a cross at the screen center. An image of a blob-like stimulus is then displayed for a short period, during which the participant is asked to make a response whether the audio label matches the true category of the blob. Finally, a feedback of the response is given after a short inter-stimulus interval followed by an inter-trial interval prior to the next trial.

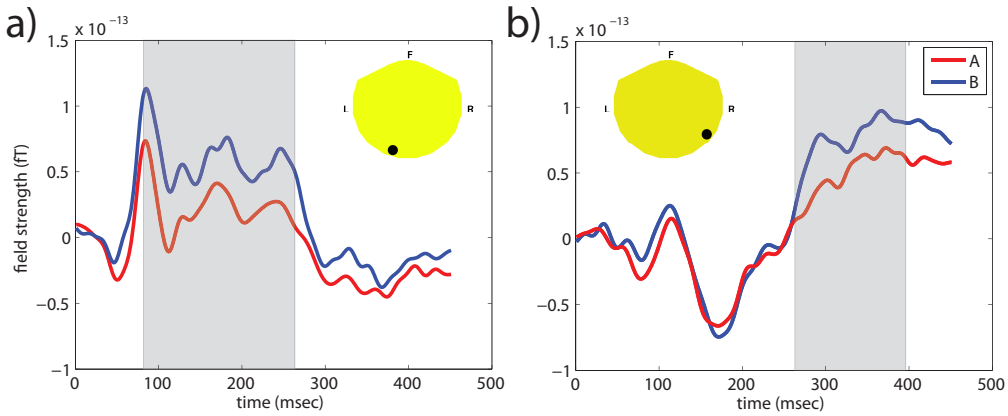

Figure 2: a) Stimulus-locked (onset at $0ms$) and baseline-removed waveforms averaged with respect to trials in each blob category (A and B) from a magnetometer located at the left occipital area (inset) of a single participant. The shaded region highlights the signal difference between the two categories. b) Waveforms similarly obtained as in a) at the right temporal area of the same participant. Note the most prominent difference occurs later in the time course.

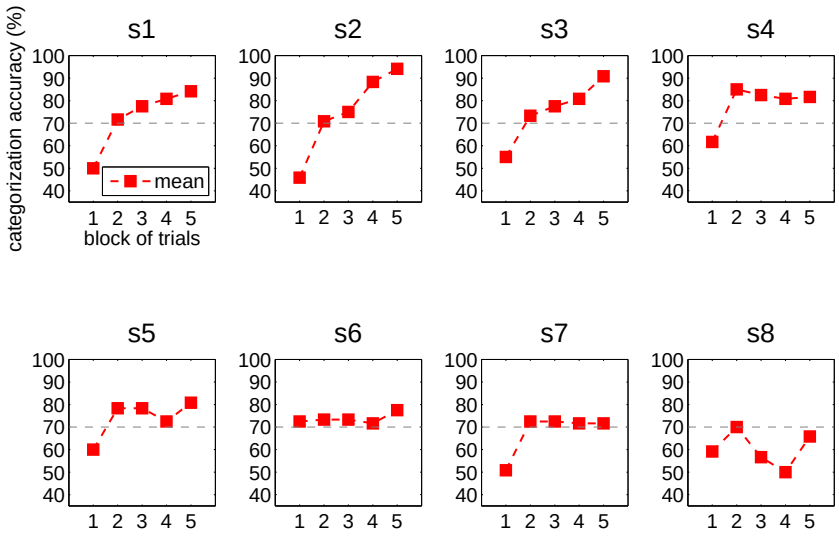

Figure 3: Categorization accuracies averaged over each block of 120 trials for all eight participants. *s*8 is the only non-learner that fails to reach a threshold of 70% accuracy (dashed line) in the final block and exhibits erratic learning behaviors.

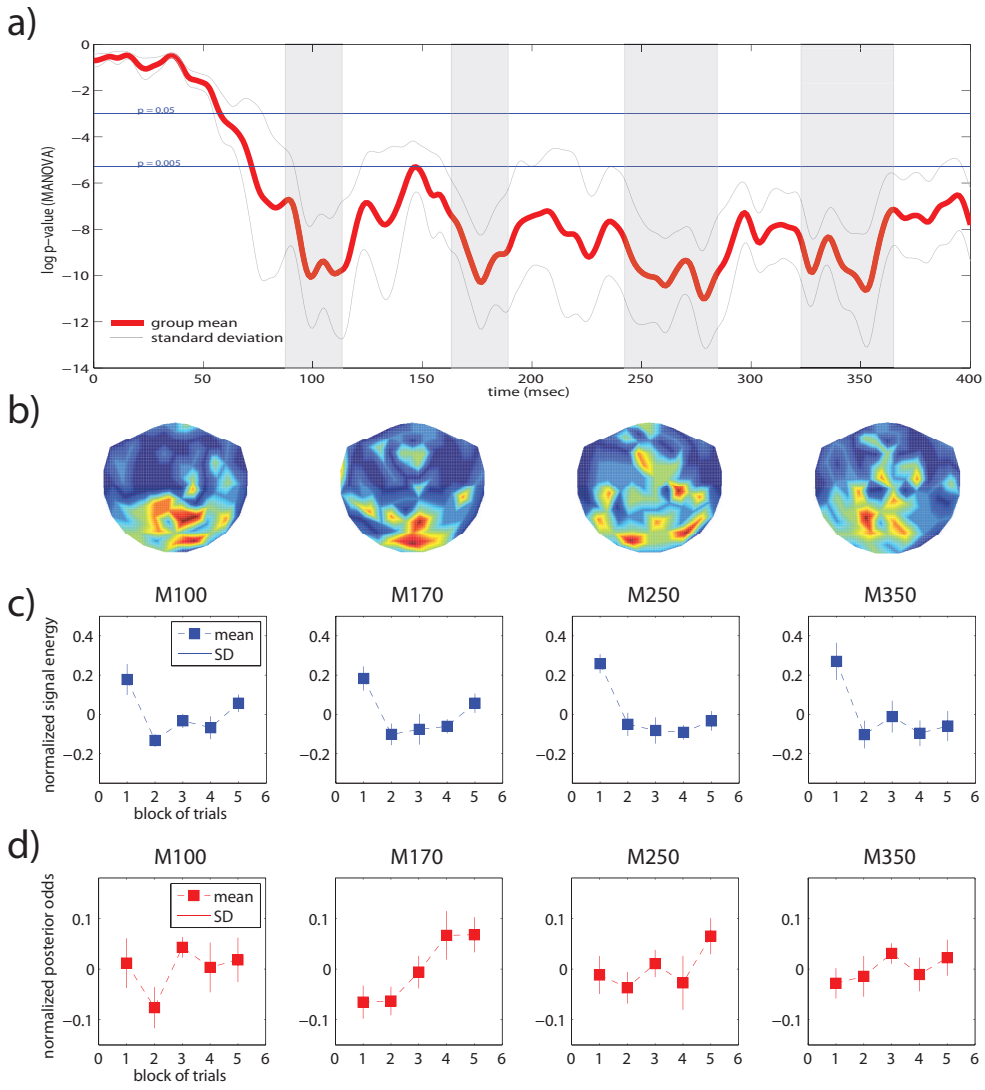

Figure 4: a) Logarithmic p-value curves obtained from MANOVA test at each millisecond after the onset of blob stimuli for the group of eight participants. The shaded regions highlight temporal processes that encode significant information about the visual categories, termed as $M100$, $M170$, $M250$ and $M350$ respectively. b) Juxtaposed sensors of interest identified at the four temporal markers at the group level—the heat map indicates the tallies at each sensor positions from the eight participants. c) Sum of squared signals (measure of *efficiency*) averaged over each block of 120 trials at the group level. The signals are first taken from the sensors of interest at each of four temporal markers and normalized by subtracting off the mean and dividing into the standard deviation within each participant and then combined. d) Log Bayesian posterior ratios (measure of *informativeness*) averaged over each block of 120 trials at the group level. The signals processed similarly as in c).

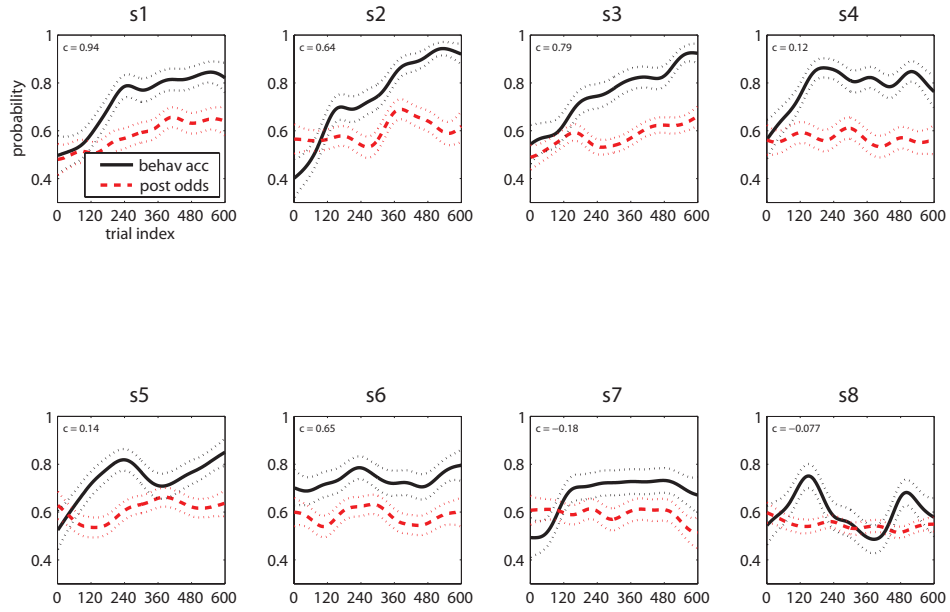

Figure 5: Trial-by-trial comparison of posterior odds (probability of successfully inferring the blob categories from neural activities) and behavioral accuracies (probability of actually categorizing correctly or learning curve). Both measures are smoothed by Gaussian density estimator of width 30 (the dotted lines correspond to 95% confidence intervals). The correlation between the neural and the behavioral odds is shown at the top left corner.