The effects of inattention on binocular rivalry between complex images: Preliminary results

Alice Forehand
ICA-3887545

Utrecht University

October 22, 2014

Abstract

Whether consciousness depends on attention is a significant question in consciousness research. Consciousness is often studied using binocular rivalry: images causing perceptual fluctuations during invariant physical stimulation. The role that attention plays in this phenomenon was debated. On one hand, the neural competition underlying rivalry appears to resolve in the preattentive stages of visual processing [33, 31, 17]. However, diverting attention has been shown to disrupt rivalry [37, 14, 3]. Analysis addressing the main hypothesis remains in future work, but initial analysis revealed several noteworthy findings. Namely, linear discriminant analysis decoded face and house categories above chance from a low frequency band of EEG data during rivalry.

1 Introduction

Binocular rivalry occurs when each eye is presented with a different image. Observers' perception alternates between the images, with one image "dominating" for a few seconds before being "suppressed" by the other. To what extent does this phenomenon depend on top-down influences such as attention? Investigations into this question have produced apparently conflicting results. The aim of this project is to resolve those conflicts.

Classical models characterize binocular rivalry as arising from interocular suppression: monocular neurons receiving the dominant image inhibit those receiving the suppressed image. The dominant neurons adapt, eventually losing the competition to the nondominant neurons, and were in turn inhibited. This "reversal" process causes perceptual shifts [2, 19], which continue indefinitely. More recent observations of phenomena such as pattern rivalry [12] have necessitated including higher visual areas in rivalry models [32].

It is now accepted that binocular rivalry involves inhibition at multiple levels of the visual hierarchy, but the role of feedback remains unclear. Research suggests that neural competition underlying binocular rivalry is resolved in the preattentive stages of visual processing, preceeding the involvement of feedback from extrastriate areas [33, 31, 17]. However, a number of recent studies have concluded that removing attention eliminates rivalry [37, 14, 3]. It is not clear how to reconcile these findings. I propose that a solution lies in the mechanisms of perceptual grouping.

According to incremental grouping theory [25, 26], two mechanisms are responsible for perceptual grouping. The first, "base-grouping," refers to preattentive feature conjunction by individual neurons and reflects the selectivity of feedforward connections. Many studies suggest that base groupings for common object categories are hardwired in the visual hierarchy, occur automatically and require little to no attention [33, 15, 13, 16, 8, 21, 7, 35, 10, 24, 6, 36]. Other objects like gratings do not have high-level tuning properties associated with them and are not automatically grouped in a feedforward fashion; they require additional horizontal and feedback connections to be grouped. This recurrent grouping mechanism has been termed "incremental grouping," which is modulated by attention and includes phenomena such as Gestalt grouping [25, 26, 10, 36].

The established research concluding that binocular rivalry requires attention has used only low-level objects as target stimuli [37, 14, 3]. Therefore, it is possible that inattention disrupts perceptual grouping of the target images rather than rivalry itself. Objects behave as wholes when competing for visual representation [5], so when they are not integrated, wholesale rivalry is eliminated or reduced to piecemeal. This explanation draws support from observations that the waves of activation during perceptual shifts between gratings are preserved in primary visual cortex when attention was diverted, but fail to propagate coherently to higher areas [14]. That is, inattention does not disrupt interocular competition, but stimulus integration, preventing rivalrous activity from moving up the visual hierarchy. If this account is true, inattention might not disrupt rivalry between high-level objects that are grouped preattentively.

This project aims to disentangle the influences of perceptual grouping and attention on binocular rivalry. Outcomes indicating that wholesale rivalry is modulated by perceptual grouping, rather than attention per se, would result in a model that coheres the apparently disparate findings outlined above. Contrary outcomes would augment the growing body of evidence that attention was necessary for wholesale rivalry by demonstrating that the effects of attention extend to high-level objects.

An EEG experiment was carried out to determine whether attention modulates binocular rivalry between images that are perceptually grouped in a feedforward manner. In this experiment, subjects were presented with rivalrous images of a face and a house, which they attended and from which their attention was diverted. These categories were chosen because of their clear neural representations in the absence of attention [15, 13, 16, 8, 21, 7, 35, 10, 24, 6, 36]. In addition to the rivalry conditions, non-rivaling faces and houses were presented for the purpose of training a linear discriminant analysis (LDA) classifier, which was the primary analysis tool.

Some analysis remains before the main hypothesis can be addressed. What has been completed so far includes LDA classification of the training data, time-frequency analysis of the training data, and further LDA classification of specific frequencies of the training and attended rivalry data. In within-condition cross-classification, the LDA decoded image categories from the full frequency spectrum of the training data. Decoding was found to be most effective in a band around 6Hz. The LDA was applied to attended binocular rivalry data (epochs time-locked to reported percept resolution) at 6Hz and decoded face and house categories above chance. Cross-classification showed these face and house representations to be stable over time. Within the training data, cross-classification also revealed patterns of inverse decoding prior to stimulus onset. These findings were promising for future analysis. The next steps in analysis include using LDA to decode the continuous attended rivalry data and the unattended rivalry data.

2 Methods

Inattention has been shown to interrupt wholesale rivalry between images of objects that were perceptually grouped through incremental (top-down) grouping, such as gratings. This experiment aims to determine the effect of inattention on rivalry between objects that were grouped through base (bottom-up) grouping. If preattentive base-grouping was sufficient to produce wholesale rivalry, then images of faces and houses, which were grouped preattentively [33, 15, 13, 16, 8, 21, 7, 35, 24], should rival even when unattended. If instead wholesale rivalry requires attention regardless of grouping mechanism, these images will not rival when unattended.

This experiment builds on work by Zhang [37], which found that removing attention abolishes rivalry between incrementally-grouped checkerboard images. I closely follow their experimental design, instead using base-grouped face and house images.

Three phases make up the procedure: a Task Training phase, an Experimental phase, and a Classifier Training phase. In the Task Training phase, subjects practiced the attentionally demanding task that they then performed in the Experimental phase. The Experimental phase consists of four conditions: Unattended Rivalry, Attended Rivalry, Unattended Replay, and Attended Replay. In the Classifier Training phase, subjects attended to a stream of stimuli to be used to train the Linear Discriminant Analysis (LDA) classifier. This classifier was the primary analysis tool, and uses linear discriminant analysis to classify the experimental data at each time point as belonging to the category "face" or "house" based on Classifier Training data gathered in the Classifier Training phase.

Attention was diverted during the Unattended conditions with a rapid serial visual presentation task (RSVP task). It was a 2-back task consisting of a letter stream superimposed on the images 1. Subjects were instructed to ignore the images and to click a mouse when a letter matched one presented two items earlier.

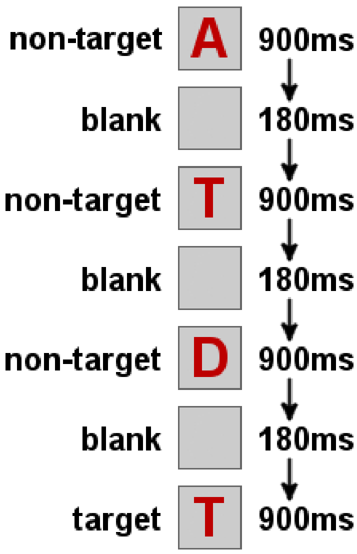


Figure 1: Portion of a 2-back task. Subjects must click the mouse when they see a letter that matches one presented two items earlier ("target"). In the first block of task training, each letter appears for a duration of 900ms followed by a blank box for 180ms. If the subject scores over 80% on the training task in one block, the letters in the next block appeared for 180ms less; if they score less than 60%, the letters in the next block appeared for 180ms longer. When a subject achieves a score between 70% and 80%, the duration of the letters in that block was used during the experimental phase. The blank between letters appears for 180ms in all blocks.

2.1 Subjects

Fifteen subjects participated in this study. All had normal or corrected-to-normal vision, as determined by a standard Snellen test. Fourteen subjects were right-handed, and ten were female. Subjects' ages averaged 29.1 years (SD 9.7 years). Subjects were compensated €8 per hour of participation.

2.2 Equipment

The images were presented on a CRT screen at a refresh rate of 100 Hz in a darkened room. All stimuli were presented through a stereoscope. The stereoscope was mounted on a frame around the stimulus presentation monitor. Two images were presented at a time, one to the left and one to the right eye. A partition was placed in the center of the screen to ensure that each eye could only see the image on its side of the monitor. The stereoscope was positioned such that it was vertically aligned with the letters. The height of the table could be adjusted such that the stereoscope and monitor were at eye level for each participant.

The data were collected using a 64-electrode BioSemi Active Two EEG system, manufactured in Amsterdam, The Netherlands. Four EOG electrodes were placed: one on either side of the eyes to measure horizontal eye movements, and one above and below the left eye to measure vertical eye movements and blinks. Two electrodes were placed on the mastoids to use as a reference. EEG data were collected at 2048 Hz on a data acquisition computer separate from the stimulus presentation computer. Subjects made responses on a mouse that was mounted at the end of the right arm of a chair. Responses made on the mouse were logged on both computers.

2.3 Stimuli

The images were presented at a visual angle of 7.9° and the letters had a visual angle of 0.93°. The letters appeared in the vertical center of the screen and the images appeared behind them with a vertical offset of 1.19°. The images were frequency tagged with dynamic random noise, based on the method described by Parkonnen [20]. For each image, a random value ranging from -100 to +100 was added to the intensity (range 0-255) of each pixel (see Figure 2). Ten copies of each image were presented in a random order throughout the rivalry and replay conditions, creating the effect of static. The images on the right alternated at a rate of 12.5 Hz; on the left, they alternated at a rate of 16.67 Hz. This remained consistent for each block of each condition.

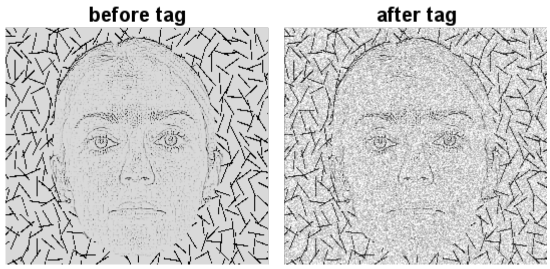


Figure 2: The face used in the experimental phase before and after the addition of random noise. For clarity, the figures here depict the images without noise, but during the experiment, all images in the Task Training and Experimental phases were noisy.

Frequency tagging has been shown to accurately track perceptual switches in binocular rivalry: during rivalry, the amplitudes of the tag frequencies modulate with a counterphase relationship [37]. These frequency tags have not been analyzed in the rest of the experiment due to time constraints.

There were three sets of images that could appear during the Task Training phase: tree → flower, drink → cake, and laptop → book. These were randomly chosen for each block. There were two sets of images that appeared during the Experimental phase: box → house and helmet → face. Images of various object categories were used to keep subjects agnostic with respect to the goal of the experiment. All stimuli were taken from color photographs. In GIMP, they were desaturated based on luminance, filtered with a Difference of Gaussians edge-detection algorithm, then sharpened. If necessary, they were distorted such that all images would take up a similar area.

2.4 Procedure

When subjects arrived, they performed a visual acuity test and then they read an overview of the experiment and sign an informed consent form. Subjects were told that the purpose of the experiment was to determine the effects of distraction on their ability to perform a difficult task. They were encouraged to focus on the task and to ignore the "distractors," i.e. the images behind the letters. The purpose of this was to give incentive to avoid attending to the images.

Next, subjects were fitted for an EEG cap and the electrodes were placed. The EEG system was tested to ensure that all electrodes were in place and operating correctly, then the stimulus presentation system was calibrated.

After this subjects completed the RSVP Task Training phase, followed by the Experimental phase. Finally, subjects completed the Classifier Training phase, during which they performed a task to train the LDA classifier, which was used to analyze the experimental data.

At the beginning of every task, the task instructions were first given orally by the experimenter and then presented (stereoscopically) on the monitor for the subject to read.

2.4.1 Stimulus Calibration

Before the experiment begins, subjects viewed identical images and adjusted the horizontal position of them with the mouse so that they were comfortably "in focus" (visually aligned) when viewed with both eyes. The position that they chose dictated the positions of the images in the rest of the experiment. Some subjects reported eye strain or difficulty maintaining alignment in later conditions and were allowed to repeat this calibration procedure before continuing.

2.4.2 Task Training Phase

After they adjust the images to a comfortable position, subjects practice the RSVP task in 4-minute blocks. The RSVP letter stream appeared in front of frequency-tagged images (see Figure 3). All aspects of stimulus presentation aside from image content were identical between the Task Training and Experimental phases.

In each block, one pair of images appeared behind the letter stream for 58s, then faded into another pair of images over a period of 4.5s. The physical transitions from one set of images to another also occurred in the Experimental phase, and gave subjects some time to focus their attention on the task; if a subject became distracted toward the beginning of the task during the Experimental phase, they would attend to an irrelevant image. The second set of images remains behind the letter stream for the rest of the block. The durations listed above were chosen to ensure than the first pair of images were in place long enough to allow subjects time to focus their attention, and to ensure a smooth transition from one set to the other.

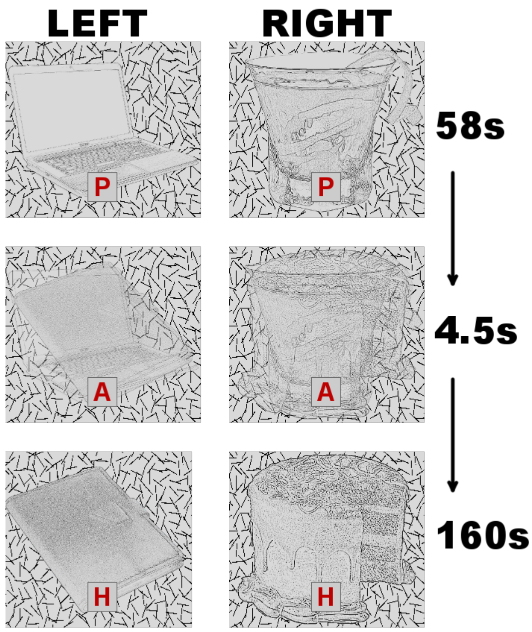


Figure 3: An example training block. One pair of images appears for 58s while the 2-back task 1 runs in front of them. The task continued while the first images fade into a second pair of images over a period of 4.5s. The second pair of images remains behind the letter stream for the remaining 160s. Subjects performed a minimum of two blocks of task training, and continued until they achieved a score between 70% and 80%.

At the beginning of each block a letter appeared in the same positions as during stimulus calibration. Subjects brought the letter into focus before continuing with the task. This ensured that subjects visually fixated properly during the task. Subjects completed a minimum of two blocks of training, and continued until they achieved a score between 70% and 80% accuracy. This range was chosen so that the task would require the subject's full attention, but would not be so difficult that they would give up.

After each block of task training, the subject's performance during that block was displayed along with feedback on how it compared with the 70%-80% goal. The score was calculated as $\frac{\text{hits} - \text{false alarms}}{\text{targets}}$. If the subject scores less than 60%, the stream of letters was slowed slightly. If they score between 60% and 70%, it was repeated at the same difficulty level. If they score between 70% and 80%, training ceased (unless this score was reached on the first block of training). If they score above 80%, the rate of presentation was increased slightly. The rate of the stream during the final block of training dictated the rate of the stream during the Experimental phase (see § 2.4.3).

2.4.3 Experimental Phase

Unattended Rivalry When subjects reached the required level of performance, they moved on to the Unattended Rivalry condition. This condition consisted of two 4-minute blocks and proceeded identically as in the training task, except for image content: the first pair of images was a box and a helmet, and the second pair was a face and a house (see Figure 4). The presence of initial set of images served to minimize the subjects' awareness of the relevant images of the face and the house.

During the rivalry conditions, the box → house images were presented to the left eye and the helmet → face images were presented to the right. In the second block they were switched.

Evaluation After the Unattended Rivalry condition, subjects completed an evaluation task to determine if they were able to identify which images appeared during Unattended Rivalry. This task, unlike the others, was not performed stereoscopically. An array of 12 images appeared on the screen in a random order. Two of these images were the face and house, and the other ten were images that had not appeared in either Task Training or Unattended Rivalry. Subjects were instructed to verbally indicate which images had been presented in the previous task.

The goal of the Unattended Rivalry condition was to keep attention diverted from the rivaling images with an attentionally demanding task. As

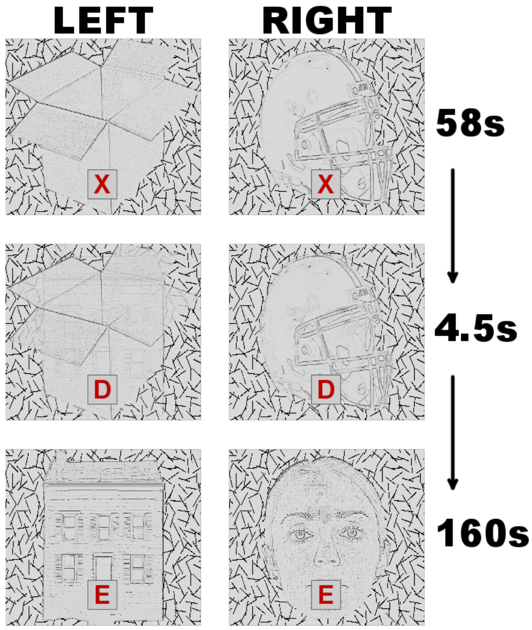


Figure 4: A schematic of the first block of the rivalry conditions. In the second block, the images were shown to the opposite eyes.

in Zhang's experiment [37], performance on the attentional manipulation task was not at ceiling (see § 3.1), so their attention was diverted to the extent that they could not further increase their score on the RSVP. Because all their attentional resources were devoted to the RSVP task, they were unable to attend to the images in the background.

Ideally, attention would be diverted to the extent that subjects would be rendered inattentionally blind to the rivaling images. The intention of the evaluation task was to determine whether this was the case. One subject was unable to report either image. Three subjects were unaware of the house but were able to report the face. Three subjects correctly guessed the house but reported uncertainty. The remaining eight subjects correctly identified both images with confidence. Although this strict inattention goal was not achieved in most subjects, their performance on the attentional manipulation task suggests that their attention was nonetheless sufficiently diverted from the rivaling images. One limitation of the present study is the absence of a dual-task condition that would ensure that th RSVP task was sufficiently attentionally demanding.

After completing the evaluation task, the experimenter asked whether the subjects attended to the images and requested qualitative descriptions of the Unattended Rivalry task. The most common report was an awareness of the presence of eyes on the screen. One subject reported that they became distracted by the images in the first block of the task. Two subjects reported perceptual switches. One subject reported no perceptual switches.

Attended Rivalry For the next task, the stimuli remained the same as in the Unattended Rivalry condition, but the task changed. Subjects were to fixate on the letters, but focus their attention on the images. They were to indicate their percepts with the mouse: hold the left mouse button when perceiving a face, hold the right mouse button when perceiving a house, and release when perception was mixed. As in the previous task, the attended rivalry task was broken into two 4-minute blocks, and the timing of stimulus presentation was identical.

Replay Conditions Replay conditions followed the rivalry conditions. The tasks were identical to those in the rivalry conditions, but the images physically transitioned such that the stimuli were the same in both eyes. That is, in addition to the physical transition from one pair of images to the other that occurred in previous conditions, the images transitioned back and forth within each pair to mimic rivalry.

First there was an Unattended Replay condition, in which subjects complete the RSVP task with images transitioning periodically in the background. Then there was an Attended Replay condition, in which the subjects visually fixate on the letters, and attended to the images, indicating (physical) perceptual switches with the mouse in the manner described above

The timing of the mouse presses and releases in the attended rivalry condition was meant to dictate the timing of stimulus presentation in the replay conditions that followed. However, due to programming and hardware errors, the timing was not correctly replicated in the replay conditions. Therefore, the analysis focuses on the rivalry conditions only.

2.4.4 Classifier Training Phase

The Classifier Training phase consists of three 8-minute blocks of an RSVP task. The stimuli for classifier training were presented binocularly, with the same image appearing in both eyes. Unlike the previous conditions, the classifier training condition did not contain any letters. In this task, images of faces, houses, and patchy face-house combinations appeared one at a time in sets of 16 (see Figure 5). Each image appeared for 480ms with 480ms (±150ms) of a fixation box in between.

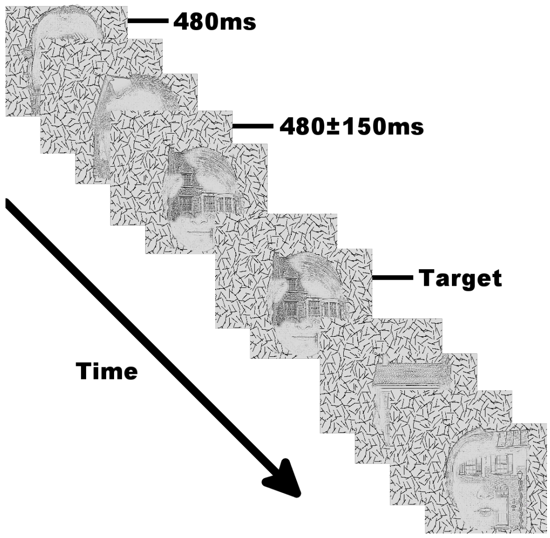


Figure 5: A section of the classifier training phase with patchy images. The same stream of images was presented to both eyes. Images of faces, houses, and patchy face-houses appeared for 480ms each with 480ms (±150ms) in between each image. In between images, a small box appeared in the center of each eye so that subjects maintain fixation. Images from each category appeared in sets of 16, with 29 sets of 16 per 8-minute block. The object categories appeared in a random order throughout each block.

The face and house images used in the testing conditions were excluded from the sets of Classifier Training stimuli. To ensure that subjects' attention remains on the images, they completed a 1-back task on the images. They were to click the mouse when an image appeared twice in a row. Each set of 16 images contained at least one and up to three repeated images. The content of each set was chosen randomly, chosen with equal probability.

2.5 Preprocessing

The preprocessing was performed using the EEGLAB toolbox for MATLAB and custom written code. The data for each condition were downsampled to 256 Hz. A highpass filter of 1 Hz was applied to remove drift and low-frequency artifacts. Large muscle artifacts were removed manually before applying independent component analysis. The ADJUST plugin [18] was used to identify components related to blinks, eye movements and generic discontinuities. Finally, the data were epoched by category when applicable. For the classifier training condition, the stimulus presentation computer sent a code to the data acquisition computer at the onset of each stimulus. For the attended rivalry condition, the mouse presses and releases sent a code to the data acquisition computer at the onset or offset of a dominant percept.

2.6 Classification

The EEG data from each condition were stored in a 2-dimensional matrix where each row was a time point and each column was an electrode (64 total). The primary analysis tool for this project was linear discriminant analysis (LDA), which was performed using MATLAB's `classify()` function. A subset of nine electrodes over the occipital cortex (`'PO7'`, `'PO3'`, `'O1'`, `'Iz'`, `'Oz'`, `'POz'`, `'PO8'`, `'PO4'`, `'O2'`; See Figure 14 in § 5 for a map of these electrodes) were found to be most effective for classification, so these alone were included in the present analysis (see Figures 16 and 17 in § 5 to compare with the results from the occipital electrodes reported in § 3). The `classify()` function classifies each row (time point) of a matrix (set of EEG data) into one of the groups given as training data (face vs. house Classifier Training data). LDA has been shown to decode complex object categories in MEG [4], and should be highly informative for this experiment because face and house categories are extracted preattentively [13, 16], and have been shown in fMRI to alternate as strongly during rivalry as during nonrivalrous stimulus alternation [33].

2.7 Time-Frequency Analysis

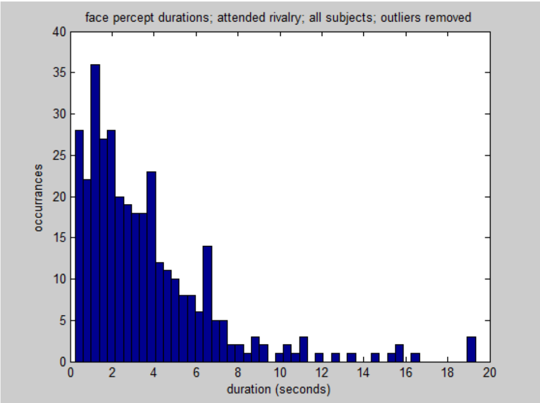
Time-frequency representations (TFRs) were calculated for each trial, and ranged from 2Hz to 100Hz in steps of 2 Hz. For low frequencies, ranging from 2Hz to 30Hz, a sliding time window of $\Delta T = 0.5s$ was used, and the data in each time window was multiplied with a Hanning taper. For high frequencies, between 30Hz and 100Hz, a multitaper approach was applied [22]. The method used sliding time windows whose sizes were inversely proportional to frequency: $\Delta T = \frac{20}{\text{frequency}}$. The resulting time windows ranged from 625ms at 32Hz to 200ms at 100Hz. Frequency smoothing was applied with filters increasing proportional to frequency: $0.1\times$frequency, resulting in a range of smoothing filters from 3.2Hz at 32Hz to 10Hz at 100Hz. These settings resulted in three orthogonal tapers at each frequency. In addition to computing total power, induced responses were isolated by subtracting the condition-specific average evoked response (ERP) waveform from each trial segment, and computing power on the resulting time courses. These induced responses did not result in above chance classification performance anywhere, however, so only total power was reported.

3 Analysis

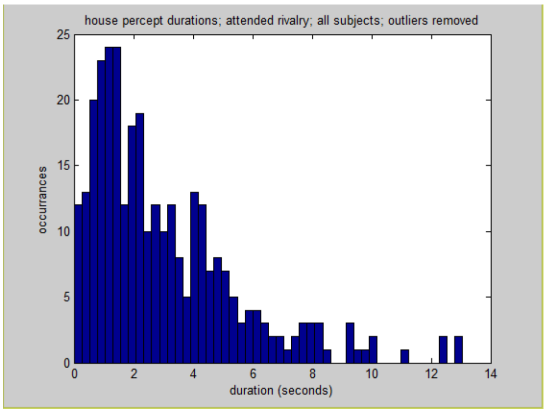
3.1 Behavior

In the unattended rivalry task, subjects performed with an average of 74.9% accuracy (SD 11.56 percentage points) on the 2-back RSVP task (see § 2.4.2 for a description of how accuracy in this task was calculated). In the attended rivalry task, percept durations were measured as the time between a button press and a button release, and transition rate was measured as the total number of button presses. If two button presses occurred within 250ms of one another, the intervening button release was assumed to be unintentional, and the release and second button press were ignored. That is, the second button press was not counted toward the transition rate measure, and the percept duration was recorded as the time between the first button press and the next release that occurred after the second button press.

Reported face percept duration averaged 4.2 seconds (SD 5.6). Reported house percept duration



(a) Face percept durations; $\mu = 3.6s, \sigma = 3.2s$



(b) House percept durations; $\mu = 3.0s, \sigma = 2.5s$

Figure 6: Subjects indicated perceiving faces and houses during the attended rivalry condition with button presses. Excluding outliers, face percepts 6a averaged 3.6 seconds, and house percepts 6b averaged 3.0 seconds

averaged 3.2 seconds (SD 3.7). With outliers removed (designated as durations longer than three standard deviations away from the median), the mean face percept duration was 3.6 seconds (SD 3.2), and the mean house percept duration was 3 seconds (SD 2.5 seconds). The average transition rate was 47.4 (SD 19.7).

3.1.1 Training Data Classification

First, a within-condition cross-classification was performed on the Classifier Training data. This cross-classification entails that the LDA classifier was applied to the test data many times, each time using a different time point from the training data

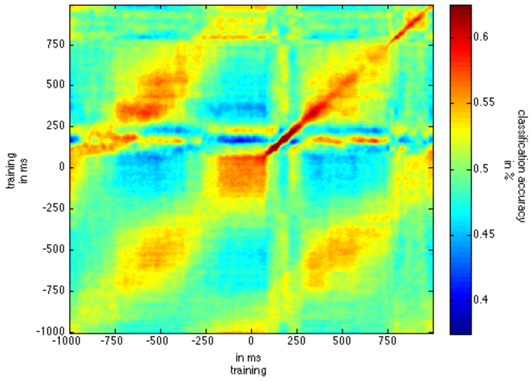


Figure 7: Within-condition classification performance averaged over all subjects. There was a region of high classification performance along the diagonal from shortly before stimulus onset (t=0ms) to stimulus offset (t=480ms). The Classifier Training data were used as both training and testing data sets to ensure that the classifier could decode the training data. This plot depicts how effectively each time point in the “training” set (the x-axis) can be used to classify each time point in the “testing” set (the y-axis). Dark red indicates performance over 60%, green indicates 50% performance, and dark blue indicates performance under 40%.There was some variation in how well each subject can be classified (see 12 in § 5), with subjects showing very strong patterns of classification (e.g. subjects 5 and 11), and others showing very weak patterns of classification (e.g. subjects 4 and 10). On average, performance was high around stimulus onset (t=0ms) and continues strongly until stimulus offset (t=480ms).

as a reference. Because this was a within-condition classification, the data from the Classifier Training phase (§ 2.4.4) were used both to train and test the LDA classifier. This was done to establish a baseline for classifier performance. For each subject, 75% of the data from each category were used to train the classifier, and the remaining 25% were used as test data. This was done four times and averaged so that the entire data-range would be used as test data. The results for each subject were then averaged (see Figure 7). The color at each coordinate indicates classification performance; that is, what percentage of trials at each time point from the test data (y-coordinate) were correctly categorized given each time point of the training trials (x-coordinate). Dark red indicates that over 60% of test trials were accurately classified. Yellow indicates that slightly over 50% of trials were accurately classified. Green indicates that 50% of trials were correctly classified–chance performance. Light blue indicates that slightly less than 50 % of trials were correctly classified. Dark blue indicates that fewer than 40% of trials were correctly classified.

Along the diagonal, the time point in the training data was used to classify the same time point in the test data. As expected, there was a line of high accuracy in dark red along this diagonal during stimulus presentation (0ms < t < 480ms). Additionally, there were two prominent blocks of above average performance: one from -200ms to 100ms, and one from 300ms to 600ms. Such blocks illustrate that neural representations were consistent over these time spans, such that all time points of the testing dataset within the block can be classified with comparable accuracy given any time point of the training dataset within that block.

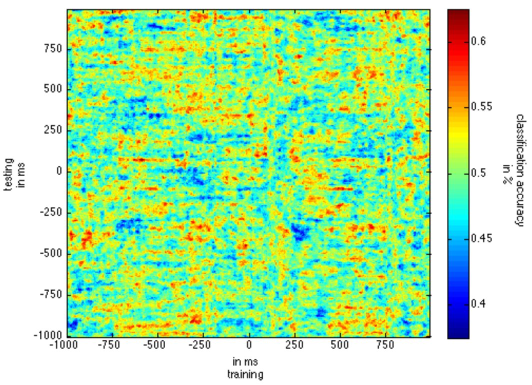


Figure 8: Classification of attended rivalry data averaged over all subjects. No blocks of above-average performance were visible, and there was no pattern as in Figure 7. No classification performance pattern appeared for individual subjects (see Figure 13 in § 5) or in aggregate for the attended rivalry condition. Some subjects show a narrow band of consistent performance during stimulus presentation, however this performance was at chance.

Aside from the main diagonal of above-chance performance, there were offset diagonals of above average classification performance on either side. These arise from the fact that images of the same category appeared one after another in sets of 16 during the Classifier Training phase (see Figure 5). Therefore, representations at one time point tend to be very similar to those offset by about 960ms, and so will be classified with similar accuracy. Another noteworthy feature of the within-condition cross-classification was the blocks of marked *below* chance classification accuracy that appeared between stimulus presentations. Points in these blocks were more often than not classified in the category opposite to that of the images that immediately precede or follow. These were discussed more in 4.3.

3.1.2 Attended Rivalry Classification

After adequate within-condition classification performance was established, the LDA classification was applied to the attended rivalry data. These data were organized into epochs, with t=0ms corresponding with the resolution of subjects' percepts as indicated by button presses. These epochs were therefore timelocked with perceptual switches, with a minimum percept duration of 500ms required for inclusion in this analysis.

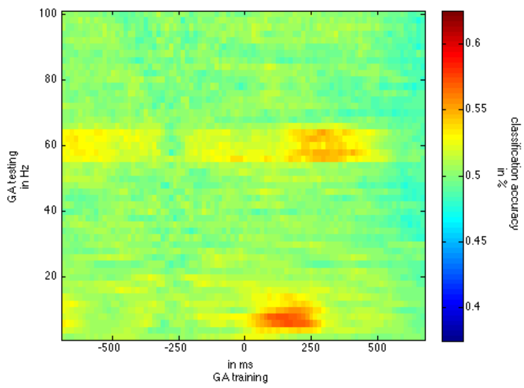
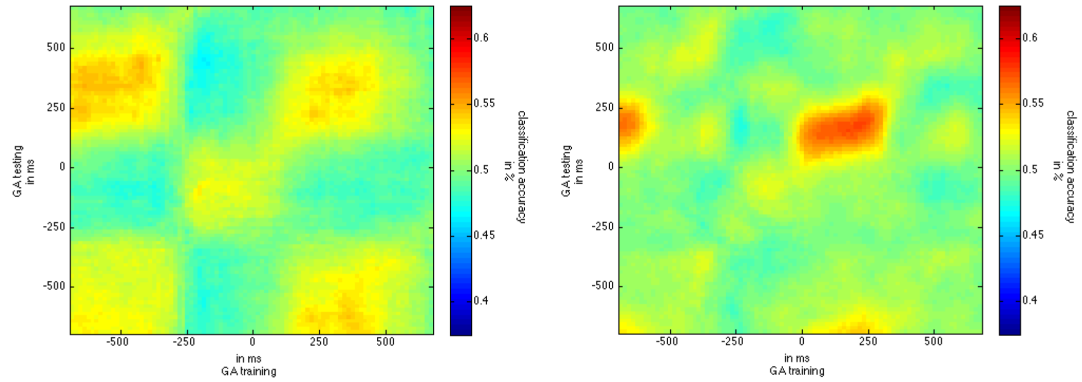


Figure 9: A time-frequency analysis of the Classifier Training data averaged over all subjects. The x-axis gives the time, and the y-axis gives frequency. For frequencies around 6Hz, the classifier performs above average between stimulus onset (t=0ms) and t=250ms. For frequencies around 60Hz, the classifier performs marginally above average for most time points, which somewhat improved performance between t=200ms and t=400ms.

Here, all the Classifier Training data were input to the classifier as the training set, and the attended rivalry data were input as the testing set. In contrast to the within-condition classification, there



(a) Within-condition classification at 60Hz averaged over all subjects.

(b) Within-condition classification at 6Hz averaged over all subjects.

Figure 10: Within-condition classification of Classifier Training data at 6Hz and 60Hz. Some subjects showed stronger patterns of classification performance at 60Hz (see Figure 15 in § 5), while others showed stronger patterns of performance at 6Hz (see Figure 18 in § 5). On average, classification at 6Hz yielded a higher performance during stimulus presentation (16b) than did classification at 60Hz (10a).

was no pattern of classification for the attended rivalry data (see Figure 8). The expected outcome of an attended rivalry classification was a block of above-chance classification accuracy during percept dominance after 0ms, but no such block appears.

This result arises from analyzing the full frequency spectrum of the data without splitting up the signal into frequency bands. While classification was apparently not possible over the full spectrum, performance may improve if the analysis were restricted to specific frequencies. To determine this, a time-frequency analysis was performed on the Classifier Training data.

3.1.3 Time-Frequency Analysis

A time-frequency analysis was conducted on the Classifier Training data (see 2.7). The classifier was trained then and tested with each frequency of the Classifier Training in a similar manner as with the full spectrum (see § 3.1.1). This, however, was not a cross-classification; rather, each time point in the test data was classified using the corresonding time point from the training data. In Figure 9, then, the horizontal lines correspond to what would appear along the diagonal in a full cross-classification at that frequency.

Classification performance was around chance at most frequencies (see Figure 9). However, two frequencies showed improved performance. A band around 60 Hz showed sustained performance marginally above average, with a patch of somewhat higher performance around 200ms $<$ t $<$ 400ms. Another band around 6Hz showed a patch during stimulus presentation of clear above-average performance. These two frequencies were examined further with a full within-condition cross-classification.

Full within-condition cross-classification at 6Hz and 60Hz revealed larger areas of high performance in some subjects (see Figure 15,18 in § 5), and reasonable performance in aggregate (see Figure 10a,16b). Interestingly, some subjects showed higher performance at 6Hz, while others showed higher performance at 60Hz. On average, classification at 6Hz performed somewhat higher overall than did classification at 60Hz, and had a higher peak performance between stimulus onset at t=0ms and t=250ms. For this reason, further analysis was performed on the 6Hz band of the attended rivalry data.

3.1.4 Attended Rivalry Classification at 6Hz

Cross-classification was repeated on the epoched attended rivalry condition at 6Hz, with much better results than classification of the full frequency spectrum (compare Figures 17b and 8). Whereas full-spectrum classification yielded noisy results with no clear regions of consistent performance, 6Hz classification resulted in some large regions of high performance for some individual subjects (see Figure 19 in § 5), and a large region of high performance during reported percepts in aggregate (see Figure 17b), from x≈-250ms to x≈300ms, and from y≈-200ms to y≈250ms. Note that the high-performance region begins before the percepts were reported at t=0ms, reflecting the establishment of percept dominance that occurs just before report. The LDA classifier was able to decode object category during attended binocular rivalry with above-average accuracy at 6Hz. Statistical varification of these results remains to be done, however tThis band appears to be a good candidate for future analysis of the unattended rivalry data.

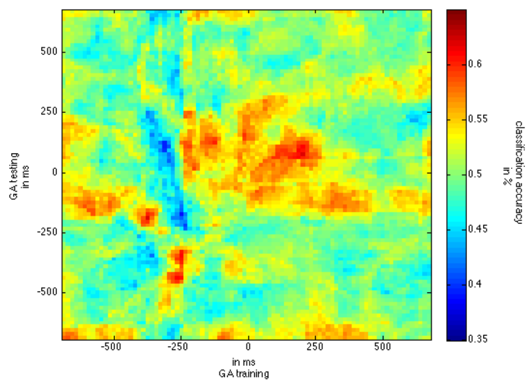


Figure 11: Classification of epoched attended rivalry data at 6Hz averaged over all subjects. Larger contiguous regions of similar performance result compared to classification of the full frequency spectrum (see Figure 8) with a clear region of high classification performance during stimulus presentation.

4 Discussion

Here, an LDA classifier decoded face and house categories above chance from EEG data (see Figure 7). A time-frequency analysis revealed that decoding was most effective around the 6Hz and 60Hz bands of the data (see Figures 9 and 10). The LDA classifier was applied to attended binocular rivalry data at 6Hz and decoded face and house categories above 50% accuracy (see Figure 17b). Cross-classification showed these face and house representations to be stable over time (see § 3.1.1; Figures 7, 10 and 17b). Within the Classifier Training data, cross-classification also revealed patterns of inverse decoding prior to stimulus onset (see Figure 7).

4.1 Category Decoding

Decoding complex object categories from patterns of neural activity has traditionally been done in fMRI. Haxby [9] decoded several object categories via correlations among distributed patterns of fMRI activity. Complex object categories have also been decoded from fMRI data during binocular rivalry [33, 29], revealing the neural markers of rivalry. Few studies have used imaging techniques other than fMRI to decode image categories during binocular rivalry. EEG/MEG studies have often employed frequency-tagging to track percept alternations [37, 20], which do not involve category-selective information, and can be used on any kind of image. Some category decoding has been achieved during binocular rivalry with MEG by analyzing event-related potentials (ERPs) [30]. Recently, a linear support vector machine (SVM) has been applied to intermittent binocular rivalry data to classify a face vs. a grating with MEG [28]. The present study was the first to decode two complex object categories with EEG, and to do so during continuous rivalry.

4.2 Time-Frequency Information

One interesting finding from the present study was the significance of the 6Hz band of the data in the success of the LDA classifier (see Figure 17b vs. Figure 8). The recent MEG SVM study [28] similarly found that applying a low-pass filter at 10Hz to the MEG data improved the performance of their classifier, suggesting that the significance of low frequencies to classifier performance was not peculiar to the present study. Further supporting this finding was a time-frequency analysis of EEG ERP dynamics for faces and other objects on a single trial basis [27]. This study found that the N170 was primarily associated with a modulation of amplitudes within the 5Hz-15Hz band. It was then quite possible that the low-frequency activity underlying the N170, which reliably distinguishes faces from other objects in ERPs, also underpins the performance of the LDA classifier here.

There was also a band around 60Hz that showed improved classifier performance. The study in [27] did not analyze frequencies above 30Hz, however a recent ECoG study found that power modulations above 50Hz were critical for distinguishing faces from checkerboard patterns [34]. It was not clear whether these same modulations would be equally critical for distinguishing between two complex object categories, but they may have contributed to the improved performance seen in the band around 60Hz (see Figure 9).

4.3 Reverse Classification

In the full-spectrum cross-classification of the Classifier Training data (see Figure 7), blocks of above-chance (yellow) classification accuracy during stimulus presentation alternated with blocks of below-chance (teal) classification accuracy between stimulus presentations. That these between-stimulus classifications were below chance rather than at chance indicates that neural representations during these times were opposite of those that occur during stimulus presentation. There were two potential explanations for why this occurs.

The first was that these regions of reverse performance result from overshoot after stimulus offset. Below chance classification performance has been observed after stimulus offset for faces, among other things [4, 1].

Another possibility relates to predictive coding theory, which characterizes visual feedback connections as carrying predictions of upcoming inputs, and feedforward connections as carrying the residual errors between those predictions and the current visual input [23, 11]. In terms of predictive coding, the observed reverse classification activity could result from the brain actively coding the inverse of the expected upcoming image. This account draws support from the fact that images from the same category appeared in sequence and the fact that classification performance becomes more reversed approaching stimulus onset.

The major difference between the two explanations was that in the first account, the classification between stimulus presentations was reversed because of the stimulus that just offset, while in the second, the classification between stimulus presentations was reversed because of the expectation of the stimulus that's about to onset. In the present study, the stimuli appeared in rapid succession, making it difficult to rule out eith explanation outright. Future work could follow the setup of the Classifier Training phase (§ 2.4.4), increasing the time between stimulus presentations past the point of there being any remaining overshoot. Then, give a visual cue at fixation to indicate an upcoming stimulus onset and see if the reverse pattern reappeared at this time. Also, because there was a 150ms jitter in the time between stimulus presentations, the Classifier Training data could be re-epoched to be timelocked with stimulus offset rather than onset, and the reverse classification performance between stimulus presentations can be compared with that from the onset-timelocked data here.

4.4 Future Work

Based on the findings of the present study, a number of future investigations are warranted. The most pressing is the analysis of the continuous attended rivalry data. The present study has analysed this data after epoching, with the effects accumulating over multiple trials per subject. Seeing these aggregate effects on a single trial basis as they appear in the continuous rivalry data will be challenging. However, based on the results in [27] showing that low frequency modulations can distinguish among object categories on a single trial basis, there is a chance that the significant low frequency activity observed in aggregate in the present study will contribute to category decoding in the continuous rivalry data.

Successfully decoding the continuous attended rivalry data will be critical to addressing the main hypothesis that neural representations of images belonging to complex object categories that occur during attended binocular rivalry will continue when attention is diverted. This is because subject reports of perceptual switches are not available during unattended rivalry, making epoching impossible. If this endeavor proves fruitful, the next step will be to analyze the unattended rivalry data in the same manner and compare the classification patterns in each condition.

Finally, the frequency tags in the experimental images will be analyzed. This analysis was critical for a direct comparison with the findings of the 2011 Zhang experiment [37]. Furthermore, if results from frequency tag analysis prove to be harmonious with the LDA results, the possibility of erroneous results will be reduced.

References

[1] S.R. Afraz and P. Cavanagh. Retinotopy of the face aftereffect. *Vision Research*, 48(1):42–54, 2008.

[2] R. Blake. A neural theory of binocular rivalry. *Psychological Review*, 96:145–167, 1989.

[3] J.W. Brascamp and R. Blake. Inattention abolishes binocular rivalry: Perceptual evidence. *Psychological Science*, 23:1159–1167, 2012.

[4] T.A. Carlson, H. Hogendoorn, R. Kanai, J. Mesik, and J. Turret. High temporal resolution decoding of object position and category. *Journal of Vision*, 11(10):9, 2011.

[5] R. Desimone and J. Duncan. Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18:193–222, 1995.

[6] V. Di Lolli. The feature-binding problem is an ill-posed problem. *Trends in Cognitive Sciences*, 16(6):317–321, 2012.

[7] J.J. Fahrenfort, T. Snijders, K. Heinen, S. van Gaal, H.S. Scholte, and V.A.F. Lamme. Neuronal integration in visual cortex elevates face category tuning to conscious face perception. *PNAS*, 109(52):21504–21509, 2012.

[8] M. Finkbeiner and R. Palermo. The role of spatial attention in nonconscious processing: A comparison of face and nonface stimuli. *Psychological Science*, 20:42–51, 2009.

[9] J.V. Haxby, M.I. Gobbini, M.L. Furey, A. Ishai, J.L. Schouten, and P. Pietrini. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293(5539):2425–?2430, 2001.

[10] S. Hochstein and M. Ahissar. View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, 36(5):791–804, 2002.

[11] P. Kok, G.J. Brouwer, and F.P. van Gerven, M.A.J. andd e Lange. Prior expectations bias sensory representations in visual cortex. *Journal of Neuroscience*, 33(41):16275?–16284, 2013.

[12] I. Kovács, T.V. Papathomas, M. Yang, and Á. Fehér. When the brain changes its mind: Interocular grouping during binocular rivalry. *PNAS*, 95(26):15508–15511, 1996.

[13] N. Lavie, R. Ro, and C. Russel. The role of perceptual load in processing distractor faces. *Psychological Science*, 14:510–515, 2003.

[14] S.-H. Lee, R. Blake, and D. Heeger. Hierarchy of cortical responses underlying binocular rivalry. *Nature Neuroscience*, 10:1048–1054, 2007.

[15] F. Li, R. VanRullen, C. Koch, and P. Perona. Rapid natural scene categorization in the near absence of attention. *PNAS*, 99(14):9596–9601, 2002.

[16] R. Marois, D. Yi, and M.M. Chun. The neural fate of consciously perceived and missed events in the attentional blink. *Neuron*, 41:465–472, 2004.

[17] M. Meng and F. Tong. Can attention selectively bias bistable perception? differences between binocular rivalry and ambiguous figures. *Journal of Vision*, 4:539–551, 2004.

[18] A. Mognon, J. Jovicich, L. Bruzzone, and M. Buiatti. Adjust: An automatic eeg artifact detector based on the joint use of spatial and temporal features. *Psychophysiology*, 48(2):229–240, 2011.

[19] C.L.E. Paffen, D. Tadin, S.F. te Pas, R. Blake, and F.A.J. Verstraten. Adaptive center-surround interactions in human vision revealed during binocular rivalry. *Vision Research*, 46:599–604, 2006.

[20] L. Parkkonen, J. Andersson, M. Hämäläinen, and R. Hari. Early visual brain areas reflect the percept of an ambiguous scene. *PNAS*, 105(51):20500–20504, 2008.

[21] M.V. Peelen, F. Li, and S. Kastner. Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature*, 460:94–97, 2009.

[22] D.B. Percival and A.T. Walden. *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques*. Cambridge University Press, Cambridge, NY, 1993.

[23] R.P.N. Rao and D.H. Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1):79–87, 1999.

[24] L. Reddy and N. Kanwisher. Category selectivity in the ventral visual pathway confers robustness to clutter and diverted attention. *Current Biology*, 17(23):2067–2072, 2007.

[25] P. Roelfsema. Cortical algorithms for perceptual grouping. *Annual Review of Neuroscience*, 29:203–227, 2006.

[26] P.R. Roelfsema and R. Houtkamp. Incremental grouping of image elements in vision. *Attention, Perception, & Psychophysics*, 73:2542–2572, 2011.

[27] G.A. Rousselet, J.S. Husk, P.J. Bennett, and A.B. Sekuler. Single-trial eeg dynamics of object and face visual processing. *NeuroImage*, 36(3):843–862, 2007.

[28] K. Sandberg, B. bahrami, R. Kanai, G.R. Barnes, M. Overgaard, and G. Rees. Early visual responses predict conscious face perception within and between subjects during binocular rivalry. *Journal of Cognitive Neuroscience*, 25(6):969–985, 2013.

[29] P. Sterzer, J.D. Haynes, and G. Rees. Fine-scale activity patterns in high-level visual areas encode the category of invisible objects. *Journal of Vision*, 8(15):10, 2008.

[30] P. Sterzer, L. Jalkanen, and G. Rees. Electromagnetic responses to invisible face stimuli during binocular suppression. *NeuroImage*, 46(3):803–808, 2009.

[31] F. Tong. Competing theories of binocular rivalry: A possible resolution. *Brain and Mind*, 2:55–83, 2001.

[32] F. Tong, M. Meng, and R. Blake. Neural bases of binocular rivalry. *Trends in Cognitive Sciences*, 10:502–511, 2006.

[33] F. Tong, K. Nakayama, J.T. Vaughan, and N. Kanwisher. Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron*, 21:753–759, 1998.

[34] N. Tsuchiya, H. Kawasaki, H. Oya, M.A. Howard, and R. Adolphs. Decoding face information in time, frequency and space from direct intracranial recordings of the human brain. *PLoS ONE*, 3(12):3892, 2009.

[35] R. VanRullen and C. Koch. Visual selective behavior can be triggered by a feed-forward process. *Journal of Cognitive Neuroscience*, 15(2):209–217, 2003.

[36] A. Wannig, L. Stanisor, and P. Roelfsema. Automatic spread of attentional response modulation along gestalt criteria in primary visual cortex. *Nature Neuroscience*, 14:1243–1244, 2011.

[37] P. Zhang, K. Jamison, S. Engel, B. He, and S. He. Binocular rivalry requires visual attention. *Neuron*, 71:362–369, 2011.

5 Supplementary Materials

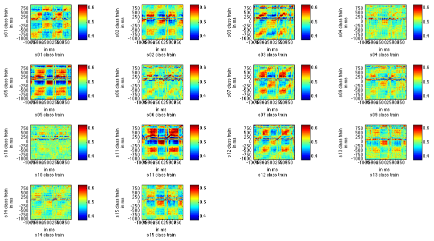


Figure 12: Within-condition classification performance for individual subjects. Some subjects show very strong (red/blue) patterns and others showing weak (green) patterns.

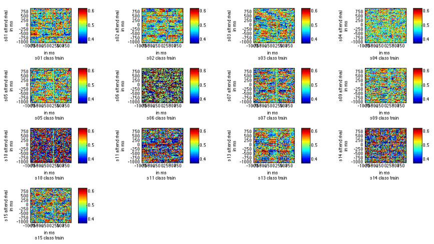


Figure 13: Classification of attended rivalry data for individual subjects. Contrasted with Figure 12, no subjects show a classification pattern.

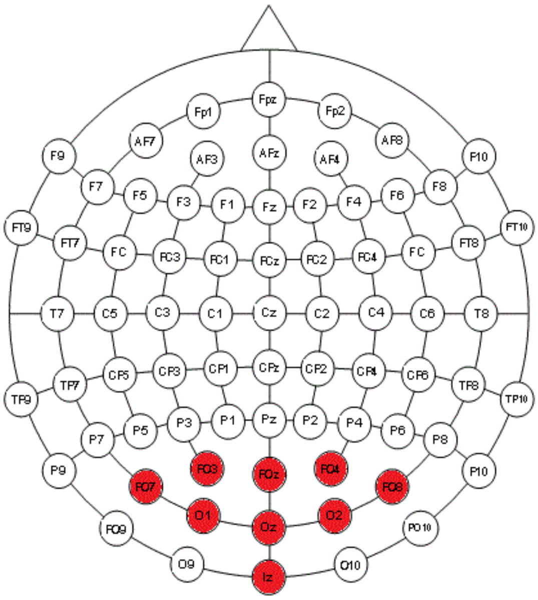


Figure 14: The occipital electrodes included in analysis

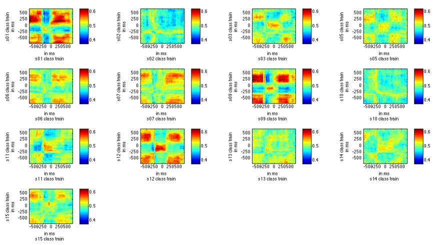
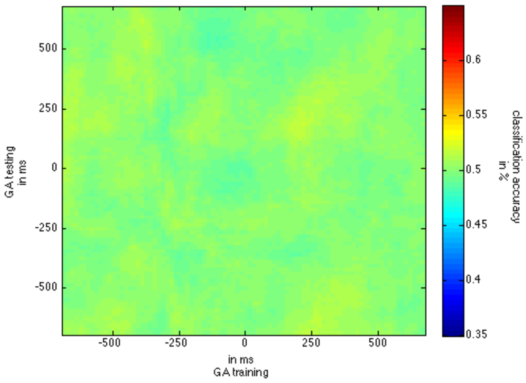
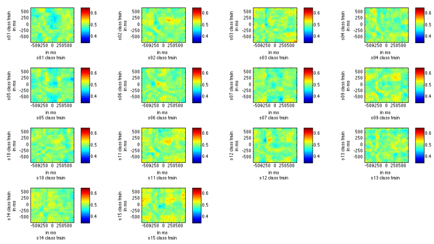


Figure 15: Within-condition classification at 60Hz for individual subjects.

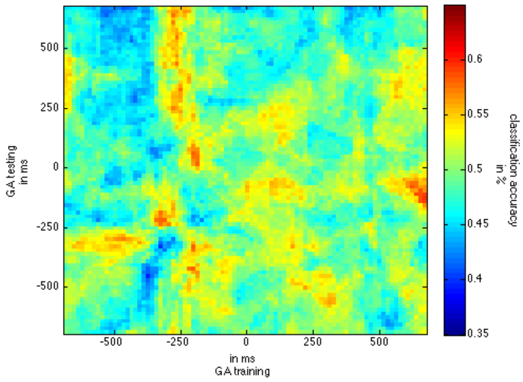


(a) Within-condition classification at 6Hz averaged over all subjects from frontal electrodes.

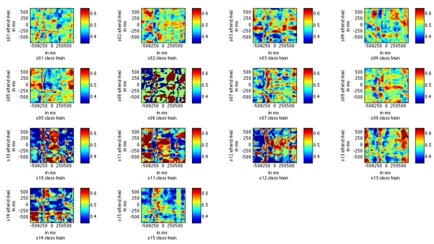


(b) Within-condition classification at 6Hz for individual subjects from frontal electrodes.

Figure 16: Within-condition classification of Classifier Training data at 6Hz from frontal electrodes. Opposed to when occipital electrodes are used (see Figure 16b), there is no clear pattern of classifier performance.



(a) Attended rivalry classification at 6Hz averaged over all subjects from frontal electrodes.



(b) Attended rivalry classification at 6Hz for individual subjects from frontal electrodes.

Figure 17: Attended rivalry classification of Classifier Training data at 6Hz from frontal electrodes. Opposed to when occipital electrodes are used (see Figure 17b), there is no clear pattern of classifier performance.

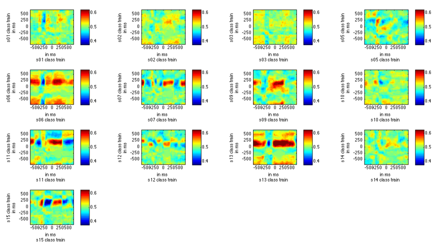


Figure 18: Within-condition classification at 6Hz for individual subjects

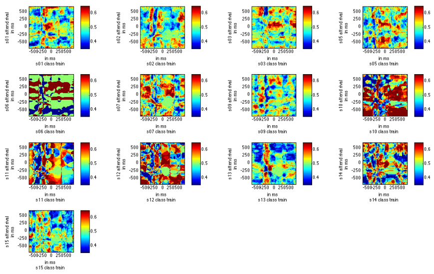


Figure 19: Classification of epoched attended rivalry data at 6Hz for individual subjects