The Role of the Primary Visual Cortex in Perceptual Suppression of Salient Visual Stimuli

Georgios A. Keliris,1 Nikos K. Logothetis,1,3 and Andreas S. Tolias1,2,4,5

1Max Planck Institute for Biological Cybernetics, 72076 Tübingen, Germany, 2Department of Neuroscience, Baylor College of Medicine, Houston, Texas 77030, 3Division of Imaging Science and Biomedical Engineering, University of Manchester, Manchester M13 9PT, United Kingdom, 4Michael E. DeBakey Veterans Affairs Medical Center, Houston, Texas 77030, and 5Department of Computational and Applied Mathematics, Rice University, Houston, Texas 77005

The role of primary visual cortex (area V1) in subjective perception has intrigued students of vision for decades. Specifically, the extent to which the activity of different types of cells (monocular versus binocular) and electrophysiological signals (i.e., local field potentials versus spiking activity) reflect perception is still debated. To address these questions we recorded from area V1 of the macaque using tetrodes during the paradigm of binocular flash suppression, where incongruent images presented dichoptically compete for perceptual dominance. We found that the activity of a minority (20%) of neurons reflect the perceived visual stimulus and these cells exhibited perceptual modulations substantially weaker compared with their sensory modulation induced by congruent stimuli. Importantly, perceptual modulations were found equally often for monocular and binocular cells, demonstrating that perceptual competition in V1 involves mechanisms across both types of neurons. The power of the local field potential (LFP) also showed moderate perceptual modulations with similar percentages of sites showing significant effects across frequency bands (18–22%). The possibility remains that perception may be strongly reflected in more elaborate aspects of activity in V1 circuits (e.g., specific neuronal subtypes) or perceptual states might have a modulatory role on more intricate aspects of V1 firing patterns (e.g., synchronization), not necessarily altering the firing rates of single cells or the LFP power dramatically.

Introduction

The use of visual stimuli that induce bistable perception has been established as a classical paradigm to identify the neural circuits subserving subjective perception (Attneave, 1971; Rock et al., 1994; Logothetis, 1999). A celebrated example of such a perceptual phenomenon is binocular rivalry (BR), involving alternations of visual perception between two different images presented dichoptically at corresponding retinal locations (DuTour, 1760; Wheatstone, 1838; Breese, 1899, 1909).

The primary visual cortex (V1) was implicated as an important candidate for the site of perceptual suppression during BR based on numerous psychophysical studies (Abadi, 1976; Cogan, 1987; Blake, 1989; Blake et al., 2006). In particular, competition between monocular channels offered a straightforward mechanism for the suppression of one of the two stimuli (Lehky, 1988). However, neurophysiological results in monkeys did not corroborate this hypothesis, but instead provided evidence for competition primarily between neurons in areas beyond V1 and specifically areas V4, V5/MT and IT (Logothetis and Schall, 1989; Leopold and Logothetis, 1996; Sheinberg and Logothetis, 1997). Moreover, a set of intriguing human psychophysical results conjectured that BR may involve competition between alternative higher-level stimulus perceptual interpretations (Diaz-Caneja, 1928; Kovács et al., 1996; Logothetis et al., 1996) similar to other bistable stimuli (e.g., necker cube) that do not involve interocular competition (Blake and Logothetis, 2002). Yet, subsequent studies using functional magnetic resonance imaging (fMRI) in humans, provided evidence that activity in V1 is robustly modulated by the subjective percept supporting the ocular competition hypothesis (Polonsky et al., 2000; Tong and Engel, 2001). These results engendered an apparent controversy between human fMRI and monkey electrophysiological recordings. Given the relationship between the BOLD signal and the LFP (Logothetis et al., 2001; Goense and Logothetis, 2008), one could speculate that the robust perceptual modulations reported in human fMRI studies may also be reflected in LFP signals. To this end, recent studies using the paradigm of generalized flash suppression (GFS) implicated that low-frequency LFPs show stronger modulations with perception compared with single-unit activity and high-frequency LFP (Wilke et al., 2006; Maier et al., 2008). Importantly though, GFS does not involve interocular competition like BR.

Here we undertook a study to characterize in detail and compare the extent to which different electrophysiological signals (spiking activity and various LFP frequency bands) are modulated by perception under conditions of interocular competition. We also tested directly the conjecture that monocular neurons in...
V1 robustly reflect perception (Lehky, 1988; Blake, 1989; Tong and Engel, 2001; Haynes and Rees, 2005). We used a variant of BR, namely binocular flash suppression (BFS) (Wolfe, 1984), while recording neural activity from V1 using tetrodes. The first electrophysiological studies using BFS were performed in anesthetized cats (Sengpiel and Blakemore, 1994; Sengpiel et al., 1995) and implicated that interocular interactions at the level of binocular neurons in V1 could provide a possible neural basis for the perceptual switches experienced during BR. Later on, BFS paradigms in awake, behaving monkeys as well as humans have been successfully used in electrophysiological experiments to study the role of higher areas in subjective perception (Sheinberg and Logothetis, 1999; Kreiman et al., 2002; Maier et al., 2007).

In our study, we recorded spiking activity from hundreds of single units and simultaneously acquired LFP signals during the dichotic presentation of orthogonal sinusoidal gratings. We find the following. (1) In agreement with previous studies (Leopold and Logothetis, 1996) only a moderate percentage of neurons (20%) in V1 is modulated in parallel with perception. The magnitude of their modulation is substantially smaller than the physical preference of these neurons. (2) Neurons showing perceptual modulations in V1 are from both binocular and monocular classes with equal probability. (3) Only moderate perceptual modulations of the power in different frequency bands of the LFP are found.

Materials and Methods

Electrophysiological recordings and surgical methods. Electrophysiological recordings were performed in two healthy adult male rhesus monkeys (Macaca mulatta) weighing 16 and 11 kg respectively (monkeys D98 and F03). All experiments were conducted with great care to ensure the well being of the animals and they were in full compliance with the guidelines of the local authorities (Regierungspäsidium) and the European Community (EUVD 6/609/EEC) for the care and use of laboratory animals. Recording chambers were positioned stereotactically over the operculum in area V1 in three hemispheres (both hemispheres in D98 and right hemisphere of F03) with the aid of high-resolution magnetic resonance anatomical imaging. These images were collected in a vertical 4.7 tesla scanner with a 40 cm bore-diameter (Biospec 47/40; Bruker Medical). The system had a 50 mT/m (180 µs rise time) actively shielded gradient coil (B-GA 26, Bruker Medical) of 26 cm diameter. A custom chair and custom system for positioning the monkeys in the magnet were used. We collected anatomical data using T1-weighted high resolution (256 × 256 × 160 real data points at 0.5 mm isotropic linear resolution) images with 3D-MDEFT (modified driven equilibrium Fourier transform) pulse sequences, with an echo time (TE) of 4 ms, repetition time (TR) of 22 ms, flip angle (FA) of 20° and four segments. The anatomical scans were done while the animals were under general anesthesia. The skull parameters were extracted using morphological methods (Paravision; Bruker Medical) and we created a 3D rendered surface (Analyze; Mayo Foundation, Rochester, NY) for designing the cranial posthead and the recording chambers to fit the skull surface. A 5-axis CNC machine (Willemin-Macodel W428) was used to build these form-specific implants that resulted in an excellent fit between the implants and the underlying skull surface. These methods have been described in detail previously (Logothetis et al., 1999).

An array of tetrodes was chronically implanted over the operculum in area V1 inside a form-specific chamber constructed from medical-grade titanium (monkey D98 left hemisphere). In both monkeys, we also recorded nonchronically from form-specific chambers implanted in the right hemispheres. The chamber of monkey D98 was made of medical-grade titanium while the chamber of monkey F03 was made of polyether ether ketone (TECAPEEK; Ensinger GmbH). All chambers were implanted under aseptic conditions under general anesthesia. Initially, the animals received subcutaneous injections of Rubinol (0.01 mg/kg) and Ketavet (15 mg/kg) and subsequently they were prepared for intubation by intravenous injections of fentanyl (0.003 mg/kg), trapanal (5 mg/kg) and lysthenol (3 mg/kg). During surgery, the animals received balanced anesthesia consisting of isoflurane 1.3%. The surgical procedures are described in detail previously (Logothetis et al., 2002). All recordings were conducted with tetrodes attached to microdrives that could be manually adjusted independently. For the chronic recordings, neural activity was recorded using a custom-built array of tetrodes (Tolias et al., 2007). The distance between nearby tetrodes was 200 µm. For the nonchronic recordings, one to four (in most sessions) manually adjustable microdrives (Crist Instrument Co.) were inserted into a custom-built grid and activity was recorded using tetrodes.

Multiunit and single-unit activity was sampled at 32 kHz, digitized (12 bits), and stored using the Cheetah data acquisition system (Neuralynx). LFP signals were recorded by filtering the raw voltage signal using analog bandpass filtering (high-pass set at 1 Hz and low-pass set at 475 Hz) and digitized at 2 kHz (12 bits). Multiunit activity was defined as the events that exceeded a predefined threshold (25 µV) of the filtered, digitized signal (analog filtering high-pass set at 600 Hz and low-pass set at 6 kHz and digitized at 32 kHz, 12 bits). Single units were isolated using a custom-built offline clustering system working on features extracted from the recorded waveforms (Tolias et al., 2007). No preselection functional criteria were applied for the neurons. Details of single-unit isolation methods have been described previously (Tolias et al., 2007).

The animals were implanted with a scleral search coil (Robinson, 1963; Judge et al., 1980) and their eye movements were monitored on-line. Data were also collected for off-line analysis using both the QNX-based data acquisition system at 200 Hz and the Cheetah data acquisition system at 2000 Hz.

Visual stimulation and behavioral paradigm. Visual stimuli were displayed using a dedicated graphics workstation (TDZ 2000; Intergraph Systems) running an OpenGL-based stimulation program. Stimuli were presented dichoptically by using a custom-made stereo with two LCD monitors at both sides running at a resolution of 1280 × 1024 and a refresh rate of 60 Hz. The behavioral aspects of the experiment were controlled using the QNX real-time operating system (QNX Software Systems Ltd).

At the beginning of each session, the mirror stereoscope was positioned in front of the monkeys’ head and two circular apertures were aligned with the animal’s eyes. These apertures served to limit the visual field to the central 15° of visual angle and prevented nasal viewing of the opposite display. In succession, we calibrated the monkeys’ eye movements using a fixation-saccade task. To ensure that the two displays were correctly aligned in front of the two eyes we used the following procedure. First we calibrated the left eye alone while the right eye display was kept blank. The animal had to fixate briefly on a central fixation spot (1°) and the fixation spot moved on the periphery randomly in eight different directions (Δp = 45°). After successful acquisition of the presented saccadic locations, the amplitude of the saccades was increased until the monkey failed, in this way mapping exactly the visible portion of the monitor. When the first eye was fully calibrated we switched to an iterative procedure between the two eyes: first the central fixation spot was presented briefly to the left eye. After the monkey acquired fixation, it was switched off and switched on in the right eye. If the monkey could not fuse (i.e., directly overlay the two stimuli) he performed a saccade to a new location. The monitoring system estimated the amplitude of his saccade by calculating the difference between the two fixations and moved the location of the right eye fixation in the opposite direction for the following trial. Usually after a few trials of calibration (typically four to five) the monkeys were able to fixate the target continuously (although it was switched between the eyes) and we concluded that they could correctly overlay (fuse) the two displays. The fixation-saccade procedure was then performed for the right eye alone with all the targets displaced according to the offset values registered in the previous eye-switching procedure. Finally, we checked the calibration with both displays on. The full calibration described above was performed additionally at the end of the experiments and sometimes between the sessions to ascertain that the two displays remained correctly aligned. During calibration, the monkey received a drop of juice at end of each trial.
After eye calibration and alignment of the displays, a coarse receptive field mapping was performed to position the stimulus for the experiments. Oriented gratings (similar to the ones used in the BFS experiments) were presented parfoveally while the monkey fixated a central target. The multiunit responses of the cells were then selected to be used in the BFS experiments. We will further refer to these two orthogonal orientations as \( \theta \) and \( \theta_{\text{orth}} \). In a subset of experiments for which more than one tetrode was used, the stimulus optimization was typically performed separately for each site.

To study the relationship between neural activity and perception, we used the paradigm of BFS. In this case, two rivalrous patterns are presented dichoptically and asynchronously to the two eyes. Under these conditions, the latter pattern dominates perceptually over the first, provided the two differ sufficiently in their structure (Wolfe, 1984). Several studies have shown that BFS is intimately related to BR at the psychophysical level. Specifically, the results of several parametric manipulations of BFS suggested a common mechanism between BR and BFS. Moreover, forward masking or simple light adaptation are not thought to underlie the perceptual suppression during BFS (Wolfe, 1984; Blake et al., 1990; Baldwin et al., 1996; Kreiman and Koch, 1999; Kreiman, 2003; Brascamp et al., 2007). A recent study (Tsuchiya et al., 2006) compared the depth of suppression for BR and BFS by using a probe detection task and found that BFS induces a transient increase of suppression up to 300 ms after the flash but then suppression becomes at the same level as in the case of BR (to avoid this transient effects of BFS in our study we excluded the first 500 ms after the flash). Based on this evidence, the authors argue that the two paradigms are based on the same mechanism of perceptual suppression. Similarly, a computational modeling study (Nichols and Wilson, 2009) demonstrated that the differences between the techniques can be attributed to the sustained versus transient stimulation rather than a difference in the mechanism of perceptual suppression. Therefore, BFS is thought to involve a temporary burst of inhibition at the onset of the flash that might be critical for the perceptual switch (Macknik and Livingstone, 1998; Macknik et al., 2000) but it is certain that BR prevails after a short delay and the subjects experience spontaneous alternations.

We used the following experimental paradigm in our experiment: after the monkey acquired fixation on a colored square target (0.2º) for 300 ms, static, sine-wave grating stimuli were presented dichoptically to the two eyes. Typically the size of the gratings was 1–2º in diameter, the spatial frequency 3–5 cycles per degree and the contrast 70%. During BFS, a grating stimulus was displayed for ~1 s monocularly to the left or the right eye followed by the onset of a second orthogonal grating (“the flash”) to the corresponding location in the other eye for another second without removal of the first grating resulting in a binocular incongruent condition (Fig. 1). The initial monocular presentation reliably biases the perception toward the second grating flashed during the binocular presentation. At the end of each successful trial, a drop of apple juice was delivered to the animal as a reward. For the physical alternation condition, the trials started in the same way with a 300 ms fixation and a monocular presentation of a grating stimulus for 1 s. Then, similar to the BFS trials, a second orthogonal grating was presented in the second eye but the first grating was removed upon presentation of the second (Fig. 1).
Importantly, this resulted in no binocular incongruence but to a simple successive presentation of the two orthogonal orientation gratings in opposite eyes. These conditions mimic the perceptual experiences of the subjects during BFS without introducing binocular conflict and therefore are termed “physical alternation.” During presentation of the stimuli, the monkey had to keep fixating within a circular window with a radius of 0.5° from the center of the colored fixation target; failure to do so resulted in abortion of the trials and no juice reward.

After the end of all recording sessions, one of the animals (D98) was trained to directly report its perception during the presentation of the stimuli by holding one of two levers. The left lever corresponded to the perception of one orientation (135°) while the right lever to the perception of the orthogonal (45°). During training, the animal was solely presented with physical alternation trials (i.e., without incongruence) and the two orthogonal orientations were switched at intervals drawn from a gamma distribution given by the following:

$$f(x; \lambda, r) = x^{r-1} e^{-\lambda x} \Gamma(r),$$

with parameters $\lambda = 8.3$ and $r = 7.2$ that were identical with the ones calculated for monkeys reporting binocular rivalry (Leopold and Logothetis, 2001).

Table 1. Numbers and percentages of significant modulations

<table>
<thead>
<tr>
<th></th>
<th>SUA</th>
<th>LFP (4–20 Hz)</th>
<th>LFP (30–90 Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>582</td>
<td>381</td>
<td>381</td>
</tr>
<tr>
<td>VR (Visually responsive of T)</td>
<td>523 (90%)</td>
<td>362 (95%)</td>
<td>362 (95%)</td>
</tr>
<tr>
<td>SM (Sensory stimulus modulation of VR)</td>
<td>371 (71%)</td>
<td>149 (41%)</td>
<td>275 (76%)</td>
</tr>
<tr>
<td>PM (Perceptual stimulus modulation of VR)</td>
<td>104 (20%)</td>
<td>79 (22%)</td>
<td>64 (18%)</td>
</tr>
<tr>
<td>PαS (Perceptual &amp; sensory of PM)</td>
<td>94 (90%)</td>
<td>55 (70%)</td>
<td>55 (86%)</td>
</tr>
<tr>
<td>xP (Only perceptual of PM)</td>
<td>10 (10%)</td>
<td>24 (30%)</td>
<td>9 (14%)</td>
</tr>
</tbody>
</table>

The absolute numbers and respective percentages of significant modulations are presented for single units (SUA (single-unit activity)) and two frequency bands of the local field potential: 4–20 Hz and 30–90 Hz. In the first row (T), the total numbers of isolated single cells and recorded LFP sites are reported. The second row (VR) presents the number (percentage) of cells/sites that showed significant visual responses. The third row (SM) presents the number of cells/sites that were responding differentially two the different congruent stimuli (sensory modulation) and the fourth row (PM) the number of cells/sites that showed differential responses under the different perceptual conditions (under the same stimulus) as a percentage of visually responsive cells/sites. In the last two rows, PαS presents the numbers of perceptually modulating cells/sites that showed, in addition, sensory modulations and xP presents the numbers of cells/sites that showed exclusively perceptual modulations.

Figure 2. Examples of single-unit modulations during BFS. A–F. In each panel, the activity of one single neuron is presented. In the first three panels (A–C), neurons from monkey D98 are shown and in the last three (D–F), neurons from monkey F03. The diagrams on the top of the panels demonstrate the sequence of stimulus presentation with a green dot denoting the percept during the binocular period (1000–2000 ms). Note that in general we used different pairs of orthogonal gratings (see Materials and Methods) but here cases where the monkey was presented with a horizontal-vertical pair are shown. In each panel, raster plots of 25 trials (rows) of each of two conditions with the same stimulation but different perceptual outcome (see Fig. 1 and Materials and Methods) are presented in the upper part in red and dark gray. At the lower part, the corresponding spike-density-functions (red and dark gray lines) are presented. Spike-density-functions were calculated using a convolution of the spike-trains with a rectangular window of 100 ms width. The shaded areas (lighter red and gray) represent SEM from a total of 100 trials per neuron. Light blue shadings at the background between 500 and 1000 and 1500 and 2000 ms denote the time windows for which we performed the statistical comparisons.
Trials started with the presentation of a fixation target for 300 ms after which one of the stimuli (pseudorandomly selected) was presented parafoveally. At the onset of the stimulus and after every stimulus switch, the animal was expected to press the correct lever within a maximum response time of 1 s or else the trial was considered incorrect and aborted. After pressing the correct lever, the animal had to keep holding the lever either up to the next stimulus switch (for which it was required to change its response) or until the end of the trial which was signified with the removal of the stimulus. If the animal responded correctly and kept its gaze within a window of ±0.5° around the fixation target for the whole interval, the trial was considered correct and the animal received a small amount of apple juice as a reward. Correct trials ended after a minimum prespecified time (typically 4 s) was crossed by the last stimulation interval, which was not interrupted until its completion (stimulation intervals were drawn from a gamma distribution as above). This resulted in trials of variable durations (depending on the duration of the last alternation period) and variable number of alternation periods. Typically, trials lasted 4–6 s and the animal had to report two to five stimulus switches. Importantly, in this stimulation scheme stimulus switches were unpredictable to the animal. When the animal reached a performance level of >95% correct responses, we gradually started presenting to it a small percentage (<25%) of incongruent stimulus trials randomly interleaved with the physical alternation trials. The incongruent stimulation trials started with a monocular presentation of one of the orthogonal gratings for 1 s after which the second orientation grating was switched on in the opposite eye without removal of the first grating. Note that this presentation is identical to the BFS presentation we have used during the recording sessions. During those sessions, the stimulus presentation lasted for 2 s (1 s monocular and 1 s binocular). For these experiments, the binocular stimulus presentation was extended to longer periods (as much as the minimum time of the physical alternation trials—typically 4 s) in contrast to the 1 s used for the BFS recordings. This was intentionally designed to allow the animal to respond to potential spontaneous perceptual reversals following the induced reversals, which we expected to happen shortly after the “flash” of the second stimulus. During the binocular stimulus presentation, the animal was allowed to press any of the two levers and to change its response from one to the other lever an arbitrary number of times. Similar to the physical alternation trials, we required that it did so within a maximum interval of 1 s, or else the trial was considered incorrect and aborted. If the animal followed the above criteria for the whole duration of the stimulation interval we considered the trial as correct and the animal was rewarded with a small amount of juice as in the physical alternation trials. 

**Statistical and data analysis.** Custom programs written in Matlab (MatWorks) were used for data analysis. Statistical significance of sensory and perceptual modulations was assessed by using a nonparametric Wilcoxon rank sum test (also referred to as the Mann–Whitney U test) that performs a twosided test of the null hypothesis that the data of two conditions are independent samples from identical continuous distributions with equal medians, against the alternative that they do not have equal medians. Since for every neuron or recording site we presented two different configurations of the stimuli according to the eye of presentation, we tested each configuration independently and then we corrected for multiple comparisons using a Bonferroni correction. We considered a neuron/site to show significant sensory/perceptual modulations when at least one of the pairs of conditions tested achieved a significance level \( p \) (Bonferroni corrected) smaller than the critical significance level \( \alpha \) (\( \alpha = 0.05 \)). For all of our comparisons, we excluded the first 500 ms of the responses to avoid effects biased to the initial transients. As a result, we always used the last 500 ms of each condition that effectively reflected the sustained part of the responses.

The preference/modulation index was computed using \( d' \) that is related to the discriminability of the two conditions and is defined as follows:

\[
d' = \frac{\mu_A - \mu_B}{\sigma}
\]

Here \( \mu_A \) and \( \mu_B \) denote the mean responses to the two conditions being compared and \( \sigma = \sqrt{\frac{\sigma^2_A + \sigma^2_B}{2}} \) is the pooled variance of the two response distributions. For our purpose, \( d' \) indices were calculated either for pairs of monocularly presented orientation gratings (referred to as sensory or physical preference \( d' \)), or under the dichoptic presentation of two incongruent orthogonal gratings each perceived at a time (referred to as perceptual preference \( d' \)). Oculatory preference was calculated by comparing the monocular responses of the preferred orientation across the two eyes. We have always used right eye as condition A therefore responses were positive for right-eye selective sites and negative for left eye selective sites. The orientation preference for each cell was defined as the \( d' \) value between the two orthogonal orientations presented to the preferred eye. Similarly, perceptual modulation was defined as the \( d' \) value between the binocular incongruent conditions with the same stimuli and different percepts (Fig. 1).

The percentages of perceptual modulations in different classes of neurons were compared by using a \( x^2 \) test for homogeneity (also referred to as contingency table analysis) by using the following formula:

\[
\chi^2 = \sum \sum \frac{(f_{ij} - \hat{f}_{ij})^2}{\hat{f}_{ij}}
\]

In this formula, \( f_{ij} \) refers to the frequency expected in a row \( i \) column \( j \) if the null hypothesis (i.e., that the percentage is independent of class) is true. We analyzed the modulations across oculatory and orientation pref-
The sensory preference (time window 500–1000 ms) was calculated for all visually responsive neurons. A corresponding perceptual preference was calculated according to which stimulus is perceived under BFS. The preference can be because of orientation or ocularity preference, or a combination of both. A corresponding sensory preference (between orthogonal gratings in opposite eyes) can be calculated via the Parks-McClellan optimal equiripple FIR filter design with an attenuation factor of 60 dB/Hz outside the cutoff frequencies. The power spectral density (PSD) of the raw LFP signals was estimated using the multitaper method (Thomson, 1982). This method uses linear or nonlinear combinations of modified periodograms to estimate the PSD. These periodograms are computed using a sequence of orthogonal tapers (windows in the frequency domain) specified from the discrete prolate spheroidal sequences. We used an adaptive nonlinear combination of seven tapers with a time-bandwidth product (NW = 4) of four for each 500 ms data segment. This resulted in spectra with independent frequency bins with a bandwidth of 8 Hz. Average spectrograms were calculated by moving 500 ms windows with 90% overlap.

The LFP power over time was calculated by bandpass filtering the raw signal in different frequency bands. Digital filters were constructed via the Parks-McClellan optimal equiripple FIR filter design with an attenuation factor of 60 dB/Hz outside the cutoff frequencies.

\[
\begin{align*}
    r_{VR} &= 0.43, P_{VR} = 4.99 \times 10^{-49} \\
    r_{PM} &= 0.62, P_{PM} = 1.87 \times 10^{-23}
\end{align*}
\]

Figure 4. A, Scatter plot of sensory versus perceptual modulation of the neuronal responses. The sensory preference (time window 500–1000 ms) was calculated for all visually responsive (VR) neurons (n = 523, black open circles) by the \( d' \) index (see Materials and Methods) between the monocular orthogonal gratings presented in the left and right eyes for each configuration. Note that this sensory preference (between orthogonal gratings in opposite eyes) can be because of orientation or ocularity preference, or a combination of both. A corresponding perceptual preference was calculated using a \( d' \) index during the binocular presentation of the two gratings (time window 1500–2000 ms). Positive values indicate that the response of the neuron was greater during presentation/perception of orientation \( \theta \) and negative values that the response was higher during presentation/perception of the orthogonal orientation \( \theta_{orth} \). Red filled circles correspond to the neurons that showed significant perceptual modulations (Wilcoxon rank sum test, \( \alpha = 0.05 \)) in at least one of the two configurations. A Pearson correlation coefficient was calculated for the whole population of visually responsive (VR) cells as well as for the subset of perceptually modulating cells (PM) and are presented with their \( p \) values at the upper left part of the plot in black and red, respectively. The inset at the lower right presents a bar plot of the mean absolute sensory [VR: 0.84 ± 0.03 (SEM), PM: 1.48 ± 0.09 (SEM)] and perceptual \( d' \) indices [VR: 0.22 ± 0.01 (SEM), PM: 0.41 ± 0.02 (SEM)] for the two populations. B, The spike density functions of four PM neurons are presented. For each neuron, responses from both stimulus configurations are presented in the left and right columns. The diagrams above each column demonstrate the stimulus used in each configuration; a horizontal grating was used to depict orientation \( \theta \) and a vertical for the orthogonal \( \theta_{orth} \). The perceptual transitions from \( \theta \) to \( \theta_{orth} \) and \( \theta_{orth} \) to \( \theta \) are depicted by magenta and green colors respectively while the transitions from right-to-left eyes or from left-to-right with solid and dotted lines, respectively. The sensory and perceptual \( d' \) are printed on the upper side of the plots for each condition compared. The cell identifications are also presented in the scatter plot in A for better visualization. Note that each cell is represented in the scatter plot with two points reflecting the two stimulus configurations from the left and right columns of B.
Results

Perceptual modulations of single-unit activity

We recorded neuronal activity from 582 single units from three hemispheres of two awake, behaving monkeys (M. mulatta). Ninety percent (n = 523) of these cells were visually responsive (280/303 for D98 and 243/279 for F03). Activity was recorded during binocular flash suppression (BFS), a behavioral paradigm that ensures robust perceptual suppression of a monocular stimulus upon presentation of a second stimulus to the other eye after a delay (Wolfe, 1984; Sheinberg and Logothetis, 1997; Maier et al., 2007). The stimuli were positioned so that they covered the classical receptive fields of the neurons (see Materials and Methods for details).

From a total of 523 visually responsive cells, 371 (71%) showed significant sensory tuning, which we also refer to as physiological preference (Wilcoxon rank-sum test, 6 = 0.05). A substantially smaller number of cells [104 neurons (20%)] exhibited significant perceptual modulation (i.e., responses modulated with subjective perception, Table 1). Moreover, only a couple of cells (three to four in each monkey) exploited perceptual modulations comparable to their physical preference (see for example Fig. 2A, D). Instead, most of the perceptually modulated cells showed substantially weaker perceptual modulations compared with their sensory preference during physical alternation (Fig. 2B, C, E, F). Specifically, the magnitude of the average perceptual modulation across these neurons was only 27% [D98: 26%, F03: 29%] compared with the sensory preferences of these cells. Across the population of all cells showing significant physical preference (see method for details) the average magnitude of the perceptual modulation was 15% [D98: 16%, F03: 14%] (Fig. 3). The magnitude of the perceptual effect was also substantially smaller as measured by the absolute value of $d'$ indices [perceptual $d'$: 0.41 ± 0.02 (SEM), sensory $d'$: 1.48 ± 0.09 (SEM)] (Fig. 4A, inset).

We also found that the sensory and perceptual $d'$ indices were significantly positively correlated (Fig. 4A, red dots; Pearson $r = 0.62, p = 1.87 \times 10^{-23}$) demonstrating that the strength of the perceptual effect depends on the strength of sensory tuning to the stimuli. Accordingly, the great majority of perceptually modulating cells (84/104) show modulations in the same direction as their physical preference. This is in contrast to other areas like V4 and V5/MT where it was previously shown that half of the perceptually modulated cells showed higher activity during the perception of their nonpreferred stimuli (Logothetis and Schall, 1989; Leopold and Logothetis, 1996). In our study, a very small number of cells (n = 10) showed significant perceptual modulations (Wilcoxon rank-sum test, 6 = 0.05) without showing significant preference in the physical alternation conditions. The rest of the perceptually modulating cells (n = 94) showed significant preference during both physical and perceptual alternations (see Table 1).

Monocularity and orientation preference

We found both monocular and binocular feature-selective cells to be modulated with perception. In addition, neurons that exhibited stronger orientation or ocularity tuning were more likely to show statistically significant perceptual modulations (Fig. 5). We calculated an orientation and an ocularity preference index for each cell, based on their responses to the two orthogonal gratings presented monocularly to each eye (see Materials and
Modulations of the local field potentials

Local field potential signals were acquired from 381 sites recorded from the two monkeys. A typical example of the raw LFP data recorded from a single site is presented in Figure 6, in which an increase in oscillatory activity in the gamma frequency range with a short delay after stimulus onset is evident. In addition, the magnitude of the gamma-band oscillations is significantly different for the two orthogonal monocular gratings (see Fig. 6 for details), indicative of orientation tuning of the gamma-band oscillations (Frien et al., 2000; Berens et al., 2008a,b). In contrast, during the dichoptic presentation representing differences only in the perception of the two stimuli, there was no obvious differ-
same two frequency bands: 30–90 Hz (age of the maximum modulations during the monocular presentation (0–1000 ms) for the expressed as differences between preferred and nonpreferred stimuli and shown as a percent-
copies of the physical alternation in a sequence of presentation of preferred and nonpreferred is depicted in the diagrams in the upper
imultitaper method (see Materials and Meth-

Before averaging, conditions were sorted to preferred (P) and nonpreferred (N) according to the responses during the monocular period of presentation. A. The average differences in the spectrograms (frequencies <100 Hz) between the preferred and nonpreferred orientations during physical alternation. B. The average differences in spectrograms for the same sites under the flash suppression conditions. C, D, G, H. Time-domain bandpass filtered averages for the physical alternation and flash suppression conditions for two different frequency bands: the gamma-band (30–90 Hz) (C, D), which showed the highest physical preferences (n = 275), and the lower frequencies (4–20 Hz) (G, H), which showed weaker preferences (n = 149). The sequence of presentation of preferred and nonpreferred is depicted in the diagrams in the upper part of C and D (the black dots denote the perceived stimulus). E, F, I, J. Average modulations expressed as differences between preferred and nonpreferred stimuli and shown as a percentage of the maximum modulations during the monocular presentation (0–1000 ms) for the same two frequency bands: 30–90 Hz (E, F) and 4–20 Hz (I, J). The dotted lines in F and J are copies of the physical alternation in E and I and are presented to provide a means of direct comparison between the flash suppression and the physical alternation modulations.

Figure 7. Population averages of the LFP signals for all sites showing physical preferences. We quantified the population mean of LFP modulations by using the difference of the average spectrograms of the preferred and nonpreferred orientations (see Materials and Methods). The average power change under the flash suppression conditions was 0.4 dB (Fig. 7B) compared with 2.1 dB under physical alternation conditions (Fig. 7A). We also analyzed perceptual modulations across different frequency bands of the LFP. The average perceptual modulation in the gamma frequency band was weak compared with the modulation during physical alternation (Fig. 7C,D). On average, the perceptual modulation (during the last 500 ms) was 23% of the modulation of the physical alternation conditions (Fig. 7E,F). The average perceptual modulations of the lower frequency band (4–20 Hz) were also very weak (Fig. 7G,H) with an amplitude of ~8% of the modulation during physical alternation (Fig. 7I,J). The magnitude of the perceptual effect in both frequency bands was also very small as measured by using the d’ indices of each LFP site (Fig. 8). Notably, the gamma frequency band showed substantially higher sensory preferences compared with the low frequencies (Fig. 8A,B).

Behavioral responses
We analyzed the behavioral responses of one animal from a total of 43 sessions [corresponding to 17,230 physical alternation trials and 4025 (24.4%)] binocular flash suppression/binocular rivalry (BFS/BR) trials] while the animal reported its perception. All
sessions were acquired after the animals’ performance stabilized over 95% correct responses to presentation of congruent stimuli.

To make sure that BFS worked behaviorally as expected, we analyzed the responses of the animal to both physical alternation and BFS/BR trials. First, we measured the response time and percentage of correct responses of the monkey to physical alternations of two orthogonal gratings (Fig. 9A, C). The monkey consistently reported the correct grating orientation in 96% of the trials with an average response time of 500 ± 3 ms (mean ± SEM). Then, we calculated the percentage of BFS trials that the monkey reported the new stimulus after the flash. The monkey reported perceiving the newly flashed stimulus in 97% of the trials (Fig. 9B), demonstrating that, under the stimulation conditions we have used, BFS works in almost all trials. The response time was slightly higher in comparison with physical alternations at 539 ± 6 ms (Fig. 9C). We kept the incongruent stimuli on the monitors for up to 5 s to record the time of spontaneous reversals after the flash. The monkey reported reversals with a mean time of 2057 ± 51 ms after the flash, which was longer than the 1000 ms recorded during the electrophysiological experiments (Fig. 9C). The minimum spontaneous reversal time was 1240 ms. The distribution of spontaneous reversals was fitted by a gamma distribution (Fig. 9C; $r = 23.6$ and $\lambda = 10.9$ Kolmogorov–Smirnov goodness-of-fit test, $p = 0.001$).

**BFS and neuronal adaptation**

An inherent potential complication of the BFS paradigm is neuronal adaptation. In contrast to BR, the history of stimulation leading to the two alternative percepts is different. This can introduce differences in the level of adaptation in the neuronal populations encoding the competing stimuli. For example, it is expected that when the preferred stimulus of a neuron is presented first then its activity will be reduced further until the end of the trial compared with the case when the preferred stimulus is presented second. This is due to differences in the length of time for adaptation between these two conditions. Therefore, to appreciate the suppressive effect of the second stimulus on the first we should compare a neuron’s activity with a condition where the first stimulus is presented alone for the whole duration of the trial. To address this issue, we estimated the level of adaptation for the stimulus presented first across the whole duration of the trial by using an exponential decay function (Fig. 10). We find that the presence of the nonpreferred stimulus (presented second in this case) introduces additional suppression compared with the level of activity predicted by a simple adaptation mechanism (Fig. 10A). Moreover, when the presentation sequence is reversed (i.e., when the preferred appears second) the activity level is also suppressed in comparison with the presentation of the preferred stimulus alone (Fig. 10B). These results, demonstrate that a simple model of activity summation that takes into account adaptation, cannot explain the activity during the binocular incongruent stimulation. Instead, nonlinear interocular interactions are necessary to account for the suppression. Such nonlinear interactions were also present at the level of single neurons. Some cells demonstrated much more suppression than it would be predicted by adaptation (supplemental Fig. 3D,F,K, available at www.jneurosci.org as supplemental material) while others exhibited a pronounced enhancement relative to the presentation of the preferred stimu-
lus alone (supplemental Fig. 3 B, C, available at www.jneurosci.org as supplemental material) that also cannot be explained by adaptation. Importantly the direction of the modulation correlated with perception.

As we have shown above, a trivial model of adaptation cannot explain the responses of V1 neurons during BFS. Nevertheless, this does not exclude more complicated types of adaptation for example affecting only the neurons encoding the perceptually dominant stimulus. Indeed, some models of BR include neuronal adaptation (dependent on the stimulus that is being perceived) as a critical component (Tsuchiya et al., 2006; van Ee, 2009; Kang and Blake, 2010) therefore adaptation likely plays an important role even in studies of classical BR.

Eye movements

One possible confound that might account for the perceptual modulations we have found is a potential difference in the distribution of eye-movements between the two perceptual conditions. To control for this, we extracted various eye-movement parameters (fixation positions, microsaccade amplitudes, microsaccade directions and microsaccade rates) and compared their distributions between the two different perceptual conditions as we did for the electrophysiological signals (for details see Materials and Methods and supplemental Fig. 1, available at www.jneurosci.org as supplemental material). We did not find a significant difference between the distributions for any of these parameters (see supplemental Fig. 2, available at www.jneurosci.org as supplemental material).

Discussion

Discovering which aspects of neural activity underlie our subjective percepts and not simply the sensory input impinging upon us has been a question that has fascinated scientists and philosophers for centuries. Specifically, the role that area V1 plays in perception has been a subject of debate (Crick and Koch, 1995; Pollen, 1995). Psychophysical, single-unit and more recently fMRI studies in primates have argued both for and against V1 activity robustly reflecting perception (Blake, 1989; Leopold and Logothetis, 1996; Logothetis et al., 1996; Polonsky et al., 2000; Tong and Engel, 2001). Here we have undertaken a comprehensive study to investi-
hierarchical levels of vision, and incorporating possible competition of both ocular and feature-selective channels. Furthermore, additional psychophysical and computational evidence supports the notion that indeed both competition between monocular channels and high-level stimulus interpretations are involved in perception during binocular competition (Dayan, 1998; Brascamp et al., 2007; Ikeda and Morotomi, 2007; Silver and Logothetis, 2007; Bhardwaj et al., 2008). However, electrophysiological evidence supporting this idea has been absent. Here we find that equal percentages of monocular and binocular cells are modulated with perception. In addition, the strength of the perceptual effect correlates with both the orientation and ocularity preferences of the neurons. These results provide the first direct electrophysiological support—at the level of the activity of single cells—for the idea that during BFS, competition does indeed involve mechanisms active across both monocular and binocular neurons. This finding is in agreement with human fMRI evidence for modulations in V1 according to the eye-of-origin of the perceived and suppressed stimuli that led those investigators to reinstate the hypothesis of ocular competition at the level of V1 (Tong and Engel, 2001; Haynes and Rees, 2005) and the lateral geniculate nucleus (Haynes et al., 2005; Wunderlich et al., 2005). Our findings however, illustrate that in addition to eye-of-origin signals, single cells in V1 contain at least equally strong signals for the perceived orientation of the stimulus. In addition, cells tuned to both orientation and ocularity showed a higher tendency to show perceptual modulations. These cells responded differentially for the two different stimulus configurations we have used depending whether the orientation eliciting the stronger response was presented in their preferred eye or the opposite. This further demonstrates that the competition does not happen exclusively for eye-of-origin or orientation signals.

Perceptual modulation of different frequency bands of the local field potential

In contrast to the electrophysiological studies in monkeys including ours (Leopold and Logothetis, 1996; Gail et al., 2004; Wilke et al., 2006), human fMRI studies have found strong effects of perceptual suppression in the activity of the primary visual cortex (Polonsky et al., 2000). Specifically, the BOLD signal during such perceptual alternations modulates almost as much when the stimuli are nonambiguously presented separately (physical modulation). The absence of large amplitude modulations in the spiking activity of single cells could potentially reside in a difference in the nature of the two signals (i.e., BOLD and spiking activity of single cells). The mod- est changes in V1 activity could reflect a modulatory input from higher visual cortices. Such a perception-related modulatory effect could be more pronounced in LFP signals that reflect somatoden- dritic integrative processes (Mitzdorf, 1987).

In the present study, we analyzed the effects of perceptual-transitions to the different bands of LFP to determine whether indeed LFP signals would reflect perception more robustly. We did not find such evidence. LFP signals during the different perceptual states showed only small modulations compared with the modulations during physical alternation of the stimuli. Our result is partially consistent with a recent study that compared directly single-unit activity, LFP and BOLD activity in behaving-macaques using the generalized flash suppression (GFS) paradigm (Maier et al., 2008). They find different magnitudes of the effect for the electrophysiological and BOLD signals thus confirming the known discrepancy (human fMRI vs monkey electrophysiology) within the same species. However, in agreement with two previously contacted studies (Gail et al., 2004; Wilke et al., 2006), they find that low-frequency LFPs (<30 Hz) show a modulation ratio substantially larger than the single-unit activity. Moreover, a recent study (Wilke et al., 2009) has demonstrated that low-frequency LFPs in the visual thalamus are critically dependent on the active engagement of the subjects in the task. Specifically, they show that low-frequency LFPs show robust perceptual modulations only if the animals actively report their percepts during GFS. These modulations are eliminated when the animals passively fixate. Therefore, the absence of stronger low-frequency LFP modulations during passive-fixation-BFS we report in our study is entirely consistent with the hypothesis that low-frequency LFPs are reflecting feedback from higher visual areas, which is stronger when the subjects are actively engaged in the task. In addition, this hypothesis can potentially explain the modulations observed in human fMRI studies for which subjects actively reported their percepts. It remains to be shown if perceptual modulations observed in human fMRI can be significantly reduced or abolished when subjects only passively fixate during the presentation of the bistable stimuli. Alternatively, it is possible that differences in the experimental paradigms trigger distinct mechanisms involved in perceptual suppression introducing therefore additional differences in the results (e.g., a potential role of center surround mechanisms in the case of GFS).

Based on the ongoing disagreement between the BOLD results and these classical neurophysiological measures (spikes and power of the LFP in different frequency bands), it is possible that simply measuring the power of the LFP in different frequency bands is not sufficient to capture processes contributing to the modulations of the BOLD signal. Neural activity from multiple sources might be generating field potential changes with inverse signs that add to zero or show minimal changes when recorded using single-point measurements. An example could be spatially disorganized dipoles with minimal or no spatial summation. In addition, anatomy suggests that top-down influences reside mainly in the supragranular layers, pointing to a clear hypothesis for layer specificity.

More detailed studies of the microcircuit organization could reveal whether and how more intricate aspects of activity patterns in V1 may be more robustly related to the mechanisms of perceptual suppression. For instance, a particular subtype of neurons or specific interactions between cells could robustly encode the percept in V1. An important direction for future research is to dissect and understand in detail the microcircuit mechanisms involved in visual perception.

References


