Reference frames for representing visual and tactile locations in parietal cortex

Marie Avillac[1], Sophie Denève[1], Etienne Olivier[2], Alexandre Pouget[3] & Jean-René Duhamel[1]

The ventral intraparietal area (VIP) receives converging inputs from visual, somatosensory, auditory and vestibular systems that use diverse reference frames to encode sensory information. A key issue is how VIP combines those inputs together. We mapped the visual and tactile receptive fields of multimodal VIP neurons in macaque monkeys trained to gaze at three different stationary targets. Tactile receptive fields were found to be encoded into a single somatotopic, or head-centered, reference frame, whereas visual receptive fields were widely distributed between eye- to head-centered coordinates. These findings are inconsistent with a remapping of all sensory modalities in a common frame of reference. Instead, they support an alternative model of multisensory integration based on multidirectional sensory predictions (such as predicting the location of a visual stimulus given where it is felt on the skin and vice versa). This approach can also explain related findings in other multimodal areas.

We acquire information about our environment though multiple sensory channels, and it is by sampling, comparing and combining multisensory signals that our brain constructs an accurate representation of space. This process, known as multisensory integration, is a widely distributed property of the nervous system of many different species. In non-human primates, input from different sensory modalities converges on association areas in the parietal[1–3], temporal[4] and frontal cortices[2,5,6] and in subcortical structures like the putamen[2] and superior colliculus (SC)[7].

When considering how multisensory representations are generated, a specific difficulty arises from the observation that in unimodal sensory areas, spatial locations are encoded differently in each modality: visual receptive fields (RFs) are anchored to the retina, auditory RFs to the head, and tactile RFs to the skin surface. How does the nervous system cope with this? A standard view is that multisensory neurons integrate information across homologous spatial locations in unimodal sensory maps and construct multisensory maps where RFs are aligned in the different modalities[8]. However, most of the relevant data come from experiments conducted in anesthetized animals, whose eyes and head are immobile and maintained in central alignment. In alert, freely moving animals, constant eye and head movements create multiple possible alignments between these primary sensory maps.

Studies that have untied eye, head and body-centered frames of reference by mapping RFs for different postures have found mixed and sometimes counterintuitive results. In the SC[9,10] and in the frontal eye field (FEF)[11], visual and auditory stimuli are reportedly brought into a common eye-centered reference frame. In the ventral premotor cortex (PMv), some bimodal visual and tactile RFs are aligned in a body-centered frame of reference[2,12–14]. Such cells seem to remap all modalities into a common frame of reference, but those cells are

often interspersed with cells with intermediate properties. For instance, most auditory RFs in the SC are intermediate between eye- and head-centered: that is, when the eyes move, the auditory RFs shift only partially with the eye[9,10]. This type of visual-auditory RF is also reported in the lateral intraparietal area (LIP)[15], and some PMv cells have partially shifting visual RFs[14].

A similar form of intermediate spatial coding is observed in the VIP[16]. VIP neurons respond to visual, somatosensory, auditory and vestibular stimuli[3,17–20]. In a study designed to investigate the reference frame used to represent visual space, it was found that, contrary to most visual areas, VIP contains neurons with head-centered and partially shifting visual RFs, in addition to cells with eye-centered RFs[16].

The role of partially shifting RFs in multisensory integration is not entirely clear. One possibility is that these partial shifts have no functional role: that is, they represent meaningless deviations from perfect remapping, which can be averaged out at the population level. Alternatively, they could contribute to the ability of neural circuits to perform optimal multisensory integration. This notion is supported by a recent neural network model which shows that partial RFs are perfectly suited to optimal multisensory integration as well as multidirectional sensory predictions, such as predicting the projected point of impact on the body of an approaching object[21,22] or, conversely, determining the visual location of a noxious tactile stimulus in order to execute a defensive movement toward it[23]. The model is based on an architecture known as a recurrent basis function network. As shown previously[22], the basis function units in this network use partially shifting RFs to perform spatially accurate multidirectional sensory predictions.

To characterize the nature of multimodal representations in VIP, and to further test the validity of recurrent basis function network as a model of multisensory integration, we mapped the visual and tactile

[1]Institut des Sciences Cognitives, Centre National de la Recherche Scientifique, 67 Boulevard Pinel, 69675 Bron, France. [2]Laboratoire de Neurophysiologie, University Catholique de Louvain, Avenue Hippocrate, 54, B-1200, Belgium. [3]Department of Brain and Cognitive Sciences, Meliora Hall, University of Rochester, Rochester, New York 14627, USA. Correspondence should be addressed to J.-R.D. (duhamel@isc.cnrs.fr).

RFs of VIP neurons. We wanted to know whether the visual and tactile RFs are in spatial correspondence for all eye positions: that is, whether VIP cells encode the same spatial location regardless of the modality being stimulated. We show that this is not always the case: a large fraction of VIP cells have visual and tactile RFs that are not always in spatial register. This is counterintuitive, as it implies that VIP cells respond maximally to different spatial locations depending on the modality being considered. This is also inconsistent with the idea that VIP neurons remap all modalities in the same frame of reference. Notably, however, multisensory units in recurrent basis function networks demonstrated the same type of RFs. These findings reinforce the notion that VIP might be part of a large recurrent network used for optimal multisensory integration as well as multidirectional sensory predictions. We argue that this approach can be applied to other multimodal areas, suggesting that the basis function approach provides a general framework to understand the neural basis of multisensory integration.

RESULTS

In two monkeys, we recorded from 63 VIP neurons (33 in monkey M, 30 in monkey N) responding significantly to either visual stimuli ($n = 11$), tactile stimuli ($n = 3$) or both ($n = 49$; Kruskall-Wallis, $P < 0.05$). Quantitative analyses were restricted to cells in which there was a well-defined RF for three different fixation points. We thus obtained 53 sets of visual RFs and 31 sets of tactile RFs. RFs for both modalities were computed for 26 neurons.

Visual RFs were mapped using arrays of briefly presented moving bars. Tactile RFs were mapped using a grid of blunted needles positioned over the monkey's face. Air puffs delivered through these needles evoked short-latency, robust responses in VIP neurons that were either sustained or phasic (**Fig. 1a,b**). RF maps (example shown in **Fig. 1c**) were then computed from spike trains evoked by air puffs at a constant pressure of 0.1 bar and can also be represented as contour maps (**Fig. 1d**).

Visual and tactile RF maps were determined for three horizontal fixation directions (0°, +18° and −18°). The amplitude of the RF displacement as a function of eye displacement was measured by cross-correlating pairs of RF maps obtained for different eye positions, and was quantified by a displacement index (see Methods). A displacement index of 1.0 corresponds to a one-to-one relationship between the amplitude of the RF and eye shifts and thus to a strict anchoring of the RF to an eye-centered frame of reference. For visual RFs, 'eye-centered' is synonymous to 'retinotopic.' A displacement index of 0.0 corresponds to a RF that does not shift at all with the eyes and thus implies a strict anchoring of the RF to a head-centered frame of reference. We use the term 'head-centered' aware that head, body and environmental coordinates cannot be distinguished in the present experiment because the head mid-axis was always aligned with the trunk mid-axis and both were fixed relative to the environment.

Reference frames for encoding visual stimuli

In agreement with previous findings[16], some neurons had a RF that shifted in space in the same direction, and by about the same amount, as the eyes (displacement index close to 1.0) and therefore could be regarded as encoding visual stimuli in an eye-centered system of coordinates (**Fig. 2a**). At the other extreme, we found cells whose RF moved little or not at all in space when the eyes changed position, demonstrating a near-perfect overlap of the RF profiles for the different fixation positions tested (**Fig. 2b**).

Across the cell population tested, these two types of RFs, eye- and head-centered, were not neatly separated. Many cells had 'partially shifting' RFs, moving in space with the eyes, but not by the same amount (**Fig. 2c**). The distribution of displacement indexes for the whole population of VIP cells (**Fig. 2d**) is characterized by a continuum from head-centered (displacement index = 0.0) to eye-centered RFs (displacement index = 1.0) coordinates. There are, however, two clear modes in this distribution close to 0 and 0.8, raising the possibility that the distribution contains only head-centered and eye-centered cells, but
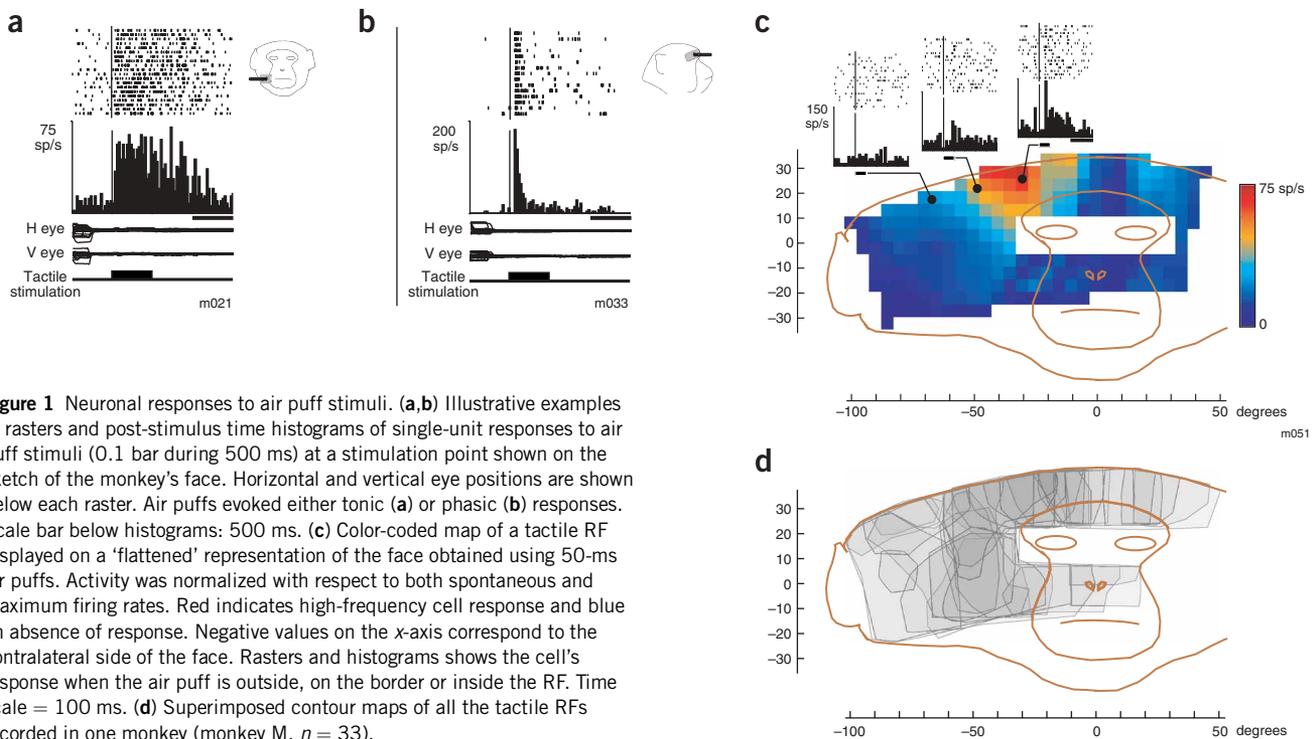
Figure 1 Neuronal responses to air puff stimuli. (**a,b**) Illustrative examples of rasters and post-stimulus time histograms of single-unit responses to air puff stimuli (0.1 bar during 500 ms) at a stimulation point shown on the sketch of the monkey's face. Horizontal and vertical eye positions are shown below each raster. Air puffs evoked either tonic (**a**) or phasic (**b**) responses. Scale bar below histograms: 500 ms. (**c**) Color-coded map of a tactile RF displayed on a 'flattened' representation of the face obtained using 50-ms air puffs. Activity was normalized with respect to both spontaneous and maximum firing rates. Red indicates high-frequency cell response and blue an absence of response. Negative values on the x-axis correspond to the contralateral side of the face. Rasters and histograms shows the cell's response when the air puff is outside, on the border or inside the RF. Time scale = 100 ms. (**d**) Superimposed contour maps of all the tactile RFs recorded in one monkey (monkey M, $n = 33$).

Visual receptive fields
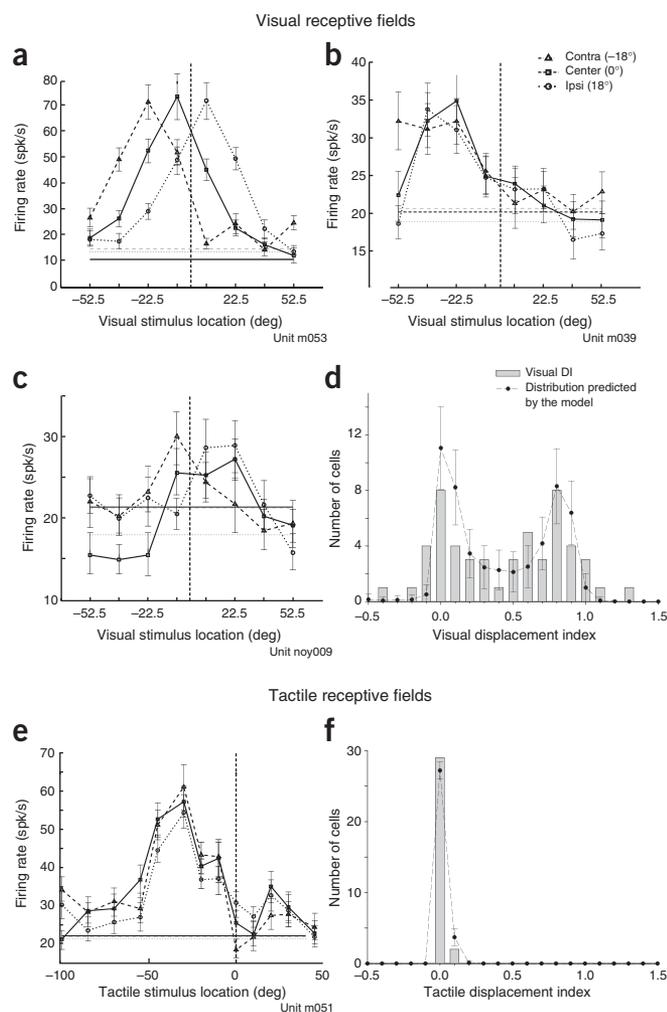
Figure 2 Visual and tactile reference frames. (**a**–**c**) RF location as a function of horizontal eye position. RFs are plotted in head-centered coordinates. Thick curves correspond to activity evoked by the stimulus for each fixation location (mean ± s.e.m.); thin, straight horizontal lines (either solid, dashed or dotted) represent the corresponding baseline activity. Negative values on the x-axis correspond to the contralateral field. **a**, **b** and **c** show examples of eye-centered (displacement index = 0.75), head-centered (displacement index = 0.15) and intermediate (displacement index = 0.46) visual RFs, respectively. (**d**) Displacement index distribution for all visual RFs ($n = 53$). The dashed line represents the distribution expected by the iterative recurrent basis function model (see text). (**e**) Horizontal extension of a representative tactile RF (displacement index = 0.00). (**f**) Displacement index distribution for all tactile RFs ($n = 31$). The dashed line represents the distribution expected by the iterative recurrent basis function model (see text).

no partially shifting RFs. We tested this hypothesis with Monte-Carlo simulations (**Supplementary Note**) and found that the observed displacement index distribution is significantly different from the expected distribution if all RFs were either head-centered or eye-centered (Kolmogorov-Smirnoff test, $P < 0.005$).

Reference frames for encoding tactile stimuli

In contrast to visual cells, all VIP neurons ($n = 31$) encoded tactile stimuli in a unique frame of reference. Shifts in eye position did not affect the tactile RF location. The RF profile shown in **Figure 2e** represents the cell's response to air puffs delivered along the upper row of the stimulation grid. In order to compute a displacement index for tactile RFs and make direct comparisons with visual RFs, we represented tactile, visual and eye positions in a common coordinate system by converting the surface coordinates of the stimulation points into spherical coordinates (see Methods). This transformation involves an arbitrary choice of reference point, which we set at the head center of gravity. This procedure was applied for the purpose of comparing shapes, not exact values of the visual and tactile RF distributions. All tactile VIP neurons had a displacement index ≤ 0.1 (**Fig. 2f**).

Visual and tactile receptive field alignment in bimodal neurons

A direct comparison of the spatial locations encoded in the tactile and visual activity of bimodal VIP neurons is difficult because the stimuli

delivered in each modality do not correspond to homologous locations in space. Nevertheless, if visual and tactile stimuli were remapped in a common frame of reference, one might expect a spatial congruence between the encoded regions of the head surface and extrapersonal visual space. The analysis presented above suggests that this hypothesis is unlikely to hold for all neurons and for all eye positions. Indeed, bimodal cells showed the same broad distribution of visual RF types, from head- to eye-centered encoding, as the general population of VIP cells (**Fig. 3**). Tactile RFs were exclusively head-centered, and therefore only a subset of bimodal neurons encoded both visual and tactile information in this reference frame (13/26 had a displacement index < 0.5 in both modalities). The actual values of tactile displacement index contained in this distribution depend on the particular reference point chosen for the angular coordinate transformation. A different reference might have produced a narrower or wider distribution, but it would still show a single mode at 0.

Naturally, cells with a head-centered RF are more likely to show congruence across different eye positions (**Fig. 4a**). We found this congruence in 85% (11/13) of bimodal cells with RFs closer to a head-centered frame of reference (visual displacement index < 0.5). In some of these congruent neurons (3/13) the overlap was only partial, as the tactile RF occupied a very lateral portion of the head, and its visual counterpart extended beyond the tangent screen used for visual stimulation. We were more surprised that two head-centered bimodal cells showed noncongruent RFs and had a contralateral tactile RF associated with an ipsilateral head-centered visual RF.
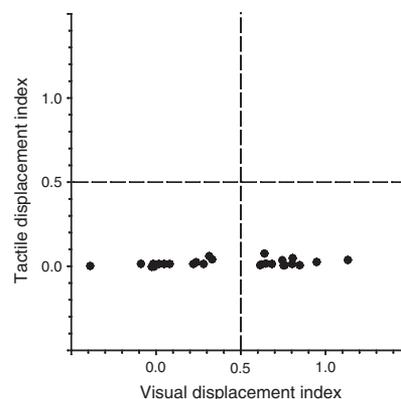

Figure 3 Distribution of the visual and tactile displacement indices for bimodal VIP cells. Only cells with well-defined RFs in both modalities were used for this analysis ($n = 26$). Each black dot corresponds to a single neuron. The reference frame of bimodal neurons is characterized by a continuum from head-centered to eye-centered coordinates for visual RFs and by the spatial invariance of tactile RFs in head-centered coordinates.
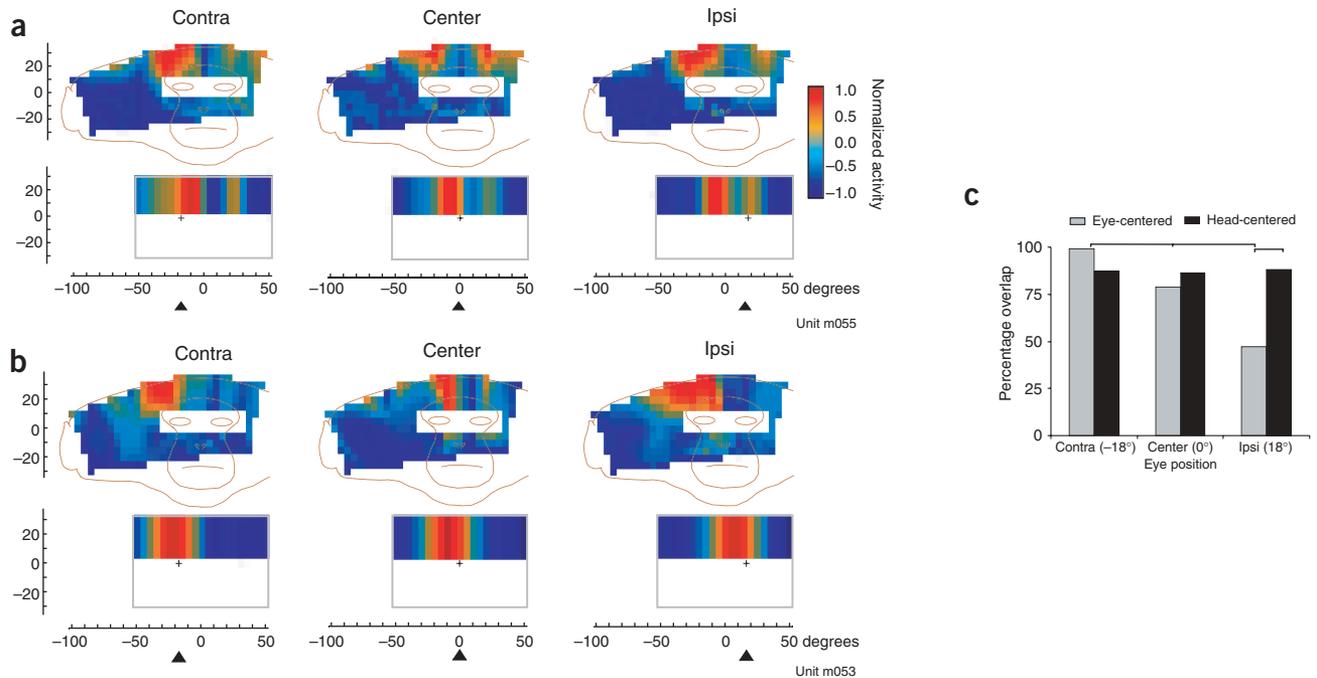
Figure 4 Effects of eye position on the tactile and visual RF alignment in bimodal neurons. (**a**) Example of a neuron with spatially congruent head-centered RFs in both sensory modalities. The visual RF (bottom) is color-coded in order to facilitate comparison with the matching tactile RF (top). The tactile RF was located above the eyebrows bilaterally and was in close correspondence with the visual RF, regardless of the eye position. Arrowheads indicate horizontal eye position. (**b**) Example of a neuron encoding visual and tactile stimuli in different coordinates. The visual RF was eye-centered (displacement index = 0.74). As a result, both the visual and tactile RFs were roughly aligned when the monkey fixated contralaterally or straight ahead but became noncongruent when the monkey fixated ipsilaterally. (**c**) Proportion of the visual RF included inside the tactile RF for head- (displacement index ≤ 0.5) and eye-centered bimodal neurons (displacement index > 0.5). Horizontal bars above histogram indicate significant post-hoc pair-wise comparisons (Holm-Sidak: $P < 0.05$).

In the case of bimodal neurons with visual displacement index > 0.5 (13/26), encoding visual stimuli in a reference frame closer to eye-centered, RF misalignment necessarily occurred for certain eye positions (**Fig. 4b**). For most of such neurons (12/13), tactile and visual RFs remained roughly aligned when the eyes pointed straight ahead or contralaterally but became noncongruent when deviated ipsilaterally. The one exception was a cell with a better alignment for fixation opposite the tactile RF.

To quantify eye position effects on RF congruence, we estimated the horizontal boundaries of visual and tactile RFs, thresholded at 50% of maximum firing rate. We calculated the proportion of the visual RF surface that was included within the tactile RF for the three fixation positions and used this as an index of bimodal RF overlap. The area of RF overlap was estimated at different eye positions for cells with head-centered (displacement index ≤ 0.5) versus eye-centered (displacement index > 0.5) RFs. The results show a mean area of overlap greater than 85% for head-centered cells, which was constant across all eye positions, whereas the amount of overlap increased from ipsilateral to contralateral eye positions for eye-centered neurons (**Fig. 4c**; reference frame by eye-position interaction: $F_{2,44} = 10.04$, $P < 0.001$).

Eye position effects on firing rate

Changes in eye position served to distinguish between eye- and head-centered encoding of spatial locations, but we found that the eye position also had a general effect on the response strength of many VIP neurons akin to a gain modulation of discharge activity across the entire RF surface. Forty percent (21/53) of neurons showed significant changes in visually evoked activity (measured at RF peak) as a function
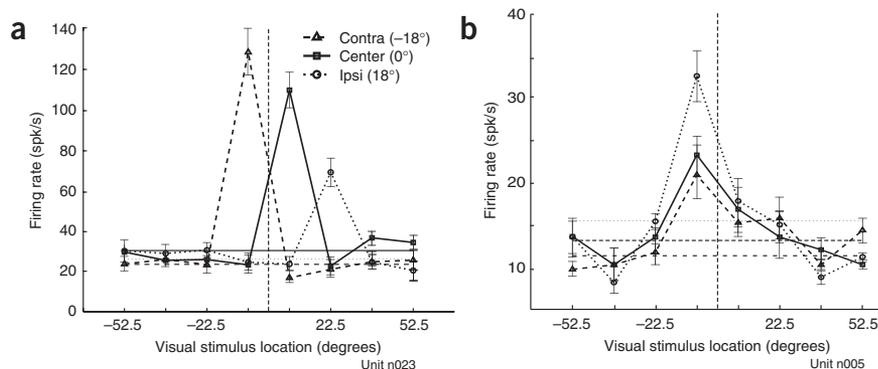
Figure 5 Effect of eye position on visual activity of VIP neurons. (**a**) Eye-centered visual RF (displacement index = 0.84) of a neuron showing changes in activity level depending on eye position. This neuron fired maximally during contralateral fixation. (**b**) Head-centered visual RF (displacement index = −0.03) of another neuron with an eye position effect. The neuron fired maximally during ipsilateral fixation. No significant eye position effect was present on the spontaneous activity in either cell. Error bar: s.e.m. All other conventions as in **Figure 2**.
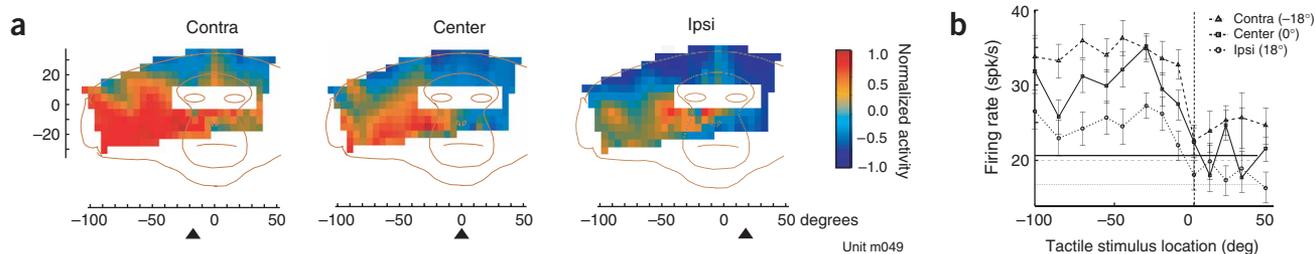
Figure 6 Effect of eye position on somatosensory activity of VIP neurons. (**a**) Color-coded tactile RF maps obtained for eye fixations at 18° contralateral, 0° and 18° ipsilateral. Arrowheads under the plots indicate fixation location; other conventions as in **Figure 1c**. Activity across the RF is maximal for contralateral fixation. (**b**) Firing rate of the same neuron for the lowest row of the tactile stimulation grid. Error bar = s.e.m.

of eye position (Kruskall-Wallis, $P < 0.05$). These effects did not simply reflect an overall change in baseline activity. Although eye position modulated background activity in 11 cells, only four showed an effect of eye position on both visually evoked and baseline activity. Eye position effects were unrelated to RF type (11/27 eye-centered, and 10/26 head-centered neurons; **Fig. 5a,b**). In most cells (81%, 17/21) this gain modulation was monotonic, but the small number of eye positions investigated precluded a test for the linearity of this effect.

We also found a significant modulation of the response to tactile stimulation in 11 out of 31 neurons (Kruskall-Wallis, $P < 0.05$). These effects were also independent of a baseline rate modulation; they were monotonic and were characterized by a gain modulation across the entire RF surface (**Fig. 6**).

Eye position effects on response latency
We examined the response latency of VIP neurons as a function of both the sensory modality and encoded reference frame. Notably, we found that cells with head-centered RFs had longer latencies than those with eye-centered RFs (**Fig. 7a**). Visual latencies ranged between 40 and 174 ms (mean latency: 86.4 ms, s.d.: 31.8 ms, $n = 53$). On average, head-centered cells (displacement index ≤ 0.5) began responding 26 ms after eye-centered cells (displacement index > 0.5; 100.5 ± 33.7 ms and 72.8 ± 23.2 ms, respectively, t-test: $P < 0.001$). The presence of an eye position effect on the discharge rate had no influence on the latency (82.3 ± 32.9 ms and 89.1 ± 31.3 ms for cells with and without a significant gain modulation, respectively; t-test: $P = 0.453$).

Tactile latencies were short and narrowly distributed (**Fig. 7b**), ranging between 10 and 62 ms (mean: 31 ± 13.7 ms, $n = 31$); they were significantly shorter than those of visual responses (t-test, $P < 0.001$). As for visual latency, the presence of an eye position effect did not affect the latency of tactile responses (33.5 ± 17.6 ms and 29.5 ± 11.2 ms for cells with and without a significant gain modulation, respectively, t-test: $P > 0.4$). We did not find a latency difference

between unimodal and bimodal neurons. Finally, visual and tactile response latencies in bimodal cells were not significantly correlated ($r = 0.31$, ns).
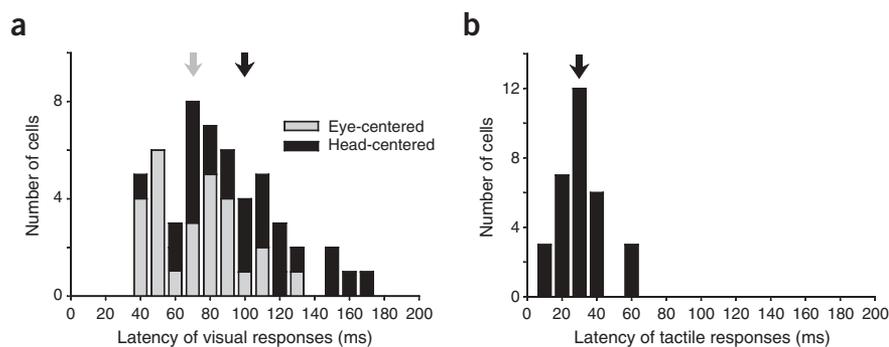
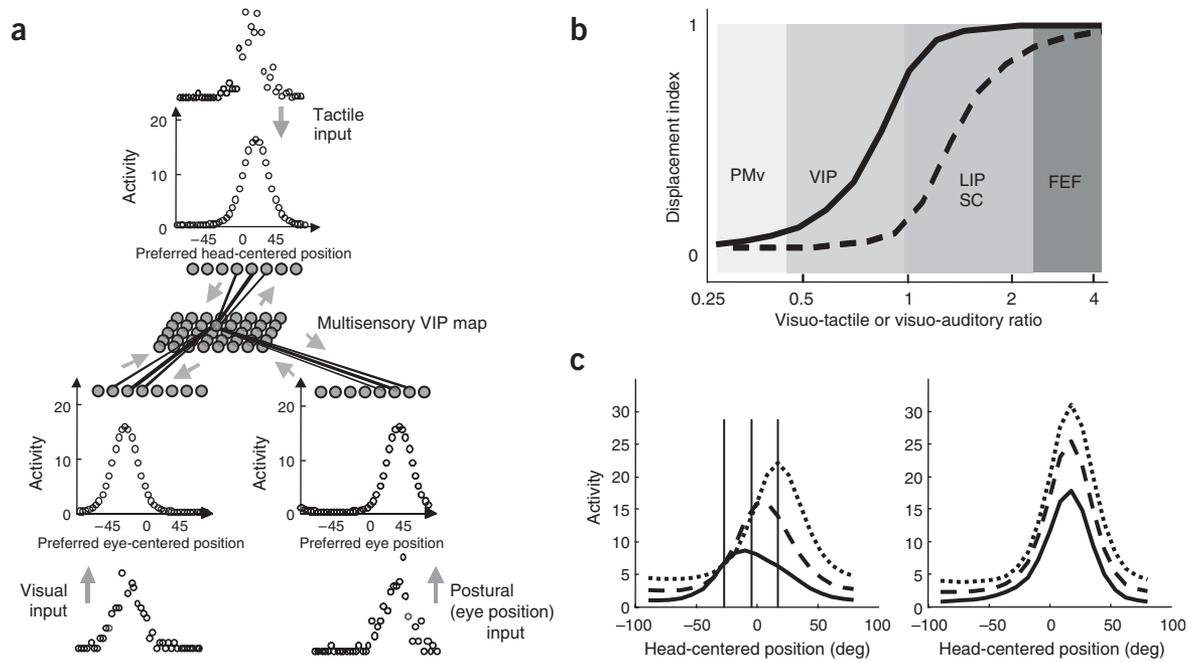### Recurrent basis function network: comparison with data
We have seen that the tactile RFs of multimodal VIP neurons are invariant in head-centered coordinates, whereas their visual RFs partially shift between the eye- and head-centered coordinates. This indicates that a large percentage of VIP neurons do use the same coordinates for their visual and tactile inputs. This result is difficult to reconcile with an intuitive view of multisensory integration according to which multimodal areas remap all their inputs in a common frame of reference. In this section, we explored whether these response properties are consistent with a recent model of optimal multisensory integration using recurrent basis function networks[22,24].

The model (**Fig. 8a**) contains three input layers: one unimodal visual layer, one unimodal tactile layer and one postural layer encoding the position of the eyes. The eye position layer is required for remapping eye-centered coordinates into head-centered coordinates and vice versa. For mathematical convenience, we used neurons with Gaussian tuning curves to eye position, but our approach would also work with sigmoidal tuning curves, which are believed to be more biologically plausible[25].

Intermediate units, also known as basis function units, combine the visual, tactile and eye position inputs and send feedback connections to the unimodal areas. Connections are set so that the feedback from VIP predicts the visual input from the tactile and postural input, and vice versa. Neurons in each layer are arranged topographically: an object at a particular location creates a hill of activity peaking at the position of the cell whose RF is centered on the object. Initialized with noisy hills, the network converges to smooth hills whose position on each layer is the most accurate estimate of object location, given noisy visual, tactile and eye position inputs.

Figure 7 Distribution of visual and tactile response latencies. (**a**) Distribution of visual response latencies for cells with predominantly head-centered (displacement index ≤ 0.5, $n = 26/53$, black) and eye-centered (displacement index > 0.5, $n = 27/53$, gray) RFs. Mean visual latencies of head-centered cells are significantly longer (100.5 ms, black arrow) than latencies of eye-centered cells (72.8 ms, gray arrow; $P < 0.001$) (**b**) Distribution of tactile response latencies (black arrow: mean latency = 31 ms).

Figure 8 Multisensory integration by multidirectional sensory prediction. (**a**) Recurrent basis function network performing optimal visuo-tactile integration (adapted from ref. 22). (**b**) displacement index of visual (solid line) and tactile (dashed line) RFs as a function of the visuo-tactile ratio; that is, the ratio between visual and tactile weights from the input layers to VIP. This visuo-tactile ratio determines the confidence given to the visual modality compared with the tactile modality in estimating object locations. Shaded areas indicate the putative range of visuo-tactile ratio in areas PMv, VIP, LIP and FEF. Note that in the case of LIP, in which no tactile responses have been reported, it is more natural to think of the x-axis as the visuo-auditory ratio. (**c**) Response of a VIP unit with a visuo-tactile ratio of 0.75, plotted as a function of the head-centered position of the object for three different eye positions. Left: visual responses, right: tactile responses. Solid, dashed and dotted lines: eye position at −20°, 0° and 20°, respectively. The vertical lines on the left panel indicate where the curves should peak if the visual RF of this unit was eye-centered. Y-axes in **a**,**b**,**c** correspond to simulated neural activity.

The resulting network can perform optimal Bayesian integration of visual and tactile signals, in the sense that it can compute the most likely location of an object, given noisy visual, tactile and postural inputs (a variation of this model in which the tactile layer is replaced by an auditory layer is described in detail in ref. 22). It can also predict the position of a visual input from its tactile position, and vice versa.

We have shown in previous work that these multidirectional sensory predictions require basis functions with partially shifting visual RFs. In the previous study, however, RFs did not correspond to the immediate response to a visual stimulus, as in our current experiment. Instead, in the previous study, the network was initialized with a visual stimulus and then allowed to relax onto self-sustained activity patterns akin to memory states. Moreover, we had considered only a network in which the strength of the visual and tactile inputs onto each basis function was exactly matched, which may not be true of actual VIP neurons. Finally, we did not investigate the relative positions of the visual and tactile RFs of the basis function units, as we have done here for VIP neurons.

Therefore, to allow comparison with our experimental data, we mapped the visual and tactile RFs of the basis function units by averaging their activity over the first ten iterations in each trial, during which the sensory input was clamped (see Methods). Moreover, we varied the visuo-tactile ratio of the basis function units: that is, the ratio of the visual weights over the tactile weights for a given basis function unit.

The results plotted in **Figure 8b** show the displacement index of the visual and tactile RFs of the basis function units as a function of their visuo-tactile ratio. We found it important, first, that the two curves are not superimposed, indicating that the visual and tactile RFs do not shift

by the same amount, except for extreme values of the visuo-tactile ratio (0.25 and 4). This shows that spatial alignment is not a prerequisite for accurate multisensory predictions, because all the networks we have simulated are spatially accurate, regardless of the value of the visuo-tactile ratio.

We found it important, second, that the intermediate units behave like VIP neurons when the visuo-tactile ratio is below 1 (that is, when the tactile inputs dominate). In that range, the tactile RFs barely shift, while the visual RFs shift partially with shift indices between 0 (head-centered) and 0.8 (close to eye-centered; **Fig. 8b**). Indeed, the model accounts well for the distribution of displacement index found in our dataset: using Monte Carlo simulations (**Supplementary Note**) we found that the distribution of visual and tactile displacement index in our data was not significantly different from the distribution predicted by the model. The best match between the model and data distributions of displacement index were found for visuo-tactile ratio randomly distributed between 0.3 and 1.2 (that is, visual weights 30% to 120% as strong as tactile weights; Kolmogorov Smirnoff test: $P > 0.32$ for visual displacement index, $P > 0.15$ for tactile displacement index; see **Fig. 2d**,**f**.)

Moreover, the gain (that is, amplitude) of the basis function unit responses is modulated by eye position for both modalities (**Fig. 8c** shows the case of visual stimulus). Those observations are notably similar to what we have reported above for VIP neurons. Therefore, our model can account for VIP responses, under the assumption that VIP is dominated by tactile inputs. This assumption is consistent with the physiology and anatomy of VIP: as we have shown here, tactile responses have very short latencies compared with visual inputs, which have longer latencies and are known to go through multiple

cortical areas before reaching VIP. This would suggest that VIP lies closer to the primary somatosensory areas than it does to the primary visual areas, which could explain why the tactile RFs are more stable then the visual ones.

The partial shift of visual RFs with the eye is a consequence of recurrent dynamics in the networks which brings the visual RFs progressively closer to a head-centered frame of reference. Counting activities over longer periods of time or later in the trial results in bringing the two curves in **Figure 8c** closer to one another. This predicts that visual cells with longer latencies have RFs closer to a head-centered frame of reference, as observed in our VIP data.

DISCUSSION
A common representation of visual and tactile space in VIP?
We find that the most counterintuitive result of this study is that visual and tactile RFs are not aligned independently of the posture in a sizeable proportion of bimodal VIP neurons. In agreement with previous findings[16], we have shown that individual neurons in area VIP encode visual locations in different frames of reference that can be eye-centered, head-centered or intermediate between the two. By contrast, tactile RFs are strictly head-centered and remain at the same location on the skin regardless of eye position. Consequently, only bimodal neurons with head-centered visual RFs show aligned visual and tactile RFs across changes in eye position. Bimodal cells with partially shifting or eye-centered visual RFs have roughly congruent tactile RFs for one particular eye position, most often when the animal looks straight ahead or ipsilaterally to the tactile RF. But this alignment is broken for other eye positions, which might suggest that visual and tactile locations are not put in correspondence, and thus not integrated, in the neurons. However, integration is an issue in cases where stimuli are truly bimodal: that is, both visual and tactile. Our stimuli were unimodal. As suggested above, the view of multisensory integration that we propose does not imply that unimodal RFs are remapped in a common frame of reference.

Gain modulation and implicit spatial representation
The position of the eyes in the orbit modulates the gain of 40% of the visual responses and of 35% of tactile responses in our sample of VIP neurons, in agreement with previous work[16,26]. Eye position effects occur across the whole continuum of eye- and head-centered neurons, on both visual and tactile-evoked responses, and in most cases they seem to reflect a monotonic function or 'gain field'. Eye position effects on visual responses are widespread in the striate and extrastriate cortex[27], V3A[28], MT and MST[29], PO/V6 (ref. 30), 7a[31] and LIP[32]. In modalities other than visual, gain fields are reported in the inferior colliculus[33] and premotor cortices[34].

In accordance with a population coding hypothesis[31–33,35–37], it has been suggested that gain modulation effects reflect the implicit representation of visual and tactile stimuli in head-centered coordinates. In fact, both eye-centered and head-centered frames of reference can be computed from partially shifting RFs that are gain modulated by eye position[22,38]. More generally, such a representation can be considered as a basis function map from which any function of the object and eye position can be computed, including non-linear sensorimotor transformations[38]. VIP projects on the ventral premotor cortex[39,40], an area involved in head and limb movements[12–14,41], and may project directly or indirectly to the SC[42,43], a structure important for the production of eye- and head- orienting movements. Gain-modulated visual and tactile responses in VIP could provide the needed intermediate representation for the computations performed in these sensorimotor areas.

Latencies of visual and tactile responses
VIP neurons respond to tactile stimuli with an average latency of about 30 ms, sometimes with a latency as short as 10 ms. Neurons in the primary somatosensory areas 3B and 1 (SI) discharge to air puffs with a latency of about 12 ms[44]: that is, 18 ms on average before VIP. By contrast, VIP visual activity begins about 86 ms after stimulus onset, close to the latency of neurons in the hierarchically related dorsal stream area MT (70–90 ms)[45,46] and about 26 ms after V1 layer 4Cα neurons (60 ms)[47]. Hence, there may be fewer synaptic steps from SI to VIP than from V1 to VIP, and given simultaneous input at the receptor surfaces, tactile signals would activate VIP bimodal cells earlier than visual signals. The significance of this latency difference is unclear, as the assumption of simultaneity is not always ecologically valid. Visual and tactile input can be well correlated in time, for example, during active head movements causing self-motion cues to impinge simultaneously upon different sensory receptors. But in the case of an object moving or being brought (for example, during feeding behavior) toward an observer's face, visual input will actually lead tactile input to area VIP.

We also found that visual latencies vary in function of the coordinates used by the cell to encode stimulus location. Visual head-centered cells in VIP had latencies significantly longer than visual eye-centered cells. This relationship between delays and frame of reference was predicted by our model: when a cell starts responding, its response reflects the current state of the recurrent network in VIP. Visual cells with the shortest latency will reflect the first iteration of the network, corresponding to a purely eye-centered input in the case of visual responses. Cell with longer latencies will reflect the state of the network after several iterations, resulting in visual receptive field that are closer to an intermediate or head-centered frame of reference.

Sensory alignment versus multidirectional sensory prediction
Our experimental and simulation results suggest that VIP neurons are part of a recurrent basis function network involved in optimal multisensory integration and multidirectional sensory predictions. The fact that the visual-tactile RFs were out of alignment for certain eye positions does not require that multidirectional sensory predictions are spatially inaccurate. In fact, it is the opposite: those response properties emerge naturally in networks performing accurate multidirectional sensory predictions.

Our results also suggest that the fact that many visual RFs partially shift when the tactile RFs are head-centered is the result of a slight dominance of the tactile modality over the visual modality in this area (**Fig. 8c**). Similarly, our model predicts that if an area is dominated by visual inputs (visuo-tactile ratio greater than 1), its neurons should have stable eye-centered RFs and partially shifting tactile RFs. The prediction is not specific to visuo-tactile neurons but also holds for any pair of modalities, such as visuo-auditory. We found it interesting that in LIP and the SC, two areas that are primarily visual, auditory-visual neurons behave as predicted: the visual RFs tends to be invariant in eye-centered coordinates, whereas the auditory RFs show a partial shift[9,15]. Finally, if one modality is strongly dominant (that is, if the visuo-tactile or the visuo-auditory ratio is below 0.25 or above 4), RFs in the two modalities will appear to be aligned in the frame of reference of the dominant modality. This could be the case in PMv, where visual RFs are described as skin-centered[14], or in the FEF, where auditory RFs are described as eye-centered[11]. This congruence of the sensory maps could be explained if the tactile modality strongly dominates in PMv, or the visual modality strongly dominates in FEF (**Fig. 8c**). It is therefore possible that all multimodal areas use basis function representations and that the degree of partial shift simply reflects the dominant sensory

modalities in a given area. This suggests that recurrent basis function networks provide a unifying framework to understand the neural basis of multisensory integration.

METHODS

Animal preparation. Two adult rhesus monkeys (*Macaca mulatta*) weighing 4 and 6.5 kg were used in this study following procedures in compliance with European Community guidelines on Animal Care. Animals were prepared for chronic recording of eye position and single neuron activity in area VIP during a single surgery performed under propofol anesthesia (10 mg kg^{-1} for induction and 0.3–0.4 mg kg^{-1} min^{-1} for maintenance). We followed standard procedures for eye position search coil[48], head restraint post and recording chamber (P5 L12) implantation.

Experimental procedures. The experiments were conducted in a dark room. The monkeys faced a tangent screen 57 cm away that spanned a $120° \times 90°$ area of the central visual field. Visual stimuli were back-projected onto the screen by means of a video projector (Electrohome Marquee 8500). Eye movements were recorded with the magnetic search coil technique (Primelec). Tactile stimuli consisted of brief air puffs. Behavioral paradigms, visual displays and data storage were under the control of a PC running the REX system (A.V. Hays, B.J. Richmond & L.M. Optican, WESCON Conference Proceedings **2**, 1–10, 1982). Single VIP neuron activity was recorded extracellularly with tungsten microelectrodes (Frederick Haer, 1–2 MΩ at 1 kHz), amplified using a Neurolog system (Digitimer) and was digitized for online spike discrimination using the MSD software (Alpha-Omega). Area VIP was identified on the basis of a set of reliable physiological properties[3,17]. Access from the lateral bank of the intraparietal sulcus (IPS) shows typical transition from simple visual and saccade-related activities characteristic of LIP to direction-selective visual responses often accompanied by direction-selective somatosensory responses on the face and the head region. Access from the medial bank of the IPS is characterized by a transition from purely hand or arm somatosensory activity to direction-selective visual and face tactile responses. All RF mapping procedures were conducted while monkeys fixated a $0.1° \times 0.1°$ spot of light which could be placed at one of three locations on the horizontal meridian: $0°$, $18°$ contralateral or $18°$ ipsilateral to the recording site. The monkeys were rewarded with a drop of liquid at the end of a visual or tactile stimulation trial, provided that the eyes remained within a $3° \times 3°$ window around the fixation spot. Tactile and visual stimulation trials were collected in separate blocks.

Visual receptive field mapping. Visual RF mapping was restricted to its horizontal extension. The mapping probe was a moving bar of white light subtending $15° \times 1.5°$ whose orientation, motion direction and elevation were optimized manually for each recorded neuron while the monkey fixated centrally. The mapping area extended from $60°$ left to $60°$ right and was divided in eight $15° \times 15°$ subregions within which the bar was swept at $100°$ s^{-1}. Each stimulation trial began with the fixation target appearing at one of the three possible locations (in randomly interleaved order). Once the animal had foveated it for 450 ms, stimuli were swept one at a time in rapid, randomized sequence for 150 ms with an interstimulus interval of 50 ms. The procedure ended when each of the eight stimulus locations had been probed at least 15 times for each fixation position.

Tactile receptive field mapping. Compressed air was delivered through a system of blunted stainless steel needles (diameter = 0.8 mm) fitted through to a plastic mask (Turbocast) that was molded under ketamine anesthesia to fit the monkey's facial morphology. The mask was positioned close to, but not in contact with, the animal's face, and a central opening allowed full view of the tangent screen. The air puff needles were distributed along three horizontal rows (brow, eye and lip levels), spaced 5 mm apart over the muzzle and forehead region and 10 mm apart on the side of the head. The arrangement of the needles allowed us to stimulate the entire contralateral portion of the face and about half of the ipsilateral side. Each needle was connected to a computer-controlled solenoid by means of Teflon tubing. Headphones continuously delivered white noise in the monkey's ears to prevent perception and localization of air puff noise (18). The mask precluded preliminary testing of the tactile responsiveness by hand, so all 39 positions of the mask were stimulated during

the mapping procedure. As for visual mapping, once the animal had foveated the fixation target, stimuli were presented one at a time in rapid, randomized sequence for 50 ms with a 190-ms interstimulus interval, and the procedure ended when each of the tactile locations had been stimulated at least 15 times. Exit air pressure at the skin surface was set at 0.1 bar. This was sufficient to elicit reliable neuronal responses without provoking saccades, eye blinks or other observable reactions. Habituation may have been facilitated by the fact that monkeys received several hundred air puffs on a near daily basis for several months. Video records of control tests using even higher puff pressures (0.2–1.5 bar) show that only one stimulation point provoked an overt response. When applied to the upper lip, air puffs evoked lip retraction, sometimes accompanied by eye blink and brow movement. Thus, the broad distribution of RF sizes and locations which included all the portions of the face (**Fig. 1d**) cannot be simple byproducts of stereotyped motor responses and must reflect the activation of underlying mechanoreceptors.

Construction of visual and tactile receptive field maps. Both visual and tactile data were initially processed in the same manner. RF profiles were constructed offline by counting the total number of spikes evoked by each point of stimulation using a temporal window adjusted to the cell's response latency. Spike count was averaged over the number of presentations and converted into mean firing rates. The horizontal extension of visual RFs is represented by an eight-point vector of mean firing rate values. Tactile receptive profiles contained more data points and are represented as a two-dimensional matrix of mean firing rate values. In order to represent eye position, tactile locations on the head and visual locations in a common spherical coordinate system, the locations of the stimulation points on the mask were converted to angular units, using a fixed reference roughly corresponding to the center of gravity of the head (4 cm behind and 1 cm below the horizontal plane of the eyes).

Cross-covariance analysis. RF displacements in conjunction with changes in eye position were analyzed using the following procedure. Briefly, interpolated receptive profiles were obtained by linear interpolation of ten new points between each original data point (bilinear interpolation in the case of tactile RF data; **Fig. 1b**). Then covariances were calculated between any two RF profiles (derived from mapping conducted during ipsilateral, central or contralateral fixations) that were systematically displaced relative to one another. The two matrices (or vectors in the case of visual RFs) were shifted column-wise (for example, along the eye displacement axis). The shift associated with the highest covariance (Δ_{RF}) was normalized with respect to the difference in eye position (Δ_{Eye}), yielding a displacement index (DI) equal to Δ_{RF}/Δ_{Eye}. Cross-covariance can be shown to provide an estimate of RF displacement that is insensitive to variations in the level of activity associated with the presence of a gain field and more robust than comparisons based on a single RF parameter such as the peak or the center of mass. RF profiles were always computed in head-centered coordinates. Thus, a spatially invariant RF in this reference frame produced a displacement index of 0.0. By contrast, a strictly eye-centered RF produced a displacement index of 1.0, indicating that the shift maximizing the covariance is equal to the difference in eye position between the two mapping conditions. For each sensory modality, this analysis yielded three displacement indices, one for each cross-covariance test (ipsilateral versus central fixation, ipsilateral versus contralateral fixation, central versus contralateral fixation), which were then averaged so that every neuron was characterized by a mean tactile displacement index and a mean visual displacement index.

Statistical analyses. Eye position effects were assessed by comparing mean baseline and stimulus-evoked discharge rate measured at the three fixation locations using non-parametric tests. Visual and tactile neuronal responses latencies were estimated using a sliding-window ANOVA procedure[49] designed to find a significant change in the firing rate of a neuron between the pre- and post-stimulus periods. Briefly, the activity measured in a fixed 100 ms–wide window before stimulus onset is compared with activity in a 20 ms–wide sliding post-stimulus windows, and response latency is defined as the point at which a probability threshold of $P \leq 0.05$ is attained.

Network simulations. Simulations were performed with the same model as in ref. 22 with one major difference. In our previous work, we used the sensory inputs to initialize the activity of the units, but we did not maintain them

thereafter. In the simulations described here, sensory inputs were clamped on the input sensory layers at each iteration. Thus, the activity of the units in the sensory layer was determined according to

$$a_i(t+1) = (1-\lambda)\frac{\left[\sum_k w_{ik}b_k(t)\right]^2}{S+\mu\sum_j\left[\sum_k w_{jk}b_k(t)\right]^2} + \lambda I_i$$

where $a_i(t+1)$ is the activity of sensory unit i at time $t+1$, $b_k(t)$ is the activity of the basis function unit k at time t, w_{ik} is the weight of the connection from the basis function unit to the sensory unit i and I_i is the sensory input onto unit i. All parameters were set as in ref. 22, except for the new parameter λ, which was set to 0.1. RFs were mapped by averaging the activity of the multisensory basis function units over the first $N = 10$ iterations. Decreasing the value of or increasing reduces the spatial discrepancy between the visual and tactile RFs but does not affect the qualitative aspects of our conclusions. The visuo-tactile dominance was adjusted by multiplying the weights of all the connections between the visual layer and the multisensory layer by a number called the 'visuo-tactile ratio'. This ratio was varied from 0.25 to 4. Displacement indices were measured using the same methods as for experimental data.

1. Hyvärinen, J. Regional distribution of functions in parietal association area 7 of the monkey. *Brain Res.* **206**, 287–303 (1981).
2. Graziano, M.S.A. & Gross, C.G. The representation of extrapersonal space: a possible role for bimodal, visual-tactile neurons. In *The Cognitive Neurosciences* (ed. M.S. Gazzaniga), 1021–1034 (MIT Press, Cambridge, Massachusetts 1994).
3. Duhamel, J.-R., Colby, C.L. & Goldberg, M.E. Ventral intraparietal area of the macaque: congruent visual and somatic response properties. *J. Neurophysiol.* **79**, 126–136 (1998).
4. Bruce, C., Desimone, R. & Gross, C.G. Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *J. Neurophysiol.* **46**, 369–384 (1981).
5. Rizzolatti, G., Scandolara, C., Matelli, M. & Gentilucci, M. Afferent properties of periarcuate neurons in macaque monkeys. I. Somatosensory responses. *Behav. Brain Res.* **2**, 125–146 (1981).
6. Graziano, M.S.A., Reiss, L.A. & Gross, C.G. A neuronal representation of the location of nearby sounds. *Nature* **397**, 428–430 (1999).
7. Wallace, M.T., Wilkinson, L.K. & Stein, B.E. Representation and integration of multiple sensory inputs in primate superior colliculus. *J. Neurophysiol.* **76**, 1246–1266 (1996).
8. Stein, B.E. & Meredith, M.A. *The Merging of the Senses* (Bradford, Cambridge, Massachusetts 1993).
9. Jay, M.F. & Sparks, D.L. Auditory receptive fields in primate superior colliculus shift with changes in eye position. *Nature* **309**, 345–347 (1984).
10. Jay, M.F. & Sparks, D.L. Sensorimotor integration in the primate superior colliculus. II. Coordinates of auditory signals. *J. Neurophysiol.* **57**, 35–55 (1987).
11. Russo, G.S. & Bruce, C.J. Frontal eye field activity preceding aurally guided saccades. *J. Neurophysiol.* **71**, 1250–1253 (1994).
12. Fogassi, L. *et al.* Space coding by premotor cortex. *Exp. Brain Res.* **89**, 686–690 (1992).
13. Fogassi, L. *et al.* Coding of peripersonal space in inferior premotor cortex (area F4). *J. Neurophysiol.* **76**, 141–157 (1996).
14. Graziano, M.S.A., Yap, G.S. & Gross, C.G. Coding of visual space by premotor neurons. *Science* **266**, 1054–1057 (1994).
15. Stricanne, B., Andersen, R.A. & Mazzoni, P. Eye-centered, head-centered, and intermediate coding of remembered sound locations in area LIP. *J. Neurophysiol.* **76**, 2071–2076 (1996).
16. Duhamel, J-R., Bremmer, F., Ben Hamed, S. & Graf, W. Spatial invariance of visual receptive fields in parietal cortex neurons. *Nature* **389**, 845–848 (1997).
17. Colby, C.L., Duhamel, J.-R. & Goldberg, M.E. Ventral intraparietal area of the macaque: anatomic location and visual response properties. *J. Neurophysiol.* **69**, 902–914 (1993).
18. Schlack, A., Sterbing-D'Angelo, S.J., Hartung, K., Hoffmann, K.P. & Bremmer, F. Multisensory space representations in the macaque ventral intraparietal area. *J. Neurosci.* **25**, 4616–4625 (2005).
19. Bremmer, F., Klam, F., Duhamel, J.-R., Ben Hamed, S. & Graf, W. Visual-vestibular interactive responses in the macaque ventral intraparietal area (VIP). *Eur. J. Neurosci.* **16**, 1569–1586 (2002).
20. Schlack, A., Hoffmann, K-P. & Bremmer, F. Interaction of linear vestibular and visual stimulation in the macaque ventral intraparietal area (VIP). *Eur. J. Neurosci.* **16**, 1877–1886 (2002).
21. Neppi-Modona, M., Auclair, D., Sirigu, A. & Duhamel, J.-R. Spatial coding of the predicted impact location of a looming object. *Curr. Biol.* **14**, 1174–1180 (2004).
22. Deneve, S., Latham, P.E. & Pouget, A. Efficient computation and cue integration with noisy population codes. *Nat. Neurosci.* **4**, 826–831 (2001).
23. Cooke, D.F., Taylor, C.S.R., Moore, T. & Graziano, M.S.A. Complex movements evoked by microstimulation of the ventral intraparietal area. *Proc. Natl. Acad. Sci. USA* **100**, 6163–6168 (2003).
24. Pouget, A., Deneve, S. & Duhamel, J.R. A computational perspective on the neural basis of multisensory spatial representations. *Nat. Rev. Neurosci.* **3**, 741–747 (2002).
25. Squatrito, S. & Maioli, M. Gaze field properties of eye position neurons in areas MST and 7a of macaque monkey. *Vis. Neurosci.* **13**, 385–398 (1996).
26. Bremmer, F., Graf, W., Ben Hamed, S. & Duhamel, J.R. Eye position encoding in the macaque ventral intraparietal area (VIP). *Neuroreport* **10**, 873–878 (1999).
27. Rosenbluth, D. & Allman, J.M. The effect of gaze angle and fixation distance on the responses of neurons in V1, V2, and V4. *Neuron* **33**, 143–149 (2002).
28. Galletti, C. & Battaglini, P.P. Gaze-dependent visual neurons in area V3A of monkey prestriate cortex. *J. Neurosci.* **9**, 1112–1125 (1989).
29. Bremmer, F., Ilg, U.J., Thiele, A., Distler, C. & Hoffmann, K.-P. Eye position effects in monkey cortex. I. Visual and pursuit related activity in extrastiate areas MT and MST. *J. Neurophysiol.* **77**, 944–956 (1997).
30. Galletti, C., Battaglini, P.P. & Fattori, P. Eye position influence on the parieto-occipital area PO(V6) of the macaque monkey. *Eur. J. Neurosci.* **7**, 2486–2501 (1995).
31. Andersen, R.A. & Mountcastle, V.B. The influence of the angle of gaze upon the excitability of the light-sensitive neurons of the posterior parietal cortex. *J. Neurosci.* **3**, 532–548 (1983).
32. Andersen, R.A. Encoding of intention and spatial location in the posterior parietal cortex. *Cereb. Cortex* **5**, 457–469 (1995).
33. Groh, J.M., Trause, A.S., Underhill, A.M., Clark, K.R. & Inati, S. Eye position influences auditory responses in primate inferior colliculus. *Neuron* **29**, 509–518 (2001).
34. Boussaoud, D., Barth, T.M. & Wise, S.P. Effect of gaze on apparent visual responses of monkey frontal cortex neurons. *Exp. Brain Res.* **93**, 423–434 (1993).
35. Brotchie, P.R., Andersen, R.A., Snyder, L.H. & Goodman, S.J. Head position signals used by parietal neurons to encode locations of visual stimuli. *Nature* **375**, 232–235 (1995).
36. Salinas, E. & Thier, P. Gain modulation: a major computational principle of the central nervous system. *Neuron* **27**, 15–21 (2000).
37. Zipser, D. & Andersen, R.A. A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature* **331**, 679–684 (1988).
38. Pouget, A. & Sejnowski, T. Spatial transformations in the parietal cortex using basis functions. *J. Cogn. Neurosci.* **9**, 222–237 (1997).
39. Luppino, G., Murata, A., Govoni, P. & Matelli, M. Largely segregated parietofrontal connections linking rostral intraparietal cortex (areas AIP and VIP) and the ventral premotor cortex (areas F5 ans F4). *Exp. Brain Res.* **128**, 181–187 (1999).
40. Lewis, J.W. & Van Essen, D.C. Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. *J. Comp. Neurol.* **428**, 112–137 (2000).
41. Gentilucci, M., Scandolara, C., Pigarev, I.N. & Rizzolatti, G. Visual responses in the postarcuate cortex (area 6) of the monkey that are independent of eye position. *Exp. Brain Res.* **50**, 464–468 (1983).
42. Lynch, J.C., Graybiel, A.M. & Lobeck, L.J. The differential projection of two cytoarchitectonic subregions of the inferior parietal lobule of macaque upon the deep layers of the superior colliculus. *J. Comp. Neurol.* **235**, 241–254 (1985).
43. Baizer, J.S., Desimone, R. & Ungerleider, L.G. Comparison of subcortical connections of inferior temporal and posterior parietal cortex in monkeys. *Vis. Neurosci.* **10**, 59–72 (1993).
44. Gardner, E.P., Hamalainen, H.A., Warren, S., Davis, J. & Young, W. Somatosensory evoked potenials (SEPs) and cortical single unit responses elicited by mechanical tactile stimuli in awake monkeys. *Electroencephalogr. Clin. Neurophysiol.* **58**, 537–552 (1984).
45. Raiguel, S.E., Xiao, D.-K., Marcar, V.L. & Orban, G.A. Response latency of macaque area MT/V5 neurons and its relationship to stimulus parameters. *J. Neurophysiol.* **82**, 1944–1956 (1999).
46. Schmolesky, M.T. *et al.* Signal timing across the macaque visual system. *J. Neurophysiol.* **79**, 3272–3278 (1998).
47. Bullier, J. & Nowak, G. Parallel versus serial processing: new vistas on the distributed organization of the visual system. *Curr. Opin. Neurobiol.* **5**, 497–503 (1995).
48. Judge, S.J., Richmond, B.J. & Chu, F.C. Implantation of magnetic search coils for measurement of eye position: an improved method. *Vision Res.* **20**, 535–538 (1980).
49. Ben Hamed, S., Duhamel, J.-R., Bremmer, F. & Graf, W. Representation of the visual field in the lateral intraparietal area of macaque monkeys: a quantitative receptive field analysis. *Exp. Brain Res.* **140**, 132–144 (2001).