Neuron
Article

Multisensory Integration in Macaque Visual Cortex Depends on Cue Reliability

Michael L. Morgan,[1] Gregory C. DeAngelis,[2,3] and Dora E. Angelaki[1,3,*]
[1]Department of Anatomy and Neurobiology, Washington University School of Medicine, St. Louis, MO 63110, USA
[2]Department of Brain and Cognitive Sciences, Center for Visual Science, University of Rochester, Rochester, NY 14627, USA
[3]These authors contributed equally to this work
*Correspondence: angelaki@pcg.wustl.edu
DOI 10.1016/j.neuron.2008.06.024

SUMMARY

Responses of multisensory neurons to combinations of sensory cues are generally enhanced or depressed relative to single cues presented alone, but the rules that govern these interactions have remained unclear. We examined integration of visual and vestibular self-motion cues in macaque area MSTd in response to unimodal as well as congruent and conflicting bimodal stimuli in order to evaluate hypothetical combination rules employed by multisensory neurons. Bimodal responses were well fit by weighted linear sums of unimodal responses, with weights typically less than one (subadditive). Surprisingly, our results indicate that weights change with the relative reliabilities of the two cues: visual weights decrease and vestibular weights increase when visual stimuli are degraded. Moreover, both modulation depth and neuronal discrimination thresholds improve for matched bimodal compared to unimodal stimuli, which might allow for increased neural sensitivity during multisensory stimulation. These findings establish important new constraints for neural models of cue integration.

INTRODUCTION

Multisensory neurons are thought to underlie the performance improvements seen when subjects integrate multiple sensory cues to perform a task. In their groundbreaking research, Meredith and Stein (1983) found that neurons in the deep layers of the superior colliculus receive visual, auditory, and somatosensory information and typically respond more vigorously to multimodal than to unimodal stimuli (Meredith et al., 1987; Meredith and Stein, 1986a, 1986b, 1996; Wallace et al., 1996). Multimodal responses are often characterized as enhanced versus suppressed relative to the largest unimodal response, or as super- versus subadditive relative to the sum of unimodal responses (see Stein and Stanford, 2008, for review).

Because the superior colliculus is thought to play important roles in orienting to stimuli, the original investigations of Stein and colleagues emphasized near-threshold stimuli that are most relevant to detection. Many subsequent explorations, including human neuroimaging studies, have focused on superadditivity and the principle of inverse effectiveness (greater response enhancement for less effective stimuli) as hallmark properties of multisensory integration (e.g., Calvert et al., 2001; Meredith and Stein, 1986b; but see Beauchamp, 2005; Laurienti et al., 2005). However, use of near-threshold stimuli may bias outcomes toward a nonlinear (superadditive) operating range in which multisensory interactions are strongly influenced by a threshold nonlinearity (Holmes and Spence, 2005). Indeed, using a broader range of stimulus intensities, the emphasis on superadditivity in the superior colliculus has been questioned (Perrault et al., 2003; Stanford et al., 2005; Stanford and Stein, 2007), and studies using stronger stimuli in behaving animals frequently find subadditive effects (Frens and Van Opstal, 1998; Populin and Yin, 2002).

In recent years, these investigations have been extended to a variety of cortical areas and sensory systems, including auditory-visual integration (Barraclough et al., 2005; Bizley et al., 2007; Ghazanfar et al., 2005; Kayser et al., 2008; Romanski, 2007; Sugihara et al., 2006), visual-tactile integration (Avillac et al., 2007), and auditory-tactile integration (Lakatos et al., 2007). These studies generally find a mixture of superadditive and subadditive effects, with some studies reporting predominantly enhancement of multimodal responses (Ghazanfar et al., 2005; Lakatos et al., 2007) and others predominantly suppression (Avillac et al., 2007; Sugihara et al., 2006). The precise rules by which neurons combine sensory signals across modalities remain unclear.

Rather than exploring a range of stimuli spanning the selectivity of the neuron, most studies of multisensory integration have been limited to one or a few points within the stimulus space. This approach may be insufficient to mathematically characterize the neuronal combination rule. A multiplicative interaction, for instance, can appear to be subadditive ($2 \times 1 = 2$), additive ($2 \times 2 = 4$), or superadditive ($2 \times 3 = 6$), depending on the magnitudes of the inputs. Because input magnitudes vary with location on a tuning curve or within a receptive field, characterizing sub-/superadditivity at a single stimulus location may not reveal the overall combination rule. We suggest that probing responses to a broad range of stimuli that evoke widely varying responses is crucial for evaluating models of multisensory neural integration.

In contrast to the concept of superadditivity in neuronal responses, psychophysical and theoretical studies of multisensory integration have emphasized linearity. Humans often integrate cues perceptually by weighted linear combination, with weights

proportional to the relative reliabilities of the cues as predicted by Bayesian models (Alais and Burr, 2004; Battaglia et al., 2003; Ernst and Banks, 2002). Although linear combination at the level of perceptual estimates makes no clear prediction for the underlying neuronal combination rule, theorists have proposed that neurons could accomplish Bayesian integration via linear summation of unimodal inputs (Ma et al., 2006). A key question is whether the neural combination rule changes with the relative reliabilities of the cues. Neurons could accomplish optimal cue integration via linear summation with fixed weights that do not change with cue reliability (Ma et al., 2006). Alternatively, the combination rule may depend on cue reliability, such that neurons weight their unimodal inputs based on the strengths of the cues.

This study addresses two fundamental questions. First, what is the combination rule used by neurons to integrate sensory signals from two different sources? Second, how does this rule depend on the relative reliabilities of the sensory cues? We addressed these issues by examining visual-vestibular interactions in the dorsal portion of the medial superior temporal area (MSTd) (Duffy, 1998; Gu et al., 2006; Page and Duffy, 2003). To probe a broad range of stimulus space, we characterized responses to eight directions of translation in the horizontal plane using visual cues alone (optic flow), vestibular cues alone, and bimodal stimuli including all 64 (8 × 8) combinations of visual and vestibular headings, both congruent and conflicting. By modeling responses to this array of bimodal stimuli, we evaluated two models for the neural combination rule, one linear and one nonlinear (multiplicative). We also examined whether the combination rule depends on relative cue reliabilities by manipulating the motion coherence of the optic flow stimuli.

RESULTS

We recorded from 112 MSTd neurons (27 from monkey J and 85 from monkey P). We characterized their heading tuning in the horizontal plane by using a virtual-reality system to present eight evenly spaced directions, 45° apart. Responses were obtained during three conditions: inertial motion alone (vestibular condition), optic flow alone (visual condition), and paired inertial motion and optic flow (bimodal condition). For the latter, we tested all 64 combinations of vestibular and visual headings, including eight congruent and 56 incongruent (cue-conflict) presentations. Monkeys were simply required to maintain fixation on a head-fixed target during stimulus presentation.

Raw responses are shown in Figure 1 for an example MSTd neuron, along with the Gaussian velocity profile of the stimulus (gray curves). Peristimulus time histograms (PSTHs) show responses to each unimodal cue at both the preferred and antipreferred headings. PSTHs are also shown for the four bimodal conditions corresponding to all combinations of the preferred and antipreferred headings for the two cues. Note that the bimodal response is enhanced when both individual cues are at their preferred values and that the bimodal response is suppressed when either cue is antipreferred. Responses to each stimulus were quantified by taking the mean firing rate over the central 1 s of the 2 s stimulus period when the stimulus velocity varied the most (dashed vertical lines in Figure 1; see also Experimental Procedures). Using other 1 s intervals, except ones at the beginning or end of the trial, led to similar results (see also Gu et al., 2007).

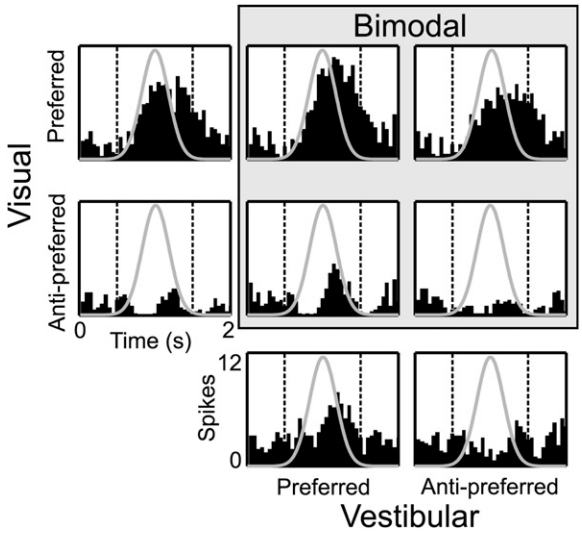

Figure 1. Peristimulus Time Histograms of Neural Responses for an Example MSTd Neuron

Gray curves indicate the Gaussian velocity profile of the stimulus. The two leftmost PSTHs show responses in the unimodal visual condition for preferred and antipreferred headings. The two bottom PSTHs represent preferred and antipreferred responses in the unimodal vestibular condition. PSTHs within the gray box show responses to bimodal conditions corresponding to the four combinations of the preferred and antipreferred headings for the two cues. Dashed vertical lines bound the central 1 s of the stimulus period, during which mean firing rates were computed.

Of the 112 cells recorded at 100% motion coherence, 44 (39%) had significant heading tuning in both unimodal conditions (one-way ANOVA, $p < 0.05$). Note that we use the term "unimodal" to refer to conditions in which visual and vestibular cues are presented in isolation. However, visual and vestibular selectivity can also be quantified by examining the visual and vestibular main effects in the responses to bimodal stimuli (by collapsing the bimodal responses along one axis or the other). When computed from bimodal responses, the percentage of vestibularly selective cells increased: 74 cells (66%) showed significant vestibular tuning in the bimodal condition (main effect of vestibular heading in two-way ANOVA, $p < 0.05$, see Figure S1 available online). Thus, the influence of the vestibular cue is sometimes more apparent when presented in combination with the visual cue, as reported in other multisensory studies (e.g., Avillac et al., 2007).

Data from two representative MSTd neurons are illustrated in Figures 2A and 2B. For each unimodal stimulus (visual and vestibular), tuning curves were constructed by plotting the mean response versus heading direction (Figure 2, marginal tuning curves). Both neurons had visual and vestibular heading preferences that differed by nearly 180° (Figure 2A: vestibular, 10°; visual, 185°; Figure 2B: vestibular, 65°; visual, 226°). Thus, both cells were classified as "opposite" (Gu et al., 2006). Note that heading directions in all conditions are referenced to physical body motion. For example, a bimodal stimulus in which both

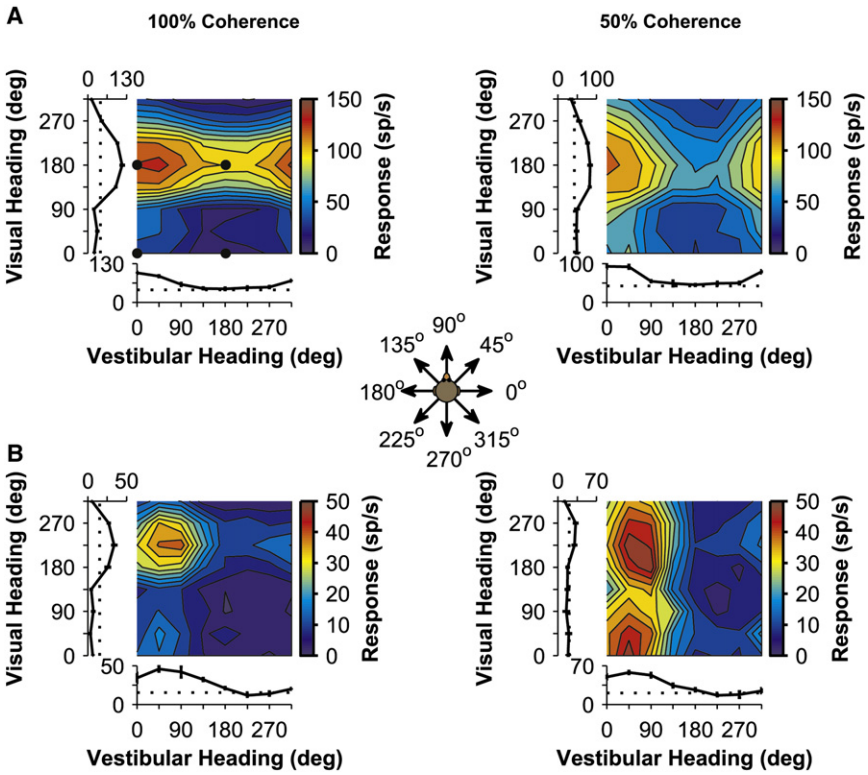

Figure 2. Examples of Tuning for Two "Opposite" MSTd Neurons

Color contour maps show mean firing rates as a function of vestibular and visual headings in the bimodal condition. Tuning curves along the left and bottom margins show mean (±SEM) firing rates versus heading for the unimodal conditions. Data collected using optic flow with 100% and 50% motion coherence are shown in the left and right columns, respectively.

(A) Data from a neuron with opposite vestibular and visual heading preferences in the unimodal conditions (same cell as in Figure 1). Black dots indicate the bimodal response conditions shown in Figure 1. Bimodal tuning shifts from visually dominated at 100% coherence to balanced at 50% coherence.

(B) Data from another "opposite" neuron. Bimodal responses reflect an even balance of visual and vestibular tuning at 100% coherence and become vestibularly dominated at 50% coherence. (Inset) A top-down view showing the eight possible heading directions (for each cue) in the horizontal plane.

visual and vestibular cues indicate rightward (0°) body motion will contain optic flow in which dots move leftward on the display screen. Distributions of the differences in direction preference between visual and vestibular conditions are shown in Figure S1 for all neurons. Neurons with mismatched preferences for visual and vestibular cues have also been seen in area VIP (Bremmer et al., 2002; Schlack et al., 2002).

For the bimodal stimuli, where each response is associated with both a vestibular heading and a visual heading, responses are shown as color contour maps with vestibular heading along the abscissa and visual heading along the ordinate (Figure 2; black dots in Figure 2A indicate the bimodal conditions corresponding to the PSTHs in Figure 1). At 100% motion coherence, bimodal responses typically reflect both unimodal tuning preferences to some degree. For the cell in Figure 2A, bimodal responses were dominated by the visual stimulus, as indicated by the horizontal band of high firing rates. In contrast, bimodal responses of the cell in Figure 2B were equally affected by visual and vestibular cues, creating a circumscribed peak centered near the unimodal heading preferences (54°, 230°).

Reducing the motion coherence of optic flow (see Experimental Procedures) altered both the unimodal visual responses and the pattern of bimodal responses for these example cells. In both cases, the visual heading tuning (tuning curve along ordinate) remained similar in shape and heading preference, but the peak-to-trough response modulation was reduced at 50% coherence. For the cell of Figure 2A, the horizontal band of high firing rate seen in the bimodal response at 100% coherence is replaced by a more discrete single peak centered around a vestibular heading of 8° and a visual heading of 180°, reflecting a more even mix of the two modalities at 50% coherence. For the cell in Figure 2B, the well-defined peak seen at 100% coherence becomes a vertical band of strong responses, reflecting a stronger vestibular influence at 50% coherence. For both cells, the vestibular contribution to the bimodal response was more pronounced when the reliability (coherence) of the visual cue was reduced.

Similar results were seen for MSTd cells with "congruent" visual and vestibular heading preferences. Figure 3 shows responses from a third example cell (with vestibular and visual heading preferences of −25° and −21°, respectively) at three coherences, 100%, 50%, and 25%. Whereas vestibular tuning remains quite constant (tuning curves along abscissa), visual responsiveness declines with coherence (tuning curves along ordinate), such that little visual heading tuning remains at 25% coherence. Bimodal responses were visually dominated at 100% coherence (horizontal band in Figure 3A) but became progressively more influenced by the vestibular cue as coherence was reduced. At 50% coherence, the presence of a clear symmetric peak suggests well-matched visual and vestibular contributions to the bimodal response (Figure 3B). As coherence was further reduced to 25%, vestibular dominance is observed, with the bimodal response taking the form of a vertical band aligned with the vestibular heading preference (Figure 3C). Data from five additional example neurons, tested at both 100% and 50% coherence, are shown in Figure S2.

In the following analyses, we quantify the response interactions illustrated by these example neurons. First, we evaluate whether weighted linear summation of unimodal responses can account for bimodal tuning. Second, we explore how the relative contributions of visual and vestibular inputs to the bimodal response change with coherence. Third, we investigate whether neuronal discrimination of heading improves when stimuli are aligned with visual and vestibular preferences of each neuron.

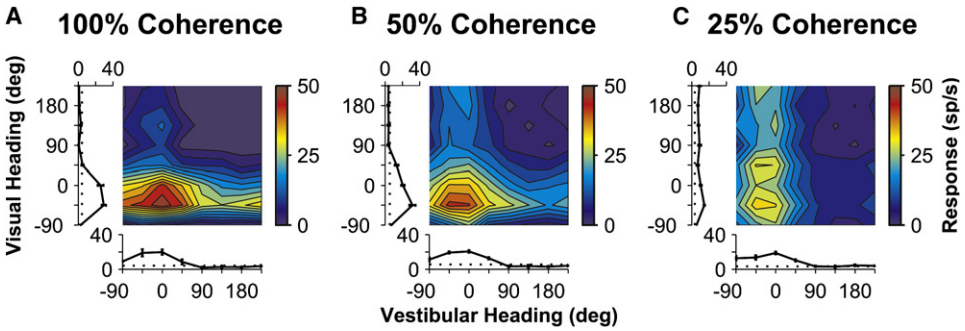

Figure 3. Data for a "Congruent" MSTd Cell, Tested at Three Motion Coherences

(A) Bimodal responses at 100% coherence are visually dominated.

(B) Bimodal responses at 50% coherence show a balanced contribution of visual and vestibular cues.

(C) At 25% coherence, bimodal responses appear to be dominated by the vestibular input.

Linearity of Cue Interactions in Bimodal Response Tuning

Approximately half of the MSTd neurons with significant visual and vestibular unimodal tuning (22 of 44 at 100% coherence, 8 of 14 at 50% coherence) had a significant interaction effect (two-way ANOVA) in the bimodal condition. This suggests that responses of many cells are well described by a linear model, whereas other neurons may require a nonlinear component. To explore this further, we compared the goodness of fit for simple linear and nonlinear interaction models (see Experimental Procedures). In the linear model, responses in the bimodal condition were fit with a weighted sum of responses from the vestibular and visual conditions. The nonlinear model included an additional term consisting of the product of the vestibular and visual responses (see Experimental Procedures). Both models provided good fits to the data, as illustrated in Figures 4A and 4B for the example congruent neuron of Figure 3A. Although the nonlinear model provided a significantly improved fit when adjusted for the additional fitting parameter (sequential F test, $p = 0.00013$), the improvement in variance accounted for (VAF) was quite modest (94.5% versus 95.7%), and the patterns of residual errors were comparable for the two fits (Figures 4C and 4D).

Across the population of MSTd neurons, the linear model often resulted in nearly as good a fit as the nonlinear model. For data collected at 100% coherence, the nonlinear model provided a significantly better fit for 16 out of 44 (36%) neurons (sequential F test, $p < 0.05$; Figure 4E, filled circles). At 50% coherence, this was true for 21% of the neurons (Figure 4E, filled triangles). With the exception of a few cells, however, the improvement in VAF due to the nonlinear term was quite modest. The median VAF for the linear and nonlinear fits were 89.1% versus 90.1% (100% coherence) and 89.2% versus 89.5% (50% coherence). Thus, linear combinations of the unimodal responses generally provide good descriptions of bimodal tuning, with little explanatory power gained by including a multiplicative component.

Subadditive Rather than Superadditive Interactions in Bimodal Responses

The weights from the best-fitting linear or nonlinear combination rule (w_{visual} and $w_{vestibular}$, Equations 1 and 2) describe the strength of the contributions of each unimodal input to the bimodal response. Visual and vestibular weights from the linear and nonlinear models were statistically indistinguishable (Wilcoxon signed rank test: $p = 0.407$ for vestibular weights, $p = 0.168$ for visual weights), so we further analyzed the weights from the linear model. The majority of MSTd cells combined cues subadditively, with visual and vestibular weights being typically less than 1 (Figures 5A and 5B). We computed 95% confidence intervals for the visual and vestibular weights and found that, at 100% coherence, none of the vestibular weights was significantly larger than 1, whereas the majority (41 of 44) of cells had vestibular weights that were significantly smaller than 1. Similarly, only three cells had visual weights significantly larger than 1, whereas 27 of 44 cells had weights significantly less than 1. We also performed a regression analysis in which the measured bimodal response was fit with a scaled sum of the visual and vestibular responses (after subtracting spontaneous activity). The regression coefficient was significantly lower than 1 for 37 of 44 cells, whereas none had a coefficient significantly larger than unity. For data obtained at 50% coherence, all 14 cells had regression coefficients significantly smaller than unity. Thus, MSTd neurons most frequently exhibited subadditive integration with occasional additivity and negligible superadditivity.

Dependence of Visual and Vestibular Weights on Relative Cue Reliabilities

We investigated how visual and vestibular weights change when the relative reliabilities of the two cues are altered by reducing motion coherence (see Experimental Procedures). It is clear from Figures 2 and 3 that the relative influences of the two cues on bimodal responses change with motion coherence. This effect could arise simply from the fact that lower coherences elicit visual responses with weaker modulation as a function of heading. Thus, one possibility is that the weights with which each neuron combines its vestibular and visual inputs remain constant and that the decreased visual influence in bimodal tuning is simply due to weaker visual responses at lower coherences. In this scenario, each neuron has a combination rule that is independent of cue reliability. Alternatively, the weights given to the vestibular and visual inputs could change with the relative reliabilities of the two cues. This outcome would indicate that the neuronal combination rule is not fixed but may change with cue

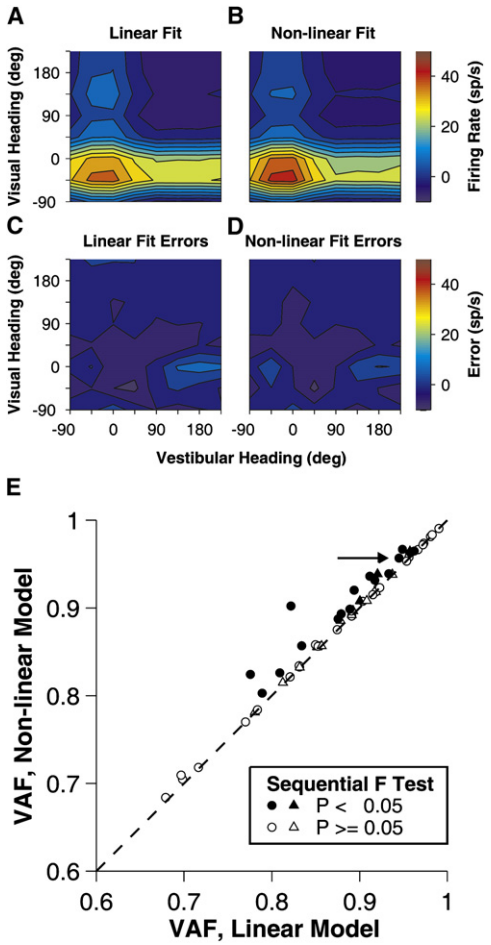

Figure 4. Fitting of Linear and Nonlinear Models to Bimodal Responses

(A–D) Model fits and errors for the same neuron as in Figure 3A. Color contour maps show fits to the bimodal responses using (A) a weighted sum of the unimodal responses and (B) a weighted sum of the unimodal responses plus their product.

(C and D) Errors of the linear and nonlinear fits, respectively.

(E) Variance accounted for (VAF) by the nonlinear fits is plotted against VAF from the linear fits. Data measured at 100% coherence are shown as circles; 50% coherence as triangles. Filled symbols represent neurons (16 of 44 neurons at 100% coherence and 3 of 14 at 50% coherence) whose responses were fit significantly better by the nonlinear model (sequential F test, $p < 0.05$).

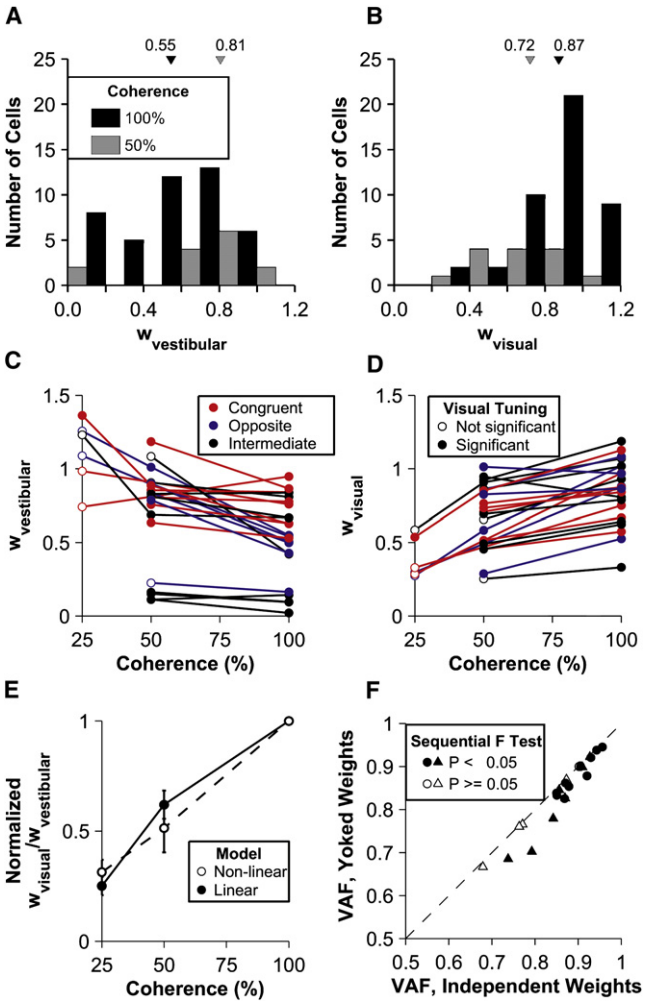

Figure 5. Dependence of Vestibular and Visual Response Weights on Motion Coherence

(A and B) Histograms of vestibular and visual weights (linear model) computed from data at 100% (black) and 50% (gray) coherence. Triangles are plotted at the medians.

(C and D) Vestibular and visual weights (linear model) are plotted as a function of motion coherence. Data points are coded by the significance of unimodal visual tuning (open versus filled circles) and by the congruency between vestibular and visual heading preferences (colors).

(E) The ratio of visual to vestibular weights (±SEM) is plotted as a function of coherence. For each cell, this ratio was normalized to unity at 100% coherence. Filled symbols and solid line: weights computed from linear model fits. Open symbols and dashed line: weights computed from nonlinear model fits.

(F) Comparison of variance accounted for (VAF) between linear models with yoked weights and independent weights. Filled data points (17 of 23 neurons) were fit significantly better by the independent weights model (sequential F test, $p < 0.05$). Of 23 total neurons, 12 with unimodal tuning that remained significant (ANOVA $p < 0.05$) at lower coherences are plotted as circles; the remainder are shown as triangles.

reliability. We therefore examined how vestibular and visual weights change as a function of motion coherence.

We used weights from the linear model fits to quantify the relative strengths of the vestibular and visual influences in the bimodal responses. Figures 5A and 5B summarize the vestibular and visual weights, respectively, for two coherence levels, 100% and 50% (black and gray filled bars; $n = 44$ and $n = 14$, respectively). As compared to 100% coherence, vestibular weights at 50% coherence are shifted toward larger values (median of 0.81 versus 0.55, one-tailed Kolmogorov-Smirnov test, $p < 0.001$), and visual weights at 50% coherence are shifted toward smaller values (median of 0.72 versus 0.87, $p = 0.037$). Thus, across the population, the influence of visual cues on bimodal responses decreased as visual reliability was reduced while simultaneously the influence of vestibular cues increased.

For neurons recorded at multiple coherences, we were able to examine how the vestibular and visual weights changed for each cell. Among 44 neurons with significant tuning in both unimodal

conditions at 100% coherence, responses were recorded at 50% and 100% coherences for 17 neurons, at 25% and 100% coherences for three neurons, and at all three coherences for three neurons. All 23 cells showed significant visual tuning at all coherences in the bimodal condition (main effect of visual cue, two-way ANOVA, $p < 0.05$). Vestibular and visual weights are plotted as a function of motion coherence for these 23 cells in Figures 5C and 5D, respectively (filled symbols indicate significant unimodal visual tuning, ANOVA, $p < 0.05$). Both weights depended significantly on coherence (ANCOVA, $p < 0.005$), but not on visual-vestibular congruency ($p > 0.05$, Figures 5C and 5D). Vestibular weights declined with increasing motion coherence whereas visual weights increased. In contrast, the constant offset term of the model (Equation 1) did not depend on coherence (ANCOVA, $p = 0.566$).

These changes in weights are further quantified in Figure 5E by computing the ratio of the weights, $w_{visual}/w_{vestibular}$, and normalizing this ratio to be 1 at 100% coherence for each neuron. The normalized weight ratio declined significantly as coherence was reduced (ANCOVA, $p << 0.001$), dropping to a mean value of 0.62 at 50% coherence and to 0.25 at 25% coherence (filled symbols, Figure 5E). Weights from the nonlinear model fits showed a very similar effect (open symbols, Figure 5E). These results demonstrate that single neurons apply different weights to their visual and vestibular inputs when the relative reliabilities of the two cues change. In other words, the neuronal combination rule is not fixed.

Although weights vary significantly with coherence across the population, one might question whether a single set of weights adequately describes the bimodal responses at all coherences for individual neurons. To address this possibility, we fit the data with an alternative model in which the visual and vestibular weights were common (i.e., yoked) across coherences. Figure 5F shows the VAF for yoked weights versus the VAF for independent weights. Data from 17 of 23 cells were fit significantly better by allowing separate weights for each coherence (sequential F test, $p < 0.05$), further demonstrating that weights of individual neurons change with cue reliability.

Because we fit the bimodal responses with a weighted sum of measured unimodal responses, one might question whether noise in the unimodal data biases the outcome. To address this possibility, we first fit the unimodal responses with a family of wrapped Gaussian functions (see Experimental Procedures and Figure S3). These fits were done simultaneously for all coherence levels, allowing the amplitude of the Gaussian to vary with coherence while the location of the peak remained fixed (see Figure S3 for details). This approach generally fit the unimodal data quite well and allowed us to use these fitted functions (rather than raw data) to model the measured bimodal responses. Results from this analysis, as summarized in Figure S4, are quite similar to those of Figure 5. The change in normalized weight ratio as a function of coherence was again highly significant (ANCOVA $p << 0.001$). Thus, analyzing our data parametrically to reduce the effect of measurement noise did not change the results appreciably.

Modulation Depth of Bimodal Tuning: Comparison with Unimodal Responses

The analyses detailed above describe how visual and vestibular cues are weighted by neurons during cue combination, but do not address how this interaction affects bimodal tuning. Does simultaneous presentation of vestibular and visual cues improve bimodal tuning compared to tuning for unimodal cues? It is logical to hypothesize that peak-to-trough modulation for bimodal stimuli should depend on both (1) the disparity in the alignment of the visual and vestibular heading stimuli and (2) each cell's visual-vestibular congruency (Gu et al., 2006; Takahashi et al., 2007). The present experiments, which include many cue-conflict stimuli, allow for an investigation of how bimodal stimuli alter the selectivity of responses in MSTd.

Consider two particular diagonals from the two-dimensional bimodal stimulus array: one corresponding to stimuli with aligned visual and vestibular headings and the other corresponding to stimuli with antialigned headings (Figure 6A). We computed tuning curves for trials when vestibular and visual stimuli were aligned (Figure 6A, magenta line) and tuning curves for trials when the visual and vestibular stimuli were antialigned (i.e., 180° opposite; Figure 6A, orange lines). We then examined how the modulation depths (maximum-minimum responses) for these bimodal response cross-sections (Figure 6B) differed from the modulation seen in the unimodal tuning curve with the strongest modulation. Based on linear weighted summation, we expect that aligned bimodal stimuli should enhance modulation for congruent neurons and reduce it for opposite neurons. In contrast, when antialigned vestibular and visual heading stimuli are paired,

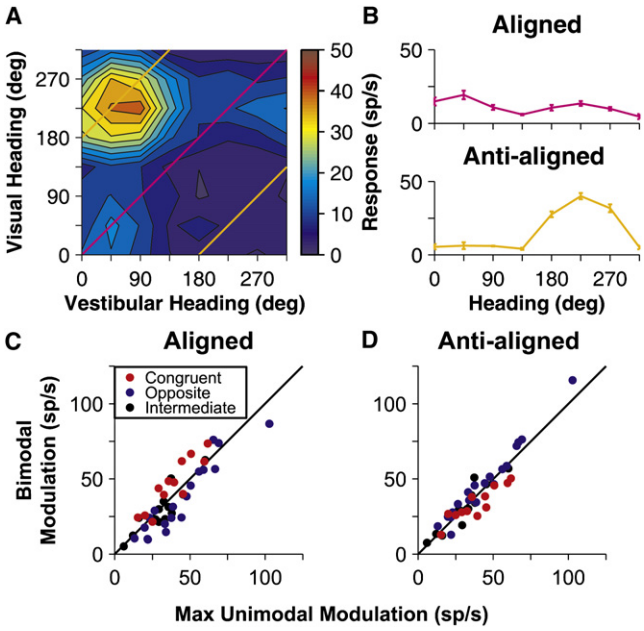

Figure 6. Modulation Depth and Visual-Vestibular Congruency

(A) Illustration of "aligned" (magenta) and "antialigned" (orange) cross-sections through the bimodal response array.

(B) Mean (±SEM) firing rates for aligned and antialigned bimodal stimuli (extracted from [A]). The antialigned tuning curve is plotted as a function of the visual heading.

(C and D) Modulation depth (maximum-minimum response) for aligned and antialigned bimodal stimuli is plotted against the largest unimodal response modulation. Color indicates visual-vestibular congruency (red, congruent cells; blue, opposite cells; black, intermediate cells).

the reverse would be true: modulation should be reduced for congruent cells and enhanced for opposite cells.

Results are summarized for aligned and antialigned stimulus pairings in Figures 6C and 6D, respectively (data shown for 100% coherence). As expected, the relationship between modulation depths for bimodal versus maximal unimodal responses depended on visual-vestibular congruency. For congruent cells, modulation depth increased for aligned bimodal stimuli (Wilcoxon signed rank test, $p = 0.005$) and decreased for antialigned stimuli ($p = 0.034$; Figures 6C and 6D, red symbols). The reverse was true for opposite cells: modulation depth decreased for aligned stimuli and increased for antialigned stimuli (aligned $p = 0.005$, antialigned $p = 0.016$; Figures 6C and 6D, blue symbols).

Comparison of Direction Discriminability for Bimodal versus Unimodal Responses

To further examine sensitivity in the bimodal versus unimodal conditions, we computed a measure of the precision with which each neuron discriminates small changes in heading direction for both the bimodal and unimodal conditions. Based on Figure 6, the largest improvement in bimodal response modulation for each cell is expected to occur when visual and vestibular headings are paired such that their angular alignment corresponds to the difference between the heading preferences in the unimodal conditions. Greater response modulation should, in turn, lead to enhanced discriminability due to steepening of the slope of the tuning curve. To test this, we computed a "matched" tuning curve by selecting the elements of the array of bimodal responses that most closely matched the difference in heading preference between the unimodal conditions. This curve corresponds to a diagonal cross-section through the peak of the bimodal response profile, and it allows us to examine discriminability around the optimal bimodal stimulus for each MSTd cell. In Figure 6A, the matched tuning curve happens to be identical to the antialigned cross-section, since the latter passes through the peak of the bimodal response profile.

Theoretical studies (Pouget et al., 1999; Seung and Sompolinsky, 1993) have shown that, for any unbiased estimator operating on the responses of a population of neurons, the discriminability (d') of two closely spaced stimuli has an upper bound that is proportional to the square root of Fisher information (see Experimental Procedures). Thus, we computed minimum discrimination thresholds derived from Fisher information for both unimodal and matched bimodal stimuli. Because we sparsely sampled heading directions (every 45°), it was necessary to interpolate the heading tuning curves by fitting them with a modified Gaussian function. We fit the vestibular, visual, and "matched" curves parametrically as shown in Figure 7A (same cell as Figure 2B). VAF from the fits across the population are shown in Figure 7B. In general, the fits were good, with median VAF values >0.9 in all three stimulus conditions. Figures 7C and 7D plot the population tuning curves for 44 MSTd neurons at 100% coherence and 14 MSTd neurons at 50% coherence. Each tuning curve was shifted to have a peak at 0° prior to averaging. Matched tuning curves tend to show greater modulation than both vestibular (100% coherence, $p << 0.001$; 50% coherence, $p = 0.0012$) and visual tuning curves (100% coherence, $p << 0.001$; 50% coherence, $p = 0.0012$). Based on these modulation differences, one may expect lower discrimination thresholds for the matched tuning curves compared to the unimodal curves.

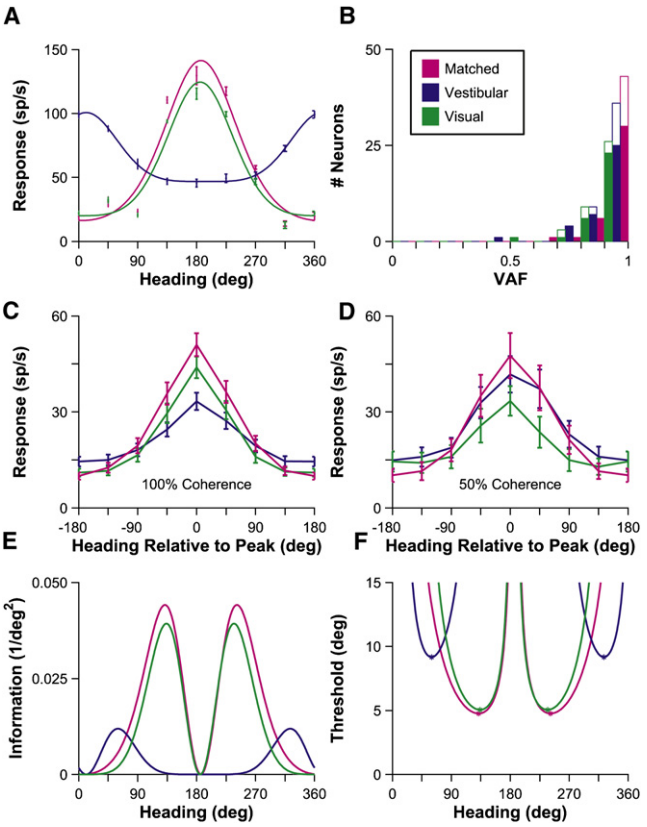

Figure 7. Fisher Information and Heading Discriminability

(A) Example wrapped Gaussian fits to vestibular, visual, and "matched" tuning curves for the neuron shown in Figure 2B. Error bars show ±SEM.

(B) Population histogram of VAF for parametric fits to vestibular (blue), visual (green), and matched (magenta) tuning curves. Filled bars denote fits to 100% coherence data; open bars to 50% coherence data.

(C and D) Population vestibular (blue), visual (green), and matched (magenta) tuning curves for 44 cells tested at 100% coherence (C) and 14 cells tested at 50% coherence (D). Individual curves were shifted to align the peaks at 0° before averaging.

(E) Fisher information (see Experimental Procedures) is plotted as a function of heading for the example neuron from (A).

(F) Discrimination threshold (derived from Fisher information) is plotted against heading for the same example neuron. The two threshold minima for each curve are shown as asterisks.

We used the parametric fits to calculate Fisher information (I_F) as illustrated in Figure 7E. We calculated the slope of the tuning curve from the modified Gaussian fit and the response variance from a linear fit to the log-log plot of variance versus mean firing rate. Assuming the criterion d' = 1, one can use Fisher information to calculate a discrimination threshold as

$$\Delta\theta = \frac{1}{\sqrt{I_F(\theta)}} \quad (7)$$

We thereby identified the lowest discrimination thresholds for the vestibular, visual, and matched tuning curves (asterisks in Figure 7F; see Experimental Procedures for details).

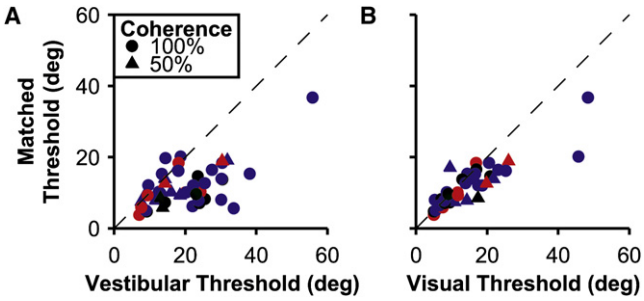

Figure 8. Heading Discrimination Thresholds Derived from Bimodal (Matched) and Unimodal Tuning Functions

Circles and triangles represent data collected at 100% and 50% coherence, respectively. Color indicates visual-vestibular congruency (red, congruent cells; blue, opposite cells; black, intermediate cells).

(A) Comparison of matched thresholds with unimodal vestibular thresholds.

(B) Comparison of matched thresholds with unimodal visual thresholds.

Figures 8A and 8B show minimum thresholds derived from the matched tuning curves plotted against minimum thresholds for the unimodal tuning curves. In both comparisons, points tend to fall below the diagonal, indicating that cue combination improves discriminability (Wilcoxon signed rank test, $p < 0.001$). Thus, when slicing along a diagonal of the bimodal response array that optimally aligns the vestibular and visual preferences, improvements in threshold are common and do not depend on visual-vestibular congruency. This is consistent with the finding that the combination rule used by MSTd neurons does not depend on congruency (Figures 5C and 5D). Thus, all MSTd neurons are potentially capable of exhibiting improved discriminability under bimodal stimulation, though congruent cells may still play a privileged role under most natural conditions in which visual and vestibular heading cues are aligned.

DISCUSSION

We have characterized the combination rule used by neurons in macaque visual cortex to integrate visual (optic flow) and vestibular signals and have examined how this rule depends on relative cue reliabilities. We found that a weighted linear model provides a good description of the bimodal responses with subadditive weighting of visual and vestibular inputs being typical. When the strength (coherence) of the visual cue was reduced, we observed systematic changes in neural weighting of the two inputs: visual weights decreased and vestibular weights increased as coherence declined. These findings establish a combination rule that can account for multisensory integration by neurons, and they provide important constraints for models of optimal (e.g., Bayesian) cue integration.

Linear Combination Rule for Bimodal Responses

Firing rates of bimodal MSTd cells were described well by weighted linear sums of the unimodal vestibular and visual responses. Addition of a nonlinear (multiplicative) term significantly improved fitting of bimodal responses for about one-third of MSTd neurons, but these improvements were very modest (difference in median VAF less than 1%). Our findings are consistent with recent theoretical studies which posit that multisensory neurons combine their inputs linearly to accomplish optimal cue integration (Ma et al., 2006). In the Ma et al. study, neurons were assumed to perform a straight arithmetic sum of their unimodal inputs, but the theory is also compatible with the possibility that neurons perform a weighted linear summation that is subadditive (A. Pouget, personal communication). In psychophysical studies, humans often combine cues in a manner consistent with weighted linear summation of unimodal estimates where the weights vary with the relative reliabilities of the cues (e.g., Alais and Burr, 2004; Battaglia et al., 2003; Ernst and Banks, 2002). Although linear combination at the level of perceptual estimates does not necessarily imply any particular neural combination rule, our findings suggest that the neural mechanisms underlying optimal cue integration may depend on weighted linear summation of responses at the single neuron level.

In their pioneering studies of the superior colliculus, Stein and colleagues emphasized superadditivity as a signature of multisensory integration (Meredith and Stein, 1983, 1986b, 1996; Wallace et al., 1996). While our findings lie in clear contrast to these studies, there are important differences that must be considered, in addition to the fact that we recorded in a different brain area. First, the appearance of the neural combination rule may depend considerably on stimulus strength. The largest superadditive effects seen in the superior colliculus were observed when unimodal stimuli were near threshold for eliciting a response (Meredith and Stein, 1986b). Recent studies have shown that interactions become more additive as stimulus strength increases (Perrault et al., 2003, 2005; Stanford et al., 2005). Note, however, that the difference in average firing rate between the "low" and "high" intensity stimuli used by Stanford et al. (2005) was less than two-fold and that responses were generally much weaker than those elicited by our stimuli. Linear summation at the level of membrane potentials (Skaliora et al., 2004) followed by a static nonlinearity (e.g., threshold) in spike generation will produce superadditive firing rates for weak stimuli (Holmes and Spence, 2005). Thus, the predominant subadditivity seen in our study may reflect the fact that we typically operate in a stimulus regime well above response threshold. Our finding of predominant subadditivity is consistent with results of other recent cortical studies that have also used suprathreshold stimuli (Avillac et al., 2007; Bizley et al., 2007; Kayser et al., 2008; Sugihara et al., 2006), as well as other cortical studies that show that superadditive interactions tend to become additive or subadditive for stronger stimuli (Ghazanfar et al., 2005; Lakatos et al., 2007).

Second, whereas most previous studies have examined multisensory integration at one or a few points within the stimulus space (e.g., visual and auditory stimuli at a single spatial location), our experimental protocol explored a broad range of stimuli, and our analysis used the responses to all stimulus combinations to mathematically characterize the neuronal combination rule. If nonlinearities such as response threshold or saturation play substantial roles, then the apparent sub-/superadditivity of responses can change markedly, depending on where stimuli are placed within the receptive field or along a tuning curve. Thus, we have examined bimodal responses in MSTd using all

possible combinations of visual and vestibular headings that span the full tuning of the neurons in the horizontal plane. This method allows us to model the combination of unimodal responses across a wide range of stimuli, including both congruent and conflicting combinations with varying efficacy. This approach, which avoids large effects of response threshold yet spans the stimulus tuning of the neurons, provides a more comprehensive means of evaluating the combination rule used by multisensory neurons.

Dependence of Weights on Cue Reliability

Human psychophysical studies show that a less-reliable cue is given less weight in perceptual estimates of multimodal stimuli (Alais and Burr, 2004; Battaglia et al., 2003; Ernst and Banks, 2002). Our findings suggest that an analogous computation may occur at the single-neuron level, since MSTd neurons give less weight to visual inputs when optic flow is degraded. It must be noted, however, that such reweighting of unimodal inputs by single neurons is not necessarily required to account for the behavioral observations. Rather, the behavioral dependence on cue reliability could be mediated by multisensory neurons that maintain constant weights on their unimodal inputs. A recent theoretical study shows that a population of multisensory neurons with Poisson-like firing statistics and fixed weights can accomplish Bayes-optimal cue integration (Ma et al., 2006). In this scheme, changes in cue reliability are reflected in the bimodal population response because of the lower responses elicited by a weaker cue, but the neural combination rule does not change.

Our findings appear contrary to the assumption of fixed weights in the theory of Ma et al. (2006). When the reliability of the optic flow cue was reduced by noise, its influence on the bimodal response diminished while the influence of the vestibular cue increased. This discrepancy may arise because our neurons exhibit firing rate changes with coherence that violate the assumptions of the theory of Ma et al. Their framework assumes that stimulus strength (coherence) multiplicatively scales all of the responses of sensory neurons. In contrast, we find that responses to nonpreferred headings often decrease at high coherence (Figure S3B). When the theory of Ma et al. (2006) takes this fact into account, it may predict weight changes with coherence similar to those that we have observed (A. Pouget, personal communication).

Our finding of weights that depend on coherence cannot be explained by a static nonlinearity in the relationship between membrane potential and firing rate, since this nonlinearity is usually expansive (e.g., Priebe et al., 2004). Such a mechanism predicts that weak unimodal visual responses, such as at low coherence, would be enhanced in the bimodal response (Holmes and Spence, 2005). In contrast, we have observed the opposite effect, where the influence of weak unimodal responses on the bimodal response is less than one would expect based on a combination rule with fixed weights. The mechanism by which this occurs is unclear, but it might reflect computations (i.e., normalization) taking place at the network level.

We cannot speak to the temporal dynamics of this reweighting, because we presented different motion coherences in separate blocks of trials. Further experiments are necessary to investigate whether neurons reweight their inputs on a trial-by-trial basis. Another important caveat is that cognitive and motivational demands placed on alert animals may affect the neuronal combination rule and the effects of variations in cue reliability. For example, the proportion of neurons showing multisensory integration in the primate superior colliculus can depend on behavioral context. Compared to anesthetized animals (Wallace et al., 1996), multisensory interactions in alert animals may be more frequent when stimuli are behaviorally relevant (Frens and Van Opstal, 1998), but are somewhat suppressed during passive fixation (Bell et al., 2003). Thus, the effects of cue reliability on weights in MSTd could be different under circumstances in which the animal is required to perceptually integrate the cues (as in Gu et al., 2008), whereas animals in this study simply maintained visual fixation.

An additional caveat is that our monkeys' eyes and heads were constrained to remain still during stimulus presentation. We do not know whether the integrative properties of MSTd neurons would be different under more natural conditions in which the eyes or head are moving, which substantially complicates optic flow on the retina. In a recent study (Gu et al., 2007), we measured responses of MSTd neurons during free viewing in darkness and found little effect of eye movements on the vestibular responses of MSTd neurons.

Modulation Depth and Discrimination Threshold

A consistent observation in human psychophysical experiments is that subjects make more precise judgments under bimodal as compared to unimodal conditions (Alais and Burr, 2004; Ernst and Banks, 2002). We see a potential neural correlate in our data, but with an important qualification. Simultaneous presentation of vestibular and visual cues can enhance the modulation depth and direction discrimination thresholds of MSTd neurons, but the finding depends on the alignment of bimodal stimuli relative to the cell's unimodal heading preferences. When the disparity between vestibular and visual stimuli matches the relative alignment of a neuron's heading preferences, modulation depth increases and the minimum discrimination threshold decreases. For congruent cells, these improvements occur when vestibular and visual headings are aligned, as typically occurs during self-motion in everyday life. For opposite cells, discriminability is enhanced under conditions where vestibular and visual cues are misaligned, which can occur during simultaneous self-motion and object-motion. Future research needs to examine whether congruent and opposite cells play distinct roles in self-motion versus object-motion perception. It also remains to be demonstrated that neuronal sensitivity improves in parallel with behavioral sensitivity when trained animals perform multimodal discrimination tasks. Results from our laboratory suggest that this is the case (Gu et al., 2008). In conclusion, our findings establish two aspects of multisensory integration. We demonstrate that weighted linear summation is an adequate combination rule to describe visual-vestibular integration by MSTd neurons, and we establish that the weights in the combination rule can vary with cue reliability. These findings should help to constrain and further define neural models for optimal cue integration.

EXPERIMENTAL PROCEDURES

Subjects and Surgery

Two male rhesus monkeys (*Macaca mulatta*) served as subjects. General procedures have been described previously (Gu et al., 2006). Each animal was outfitted with a circular molded plastic ring anchored to the skull with titanium T-bolts and dental acrylic. For monitoring eye movements, each monkey was implanted with a scleral search coil. The Institutional Animal Care and Use Committee at Washington University approved all animal surgeries and experimental procedures, which were performed in accordance with National Institutes of Health guidelines. Animals were trained to fixate on a central target for fluid rewards using operant conditioning.

Vestibular and Visual Stimuli

A 6 degree-of-freedom motion platform (MOOG 6DOF2000E; Moog, East Aurora, NY) was used to passively translate the animals along one of eight directions in the horizontal plane (Figure 2, inset), spaced 45° apart. Visual stimuli were projected onto a tangent screen, which was affixed to the front surface of the field coil frame, by a three-chip digital light projector (Mirage 2000; Christie Digital Systems, Cypress, CA). The screen measured 60 × 60 cm and was mounted 30 cm in front of the monkey, thus subtending ~90° × 90°. Visual stimuli simulated translational movement along the same eight directions through a three-dimensional field of stars. Each star was a triangle that measured 0.15 cm × 0.15 cm, and the cloud measured 100 cm wide by 100 cm tall by 40 cm deep at a star density of 0.01 per cm^3. To provide stereoscopic cues, the dot cloud was rendered as a red-green anaglyph and viewed through custom red-green goggles. The optic flow field contained naturalistic cues mimicking translation of the observer in the horizontal plane, including motion parallax, size variations, and binocular disparity.

Electrophysiological Recordings

We recorded action potentials extracellularly from two hemispheres in two monkeys. In each recording session, a tungsten microelectrode was passed through a transdural guide tube and advanced using a micromanipulator. An amplifier, eight-pole band-pass filter (400–5000 Hz), and dual voltage-time window discriminator (BAK Electronics, Mount Airy, MD) were used to isolate action potentials from single neurons. Action potential times and behavioral events were recorded with 1 ms accuracy by a computer. Eye coil signals were low-pass filtered and sampled at 250 Hz.

Magnetic resonance image (MRI) scans and Caret software analyses, along with physiological criteria, were used to guide electrode penetrations to area MSTd (Gu et al., 2006). Neurons were isolated while presenting a large field of flickering dots. In some experiments, we further advanced the electrode tip into the lower bank of the superior temporal sulcus to verify the presence of neurons with middle temporal (MT) area response characteristics (Gu et al., 2006). Receptive field locations changed as expected across guide tube locations based on the known topography of MT (Albright and Desimone, 1987; Desimone and Ungerleider, 1986; Maunsell and Van Essen, 1987; Van Essen et al., 1981).

Experimental Protocol

We measured neural responses to eight heading directions evenly spaced every 45° in the horizontal plane. Neurons were tested under three experimental conditions. (1) In vestibular trials, the monkey was required to maintain fixation on a central dot on an otherwise blank screen while being translated along one of the eight directions. (2) In visual trials, the monkey saw optic flow simulating self-motion (same eight directions) while the platform remained stationary. (3) In bimodal trials, the monkey experienced both translational motion and optic flow. We paired all eight vestibular headings with all eight visual headings for a total of 64 bimodal stimuli. Eight of these 64 combinations were congruent, meaning that visual and vestibular cues simulated the same heading. The remaining 56 cases were cue-conflict stimuli. This relative proportion of congruent and cue-conflict stimuli was adopted purely for the purpose of characterizing the neuronal combination rule and was not intended to reflect ecological validity. Each translation followed a Gaussian velocity profile. It had a duration of 2 s, an amplitude of 13 cm, a peak velocity of 30 cm/s, and a peak acceleration of ~0.1 × g (981 cm/s^2).

These three stimulus conditions were interleaved randomly along with blank trials with neither translation nor optic flow. Ideally, five repetitions of each unique stimulus were collected for a total of 405 trials. Experiments with fewer than three repetitions were excluded from analysis. When isolation remained satisfactory, we ran additional blocks of trials with the coherence of the visual stimulus reduced to 50% and/or 25%. Motion coherence was lowered by randomly relocating a percentage of the dots on every subsequent video frame. For example, we randomly selected one quarter of the dots in every frame at 25% coherence and updated their positions to new positions consistent with the simulated motion while the other three-quarters of the dots were plotted at new random locations within the 3D cloud. Each block of trials consisted of both unimodal and bimodal stimuli at the corresponding coherence level. When a cell was tested at multiple coherences, both the unimodal vestibular tuning and the unimodal visual tuning were independently assessed in each block.

Trials were initiated by displaying a 0.2° × 0.2° fixation target on the screen. The monkey was required to fixate for 200 ms before the stimulus was presented and to maintain fixation within a 3° × 3° window for a liquid reward. Trials in which the monkey broke fixation were aborted and discarded.

Data Analysis

Using Matlab (Mathworks, Natick, MA), we first computed the mean firing rate during the middle 1 s of each 2 s trial (Gu et al., 2006). Subsequently, responses were averaged across stimulus repetitions to compute mean firing rates. One-way ANOVA was used to assess the significance of tuning in the (unimodal) vestibular and visual conditions. For cells with significant tuning, vestibular and visual heading preferences were calculated using the vector sum of mean responses. We classified each cell as congruent, intermediate, or opposite based on the difference between its vestibular and visual heading preferences. Cells having preferences aligned within 60° were classified as congruent, and cells whose alignments differed by more than 120° were classified as opposite, with intermediate cells falling between these two conditions (Fetsch et al., 2007).

Modeling of Bimodal Responses

For the bimodal condition, we arranged the responses into two-dimensional arrays indexed by the vestibular and visual headings (e.g., color contour maps in Figures 2 and 3). Two-way ANOVA was used to compute the significance of vestibular and visual tuning (main effects) in the bimodal responses, as well as their interaction. A significant interaction effect indicates nonlinearities in the bimodal responses.

To further explore the linearity of visual-vestibular interactions, we fit the data using linear and nonlinear models. For these fits, responses in the three conditions were defined as the mean responses minus the average spontaneous activity measured in the blank trials. For the linear model, bimodal responses were fit by a linear combination of the corresponding vestibular and visual responses.

$$r_{bimodal}(\theta,\varphi) = w_{vestibular} r_{vestibular}(\theta) + w_{visual} r_{visual}(\varphi) + C. \quad (1)$$

In this equation, $r_{bimodal}$ is the predicted response for the bimodal condition, and $r_{vestibular}$ and r_{visual} are the responses in the vestibular and visual unimodal conditions, respectively. Angles θ and φ represent vestibular and visual stimulus directions. Model weights $w_{vestibular}$ and w_{visual} and the constant C were chosen to minimize the sum of squared errors between predicted and measured bimodal responses. In addition, responses were also fit by the following equation that includes a multiplicative nonlinearity:

$$r_{bimodal}(\theta,\varphi) = w_{vestibular} r_{vestibular}(\theta) + w_{visual} r_{visual}(\varphi) + w_{product} r_{vestibular}(\theta) \times r_{visual}(\varphi) + C, \quad (2)$$

where $w_{product}$ is the weight on the multiplicative interaction term.

For each fit, the VAF was computed as

$$VAF = 1 - \frac{SSE}{SST} \quad (3)$$

where SSE is the sum of squared errors between the fit and the data, and SST is the sum of squared differences between the data and the mean of the data.

As the number of free parameters was different between the linear and nonlinear models, the statistical significance of the nonlinear fit over the linear fit was assessed using a sequential F test. A significant outcome of the sequential F test ($p < 0.05$) indicates that the nonlinear model fits the data significantly better than the linear model.

Visual and Vestibular Weights and Cue Reliability

In our main analysis, weights $w_{vestibular}$ and w_{visual} were computed separately for each motion coherence ("independent weights" model). We also examined whether a set of fixed weights for each cell is sufficient to explain the data at all coherences. For the latter pair of models (both linear and nonlinear variants), weights $w_{vestibular}$ and w_{visual} were common across coherences ("yoked weights" model). Note that for both the independent weights and yoked weights models, parameter *C* was allowed to vary with coherence. Thus, to fit the data across three coherences, the linear model would have nine free parameters when weights are independent and five free parameters when weights are yoked. VAF was used to quantify whether fits were better using the model with independent weights versus the model with yoked weights. In addition, the sequential F test was used to assess whether allowing model weights to vary with coherence provides significantly improved fits.

Unimodal versus Bimodal Response Tuning and Discriminability

To examine how cue combination alters bimodal responses, we computed tuning curves for two specific cross-sections though the array of bimodal responses. Using the main diagonal of the bimodal response array (Figure 6A, magenta), we computed a tuning curve for aligned (i.e., congruent) vestibular and visual heading stimuli. We followed a similar procedure to obtain tuning curves for antialigned vestibular and visual headings (i.e., 180° opposite; Figure 6A, orange). For each of these tuning curves, we computed modulation depth as the maximum mean response minus the minimum mean response. We then compared the best unimodal modulation depth to the modulation depth in the aligned and antialigned cross-sections for congruent, intermediate, and opposite cells.

We also derived a bimodal tuning curve along a single diagonal that was optimized for each cell to yield near-maximal bimodal responses. The difference in heading preference between the visual and vestibular conditions, which is used to define the congruency of each cell, also specifies how disparate the visual and vestibular heading stimuli should be to produce maximum response modulation. From the array of bimodal responses, we constructed a "matched" tuning curve by selecting the diagonal that matched most closely this difference in unimodal heading preferences. For vestibular and visual heading preferences within 22.5° of each other, the main diagonal would be selected. For vestibular and visual heading preferences that differ by 22.5° to 67.5°, the matched tuning curve would be derived from the diagonal along which the vestibular and visual headings differ by 45°. Selected this way, the matched tuning curve is a diagonal cross-section through the peak of the bimodal response profile. After shifting each of the vestibular, visual, and matched tuning curves to have a peak at 0°, we averaged across the population of neurons recorded at each motion coherence to construct population tuning curves.

We then used Fisher information to quantify the maximum discriminability that could be achieved at any point along the matched tuning curve and the unimodal curves. Because heading tuning was sampled coarsely, we interpolated the data to high spatial resolution by fitting the curves with a modified wrapped Gaussian function (Fetsch et al., 2007),

$$R(\theta) = A_1 \cdot \left[e^{\frac{-2\times(1-\cos(\theta-\theta_0))}{(\sigma\cdot\kappa)^2}} + A_2 \cdot e^{\frac{-2\times(1-\cos(\theta-\theta_0-\pi))}{\sigma^2}} \right] + R_0, \quad (4)$$

where θ_0 is the angular location of the peak response, σ is the tuning width, A_1 is the amplitude, and R_0 is the baseline response. The second term with amplitude A_2 is necessary to fit the tuning of a few MSTd cells that show a second response peak 180° out of phase with the first peak (with κ determining the relative widths of the two peaks; see Fetsch et al., 2007, for details). Goodness-of-fit was quantified using the VAF. Only fits having a full-width larger than 45° were used further to avoid situations in which the width of the peak (and hence the slope) was not well constrained by the data.

From these fitted tuning curves, we computed Fisher information using the derivative of the fits, R', and the variance of the responses, σ^2.

$$I_F(\theta) = \frac{R'(\theta)^2}{\sigma(\theta)^2} \quad (5)$$

The variance at each point along the fitted tuning curve was estimated from a linear fit to a log-log plot of response variance versus mean response for each cell. From the Fisher information, we computed an upper bound on discriminability (Nover et al., 2005).

$$d'(\theta) = \Delta\theta\sqrt{I_F(\theta)} \quad (6)$$

For the criterion d' = 1, the threshold for discrimination is

$$\Delta\theta = \frac{1}{\sqrt{I_F(\theta)}} \quad (7)$$

We found the minimum discrimination threshold for each of the visual, vestibular and matched curves. For our population of cells, we plotted the minimum thresholds from the matched tuning curves against the minimum thresholds from each of the two unimodal tuning curves.

SUPPLEMENTAL DATA

The Supplemental Data include figures and can be found with this article online at http://www.neuron.org/cgi/content/full/59/4/662/DC1/.

ACKNOWLEDGMENTS

We thank Amanda Turner and Erin White for excellent monkey care and training. We thank Alexandre Pouget for helpful comments on the manuscript. This work was supported by NIH EY017866 and DC04260 (to D.E.A.) and NIH EY016178 (to G.C.D.).

Accepted: June 22, 2008
Published: August 27, 2008

REFERENCES

Alais, D., and Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. Curr. Biol. *14*, 257–262.

Albright, T.D., and Desimone, R. (1987). Local precision of visuotopic organization in the middle temporal area (MT) of the macaque. Exp. Brain Res. *65*, 582–592.

Avillac, M., Ben Hamed, S., and Duhamel, J.R. (2007). Multisensory integration in the ventral intraparietal area of the macaque monkey. J. Neurosci. *27*, 1922–1932.

Barraclough, N.E., Xiao, D., Baker, C.I., Oram, M.W., and Perrett, D.I. (2005). Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. J. Cogn. Neurosci. *17*, 377–391.

Battaglia, P.W., Jacobs, R.A., and Aslin, R.N. (2003). Bayesian integration of visual and auditory signals for spatial localization. J. Opt. Soc. Am. A Opt. Image Sci. Vis. *20*, 1391–1397.

Beauchamp, M.S. (2005). Statistical criteria in FMRI studies of multisensory integration. Neuroinformatics *3*, 93–113.

Bell, A.H., Corneil, B.D., Munoz, D.P., and Meredith, M.A. (2003). Engagement of visual fixation suppresses sensory responsiveness and multisensory integration in the primate superior colliculus. Eur. J. Neurosci. *18*, 2867–2873.

Bizley, J.K., Nodal, F.R., Bajo, V.M., Nelken, I., and King, A.J. (2007). Physiological and anatomical evidence for multisensory interactions in auditory cortex. Cereb. Cortex *17*, 2172–2189.

Bremmer, F., Klam, F., Duhamel, J.R., Ben Hamed, S., and Graf, W. (2002). Visual-vestibular interactive responses in the macaque ventral intraparietal area (VIP). Eur. J. Neurosci. *16*, 1569–1586.

Calvert, G.A., Hansen, P.C., Iversen, S.D., and Brammer, M.J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. Neuroimage *14*, 427–438.

Desimone, R., and Ungerleider, L.G. (1986). Multiple visual areas in the caudal superior temporal sulcus of the macaque. J. Comp. Neurol. *248*, 164–189.

Duffy, C.J. (1998). MST neurons respond to optic flow and translational movement. J. Neurophysiol. *80*, 1816–1827.

Ernst, M.O., and Banks, M.S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. Nature *415*, 429–433.

Fetsch, C.R., Wang, S., Gu, Y., Deangelis, G.C., and Angelaki, D.E. (2007). Spatial reference frames of visual, vestibular, and multimodal heading signals in the dorsal subdivision of the medial superior temporal area. J. Neurosci. *27*, 700–712.

Frens, M.A., and Van Opstal, A.J. (1998). Visual-auditory interactions modulate saccade-related activity in monkey superior colliculus. Brain Res. Bull. *46*, 211–224.

Ghazanfar, A.A., Maier, J.X., Hoffman, K.L., and Logothetis, N.K. (2005). Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. J. Neurosci. *25*, 5004–5012.

Gu, Y., Watkins, P.V., Angelaki, D.E., and DeAngelis, G.C. (2006). Visual and nonvisual contributions to three-dimensional heading selectivity in the medial superior temporal area. J. Neurosci. *26*, 73–85.

Gu, Y., DeAngelis, G.C., and Angelaki, D.E. (2007). A functional link between area MSTd and heading perception based on vestibular signals. Nat. Neurosci. *10*, 1038–1047.

Gu, Y., Angelaki, D.E., and DeAngelis, G.C. (2008). Neural correlates of multisensory cue integration in macaque area MSTd. Nat. Neurosci., in press.

Holmes, N.P., and Spence, C. (2005). Multisensory integration: space, time and superadditivity. Curr. Biol. *15*, R762–R764.

Kayser, C., Petkov, C.I., and Logothetis, N.K. (2008). Visual modulation of neurons in auditory cortex. Cereb. Cortex *18*, 1560–1574.

Lakatos, P., Chen, C.M., O'Connell, M.N., Mills, A., and Schroeder, C.E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. Neuron *53*, 279–292.

Laurienti, P.J., Perrault, T.J., Stanford, T.R., Wallace, M.T., and Stein, B.E. (2005). On the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies. Exp. Brain Res. *166*, 289–297.

Ma, W.J., Beck, J.M., Latham, P.E., and Pouget, A. (2006). Bayesian inference with probabilistic population codes. Nat. Neurosci. *9*, 1432–1438.

Maunsell, J.H., and Van Essen, D.C. (1987). Topographic organization of the middle temporal visual area in the macaque monkey: representational biases and the relationship to callosal connections and myeloarchitectonic boundaries. J. Comp. Neurol. *266*, 535–555.

Meredith, M.A., and Stein, B.E. (1983). Interactions among converging sensory inputs in the superior colliculus. Science *221*, 389–391.

Meredith, M.A., and Stein, B.E. (1986a). Spatial factors determine the activity of multisensory neurons in cat superior colliculus. Brain Res. *365*, 350–354.

Meredith, M.A., and Stein, B.E. (1986b). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. J. Neurophysiol. *56*, 640–662.

Meredith, M.A., and Stein, B.E. (1996). Spatial determinants of multisensory integration in cat superior colliculus neurons. J. Neurophysiol. *75*, 1843–1857.

Meredith, M.A., Nemitz, J.W., and Stein, B.E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. J. Neurosci. *7*, 3215–3229.

Nover, H., Anderson, C.H., and DeAngelis, G.C. (2005). A logarithmic, scale-invariant representation of speed in macaque middle temporal area accounts for speed discrimination performance. J. Neurosci. *25*, 10049–10060.

Page, W.K., and Duffy, C.J. (2003). Heading representation in MST: sensory interactions and population encoding. J. Neurophysiol. *89*, 1994–2013.

Perrault, T.J., Jr., Vaughan, J.W., Stein, B.E., and Wallace, M.T. (2003). Neuron-specific response characteristics predict the magnitude of multisensory integration. J. Neurophysiol. *90*, 4022–4026.

Perrault, T.J., Jr., Vaughan, J.W., Stein, B.E., and Wallace, M.T. (2005). Superior colliculus neurons use distinct operational modes in the integration of multisensory stimuli. J. Neurophysiol. *93*, 2575–2586.

Populin, L.C., and Yin, T.C. (2002). Bimodal interactions in the superior colliculus of the behaving cat. J. Neurosci. *22*, 2826–2834.

Pouget, A., Deneve, S., Ducom, J.C., and Latham, P.E. (1999). Narrow versus wide tuning curves: What's best for a population code? Neural Comput. *11*, 85–90.

Priebe, N.J., Mechler, F., Carandini, M., and Ferster, D. (2004). The contribution of spike threshold to the dichotomy of cortical simple and complex cells. Nat. Neurosci. *7*, 1113–1122.

Romanski, L.M. (2007). Representation and integration of auditory and visual stimuli in the primate ventral lateral prefrontal cortex. Cereb. Cortex *17* (*Suppl 1*), i61–i69.

Schlack, A., Hoffmann, K.P., and Bremmer, F. (2002). Interaction of linear vestibular and visual stimulation in the macaque ventral intraparietal area (VIP). Eur. J. Neurosci. *16*, 1877–1886.

Seung, H.S., and Sompolinsky, H. (1993). Simple models for reading neuronal population codes. Proc. Natl. Acad. Sci. USA *90*, 10749–10753.

Skaliora, I., Doubell, T.P., Holmes, N.P., Nodal, F.R., and King, A.J. (2004). Functional topography of converging visual and auditory inputs to neurons in the rat superior colliculus. J. Neurophysiol. *92*, 2933–2946.

Stanford, T.R., and Stein, B.E. (2007). Superadditivity in multisensory integration: putting the computation in context. Neuroreport *18*, 787–792.

Stanford, T.R., Quessy, S., and Stein, B.E. (2005). Evaluating the operations underlying multisensory integration in the cat superior colliculus. J. Neurosci. *25*, 6499–6508.

Stein, B.E., and Stanford, T.R. (2008). Multisensory integration: current issues from the perspective of the single neuron. Nat. Rev. Neurosci. *9*, 255–266.

Sugihara, T., Diltz, M.D., Averbeck, B.B., and Romanski, L.M. (2006). Integration of auditory and visual communication information in the primate ventrolateral prefrontal cortex. J. Neurosci. *26*, 11138–11147.

Takahashi, K., Gu, Y., May, P.J., Newlands, S.D., DeAngelis, G.C., and Angelaki, D.E. (2007). Multimodal coding of three-dimensional rotation and translation in area MSTd: comparison of visual and vestibular selectivity. J. Neurosci. *27*, 9742–9756.

Van Essen, D.C., Maunsell, J.H., and Bixby, J.L. (1981). The middle temporal visual area in the macaque: myeloarchitecture, connections, functional properties and topographic organization. J. Comp. Neurol. *199*, 293–326.

Wallace, M.T., Wilkinson, L.K., and Stein, B.E. (1996). Representation and integration of multiple sensory inputs in primate superior colliculus. J. Neurophysiol. *76*, 1246–1266.