

Responses of Prefrontal Multisensory Neurons to Mismatching Faces and Vocalizations

Maria M. Diehl¹ and Lizabeth M. Romanski²

¹Department of Psychiatry, University of Puerto Rico School of Medicine, San Juan, Puerto Rico 00936 and ²Department of Neurobiology & Anatomy, University of Rochester School of Medicine and Dentistry, Rochester, New York 14642

Social communication relies on the integration of auditory and visual information, which are present in faces and vocalizations. Evidence suggests that the integration of information from multiple sources enhances perception compared with the processing of a unimodal stimulus. Our previous studies demonstrated that single neurons in the ventrolateral prefrontal cortex (VLPFC) of the rhesus monkey (*Macaca mulatta*) respond to and integrate conspecific vocalizations and their accompanying facial gestures. We were therefore interested in how VLPFC neurons respond differentially to matching (congruent) and mismatching (incongruent) faces and vocalizations. We recorded VLPFC neurons during the presentation of movies with congruent or incongruent species-specific facial gestures and vocalizations as well as their unimodal components. Recordings showed that while many VLPFC units are multisensory and respond to faces, vocalizations, or their combination, a subset of neurons showed a significant change in neuronal activity in response to incongruent versus congruent vocalization movies. Among these neurons, we typically observed incongruent suppression during the early stimulus period and incongruent enhancement during the late stimulus period. Incongruent-responsive VLPFC neurons were both bimodal and nonlinear multisensory, fostering their ability to respond to changes in either modality of a face-vocalization stimulus. These results demonstrate that ventral prefrontal neurons respond to changes in either modality of an audiovisual stimulus, which is important in identity processing and for the integration of multisensory communication information.

Key words: communication; faces; multisensory integration; prefrontal cortex; primate; vocalizations

Introduction

The integration of auditory cues from the voice and visual cues from the face enhance our ability to communicate effectively and enable us to make appropriate behavioral responses. Congruent multisensory information enhances stimulus perception (Meredith and Stein, 1983; Laurienti et al., 2004; Shams and Seitz, 2008), whereas mismatching or incongruent multisensory information can alter our perception, as exemplified by the McGurk effect (McGurk and MacDonald, 1976). In this paradigm, the simultaneous presentation of an incongruent auditory phoneme /ba/ with a visual phoneme /ga/ results in the perception of a different phoneme /da/. This audiovisual (AV) illusion demonstrates the strength and certainty of multisensory integration during speech processing.

Neuroimaging studies have examined brain regions activated by congruent and incongruent AV stimuli to determine the network

involved in multisensory integration. Many studies have used AV speech stimuli and dynamic movies (Calvert, 2001; Homae et al., 2002; Jones and Callan, 2003; Miller and D'Esposito, 2005; Ojanen et al., 2005; van Atteveldt et al., 2007), whereas other studies have used nonspeech AV stimuli to examine the effect of stimulus congruency on AV integration (Hein et al., 2007; Naumer et al., 2009). Across these studies, ventral prefrontal regions were activated during the processing of incongruent stimuli.

Electrophysiological recordings have examined multisensory integration of congruent stimuli in the superior colliculus, temporal lobe, and prefrontal cortex (for review, see Stein and Stanford, 2008), where both enhanced and suppressed multisensory responses were observed. Early studies in the cat superior colliculus revealed that both temporal and spatial factors dramatically influenced neuronal response magnitude when combining visual, auditory, and somatosensory cues (Meredith and Stein, 1986; Meredith et al., 1987). Single-unit recordings in nonhuman primates have demonstrated multisensory responses in the superior temporal sulcus (STS; Barraclough et al., 2005) and auditory cortex (Ghazanfar et al., 2005).

Work in our laboratory and others have shown that ventrolateral prefrontal cortex (VLPFC) neurons respond to complex visual and auditory stimuli including faces and vocalizations (Wilson et al., 1993; O'Scalaidhe et al., 1997; Romanski and Goldman-Rakic, 2002; Romanski et al., 2005; Tsao et al., 2008) and integrate these stimuli (Sugihara et al., 2006). VLPFC neurons exhibit multisensory enhancement or suppression to con-

Received Dec. 10, 2013; revised June 10, 2014; accepted July 6, 2014.

Author contributions: M.M.D. and L.M.R. designed research; M.M.D. performed research; M.M.D. and L.M.R. analyzed data; M.M.D. and L.M.R. wrote the paper.

This work was supported by the National Institutes of Health DC004845 (L.M.R.), Center for Navigation and Communication Sciences (P30) DC 005409, and Center for Visual Sciences (P30) EY001319. We thank Mark Diltz, Jaewon Hwang, John Housel, and Christopher Louie for technical assistance and Bethany Plakke for critical comments.

The authors declare no competing financial interests.

Correspondence should be addressed to Lizabeth M. Romanski, PhD, Associate Professor, Department of Neurobiology and Anatomy, University of Rochester School of Medicine, Rochester, NY 14642. E-mail: Liz_romanski@urmc.rochester.edu.

DOI:10.1523/JNEUROSCI.5168-13.2014

Copyright © 2014 the authors 0270-6474/14/3411233-11\$15.00/0

gruent AV stimuli when compared with the best unimodal response (Sugihara et al., 2006). We have recently shown that VLPFC neurons alter their responses to temporally asynchronous face-vocalization stimuli in which 59% of cells responded with decreased firing to asynchronous compared with synchronous stimuli (Romanski and Hwang, 2012). In the current study, we asked whether prefrontal neurons detect semantically incongruent face-vocalization stimuli by exhibiting differences in neuronal activity between congruent and incongruent AV stimuli. We therefore mismatched the video and audio tracks of species-specific vocalizations to create semantically incongruent stimuli that were not temporally asynchronous and compared single-unit VLPFC responses across these conditions. Our results indicate that a small population of VLPFC multisensory neurons alters its responses to incongruent face-vocalization movies, confirming the importance of VLPFC in face-voice integration.

Materials and Methods

Subjects and surgical procedures. Extracellular recordings were performed in two rhesus monkeys (*Macaca mulatta*): one female (5.0 kg) and one male (10.5 kg). All procedures were in accordance with the NIH Guidelines for the Care and Use of Laboratory Animals and the University of Rochester Care and Use of Animals in Research committee. Before recordings, a titanium head post was surgically implanted to allow for head restraint, and a titanium recording cylinder was placed over the VLPFC as defined anatomically by Preuss and Goldman-Rakic (1991) and physiologically by Romanski and Goldman-Rakic (2002). A single tungsten electrode was lowered into the VLPFC, and neurons were recorded while the monkey performed behavioral tasks. We recorded each stable unit that we encountered while lowering the electrode into the brain. Neurons were recorded from the left hemisphere of both animals and the right hemisphere of the male subject.

Apparatus and stimuli. Recordings were performed in a sound-attenuated room lined with Sonex (Acoustical Solutions). Auditory stimuli (65–75 dB SPL) were presented via a pair of Audix PH5-vs speakers (frequency response ± 3 dB, 75–20,000 Hz) located on either side of a video monitor positioned 29 inches from the monkey's head and centered at the level of the monkey's eyes. Stimuli were created from vocalization movies of animals in our home colony and from a library of monkey vocalization movies obtained by L.M.R. A large variety of movies allowed for the use of both familiar and unfamiliar callers as well as ensuring the use of multiple exemplars of affiliative and agonistic vocalizations across stimulus sets. Movie clips were edited using Adobe Premiere (Adobe Systems), Jasc Animation Studio (Corel), and several custom and shareware programs. Auditory and visual components of the movie clips were separated into wav and mpeg tracks for processing. The visual track was edited to remove extraneous and distracting elements from the viewing frame. The audio track was also filtered to eliminate background noise if present using MATLAB (MathWorks) and SIGNAL (Engineering Design). The audio and visual tracks were recombined for presentation during recordings. Movies subtended 8–10° of visual angle. Eight stimuli were presented within each testing block: two auditory vocalizations (A1 and A2), two silent movies (V1 and V2), and four combined AV stimuli. For the AV stimuli, two were congruent (A1V1 and A2V2), and two were incongruent (A2V1 and A1V2; see Fig. 1). Congruent vocalization movies consisted of naturally occurring AV tracks of a rhesus macaque vocalizing and differed in acoustic morphology, semantic meaning, valence, and in some testing blocks, caller identity. Each block included one coo and one agonistic vocalization. Coo vocalizations are positive or neutral in valence and are elicited during friendly approach, individual out of group contact, finding low-value foods, or grooming. Agonistic vocalizations (barks, growls, pant threats, and screams) are negative in valence and are elicited during threatening or adverse interactions. AV vocalizations were semantically distinct and have been previously characterized as being polar opposites in emotional valence (Gouzoules et al., 1984; Hauser and Marler, 1993). The use of AV

vocalization movies that differed greatly in acoustic features, semantic meaning, and identity, as well as emotional valence increased the possibility of capturing a larger number of responsive cells given the selectivity of VLPFC neurons (O'Scalaidhe et al., 1997; Romanski et al., 2005).

Incongruent vocalization movies were created by combining the auditory track of one movie with the video track of another movie. To create incongruent stimuli that were properly aligned in time, specific exemplars were chosen that had a similar onset and duration. We then aligned these mismatching vocalizations with the sound envelope of the original vocalization. Sometimes it was necessary to add silence to the beginning of the mismatching vocalization to match it to the timing of the original vocalization. In this manner, the stimuli were incongruent but not asynchronous. The average duration of each movie was 819.2 ms.

Task and experimental procedures. Monkeys were seated in front of the monitor and speakers. A white fixation point (0.15 × 0.15 degrees) was presented in the center of the monitor. Monkeys were required to fixate the point for 500 ms after which one of the eight stimuli appeared while the monkey maintained fixation within a 10 × 10 degree window containing the stimulus. Each testing block included two congruent vocalization movies, their separate auditory and visual components, and two incongruent movie clips (Fig. 1). Each stimulus was randomly presented 8–12 times, yielding a total of 64–96 trials per testing block. During trials where the auditory vocalization was presented alone, the white fixation point remained on the screen. A drop of juice was delivered 250 ms after the end of the stimulus period. Loss of fixation within the 10 × 10 degree window aborted the trial, and these trials were not included in the analysis. Single- and multi-unit activity was recorded and discriminated on-line using Plexon. Behavioral task and stimuli were presented using CORTEX (NIMH). Neuronal data were analyzed off-line using MATLAB and SPSS (IBM).

The location of the recording cylinder in the male subjects' right prefrontal cortex is shown in Figure 6. We recorded VLPFC neurons in the male's left ($N = 199$) and right ($N = 158$) hemispheres and in the left hemisphere of the female subject ($N = 240$).

Data analysis. Cells in the ventral prefrontal cortex were recorded during the movie presentation task and responses were analyzed off-line. Cells with firing rates lower than 0.5 Hz (assessed in all task epochs and visually inspected) or cells in which recordings were incomplete were removed from the analysis. We analyzed the response to the auditory, visual, and combined AV stimuli using a two-way MANOVA (with factors Auditory and Visual) across two time bins of neuronal activity (an early period of 0–400 ms and a late period of 401–800 ms). We selected these time bins based on the known latency of VLPFC neurons (Romanski and Hwang, 2012), the average duration of each stimulus, and as a means to compare early versus late neuronal responses to the stimuli. Analyses with different bin widths showed that 2 × 400 ms bins allowed us to detect a larger number of multisensory responsive cells than other bin widths.

We performed a two-way MANOVA on each cell for both movie stimuli tested. Spontaneous activity was measured during a 500 ms period, which preceded the onset of the fixation stimulus and converted to a spike rate. The two-way MANOVA model, which assessed the responses of neurons to the auditory vocalization (A), the visual movie (V), or combined audiovisual movie (AV), is given by the following:

$$r = \mu + \alpha_i + \beta_j + \delta_{i,j} + \sigma,$$

where r is the response of the neuron in an individual trial; α_i and β_j refer to the main effects of A and V, respectively; μ is the intercept; and σ is a Gaussian random variable. $\delta_{i,j}$ is the interaction term (A*V), which tests the null hypothesis that the response in the multisensory condition (AV) is the sum of the responses to the corresponding unimodal stimuli (A and V).

With this analysis, we characterized task-responsive cells as follows: unimodal auditory if they had a significant main effect of auditory stimuli (A), but not visual stimuli (V); unimodal visual if they had a significant main effect of V, but not A; linear multisensory, if they had a significant main effect of A and a main effect of V; nonlinear multisensory if they had a significant interaction effect (A*V). Cells that had a

main effect of A and V and no interaction effect were considered linear multisensory since the multisensory response could be explained as a linear sum of the two unimodal responses. It is important to note that the interaction term A^*V does not test for a significant response in the multisensory condition with respect to baseline, but rather tests the null hypothesis that the response in the multisensory condition is equal to the linear sum of the responses in each of the unimodal conditions. Linear multisensory neurons that had a main effect of both A and V are referred to as bimodal multisensory neurons (Sugihara et al., 2006).

For nonlinear multisensory neurons, means were assessed for multisensory suppression or enhancement with the calculation of a modified multisensory index (MMI), which quantifies the response magnitude during the AV stimulus presentation in relation to the response to the best unimodal stimulus (Sugihara et al., 2006; adapted from Meredith and Stein, 1986). The MMI is defined as follows:

$$MMI = \frac{[r_{AV} - \max(r_A, r_V)]}{[r_{AV} + \max(r_A, r_V)]}$$

where r_{AV} , r_A , and r_V are the mean firing rates to the AV, Auditory (A), and Visual (V) conditions, respectively. The value calculated is between -1 and 1 , where any value less than zero is classified as a suppressed multisensory response and any value greater than zero is classified as an enhanced multisensory response.

We next examined the multisensory time course of suppressed or enhanced nonlinear multisensory neurons using the procedure from Perrodin et al. (2014). For each nonlinear multisensory unit, we computed the $AV - (A + V)$ peristimulus time histogram (PSTH) difference vector between the AV and the summed unimodal auditory and unimodal visual responses. We computed this separately for suppressed and enhanced responses. It was then normalized by converting to SDs from baseline. The onset of the multisensory effect was defined for each unit, as the first time point when the difference score was >2 SDs of baseline for at least 10 ms.

The responses of VLPFC neurons to the incongruent AV stimuli were examined in several ways. First, we restricted our analyses to only those cells that were multisensory and displayed a nonlinear or bimodal response to AV stimuli with the two-way MANOVA described above. On this subset of cells, we performed a one-way MANOVA across the same early and late time bins (0–400 ms and 401–800 ms) by AV condition (A1V1, A2V2, A2V1, and A1V2). *Post hoc* Tukey HSD pairwise comparisons were used to examine the differences in neuronal activity between congruent and incongruent stimuli. We categorized neurons as incongruent suppressed if the response was decreased relative to the congruent stimulus and incongruent enhanced if the response was increased relative to the congruent stimulus.

The spike density function (SDF) graphs for single-cell examples were generated using the *ksdensity* MATLAB function, which convolved the spike train with a kernel smoothing function and a bin width which ranged from 20 to 30. For groups of cells, we computed a population SDF by normalizing the firing rates of neurons then averaging the responses in MATLAB.

The two congruent vocalization movies included two call types that differed in both semantic meaning and emotional valence: a positively valenced coo vocalization (A1V1) and a negatively valenced vocalization (A2V2; either a scream or an aggressive vocalization movie). Therefore, the differences in neuronal activity between congruent and incongruent stimuli might be related to a change in stimulus valence or an alteration in the behavioral meaning of this new mixed call type that occurs when switching the video or audio tracks. We therefore categorized cells according to this call type/valence change (e.g., affiliative to agonistic: A1V1 vs A2V1 or A1V2; agonistic to affiliative: A2V2 vs A1V2 or A2V1) that demonstrated significant changes in firing rate between congruent and incongruent stimuli.

The change in stimulus modality is another feature that could explain the differences in neuronal activity between congruent and incongruent vocalization movies. For each testing set, both incongruent vocalization movies were created by mismatching the auditory and visual tracks of the congruent vocalization movies. As a result, the differences between con-

gruent and incongruent stimuli can be perceived as a modality change of one component in the AV stimulus—either an auditory vocalization change or a visual face change. For neurons demonstrating significant changes in firing rate to incongruent stimuli, we categorized the difference as a vocalization change or a face change (e.g., vocalization change: A1V1 vs A2V1 or A2V2 vs A1V2; face change: A1V1 vs A1V2 or A2V2 vs A2V1).

Results

Multisensory responses of VLPFC neurons

A total of 597 neurons were isolated and tested during the movie presentation task. We examined the responses to each of the movies and their components (Movie 1: A1, V1, and A1V1; Movie 2: A2, V2, and A2V2) with a two-way MANOVA (for each movie set, see Fig. 1) across the early and late time bins of the stimulus period. Of 597 VLPFC neurons, 447 were responsive (75%) due to either a main effect of A or V or both, or an interaction effect of A^*V ($p < 0.05$), for one or both movie sets. As shown in Figure 2, 15% of cells had a main effect of auditory only (auditory unimodal, $n = 69$ cells), and 15% of cells had a main effect of visual only (visual unimodal, $n = 66$ cells). Over two-thirds of the responsive neurons (70%, $n = 312/447$ cells) were characterized as multisensory in this analysis: they were either bimodal (14% of the total responsive population) and had a significant main effect of both auditory and visual unimodal stimuli or they were nonlinear (56% of the total responsive population) and had a significant interaction effect (A^*V), as described above. Of the 312 multisensory cells, 195 were classified as bimodal, 248 had a significant interaction effect and were classified as nonlinear, and 132 neurons were both bimodal and nonlinear (main effect of A, V, and a significant interaction A^*V). Within the nonlinear multisensory neurons, 34% ($n = 85/248$) demonstrated a multisensory response to both movies, whereas 66% ($n = 163/248$) were selective, with a significant multisensory response to only one of the two movies. This finding supports previous claims that multisensory responses are selective and dependent upon the features of the combined stimuli (Meredith and Stein, 1983; Sugihara et al., 2006). It also suggests that the number of multisensory neurons for this study is an underestimate due to the finite nature of our testing set.

Upon examining the nonlinear multisensory cells ($n = 248$), we noted that significantly more cells exhibited multisensory suppression across both movies and both stimulus periods compared with multisensory enhancement (62 vs 29%, $p < 0.001$, paired *t* test). Figure 3, A–C, shows examples of three multisensory VLPFC neurons. Figure 3A depicts a VLPFC neuron that exhibited multisensory suppression during one of the monkey vocalization movies (AV) and its individual components (A and V). In this example, the neuron showed a significantly increased response to the auditory vocalization (A), no response to the silent movie (V), and a suppressed response to the AV movie stimulus compared with the response to A. In the two-way MANOVA, the interaction of A^*V was significant for the early stimulus period ($p = 0.003$) and the MMI indicated that it was suppressed relative to the unimodal auditory response. In contrast, the neuron shown in Figure 3B is an example of multisensory enhancement, with a significant main effect of V and an interaction effect (A^*V ; $p = 0.003$) that resulted in an increased response during the presentation of the AV compared with V stimuli. Figure 3C is an example of a bimodal multisensory neuron with two strikingly distinct patterns of activity during A and V ($p < 0.0001$ for A and V in both time bins). Interestingly, this neuron also demonstrated a nonlinear multisensory response as

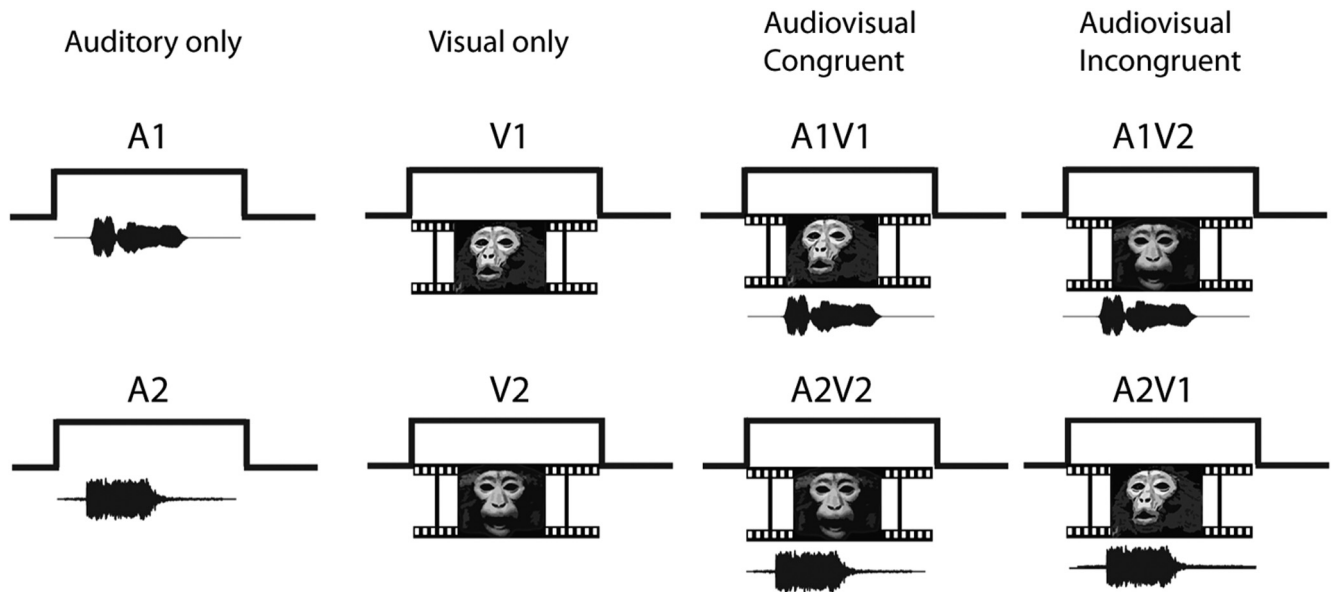


Figure 1. Audiovisual movie presentation task. Audio and video segments of monkey vocalization movies were presented separately and combined. Each testing set contained two different vocalization movies of differing valence. Each cell was tested with eight stimuli: two vocalizations (Auditory only: A1, A2), two silent movies (Visual only: V1, V2), two congruent audiovisual movies (A1V1, A2V2), and two incongruent audiovisual movies created by switching the audio and visual components between the same movies (A1V2 and A2V1).

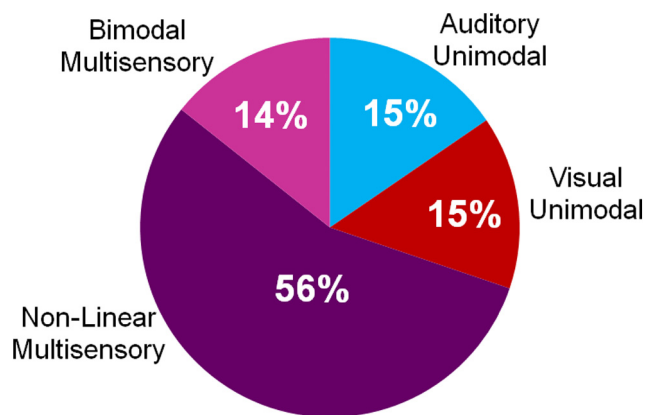


Figure 2. Percentage of VLPFC neurons with responses to auditory, visual, and audiovisual monkey vocalization movies. A total of 447 neurons responded to auditory unimodal, visual unimodal, or audiovisual vocalization stimuli based on a two-way MANOVA ($p < 0.05$). Sixty-nine of 447 were classified as auditory unimodal (blue; 15%), 66/447 were classified as visual unimodal (red; 15%), and 312/447 (75%) were classified as multisensory. Fifty-six percent of these were nonlinear multisensory (248/447; dark purple) and 14% were bimodal multisensory (64/447).

shown for the AV stimulus: the evoked response reflects V only, as if A is suppressed by V, resulting in a significant interaction effect ($p < 0.0001$). These results confirm previous findings of multisensory interactions in ventral prefrontal cortex and demonstrate the heterogeneity of multisensory VLPFC responses to the presentation of faces and vocalizations (Sugihara et al., 2006).

We performed an analysis on the time course of multisensory modulation in VLPFC neurons separately for multisensory enhanced and suppressed neurons. In this analysis, we computed a difference PSTH vector $AV - (A + V)$ and for each cell determined when this difference was significant from baseline for at least 10 ms. Assessment of $n = 35$ multisensory neurons, which showed suppression in the early stimulus period, had a multisensory onset on average of 129 ms (± 19 ms SEM). In contrast, the onset of the multisensory modulation in neurons with multisensory

enhancement in the early stimulus period ($n = 23$) was 293 ms (± 45 ms SEM). Comparison of these onset times showed a significant difference between the early onset in suppressed cells compared with the later onset in enhanced cells ($p < 0.001$, Student's t test).

VLPFC responses to incongruent movie stimuli

In an effort to understand the essential components and limits of integration, we investigated the effects of incongruent stimuli on VLPFC neurons. While all cells were tested with congruent and incongruent stimuli, our analysis of incongruence focused on neurons that were defined as multisensory to congruent AV stimuli. For the 312 multisensory neurons, we performed a one-way MANOVA on the four AV conditions (A1V1, A2V2, A2V1, and A1V2) to determine whether there was a significant difference among the congruent and incongruent movies. For this analysis, 32 of the 312 multisensory cells (10%) had a response to one or more of the incongruent movies that significantly differed from the response to the congruent movies (one-way MANOVA, $p < 0.05$, *post hoc* Tukey). When comparing neuronal activity between congruent and incongruent stimuli, 18 neurons showed suppressed activity during one or both of the incongruent movies compared with the congruent movies. Moreover, 16 neurons showed enhanced activity during one or both of the incongruent movies compared with the congruent movies. One neuron showed suppressed activity during the early stimulus period (A1V1 vs A2V1, *post hoc* Tukey, $p = 0.038$) but enhanced activity during the late stimulus period (A2V2 vs A1V2, *post hoc* Tukey, $p = 0.002$). Another neuron showed suppressed activity between A2V2 and A1V2 (*post hoc* Tukey, $p = 0.027$) but enhanced activity between A1V1 and A2V1 (*post hoc* Tukey, $p = 0.015$) both during the late stimulus period. Figure 4 illustrates two representative VLPFC neurons with differential activity between congruent and incongruent AV vocalization movies. The multisensory neuron in Figure 4A demonstrated incongruent suppression: the incongruent stimulus (A2V1, auditory scream vocalization + visual coo face) decreased neuronal firing compared with the congruent A1V1 coo vocalization movie ($p = 0.038$, early stimulus

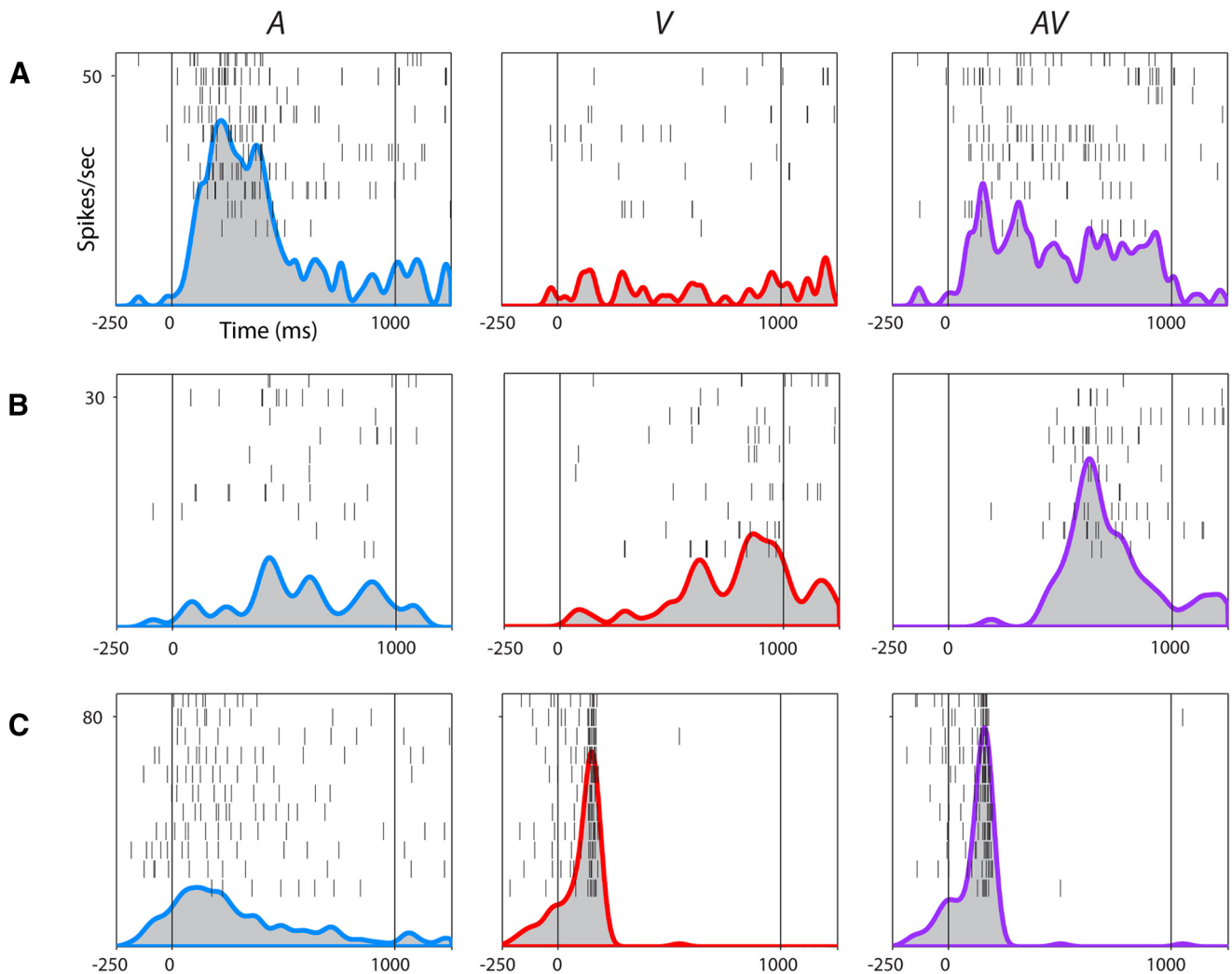


Figure 3. Multisensory responses of VLPFC neurons. Rasters and SDFs of three single cells recorded in VLPFC in **A–C** to the vocalization (A, blue SDF), silent movie (V, red SDF), and congruent face-vocalization movie (AV, purple SDF). Stimulus onset begins at 0 ms. Cell A is an example of a nonlinear multisensory neuron demonstrating a significant interaction between the auditory and visual stimuli and exhibiting multisensory suppression ($p = 0.003$, early stimulus period). Cell B is another nonlinear multisensory neuron demonstrating a significant interaction between auditory and visual stimuli and exhibiting multisensory enhancement ($p = 0.046$, late phase of stimulus period). Cell C is a bimodal multisensory neuron with neuronal activity during the auditory and visual stimuli that were significantly different from each other ($p < 0.0001$, both time bins of the stimulus period). This cell also had a nonlinear multisensory response to the AV stimulus, which could not be predicted from the linear sum of the two unimodal responses ($p < 0.0001$, both time bins of the stimulus period).

period). In Figure 4B, a different VLPFC multisensory neuron demonstrated incongruent enhancement: the incongruent stimulus (A2V1, auditory scream vocalization + visual coo face) increased neuronal firing compared with the congruent A2V2 scream vocalization movie ($p = 0.02$, late stimulus period). In Figure 4, A and B, the unimodal stimulus that contributed to the incongruent response is shown on the far right. Figure 4C shows the population response of neurons, which showed incongruent suppression during the early stimulus period (gray SDF) compared with the congruent stimulus (purple SDF; $n = 10$).

In addition to the neuron's change in response magnitude, other factors such as response latency may also change between congruent and incongruent stimuli, and this has been demonstrated in VLPFC responses to asynchronous stimuli (Romanski and Hwang, 2012). For example, a neuron may process incongruent AV stimuli more slowly than congruent AV stimuli, resulting in an early suppressed response to incongruent stimuli. We therefore calculated the number of enhanced or suppressed incongruent responses that occurred during the early or late phases of the stimulus period. Figure 5 shows the number of

VLPFC multisensory neurons with significantly enhanced or suppressed neuronal activity to the incongruent AV stimuli compared with the congruent AV stimuli. More neurons ($n = 12$) demonstrated suppressed incongruent activity in the early stimulus period (0–400 ms from onset), whereas only six neurons had suppressed incongruent activity during the late stimulus period (black bars; 401–800 ms from onset, $p < 0.05$, *post hoc* Tukey). Conversely, we observed 6 incongruent enhanced cells in the early phase and 12 incongruent enhanced cells in the late phase. These proportions include the two cells previously mentioned that showed both incongruent suppression and enhancement, as well as two additional neurons that showed enhancement in both early and late stimulus periods. A comparison of incongruent-suppressed and incongruent-enhanced cells showed that there was a trend toward an association of suppressed responses during the early and late phases of the stimulus period (Fisher's exact test, $p = 0.0943$).

We explored the relationship between incongruent suppression and enhancement with a given neuron's best unimodal response since it is possible that the unimodal responses may

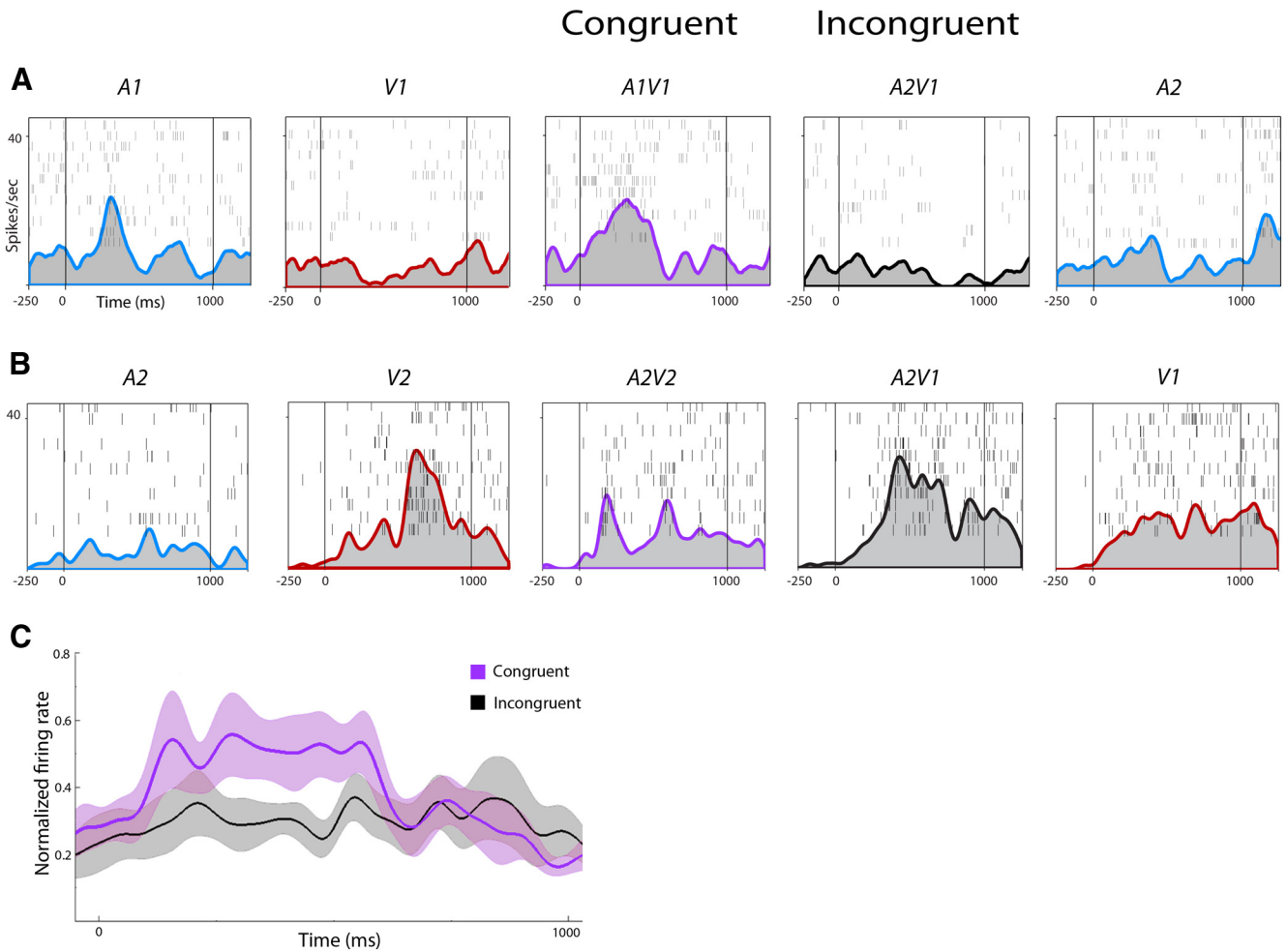


Figure 4. Multisensory neurons responsive to incongruent AV stimuli. Rasters and SDFs of two single cells during the presentation of auditory (blue SDFs), visual (red SDFs), congruent (purple SDFs), and incongruent (black SDFs) stimuli. Same conventions as in Figure 3. **A**, Cell A exhibited multisensory enhancement when the two congruent unimodal stimuli are combined in A1V1 (A*V interaction, $p < 0.05$, early stimulus period, two-way MANOVA; MMI = 0.181). However, it shows incongruent suppression to the A2V1 compared with the congruent A1V1 stimulus (purple vs black SDFs, $p < 0.05$, early stimulus period, one-way MANOVA, *post hoc* Tukey). **B**, Cell B was both bimodal and nonlinear multisensory (A2, $p = 0.003$, late stimulus period; V2, $p < 0.0001$, late stimulus period; A*V interaction, $p = 0.046$, early stimulus period) and demonstrated multisensory suppression during A1V1 compared with V1 (MMI = -0.051 , early stimulus period; MMI = -0.213 , late stimulus period). Cell B demonstrated incongruent enhancement to the A2V1 compared with the congruent A2V2 stimuli (purple vs black SDFs, $p = 0.02$, late stimulus period, one-way MANOVA, *post hoc* Tukey). **C**, Population response of VLPFC multisensory neurons that exhibited suppression to an incongruent stimulus ($n = 10$, gray SDF, with SEM) compared with the congruent response (purple SDF with SEM).

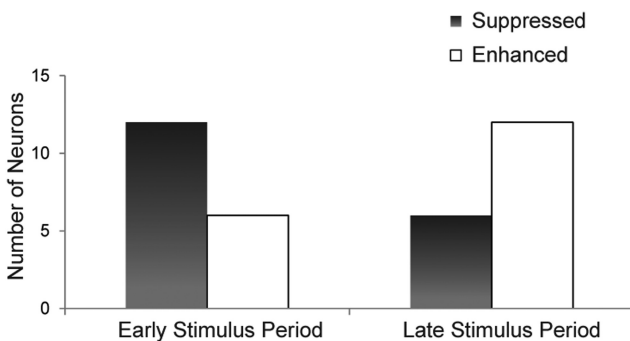


Figure 5. Multisensory neurons demonstrate incongruent enhanced or suppressed neuronal activity. Number of multisensory neurons with suppressed (black) or enhanced (gray) activity to the incongruent movie stimuli compared with the congruent vocalization movie stimuli across early and late stimulus periods (one-way MANOVA with *post hoc* Tukey HSD, $p < 0.05$).

explain the incongruent responses. Twenty-one of 32 incongruent neurons had a significant main effect of both auditory and visual unimodal stimuli in the two-way MANOVAs, while the remaining cells had a main effect of either auditory or visual

stimuli. *Post hoc* comparisons between congruent and incongruent stimuli showed that incongruent responses occurred when the modality for which the neuron had a significant main effect was changed (27/32 cells). For example, if a neuron had a main effect of auditory, it would show a significant change between A1V1 versus A2V1 and A2V2 versus A1V2—when there was a vocalization change (Fig. 4A). Thus, prefrontal neurons are likely computing incongruence based on a comparison of each unimodal component stimulus in reference to the components that make up the congruent stimulus.

Location of multisensory and incongruent responses in VLPFC

A schematic of one of the recording cylinders in VLPFC is shown in Figure 6A. Location of responsive cells (two-way MANOVA, $p < 0.05$) are shown from both hemispheres of the male subject. The locations of unimodal auditory, unimodal visual, multisensory, and incongruent-responsive cells are depicted in Figure 6B. As previously reported in Sugihara et al. (2006), visual unimodal cells were located throughout our recording area (red open squares), auditory unimodal cells were localized to anterolateral

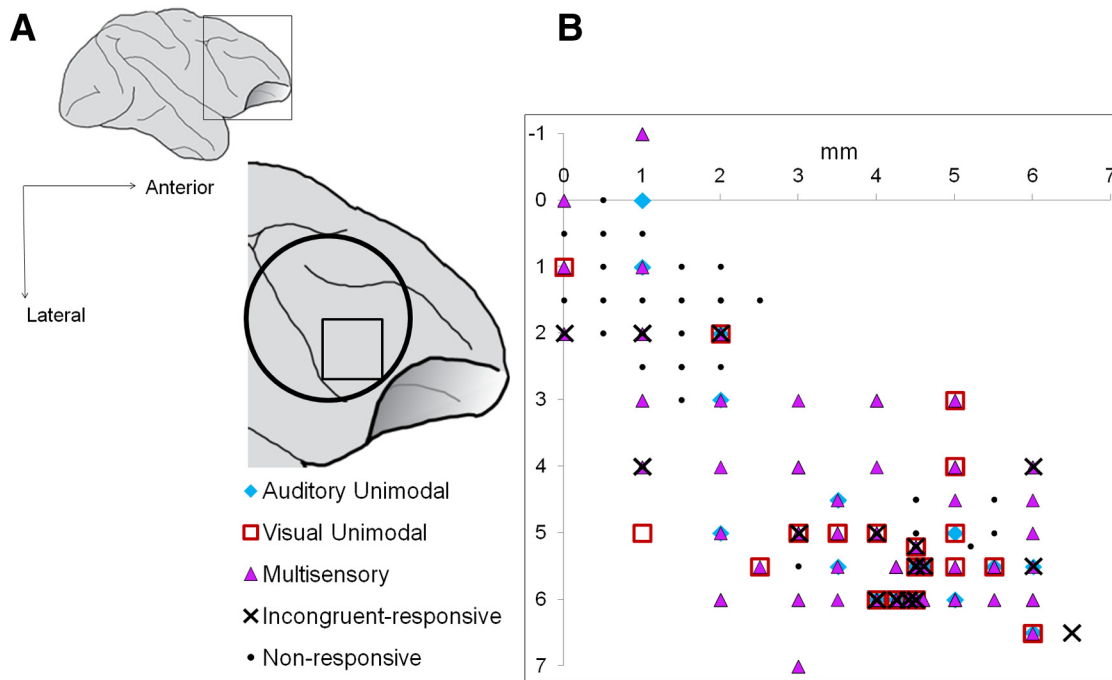


Figure 6. Location of auditory unimodal, visual unimodal, multisensory, and incongruent-responsive multisensory neurons in VLPFC. **A**, Lateral schematic of the macaque brain and the approximate location of the recording chamber over VLPFC (open circle). All recordings were made in the anterolateral quadrant of the recording chamber (open square). **B**, Auditory unimodal (blue diamonds), visual unimodal (red open squares), multisensory (purple triangles), and incongruent-responsive multisensory neurons (black Xs) are shown. Location of responsive cells is shown from both hemispheres of the male subject (two-way MANOVA, $p < 0.05$).

portions of our recording area (blue diamonds), and multisensory cells were also largely localized to anterolateral portions (purple triangles). Because enhanced or suppressed incongruent-responsive cells were selected from the multisensory cells, these too were also localized to anterolateral portions, with only two outliers shown. All recordings were made in the anterolateral quadrant of the recording chamber (black open square).

We recorded from both hemispheres in the male subject. Nineteen incongruent-responsive cells were observed out of 69 multisensory cells in the left hemisphere, while nine incongruent-responsive cells were observed out of 68 multisensory cells in the right hemisphere. Additional subjects and controls are needed to investigate any potential interhemispheric differences.

Incongruent responses related to call type/valence

Another difference between the neuronal response magnitudes of congruent and incongruent movies involves stimulus valence or call type; switching the auditory or visual track of a congruent movie could alter the semantic or emotional perception of the incongruent movie. For example, when comparing the affiliative coo vocalization movie (A1V1) to either incongruent movie (A2V1 or A1V2), one might perceive the difference between these stimuli as a shift from a completely positive valence to a partially negative valence. In this case, one of the affiliative stimulus components (face or vocalization) is switched to an agonistic stimulus component. Each testing set included one affiliative coo vocalization (A1V1) and one agonistic vocalization (either a submissive scream or an aggressive call, A2V2). We were therefore interested in whether multisensory neurons showed enhanced or suppressed incongruent activity on the basis of this alteration. The bar graph in Figure 7A shows the number of multisensory neurons with significantly incongruent suppressed or enhanced activity when compared with the activity during the affiliative congruent movie or the agonistic congruent movie. Since most

cells had a significant effect in only one time bin, we collapsed across both early and late time bins to determine whether this change specifically evoked incongruent enhanced or suppressed activity.

Of the 32 incongruent-responsive multisensory neurons, 7 showed incongruent suppression when comparing the firing rate between the affiliative congruent movie and the incongruent movies containing an agonistic stimulus component (Fig. 7A, left black bar). In contrast, 14 showed incongruent suppression when comparing the firing rate between the agonistic congruent movie and the incongruent movies containing an affiliative stimulus component (Fig. 7A, right black bar). For neurons that demonstrated incongruent enhancement related to changes in the call type/valence, we noted similar proportions: 11 were observed when comparing the firing rate between the affiliative congruent movie and the incongruent movies containing an agonistic stimulus component (Fig. 7A, left gray bar) and 12 were observed when comparing the firing rate between the agonistic congruent movie and incongruent movies containing an affiliative stimulus component (Fig. 7A, right gray bar). These proportions included the two cells with incongruent suppression and enhancement, two cells with incongruent enhancement during both stimulus periods, and four cells with incongruent enhancement and four cells with incongruent suppression for both call type changes. A Fisher's exact test revealed no association of suppressed or enhanced incongruent responses with the change in call type/valence of a stimulus ($p = 0.373$).

Incongruent responses related to modality context

Since we created the incongruent movies by switching one modality at a time, we could also determine whether a particular modality was driving an incongruent effect across the population of cells. For example, we can consider the difference between the congruent A1V1 stimulus to the incongruent A2V1 stimulus as

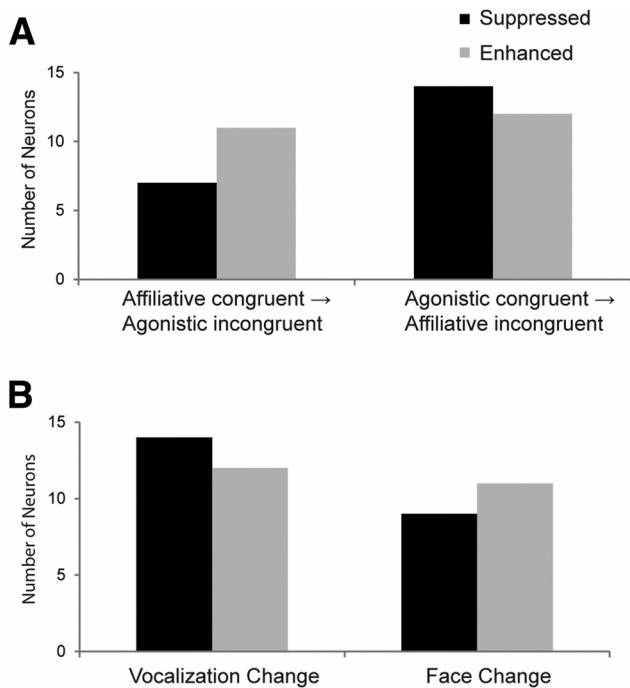


Figure 7. Multisensory neurons related to change in call type/valence and stimulus modality. **A**, Number of multisensory neurons with significantly suppressed (black) or enhanced (gray) neuronal firing to incongruent stimuli based on whether the congruent vocalization movie was affiliative (left bars) or agonistic (right bars). **B**, Number of multisensory neurons with significantly suppressed (black) or enhanced (gray) neuronal firing to the incongruent stimuli occurring during a vocalization change or a face change.

an auditory change, whereas the difference between the A1V1 and A1V2 stimuli can be considered a visual change. Because we found that 21/32 incongruent-responsive cells were bimodal and had a main effect of auditory and main effect of visual stimuli in our two-way MANOVA, we reasoned that these neurons might detect a change in either modality. Five incongruent-responsive cells had only a main effect of auditory and were nonlinear while three neurons had a main effect of visual and were nonlinear. Three neurons were nonlinear with no significant response to unimodal auditory or visual stimuli. Figure 7B shows the proportion of multisensory cells with significantly incongruent suppressed or enhanced activity when a vocalization or face change occurred. Since most cells had a significant effect in only one bin, we again collapsed across both early and late time bins for this analysis. There were also several cells with incongruent suppression or enhancement across both modality changes. We noted 14 neurons demonstrating incongruent suppression that was due to a vocalization change (Fig. 7B, left black bar), whereas only 9 neurons demonstrated incongruent suppression due to a face change (Fig. 7B, right black bar). Twelve incongruent enhanced neurons were due to a vocalization change (Fig. 7B, left gray bar) and 11 were due to a face change (Fig. 7B, right gray bar). There was no association of suppression or enhancement across the population with the type of modality change (Fisher's exact test $p = 0.767$). As shown above, incongruent responses were dependent on a neuron's response to each stimulus component (Fig. 4).

Discussion

We explored multisensory integration by testing neurons in the VLPFC of macaque monkeys with congruent (matching) or incongruent (mismatching) face-vocalization movies. Of the neurons recorded in this study, 70% exhibited multisensory profiles,

a proportion greater than shown previously (46% in Sugihara et al., 2006), which is most likely due to the additional testing sets used to determine multisensory responses. These included bimodal neurons that responded to both unimodal auditory and unimodal visual stimuli as well as nonlinear multisensory neurons with a significant response to the combined AV stimulus that differed from the linear combination of the unimodal stimuli. These nonlinear cells typically exhibited multisensory suppression in response to combined AV stimuli compared with the response of the best unimodal stimulus. Furthermore, many of the multisensory neurons were selectively responsive to only one of the two movies. This confirms the idea that multisensory responses are a product of the stimuli, not a quality of the cell, so that single units may show enhancement for one pair of congruent stimuli but suppression for another. Finally, a subset of multisensory neurons (10%) exhibited a change in neuronal response when a stimulus was made incongruent by changing one of the stimulus components. Most incongruent-responsive cells were bimodal multisensory, fostering their ability to respond to changes in either modality of a face-vocalization stimulus.

Studies have suggested that the timing of incoming afferents is crucial to the magnitude of the integrated response. Inputs arriving during the optimal phase of neuronal oscillations evoke enhanced responses, whereas inputs arriving during opposite phases are suppressed (Lakatos et al., 2005, 2007, 2009; Schroeder et al., 2008). Determining the optimal factors in a spatial domain is feasible, whereas ascertaining the optimal factors for integration in the object domain is more difficult. Presumably, evolutionary and behavioral constraints have merged congruent facial and vocal information, resulting in optimal neuronal processing of multisensory communication information. Nonetheless, even when congruent facial and vocal stimuli are temporally aligned, multisensory suppression often occurs in single units, as we have demonstrated in the current study and in a previous study (Sugihara et al., 2006). These results contrast with studies where temporally and spatially aligned stimuli elicited the greatest magnitude response in the cat superior colliculus (Meredith and Stein, 1986; Meredith et al., 1987). It remains possible that a specific face-vocalization combination would elicit this "optimal" response in VLPFC, but determining the specific combination would require a larger stimulus set.

Brain regions associated with stimulus incongruence

While many neuroimaging studies have focused on changes in the STS during multisensory integration (Calvert, 2001; Jones and Callan, 2003; Miller and D'Esposito, 2005; van Atteveldt et al., 2007; Watson et al., 2013), several studies using incongruent AV stimuli have observed changes in the activity of the inferior frontal gyrus (IFG; Ojanen et al., 2005; Hein et al., 2007; Noppeney et al., 2008; Naumer et al., 2009), a human brain region similar to the macaque VLPFC (Petrides and Pandya, 2002; Romanski, 2012). Noppeney et al. (2008) reported increased IFG activation to incongruent versus congruent words and sounds during a semantic categorization task. The signal increase was negatively correlated with subjects' reaction time and positively correlated with their accuracy, suggesting the importance of the IFG in detecting mismatching stimuli during semantic processing. Increased IFG activation occurs even with mismatching non-speech stimuli (Hein et al., 2007; Naumer et al., 2009). In comparison, VLPFC neurons demonstrated multisensory suppression in 46% of units and multisensory enhancement in 28% in response to incongruent vocalization movies. The increased blood flow observed in neuroimaging studies may reflect in-

creased inhibitory and excitatory synaptic activity that would account for multisensory suppression of single VLPFC neurons observed in the current study. Among the few studies examining the influence of incongruent stimuli on single cells in the nonhuman primate, most have focused on the STS and auditory cortex (Barracough et al., 2005; Ghazanfar et al., 2005; Dahl et al., 2010; Kayser et al., 2010; Chandrasekaran et al., 2013). These studies agree with our results and demonstrate that temporal lobe neurons also show suppression to incongruent stimuli. Thus, multisensory suppression for semantically incongruent stimuli at the single-unit level is not at odds with increased activation observed in neuroimaging studies.

Mechanisms for incongruence

Under natural conditions when a face and voice match, an individual is easily recognized. But if the face and voice are incongruent or a sensory channel is unreliable, the output signal from prefrontal neurons is altered and stimulus recognition is disrupted. Our findings revealed that more VLPFC neurons responded to incongruent stimuli with decreased firing (suppression) when compared with congruent stimuli, especially during the early stimulus period (Fig. 5). It is attractive to view incongruent suppression as a decrease in information from VLPFC due to mismatching information. This same mechanism was proposed for the neuronal suppression that occurs in VLPFC for asynchronous stimuli (Romanski and Hwang, 2012). However, suppression can occur with congruent stimuli, making it unlikely that suppression would signal a loss of information. Furthermore, it has been suggested that suppression may actually yield more information than enhancement during integration (Kayser et al., 2010).

We also observed multisensory neurons with incongruent enhancement, which occurred more often during the late stimulus period (Fig. 5). At a population level, this late response could reflect additional cognitive processing of the mismatching stimuli and the accumulation of additional information needed to decipher the confused message. Late enhancement during incongruent stimuli could also explain the increased IFG activity to incongruent stimuli in neuroimaging studies (Ojanen et al., 2005; Hein et al., 2007; Noppeney et al., 2008; Naumer et al., 2009).

Differences in VLPFC responses to congruent and incongruent multisensory stimuli might also be due to integration of different afferents by VLPFC neurons. VLPFC receives diverse anatomical inputs from areas that process both unimodal (Webster et al., 1994; Hackett et al., 1999; Romanski et al., 1999a) as well as multisensory information (Seltzer and Pandya, 1989; Barbas, 1992; Romanski et al., 1999b). We have previously hypothesized that VLPFC may integrate multiple unimodal afferents *de novo* or process already integrated multisensory afferents (from STS) or a combination of both (Sugihara et al., 2006) during social communication. Therefore, distinct combinations of these different VLPFC afferents could explain suppression or enhancement observed in the current study.

Effects of call type/valence on VLPFC responses

In humans, fearful voices are perceived as less fearful when presented with a happy face under different task demands (Vroomen et al., 2001), and congruent valence information is more easily recognized during categorization of face–voice stimuli (de Gelder and Vroomen, 2000; Kreifelts et al., 2007). Our incongruent vocalization movies were created by mixing stimulus components from an affiliative coo and an agonistic scream or bark. Although our study was not designed to specifically assess changes in emo-

tional valence, we did note response magnitude changes when an affiliative face or vocalization replaced a component of the agonistic vocalization (Fig. 7). In mismatching the faces and vocalizations, semantic meaning and emotional valence were likely altered.

Effects of stimulus modality on VLPFC responses

Though most incongruent neurons were bimodal, we observed a dissociation between neurons affected by face or vocalization changes when comparing the differences between incongruent suppressed and enhanced activity (Fig. 7). There was a trend such that neurons exhibited incongruent suppression more frequently to auditory (vocalization) changes. Similarly, when Dahl et al. (2010) switched the auditory track between AV scenes, it altered the timing and magnitude of firing in STS neurons.

Moreover, our single prefrontal cells are not necessarily detecting a semantic mismatch between the congruent and incongruent stimuli, rather they are responding to a modality change between the incongruent and congruent movies. The process of detecting a semantic mismatch within an AV stimulus likely depends on a much larger network of cells and regions. Our findings support the idea that auditory information can influence visual perception, just as it has been shown that visual information affects auditory perception (McGurk and MacDonald, 1976), especially during face and vocalization processing.

Our results highlight the importance of prefrontal neurons in the perception and integration of face and vocal stimuli and are pertinent to understanding the disruption of sensory integration in social communication disorders including autism spectrum disorders (ASD) and schizophrenia. Individuals with ASD have difficulty integrating multisensory information (Iarocci and McDonald, 2006; Collignon et al., 2013), including facial expressions, vocal stimuli, and gestures (Dawson et al., 2004; Silverman et al., 2010; Charbonneau et al., 2013; Lerner et al., 2013). Additionally, patients diagnosed with schizophrenia may experience heightened multisensory integration (de Jong et al., 2010; Stone et al., 2011). In both ASD and schizophrenia, ventral prefrontal regions were disrupted in different working memory tasks involving speech and language (Wang et al., 2004; Akechi et al., 2010; Davies et al., 2011; Ishii-Takahashi et al., 2013; Eich et al., 2014; Marumo et al., 2014). Thus, continued research to delineate the neural mechanisms of multisensory integration can shed light on the neuronal basis of these disorders.

References

- Akechi H, Senju A, Kikuchi Y, Tojo Y, Osanai H, Hasegawa T (2010) The effect of gaze direction on the processing of facial expressions in children with autism spectrum disorder: an ERP study. *Neuropsychologia* 48: 2841–2851. CrossRef Medline
- Barbas H (1992) Architecture and cortical connections of the prefrontal cortex in the rhesus monkey. *Adv Neurol* 57:91–115. Medline
- Barracough NE, Xiao D, Baker CI, Oram MW, Perrett DI (2005) Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. *J Cogn Neurosci* 17:377–391. CrossRef Medline
- Calvert GA (2001) Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cereb Cortex* 11:1110–1123. CrossRef Medline
- Chandrasekaran C, Lemus L, Ghazanfar AA (2013) Dynamic faces speed up the onset of auditory cortical spiking responses during vocal detection. *Proc Natl Acad Sci U S A* 110:E4668–E4677. CrossRef Medline
- Charbonneau G, Bertone A, Lepore F, Nassim M, Lassonde M, Mottron L, Collignon O (2013) Multilevel alterations in the processing of audiovisual emotion expressions in autism spectrum disorders. *Neuropsychologia* 51:1002–1010. CrossRef Medline
- Collignon O, Charbonneau G, Peters F, Nassim M, Lassonde M, Lepore F,

- Mottron L, Bertone A (2013) Reduced multisensory facilitation in persons with autism. *Cortex* 49:1704–1710. [CrossRef Medline](#)
- Dahl CD, Logothetis NK, Kayser C (2010) Modulation of visual responses in the superior temporal sulcus by audio-visual congruency. *Front Integr Neurosci* 4:10. [CrossRef Medline](#)
- Davies MS, Dapretto M, Sigman M, Sepeta L, Bookheimer SY (2011) Neural bases of gaze and emotion processing in children with autism spectrum disorders. *Brain Behav* 1:1–11. [CrossRef Medline](#)
- Dawson G, Webb SJ, Carver L, Panagiotides H, McPartland J (2004) Young children with autism show atypical brain responses to fearful versus neutral facial expressions of emotion. *Dev Sci* 7:340–359. [CrossRef Medline](#)
- de Gelder B, Vroomen J (2000) The perception of emotions by ear and by eye. *Cogn Emot* 14:289–311. [CrossRef](#)
- de Jong JJ, Hodiament PP, de Gelder B (2010) Modality-specific attention and multisensory integration of emotions in schizophrenia: reduced regulatory effects. *Schizophr Res* 122:136–143. [CrossRef Medline](#)
- Eich TS, Nee DE, Insel C, Malapani C, Smith EE (2014) Neural correlates of impaired cognitive control over working memory in schizophrenia. *Biol Psychiatry* 76:146–153. [CrossRef Medline](#)
- Ghazanfar AA, Maier JX, Hoffman KL, Logothetis NK (2005) Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *J Neurosci* 25:5004–5012. [CrossRef Medline](#)
- Gouzoules S, Gouzoules H, Marler P (1984) Rhesus monkey (*Macaca mulatta*) screams: representational signalling in the recruitment of agonistic aid. *Anim Behav* 32:182–193. [CrossRef](#)
- Hackett TA, Stepniewska I, Kaas JH (1999) Prefrontal connections of the parabelt auditory cortex in macaque monkeys. *Brain Res* 817:45–58. [CrossRef Medline](#)
- Hauser MD, Marler P (1993) Food-associated calls in rhesus macaques (*Macaca mulatta*): I. socioecological factors. *Behav Ecol* 4:194–205. [CrossRef](#)
- Hein G, Doehrmann O, Müller NG, Kaiser J, Muckli L, Naumer MJ (2007) Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. *J Neurosci* 27:7881–7887. [CrossRef Medline](#)
- Homae F, Hashimoto R, Nakajima K, Miyashita Y, Sakai KL (2002) From perception to sentence comprehension: the convergence of auditory and visual information of language in the left inferior frontal cortex. *Neuroimage* 16:883–900. [CrossRef Medline](#)
- Iarocci G, McDonald J (2006) Sensory integration and the perceptual experience of persons with autism. *J Autism Dev Disord* 36:77–90. [CrossRef Medline](#)
- Ishii-Takahashi A, Takizawa R, Nishimura Y, Kawakubo Y, Kuwabara H, Matsubayashi J, Hamada K, Okuhata S, Yahata N, Igarashi T, Kawasaki S, Yamasue H, Kato N, Kasai K, Kano Y (2013) Prefrontal activation during inhibitory control measured by near-infrared spectroscopy for differentiating between autism spectrum disorders and attention deficit hyperactivity disorder in adults. *Neuroimage Clin* 4:53–63. [CrossRef Medline](#)
- Jones JA, Callan DE (2003) Brain activity during audiovisual speech perception: an fMRI study of the McGurk effect. *Neuroreport* 14:1129–1133. [CrossRef Medline](#)
- Kayser C, Logothetis NK, Panzeri S (2010) Visual enhancement of the information representation in auditory cortex. *Curr Biol* 20:19–24. [CrossRef Medline](#)
- Kreifelts B, Ethofer T, Grodd W, Erb M, Wildgruber D (2007) Audiovisual integration of emotional signals in voice and face: an event-related fMRI study. *Neuroimage* 37:1445–1456. [CrossRef Medline](#)
- Lakatos P, Shah AS, Knuth KH, Ulbert I, Karmos G, Schroeder CE (2005) An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *J Neurophysiol* 94:1904–1911. [CrossRef Medline](#)
- Lakatos P, Chen CM, O'Connell MN, Mills A, Schroeder CE (2007) Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron* 53:279–292. [CrossRef Medline](#)
- Lakatos P, O'Connell MN, Barczak A, Mills A, Javitt DC, Schroeder CE (2009) The leading sense: supramodal control of neurophysiological context by attention. *Neuron* 64:419–430. [CrossRef Medline](#)
- Laurienti PJ, Kraft RA, Maldjian JA, Burdette JH, Wallace MT (2004) Semantic congruence is a critical factor in multisensory behavioral performance. *Exp Brain Res* 158:405–414. [Medline](#)
- Lerner MD, McPartland JC, Morris JP (2013) Multimodal emotion processing in autism spectrum disorders: an event-related potential study. *Dev Cogn Neurosci* 3:11–21. [CrossRef Medline](#)
- Marumo K, Takizawa R, Kinou M, Kawasaki S, Kawakubo Y, Fukuda M, Kasai K (2014) Functional abnormalities in the left ventrolateral prefrontal cortex during a semantic fluency task, and their association with thought disorder in patients with schizophrenia. *Neuroimage* 85:518–526. [CrossRef Medline](#)
- McGurk H, MacDonald J (1976) Hearing lips and seeing voices. *Nature* 264:746–748. [CrossRef Medline](#)
- Meredith MA, Stein BE (1983) Interactions among converging sensory inputs in the superior colliculus. *Science* 221:389–391. [CrossRef Medline](#)
- Meredith MA, Stein BE (1986) Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Res* 365:350–354. [CrossRef Medline](#)
- Meredith MA, Nemitz JW, Stein BE (1987) Determinants of multisensory integration in superior colliculus neurons. I. temporal factors. *J Neurosci* 7:3215–3229. [Medline](#)
- Miller LM, D'Esposito M (2005) Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. *J Neurosci* 25:5884–5893. [CrossRef Medline](#)
- Naumer MJ, Doehrmann O, Müller NG, Muckli L, Kaiser J, Hein G (2009) Cortical plasticity of audio-visual object representations. *Cereb Cortex* 19:1641–1653. [CrossRef Medline](#)
- Noppeney U, Josephs O, Hocking J, Price CJ, Friston KJ (2008) The effect of prior visual information on recognition of speech and sounds. *Cereb Cortex* 18:598–609. [CrossRef Medline](#)
- Ojanen V, Möttönen R, Pekkola J, Jääskeläinen IP, Joensuu R, Autti T, Sams M (2005) Processing of audiovisual speech in Broca's area. *Neuroimage* 25:333–338. [CrossRef Medline](#)
- O'Scalaidhe SP, Wilson FA, Goldman-Rakic PS (1997) Areal segregation of face-processing neurons in prefrontal cortex. *Science* 278:1135–1138. [CrossRef Medline](#)
- Perrodin C, Kayser C, Logothetis NK, Petkov CI (2014) Auditory and visual modulation of temporal lobe neurons in voice-sensitive and association cortices. *J Neurosci* 34:2524–2537. [CrossRef Medline](#)
- Petrides M, Pandya DN (2002) Comparative cytoarchitectonic analysis of the human and the macaque ventrolateral prefrontal cortex and corticocortical connection patterns in the monkey. *Eur J Neurosci* 16:291–310. [CrossRef Medline](#)
- Preuss TM, Goldman-Rakic PS (1991) Ipsilateral cortical connections of granular frontal cortex in the strepsirrhine primate galago, with comparative comments on anthropoid primates. *J Comp Neurol* 310:507–549. [CrossRef Medline](#)
- Romanski LM (2012) Integration of faces and vocalizations in ventral prefrontal cortex: implications for the evolution of audiovisual speech. *Proc Natl Acad Sci U S A* 109 [Suppl 1]:10717–10724. [CrossRef Medline](#)
- Romanski LM, Goldman-Rakic PS (2002) An auditory domain in primate prefrontal cortex. *Nat Neurosci* 5:15–16. [CrossRef Medline](#)
- Romanski LM, Hwang J (2012) Timing of audiovisual inputs to the prefrontal cortex and multisensory integration. *Neuroscience* 214:36–48. [CrossRef Medline](#)
- Romanski LM, Bates JF, Goldman-Rakic PS (1999a) Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. *J Comp Neurol* 403:141–157. [CrossRef Medline](#)
- Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP (1999b) Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat Neurosci* 2:1131–1136. [CrossRef Medline](#)
- Romanski LM, Averbach BB, Diltz M (2005) Neural representation of vocalizations in the primate ventrolateral prefrontal cortex. *J Neurophysiol* 93:734–747. [Medline](#)
- Schroeder CE, Lakatos P, Kajikawa Y, Partan S, Puce A (2008) Neuronal oscillations and visual amplification of speech. *Trends Cogn Sci* 12:106–113. [CrossRef Medline](#)
- Seltzer B, Pandya DN (1989) Frontal lobe connections of the superior temporal sulcus in the rhesus monkey. *J Comp Neurol* 281:97–113. [CrossRef Medline](#)
- Shams L, Seitz AR (2008) Benefits of multisensory learning. *Trends Cogn Sci* 12:411–417. [CrossRef Medline](#)
- Silverman LB, Bennetto L, Campana E, Tanenhaus MK (2010) Speech-and-gesture integration in high functioning autism. *Cognition* 115:380–393. [CrossRef Medline](#)

- Stein BE, Stanford TR (2008) Multisensory integration: current issues from the perspective of the single neuron. *Nat Rev Neurosci* 9:255–266. [CrossRef Medline](#)
- Stone DB, Urrea LJ, Aine CJ, Bustillo JR, Clark VP, Stephen JM (2011) Unisensory processing and multisensory integration in schizophrenia: a high-density electrical mapping study. *Neuropsychologia* 49:3178–3187. [CrossRef Medline](#)
- Sugihara T, Diltz MD, Averbeck BB, Romanski LM (2006) Integration of auditory and visual communication information in the primate ventrolateral prefrontal cortex. *J Neurosci* 26:11138–11147. [CrossRef Medline](#)
- Tsao DY, Schweers N, Moeller S, Freiwald WA (2008) Patches of face-selective cortex in the macaque frontal lobe. *Nat Neurosci* 11:877–879. [CrossRef Medline](#)
- van Atteveldt NM, Formisano E, Blomert L, Goebel R (2007) The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cereb Cortex* 17:962–974. [Medline](#)
- Vroomen J, Driver J, de Gelder B (2001) Is cross-modal integration of emotional expressions independent of attentional resources? *Cogn Affect Behav Neurosci* 1:382–387. [CrossRef Medline](#)
- Wang AT, Dapretto M, Hariri AR, Sigman M, Bookheimer SY (2004) Neural correlates of facial affect processing in children and adolescents with autism spectrum disorder. *J Am Acad Child Adolesc Psychiatry* 43:481–490. [CrossRef Medline](#)
- Watson R, Latinus M, Noguchi T, Garrod O, Crabbe F, Belin P (2013) Dissociating task difficulty from incongruence in face-voice emotion integration. *Front Hum Neurosci* 7:744. [CrossRef Medline](#)
- Webster MJ, Bachevalier J, Ungerleider LG (1994) Connections of inferior temporal areas TEO and TE with parietal and frontal cortex in macaque monkeys. *Cereb Cortex* 4:470–483. [CrossRef Medline](#)
- Wilson FA, Scalaidhe SP, Goldman-Rakic PS (1993) Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science* 260:1955–1958. [CrossRef Medline](#)