Behavioral/Cognitive

Prefrontal Neuronal Responses during Audiovisual Mnemonic Processing

Jaewon Hwang[1] **and Lizabeth M. Romanski**[2]

[1]Zanvyl Krieger Mind/Brain Institute, Johns Hopkins University, Baltimore, Maryland 21218 and [2]Department of Neurobiology and Anatomy, University of Rochester School of Medicine, Rochester, New York 14642

During communication we combine auditory and visual information. Neurophysiological research in nonhuman primates has shown that single neurons in ventrolateral prefrontal cortex (VLPFC) exhibit multisensory responses to faces and vocalizations presented simultaneously. However, whether VLPFC is also involved in maintaining those communication stimuli in working memory or combining stored information across different modalities is unknown, although its human homolog, the inferior frontal gyrus, is known to be important in integrating verbal information from auditory and visual working memory. To address this question, we recorded from VLPFC while rhesus macaques (*Macaca mulatta*) performed an audiovisual working memory task. Unlike traditional match-to-sample/nonmatch-to-sample paradigms, which use unimodal memoranda, our nonmatch-to-sample task used dynamic movies consisting of both facial gestures and the accompanying vocalizations. For the nonmatch conditions, a change in the auditory component (vocalization), the visual component (face), or both components was detected. Our results show that VLPFC neurons are activated by stimulus and task factors: while some neurons simply responded to a particular face or a vocalization regardless of the task period, others exhibited activity patterns typically related to working memory such as sustained delay activity and match enhancement/suppression. In addition, we found neurons that detected the component change during the nonmatch period. Interestingly, some of these neurons were sensitive to the change of both components and therefore combined information from auditory and visual working memory. These results suggest that VLPFC is not only involved in the perceptual processing of faces and vocalizations but also in their mnemonic processing.

Key words: **faces; macaque; monkey; multisensory; vocalization; working memory**

Introduction

Communication is a multisensory phenomenon (McGurk and MacDonald, 1976; Campanella and Belin, 2007; Ghazanfar et al., 2010). We employ vocal sounds, mouth movements, facial motions, and hand/body gestures, when speaking to one another. To comprehend these communication signals, we need to retain many auditory and visual cues in memory and integrate them, while we retrieve their referents from knowledge stores. Thus, a communication circuit must include brain regions that receive auditory and visual inputs and are capable of processing, remembering, and integrating complex audiovisual information.

The joint processing of auditory and visual information has been shown to take place in a number of brain regions including the superior colliculus, temporal cortex, and the frontal lobes (Stein and Stanford, 2008; Murray and Wallace, 2012). Neuroimaging studies have especially noted activations in the inferior frontal gyrus (IFG) during the processing of audiovisual speech stimuli. The human IFG, including Broca's area, is activated not only when auditory and visual verbal materials are processed (Calvert et al., 2001; Homae et al., 2002; Jones and Callan, 2003; Miller and D'Esposito, 2005; Ojanen et al., 2005; Noppeney et al., 2010; Lee and Noppeney, 2011), but also when they are stored in memory for further manipulation, which suggests IFG plays a role in working memory (WM) of communication stimuli (Paulesu et al., 1993; Schumacher et al., 1996; Crottaz-Herbette et al., 2004; Rämä and Courtney, 2005).

More specific investigations of neuronal activity during the processing of communication stimuli have been performed in nonhuman primates who also use face and vocal stimuli in their social interactions. One region involved in audiovisual integration in nonhuman primates is the ventrolateral prefrontal cortex (VLPFC), which includes areas 12/47 and 45 and is homologous with the human IFG (Petrides and Pandya, 1988). Single-unit recordings have identified "face cells" in VLPFC (Wilson et al., 1993; O Scalaidhe et al., 1997; Scalaidhe et al., 1999) and neurons that are responsive to species-specific vocalizations (Romanski and Goldman-Rakic, 2002; Romanski et al., 2005). Many VLPFC neurons are multisensory and are responsive to vocalizations and the corresponding facial gesture presented simultaneously (Sugihara et al., 2006; Romanski and Hwang, 2012; Diehl and Romanski, 2014). However, further studies are needed to elucidate the individual and ensemble activity that occurs in VLPFC when face

Received April 1, 2014; revised Oct. 29, 2014; accepted Nov. 23, 2014.

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

This work was supported by National Institutes of Health Grants DC04845 (L.M.R.) and The Schmitt Program on Integrative Brain Research.

The authors declare no competing financial interests.

Correspondence should be addressed to Lizabeth M. Romanski, Department of Neurobiology and Anatomy, University of Rochester School of Medicine, Box 603, 601 Elmwood Ave, Rochester, NY 14642. E-mail: Liz_romanski@urmc.rochester.edu.

DOI:10.1523/JNEUROSCI.1328-14.2015

 0270-6474/15/350960-12$15.00/0

and vocal stimuli are remembered and integrated during WM processing.

Previous studies have illustrated the importance of the prefrontal cortex in WM and have postulated a role for VLPFC in nonspatial WM (Wilson et al., 1993; Nee et al., 2013; Plakke et al., 2013a; Plakke and Romanski, 2014), but few studies have addressed how neurons retain combined stimuli, such as a vocalization and a facial expression. In the current study, we therefore examined neuronal activity in the primate VLPFC during a WM task where both auditory and visual stimuli are the memoranda. Our neurophysiological results indicate that VLPFC neurons are recruited during WM processing of faces and vocalizations, with some cells activated by changes in either face or vocal information while others are multisensory and attend to both stimuli.

Materials and Methods

Subjects and apparatus. Two female rhesus monkeys (*Macaca mulatta*) were used (Monkey P and T; 6.7 and 6.2 kg). All procedures conformed to the guidelines of National Institutes of Health and were approved by the University of Rochester Care and Use of Animals in Research Committee. Before training, a titanium head post was surgically implanted on the skull of each subject for head fixation. During training, the subject sat in a primate chair with the head fixed in a sound-attenuated room. Visual stimuli were presented on a computer monitor (NEC MultiSync LCD1830, 1280 × 1024, 60 Hz), which was at 75 cm distance from the eyes. Auditory stimuli were presented via two speakers (Yamaha MSP5; frequency response 50–40 kHz) placed on either side of the monitor at the height of the subject's head. Eye position was continuously monitored with an infrared pupil monitoring system (ISCAN). Behavioral data (eye position and button press) were collected on a PC via PCI interface boards (NI PCI-6220 and NI PCI-6509; National Instruments). The timing of stimulus presentation and reward delivery was controlled with in-house C++ software, which was built based on Microsoft DirectX technologies.

Stimuli. Stimuli were short movie clips of vocalizing monkeys filmed in our own colony. The video was captured at 30 fps with a size of 320 × 240 pixels (6.8 × 5.1° in visual angle), and the audio was recorded with a 48 kHz sampling rate and 16-bit resolution. Once the video and audio tracks of the movie clips were digitized, they were processed with VirtualDub (virtualdub.org) and GoldWave software (GoldWave). Each movie clip was shortened so that only the relevant vocalization was presented with the accompanying facial gesture (Fig. 1*A*). The length of the video tracks was 467–1367 ms (892 ms on average) and that of the audio ranged from 145 to 672 ms (308 ms on average). Sound pressure level of auditory stimuli was adjusted to 65–75 dB (35–112 mPa) at the level of the subject's ear.

Vocalization stimuli were also imported into MATLAB (The MathWorks) and processed to create noise sound stimuli, which were used to test the auditory discrimination of our subjects (see below). To create them, we extracted the envelope of the vocalization with the Hilbert transform and applied it on white Gaussian noise of which the frequency band was limited to be similar to that of the vocalization. The resulting noise sound was normalized in root mean square amplitude to the original vocalization. Therefore, these noise stimuli had different power spectra from the original vocalizations, but similar temporal features.

Task. The subjects were trained to perform an audiovisual nonmatch-to-sample task (Fig. 1*B*). They were required to remember a movie clip presented during the sample period and detect a nonmatching movie clip in subsequent stimulus presentations. When they successfully detected the nonmatch and indicated it by pressing a button located on the front panel of the chair, they were rewarded with juice after 0.5 s. The subject initiated a trial by fixating for 1 s on a red square presented at the center of the screen. Then, the sample stimulus (i.e., an audiovisual face-vocalization movie) was presented, followed by a 1 s delay period. In half the trials, the second stimulus presented after the delay was the nonmatch ("S2 nonmatch"). In the other half of the trials, the sample stimulus and the delay period were repeated before the nonmatch, making the occurrence of the nonmatch unpredictable. The subjects were required to withhold the button press for this repeated sample stimulus (i.e., match) until the nonmatch stimulus was finally presented ("S3 nonmatch").

In each session of training, we selected a pair of audiovisual movies for the sample stimuli (A1V1 and A2V2) and created their nonmatches by interchanging the audio and video tracks of these two movies (Fig. 2). Since each movie clip had an auditory component (A*n*) and a visual component (V*n*), the exchange could occur between the audio tracks (audio nonmatch), the video tracks (video nonmatch) or both audio and video tracks (AV nonmatch). For example, when the sample stimulus was A1V1, its audio, video, and AV nonmatches were A2V1, A1V2, and A2V2, respectively. Thus, the nonmatching audiovisual stimuli consisted of incongruent (A2V1, A1V2) or congruent (A2V2) face-vocalization movies. To create these audiovisual nonmatch stimuli, we carefully chose two movies of different vocalization call types in which vocalizations and the mouth motion were similar in length. For the incongruent nonmatch stimuli, we aligned the onset time of the nonmatching vocalization to that of the original vocalization so that the subjects could not use temporal asynchrony as a nonmatch cue.

These three types of nonmatch stimuli were used for the neurophysiological recordings, which is the main study of this paper. In addition, we performed a series of behavioral experiments to test the auditory discrimination ability of our subjects. In these behavioral experiments, we used another nonmatch type in which the vocalizations of the sample movies were replaced with band-limited white noise stimuli that we created (noise nonmatch; Fig. 2). These noise stimuli were easy to discriminate from vocalizations due to their distinct power spectra and therefore helped us to determine whether our subjects had any difficulties in general auditory processing.

Throughout the experiments, we used four pairs, or eight movie clips, and alternated them from session to session. For each testing or recording session, the two vocalization movies that were paired differed in call type and gender (e.g., female affiliative call vs male agonistic call) to make vocalization discrimination easier. Subjects were allowed a total of 900 ms plus the duration of the movie stimulus to press the button during the nonmatch period. If they did not respond during this period, the trial was aborted without reward and the next trial began. If the subject broke eye fixation during stimulus presentation or pressed the button before the nonmatch period, that trial was aborted immediately. Unrewarded trials were repeated but were randomized again with the remaining trials so that the subject could not guess the conditions of the next trial. All trial conditions were presented in a pseudorandom fashion and counterbalanced across trials.

Single-unit recording. After training in the task was complete, a titanium recording cylinder (19 mm inner diameter) was implanted over VLPFC (centered 29–30 mm anterior to the interaural line and 20–21 mm lateral to midline on skull). The recording cylinder was angled 30° to the vertical to maximize an orthogonal approach to VLPFC, areas 12/47 and 45 (Preuss and Goldman-Rakic, 1991). Recordings were made in both hemispheres of Monkey P and the right hemisphere of Monkey T while they performed the task. During recordings, one or two glass-coated tungsten electrodes (impedance 1–2 MΩ; Alpha Omega) were lowered by motorized microdrives (Nan Instruments) to the target areas. Single-unit activity was discriminated and collected with a signal processing system (RX5-2; Tucker-Davis Technologies). Electrode trajectories were confirmed with MRI for both subjects and later with histology for Monkey P. The MRI image slices were traced with NIH MIPAV software (mipav.cit.nih.gov) and were reconstructed to a 3D model with MATLAB to plot the recording sites (Fig. 5).

Analysis of behavioral data. The percentage of rewarded trials (i.e., the success rate) was calculated as follows. In this task, the subjects typically made two types of errors: (1) not detecting the S2 nonmatch stimulus (missed-press error) and (2) pressing the button during the match stimulus (wrong-press error). The other types of errors (pressing the button during the sample or delay period and missing the S3 nonmatch stimulus) accounted for only 0.39% of the total number of trials and were not considered in the analysis. When subjects made a wrong-press error, the trial was aborted before the nonmatch stimulus was presented, so the error was not attributed to a particular nonmatch condition. Therefore,

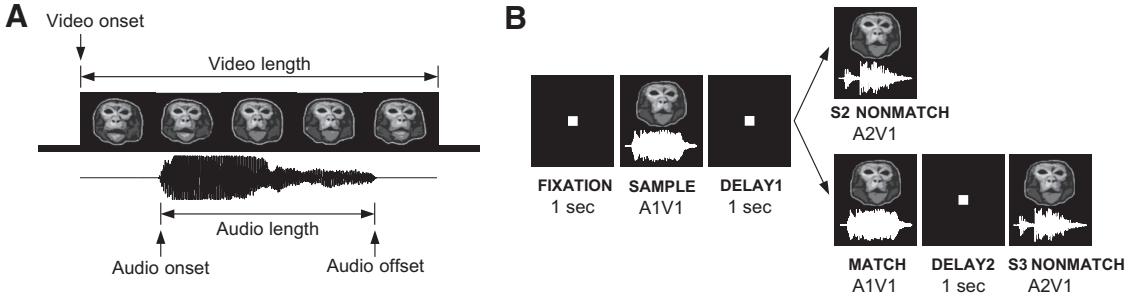

Figure 1. Audiovisual nonmatch-to-sample task. ***A***, A face-vocalization movie stimulus used in the audiovisual nonmatch-to-sample task is shown in this schematic representation. Note that in these movie stimuli, the auditory component (vocalization) was always preceded by the visual component (face movie). The length of the video tracks ranged from 467 to 1367 ms (892 ms on average) and that of the audio tracks from 145 to 672 ms (308 ms on average). ***B***, Schematic of the audiovisual nonmatch-to-sample task. A face-vocalization movie was presented as the sample stimulus and the subject was required to remember the auditory and visual components (vocalization and accompanying facial gesture) across the following delays and to detect the change of any component in subsequent stimulus presentations with a button press. In half the trials, the nonmatching stimulus was presented as the second stimulus (S2 nonmatch) and, in the other half of the trials, a matching stimulus intervened and the nonmatch occurred as the third stimulus (S3 nonmatch). The example shown here depicts an audio nonmatch trial where only the vocalization changed in the nonmatch stimulus (A1V1→A2V1) but the face component remained the same.

we divided the number of wrong-press errors that occurred in the trials of the same sample stimulus by the number of the nonmatch conditions and assigned each portion evenly to compute the success rate for each nonmatch condition. The trials aborted due to breaking eye fixation (1.61% of total trials) were not included in the calculation of the success rates.

The reaction time (RT) was defined as time from the video onset to the button-press response. However, for the audio nonmatch conditions, the RT was also calculated from the audio onset to estimate the processing time of auditory mismatch information.

Analysis of neural data. To separate activity related to the button response from stimulus-related activity during the nonmatch period, we used a 700 ms window during the nonmatch stimulus, which was long enough to capture the neural response to the late audio components, but shorter than the fastest mean RT of our subjects (Fig. 3*B*). In addition, we tested button press-related activation by comparing activity between the ±50 ms window from button press and another 100 ms window preceding it. For the neurons that had significant activity related to the button press (17 neurons; t test, $p < 0.05$), we used a window 150 ms shorter than the fastest mean RT among the nonmatch conditions instead.

To analyze neuronal responses during the nonmatch period, we applied the following regression model. This model included a set of dummy variables corresponding to the auditory and visual components of the nonmatch stimulus, as well as the nonmatch type (i.e., audio, video, and AV nonmatches).

$$SR = a_0 + a_1 S + a_2 A + a_3 V + a_4 NM_{\text{A-AV}} + a_5 NM_{\text{V-AV}}, \quad (1)$$

where SR denotes the spike rate during the nonmatch period, and a_0–a_5, regression coefficients. S indicates the sample stimulus type (0, A1V1; 1, A2V2), and A and V indicate the auditory component (0, A1; 1, A2) and the visual component (0, V1; 1, V2) of the nonmatch stimulus, respectively. Since the nonmatch type was a categorical variable that had three classes (audio, video, and AV nonmatches), two dummy variables, $NM_{\text{A-AV}}$ and $NM_{\text{V-AV}}$, were used to represent it. $NM_{\text{A-AV}}$ measured the differential effect between the audio and AV nonmatches (0, AV nonmatch; 1, audio nonmatch) and $NM_{\text{V-AV}}$, between the video and AV nonmatches (0, AV nonmatch; 1, video nonmatch).

The significance of the nonmatch type (i.e., the combined effect of $NM_{\text{A-AV}}$ and $NM_{\text{V-AV}}$) was tested with a partial F test. To compute the marginal variance explained by $NM_{\text{A-AV}}$ and $NM_{\text{V-AV}}$, we also applied the following reduced model, which did not include $NM_{\text{A-AV}}$ and $NM_{\text{V-AV}}$.

$$SR = a_0 + a_1 S + a_2 A + a_3 V. \quad (2)$$

Then, the F statistic was computed as follows:

$$F^* = \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F}, \quad (3)$$

where $SSE(R)$ and $SSE(F)$ refer to the sum of squared errors for the reduced model (Eq. 2) and the full model (Eq. 1), respectively, and df_R and df_F indicate the degrees of freedom associated with the reduced model and the full model.

For those neurons that showed a significant effect of the nonmatch type, the effects of the auditory and visual component changes were tested further with another regression model:

$$SR = a_0 + a_1 A_\Delta + a_2 V_\Delta + a_3 A_\Delta V_\Delta, \quad (4)$$

where SR denotes the spike rate during the match or the nonmatch period, and a_0–a_3, regression coefficients. A_Δ and V_Δ indicated whether the auditory and visual components of the nonmatch were changed from the sample movie or not, respectively (0, unchanged; 1, changed). $A_\Delta V_\Delta$ was an interaction term. This model is equivalent to a two-way ANOVA model with A_Δ and V_Δ as factors. Note that the auditory component change (A1→A2 or A2→A1) occurred during the nonmatch period in both audio and AV nonmatch conditions and the visual component change (V1→V2 or V2→V1), in the video and AV nonmatch conditions. Neither of them occurred during the match period.

After the models were fit, t tests were performed to determine the statistical significance of each regression coefficient. The effects of some independent variables in the regression models were compared based on the standardized regression coefficients (SRCs). The SRC of an independent variable is defined as $a_i (s_i/s_d)$, in which a_i denotes the raw regression coefficient of the independent variable and s_i and s_d the SDs of the independent variable and the dependent variable, respectively.

Results

Behavioral performance in the audiovisual nonmatch-to-sample task

The subjects were trained with an audiovisual nonmatch-to-sample task in which an audiovisual face-vocalization movie was presented as the sample and the subjects detected a nonmatching stimulus (Fig. 1). The nonmatch conditions included audiovisual stimuli in which the vocalization component of the face-vocalization track had been replaced (audio nonmatch), the facial gesture video track had been replaced (video nonmatch), or both the face and vocalization components of the sample movie had been replaced (AV nonmatch; Fig. 2). We ran 168 sessions of this task (83 sessions for Monkey P and 85 for Monkey T) and analyzed behavioral performance by examining the success rate and the reaction time.

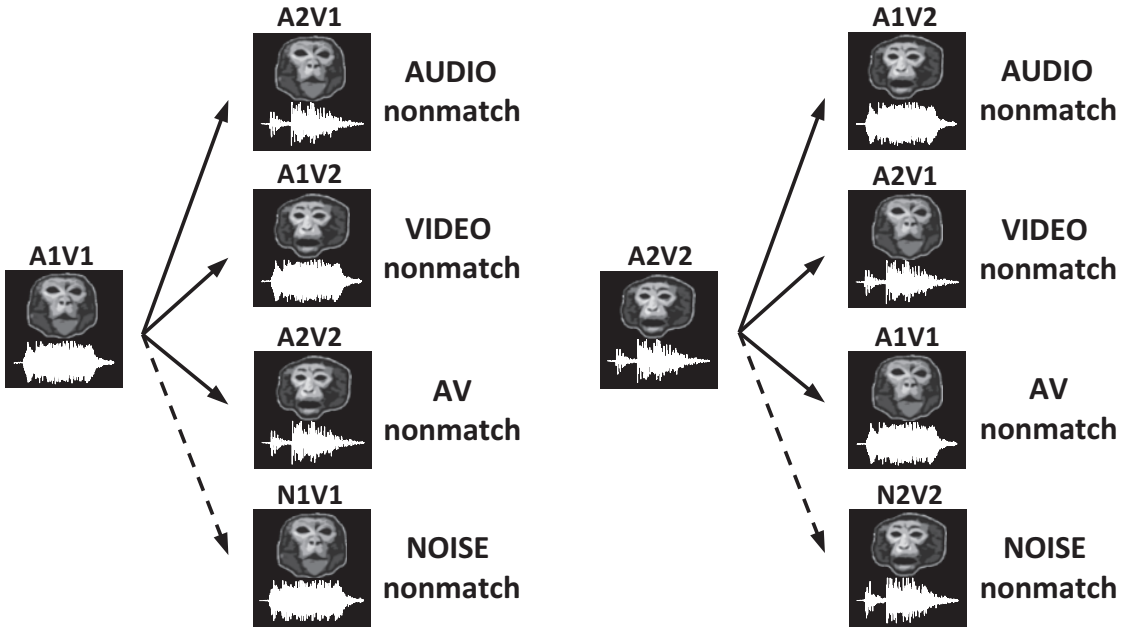

Figure 2. Nonmatch types. The types of nonmatch conditions that occurred in the audiovisual nonmatch-to-sample task are illustrated with the sample vocalization movie (A1V1 or A2V2). In the neurophysiology recording experiments, there were three types of nonmatch stimuli (audio, video, and AV nonmatches), which were created by interchanging the auditory (A*n*) and visual (V*n*) components between the two sample vocalization movies (A1V1 and A2V2). A fourth nonmatch stimulus (noise nonmatch) was used in the behavioral studies and was created by replacing the vocalization with a noise sound stimulus that had the same temporal envelope as the sample.

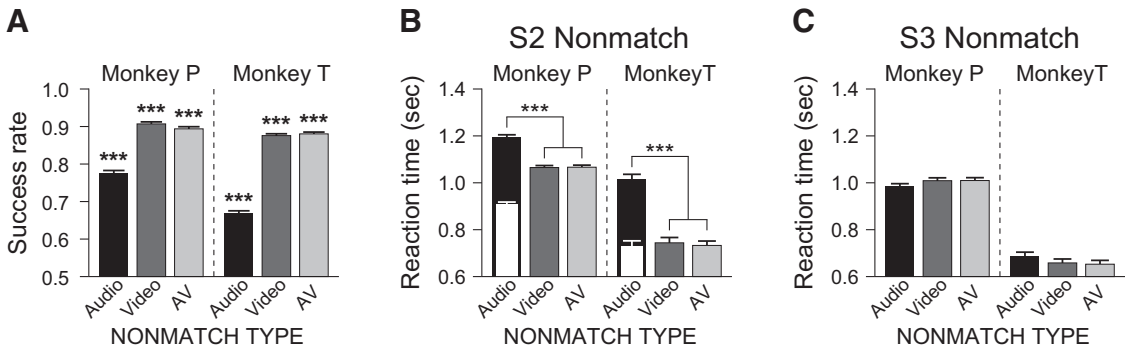

Figure 3. Behavioral performance. ***A***, The performance of each subject is shown as percentage correct by each nonmatch trial type (audio, black; video, dark gray; audiovisual or AV, light gray) during the neurophysiological recordings. ***B***, ***C***, The reaction times (RTs) are shown for trials when the nonmatch occurred as the second stimulus (S2 nonmatch) and as the third stimulus after a match stimulus (S3 nonmatch), respectively. The white bars in ***B*** are the RTs recalculated from the audio onset. Error bars indicate SEM; ***$p < 0.001$.

Both subjects performed the task above chance level in all conditions (χ^2 test, $p < 0.001$; Fig. 3*A*). Comparing between conditions, we found that their success rates in the audio nonmatch condition were significantly lower than the other nonmatch conditions (χ^2 test, $p < 0.001$). Although previous studies have shown that performance of auditory discrimination by nonhuman primates is reduced compared with visual discrimination (Goldman and Rosvold, 1970; Fritz et al., 2005), it was important for us to confirm that this was not due to a general deficiency in acoustic processing or insufficient training for auditory discrimination. Therefore, we tested the auditory discrimination ability of our subjects with a second discrimination task in which the audio component of the face-vocalization stimuli was replaced with not only a different vocalization but also a noise sound that is easier to distinguish (noise nonmatch; see Materials and Methods). This discrimination test was performed in an identical manner as the main study. Monkey P was tested with audio, AV, and noise nonmatch conditions for 86 sessions, and Monkey T, with audio, video, AV, and noise nonmatch conditions for 31 sessions. The results were comparable to those of the main study (Fig. 4). Both subjects performed the task well above chance level in all nonmatch conditions (χ^2 test, $p < 0.001$). As in the main study, the performance in the audio nonmatch condition (i.e., vocalization change) was lower (χ^2 test, $p < 0.001$), but the success rate in the noise nonmatch condition was higher than that in the audio nonmatch (χ^2 test, $p < 0.001$) and as good as the performance in the video and AV nonmatch conditions in both subjects. Both the audio and noise nonmatch conditions require detection of a change in the auditory component of the audiovisual stimulus. If our subjects were impaired in general auditory processing or did not know how to respond to the auditory component change, the same low success rate should have been observed in both noise nonmatch and audio nonmatch conditions. Since this was not the case, we concluded that the low performance in the audio nonmatch condition might simply reflect the difficulty in discriminating between the vocalization stimuli. In fact, these results are similar to a recent study that examined auditory discrimination in monkeys with a variety of sound types and found that

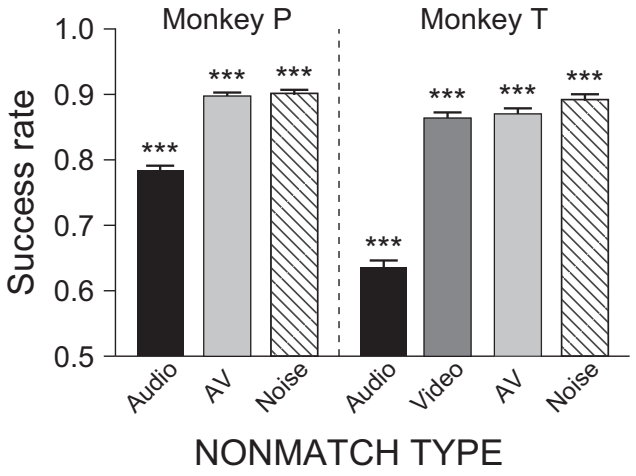

Figure 4. Performance in the behavioral task with the noise nonmatch stimulus. The auditory discrimination performance of both subjects was better with stimuli that were acoustically dissimilar to the vocalization (i.e., noise nonmatch, striped bars), compared with the vocalization–vocalization discrimination in the audio nonmatch (black bars). Error bars indicate SEM; $^{***}p < 0.001$.

noise or pure tones were discriminated from monkey vocalizations better than other sound types (Scott et al., 2013).

We also analyzed the RT of the subjects. During the S2 nonmatch period, the RTs were significantly different between the subjects and across the nonmatch conditions (two-way ANOVA; Fig. 3*B*), which was unexpected. First, the RTs of Monkey P were longer than those of Monkey T in all nonmatch conditions (*F* test, $p < 0.001$). In our task, the subjects were not required to position their responding hands at a particular location at the beginning of a trial. We observed that Monkey P retracted its hand from the button after making a response and reached out again in the next trial, whereas Monkey T held the hand near the button at all times. Therefore, Monkey P's longer RTs were partly due to longer arm travel distance. Second, the RT in the audio nonmatch condition was significantly slower than those in the other two nonmatch conditions (*F* test, $p < 0.001$). This could be due to the fact that, in macaque vocalizations, the auditory component (vocalization) naturally lags behind the onset of the visual component (mouth movement) by hundreds of milliseconds (160–404 ms, 283 ms on average in our stimuli; Fig. 1*A*), as discussed in previous studies (Ghazanfar et al., 2005; Chandrasekaran and Ghazanfar, 2009). In other words, subjects might be slower in detecting the audio change in the audio nonmatch condition simply because the vocalization occurred later than the video. To fairly compare the time taken from stimulus changes to button responses, we recalculated the RT from audio onset for the audio nonmatch condition (Fig. 3*B*, white bars). Compared with this recalculated RT, Monkey P actually took longer to respond in the video and AV nonmatch conditions than it did to the audio change in the audio nonmatch condition (*t* test, $p < 0.001$) and Monkey T showed no difference (*t* test, $p > 0.69$). This suggests that the slow RT in the audio nonmatch condition could be due to the relatively late audio onset rather than the longer processing of the audio mismatch.

The RTs in the S3 nonmatch period were shorter than those in the S2 nonmatch period overall (*t* test, $p < 0.001$; Fig. 3*C*). Furthermore, there was no significant difference in the RT across the nonmatch conditions such as was observed in the S2 nonmatch period (*F* test, $p = 0.92$). This is not surprising since the third stimulus was always a nonmatch and thus required a button press. In fact, in our task, the match/nonmatch decision was required only for the second stimulus that was presented after the sample. The second delay and S3 nonmatch were used so that a behavioral response would be required on every trial. Because a decision was only required for the second stimulus and the forthcoming behavioral response was predictable from the second delay, the neural activity during the S3 nonmatch period will not be considered in the current paper but will be discussed in a separate manuscript.

Neurophysiological responses of VLPFC neurons to face-vocalization stimuli during the sample and delay period

We recorded the activity of 215 neurons from VLPFC (114 from Monkey P, 101 from Monkey T; Fig. 5), while the subjects performed the task with audio, video, and AV nonmatch conditions (Fig. 6*A*). First, we assessed the responsiveness of VLPFC neurons during our audiovisual task by comparing the activity during the sample and first delay periods with the spontaneous baseline firing rate. The baseline firing rate was estimated from a 500 ms period before the sample stimulus onset. Some neurons not only responded differently from baseline during the sample period, but also maintained differential activity over the first delay period (Fig. 6*B*). This delay period activity is likely related to the maintenance of the sample stimulus in memory, as previously suggested by sustained activity during the delay period in other working memory tasks (Fuster, 1973; Rosenkilde et al., 1981; Miller et al., 1996; Fuster et al., 2000; Freedman et al., 2001; Plakke et al., 2013b). In total, 58.1% (= 125/215) of the recorded neurons showed significantly different activity during the sample period compared with baseline as did the same proportion of neurons (58.1%) during the first delay period (*t* test, $p < 0.05$; Fig. 7*A*). The proportions of the neurons that had greater or smaller activity than baseline were similar to each other during the sample period [52.8% (= 66/125) and 47.2% (= 59/125), respectively]. During the first delay period, however, 37.6% (= 47/125) showed elevated activity and 62.4% (= 78/125) showed reduced activity (χ^2 test, $p < 0.05$). We also compared the sample and delay activity between the two sample audiovisual movies to determine whether selectivity played a role in the responses. Such differential responses were found in 23.7% (= 51/215) of neurons during the sample period and 12.1% (= 26/215) during the first delay period, indicating some selectivity even with a small stimulus set (*t* test, $p < 0.05$; Fig. 7*B*).

We hypothesized that activity related to working memory would be diminished during the second delay period, since the next stimulus was always a nonmatch and there was no need for the subjects to keep the sample stimulus in memory to achieve a correct response. As expected, the proportion of neurons with elevated/suppressed activity decreased from 58.1% (= 125/215; first delay period) to 43.7% (= 94/215; second delay period). In addition, only 40.4% (= 19/47) of the neurons with elevated activity and 65.4% (= 51/78) with suppressed activity during the first delay showed the same behavior during the second delay. The percentage of neurons with stimulus-selective activity during the second delay period was 14.0% (= 30/215), which was not significantly different from that during the first delay period.

VLPFC activity related to working memory during stimulus comparison

Previous working memory studies that were performed with visual or auditory stimuli found that neurons in the lateral prefrontal cortex responded differently to the test stimulus depending on whether it was a match or a nonmatch (Miller et al., 1991, 1996;

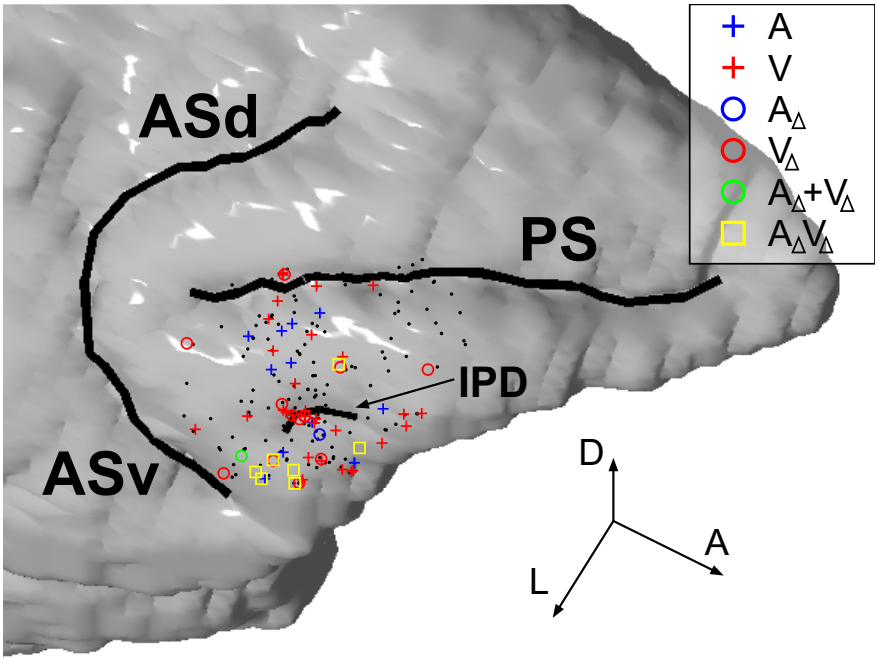

Figure 5. The locations of neurons that had significant activity in the audiovisual task are shown on a reconstruction of Monkey T's brain based on MRI. The neurons recorded from Monkey P were overlaid onto the reconstruction based on histologically confirmed anatomical landmarks. Color-coded dots are shown where activity was significantly modulated by the auditory component (A; A1 or A2) or the visual component (V; V1 or V2) of the vocalization movies (Eq. 1) and the auditory component change (A_Δ; A1⟶A2 or A2⟶A1) or the visual component change (V_Δ; V1⟶V2 or V2⟶V1) in Equation 4. $A_\Delta + V_\Delta$ indicates significant main effects of both A_Δ and V_Δ, and $A_\Delta V_\Delta$, a significant interaction. Black dots are neurons with no such effects. The three arrow lines indicate the directions of the anteroposterior (A), dorsoventral (D), and mediolateral (L) axes, respectively, and their lengths correspond to 5 mm each. PS, principal sulcus; ASd, dorsal arcuate sulcus; ASv, ventral arcuate sulcus; IPD, inferior prefrontal dimple.

Plakke et al., 2013b). Some neurons responded more strongly to stimuli that matched the sample than when the same stimuli were nonmatching, referred to as match enhancement, and some responded less to matching than nonmatching stimuli, referred to as match suppression. These match/nonmatch-related modulations indicate that the neurons are involved in comparing the match and nonmatch stimuli with the remembered sample. To see how many of our VLPFC neurons responded differently to the match and nonmatch stimuli, we first compared the activity of the neurons between the match period and the AV nonmatch period in which the identical stimulus was presented. We tested all recorded neurons with a two-way ANOVA, using the stimulus type (A1V1 or A2V2) and the task epochs (match or nonmatch) as the factors. Thirty-six (16.7%) neurons were stimulus selective, 34 (15.8%) neurons distinguished between match and nonmatch stimuli, and 10 (4.7%) neurons showed a significant interaction between the factors (F test, $p < 0.05$). Among those 34 neurons that responded differently to match and nonmatch stimuli, most (29/34) showed a reduction in the neuronal response to match stimuli while a few (5/34) showed enhancement.

In addition to the AV nonmatch stimuli (A1V1 or A2V2), however, our task also includes the audio and video nonmatch stimuli that have one component in common with the sample but are still considered nonmatching. Therefore, we must consider whether a neuron responded to the audio and video nonmatches in the same way as it did to the AV nonmatch, to determine whether the neuron differentiated matches and nonmatches. For that reason, we excluded neurons from this analysis that showed a significantly different response in the audio or video nonmatch condition compared with the AV nonmatch condition. This left 21 neurons that distinguished between matching and nonmatching stimuli. Among those 21 neurons, 4 showed match enhancement and 17 showed match suppression. An example neuron with match suppression is shown in Figure 6. This neuron responded strongly to all three nonmatch types, regardless of the modality of the changed component (Fig. 6*C*). This effect was not related to the button-press response, since activity related to the button response was excluded from the analysis window (see Materials and Methods) and the suppression in the match period occurred much earlier than the time of button press in this neuron (Fig. 6*D*).

We further examined how the activity of these VLPFC neurons is correlated with the match and nonmatch decisions of our subjects, by comparing the neuronal response on success and error trials. Our subjects sometimes mistook a match stimulus for a nonmatch and pressed the button (wrong-press error) or they did not press the button because they incorrectly judged the test stimulus to be a match when it was actually a nonmatch (missed-press error). In those error trials, activity of some neurons was modulated not by the actual matching/nonmatching status of the incoming stimuli, but by the subjective judgment, or decision, of the monkeys. Therefore, even though the presented stimuli were identical, the neuronal activity was different between wrong-press errors and match responses and so was the activity between missed-press errors and nonmatch responses (Fig. 6*E*). Among the 21 neurons with match enhancement or match suppression, 6 were excluded due to insufficient number of error trials, but 9 of the remaining 15 showed significantly different responses between success and error trials during the match and nonmatch periods (t test; $p < 0.05$).

Response of VLPFC neurons to the different nonmatch types

We were most interested in whether and how VLPFC neurons responded to the different nonmatch conditions. Many neurons modulated their activity according to one or more variables in our regression model, which allowed us to differentiate effects of particular stimuli, or the auditory and visual components of the audiovisual movie stimuli as well as effects of task variables such as the nonmatch type (Eq. 1; see Materials and Methods). In some neurons, the activity during the nonmatch period was simply modulated by a particular sensory component of the nonmatch stimuli. For example, the neuron shown in Figure 8*A* increased its activity whenever the presented stimulus included the face movie from the second sample movie, V2. The fact that the neuron is responsive to a visual component is confirmed by grouping and comparing trials according to the component included in the match/nonmatch stimuli (Fig. 8*B*). There was no difference in response between conditions that contained A1 and those that included A2, but there was a significant increase for those conditions that included V2, compared with the ones that included V1. From the regression analysis, 30 (14.0%) neurons had a signifi-

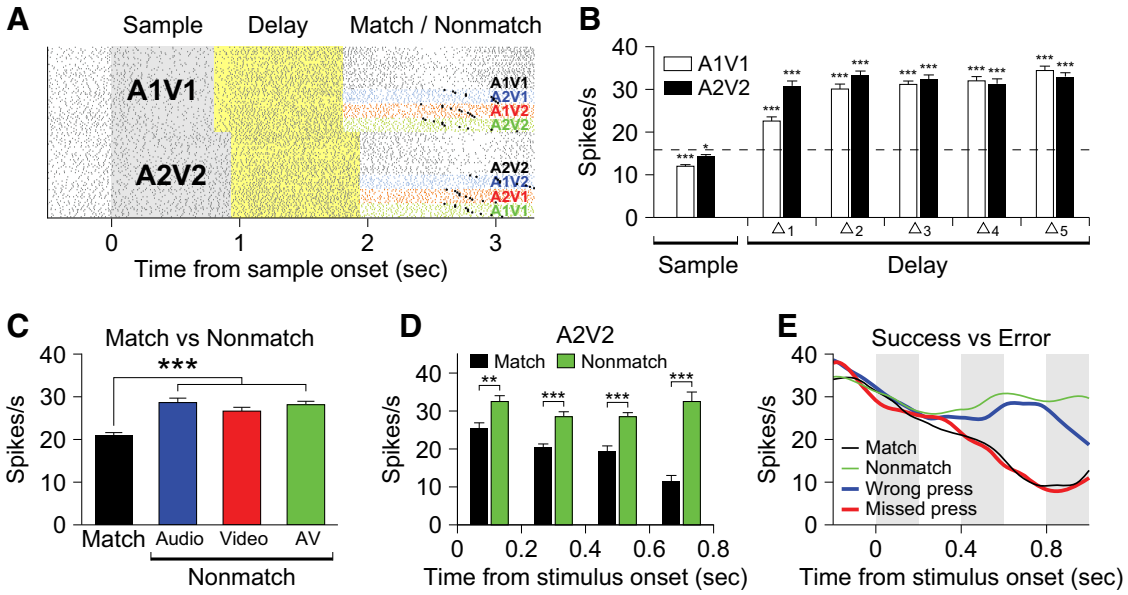

Figure 6. An example neuron with sustained delay activity and match suppression. ***A***, A raster plot of the activity of a single neuron during the audiovisual nonmatch-to-sample task. Each raster line corresponds to a single trial and each tick represents a single spike. Trials are regrouped according to match/nonmatch conditions and color coded as follows: black, sample/match; blue, audio nonmatch; red, video nonmatch; green, AV nonmatch. The asterisks on the rasters of nonmatch trials indicate the time of button presses. ***B***, Sustained delay activity. The activity of the same neuron during the sample and delay periods is compared with the baseline activity (the dotted line). The size of each bin in the delay period (Δ_n) is 200 ms. ***C***, The response of this same neuron is significantly decreased to the match stimulus compared with the nonmatch stimuli, while there is no difference among the nonmatch stimuli. The bars are color coded with the same scheme as in ***A***. ***D***, The same neuron's response to stimulus A2V2 when it appeared as a match (black bars) and elicited suppression compared with the response as a nonmatch stimulus (green bars). The result for A1V1 was similar except that the match activity decreased after 200 ms. ***E***, Comparison of neural activity between success and error trials in this single neuron. Neural activity during missed nonmatch responses (missed-press errors) is similar to correct match responses where the button press is withheld. Conversely, wrong presses during the match period resemble correct button presses during the nonmatch stimulus. Error bars indicate SEM; $***p < 0.001$; $**p < 0.01$; $*p < 0.05$.

cant modulation by visual components (V*n*) and 12 (5.6%) by auditory components (A*n*; *t* test, $p < 0.05$; Fig. 9). The proportions of these two neuron groups were significantly different (χ^2 test, $p < 0.01$), indicating that more neurons were responsive to the visual than the auditory components of the vocalization movies.

The activity of some neurons was related to the nonmatch type. For example, the neuron shown in Figure 8*C* increased its activity during the nonmatch period for the video nonmatch (red) and AV nonmatch (green) conditions. Another example neuron shown in Figure 8*E* exhibited an increase for the audio nonmatch (blue) conditions only. Note that this nonmatch type effect was not stimulus specific: it was not dependent on either a particular sample stimulus or a particular auditory or visual component (Fig. 9) but upon the rule of which component of the nonmatch stimulus had changed from the sample. We identified 32 neurons with a significant effect of the nonmatch type (*F* test, $p < 0.05$; Eq. 3).

We tested these nonmatch-type neurons further with another regression model to investigate how their activity was modulated by the change of each sensory component in the nonmatch stimuli (Eq. 4; see Materials and Methods). This analysis clearly revealed that some neurons responded to the change of one sensory component while others were modulated by particular combinations of the auditory and visual changes. For example, the activity of the neuron in Figure 8*C* was enhanced when the visual component changed from the sample stimulus in the video and AV nonmatch conditions (V1→V2 or V2→V1). However, this activity modulation was unrelated to the auditory component change, since the response elicited by the neuron was not different between both nonmatch conditions, whether the auditory component was changed (AV nonmatch) or not (video nonmatch; Fig. 8*D*). On the other hand, the neuron in Figure 8*E* increased its activity when the auditory component switched but the visual component remained unchanged (A1V1→A2V1 or A2V2→A1V2; Fig. 8*F*), which cannot be accounted for by either auditory component change or visual component change alone. These results indicate that some VLPFC neurons are sensitive to the change of one sensory component from the sample stimulus (unisensory component change neuron), while other neurons are sensitive to the changes of both auditory and visual components

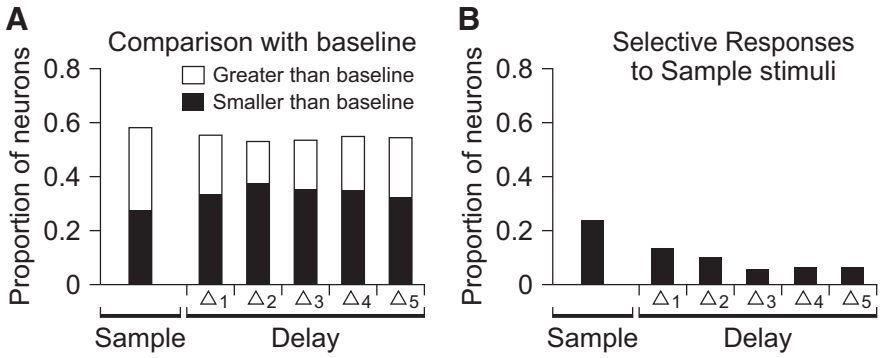

Figure 7. Neurons significantly active during the sample and delay period. ***A***, The proportions of neurons in which activity was significantly different from spontaneous activity during the sample and the delay period (*t* test, $p < 0.05$). The size of each bin in the delay period (Δ_n) is 200 ms. ***B***, The proportions of neurons that showed selective activity in either the sample or the delay period for one of the two sample movies tested (*t* test, $p < 0.05$).

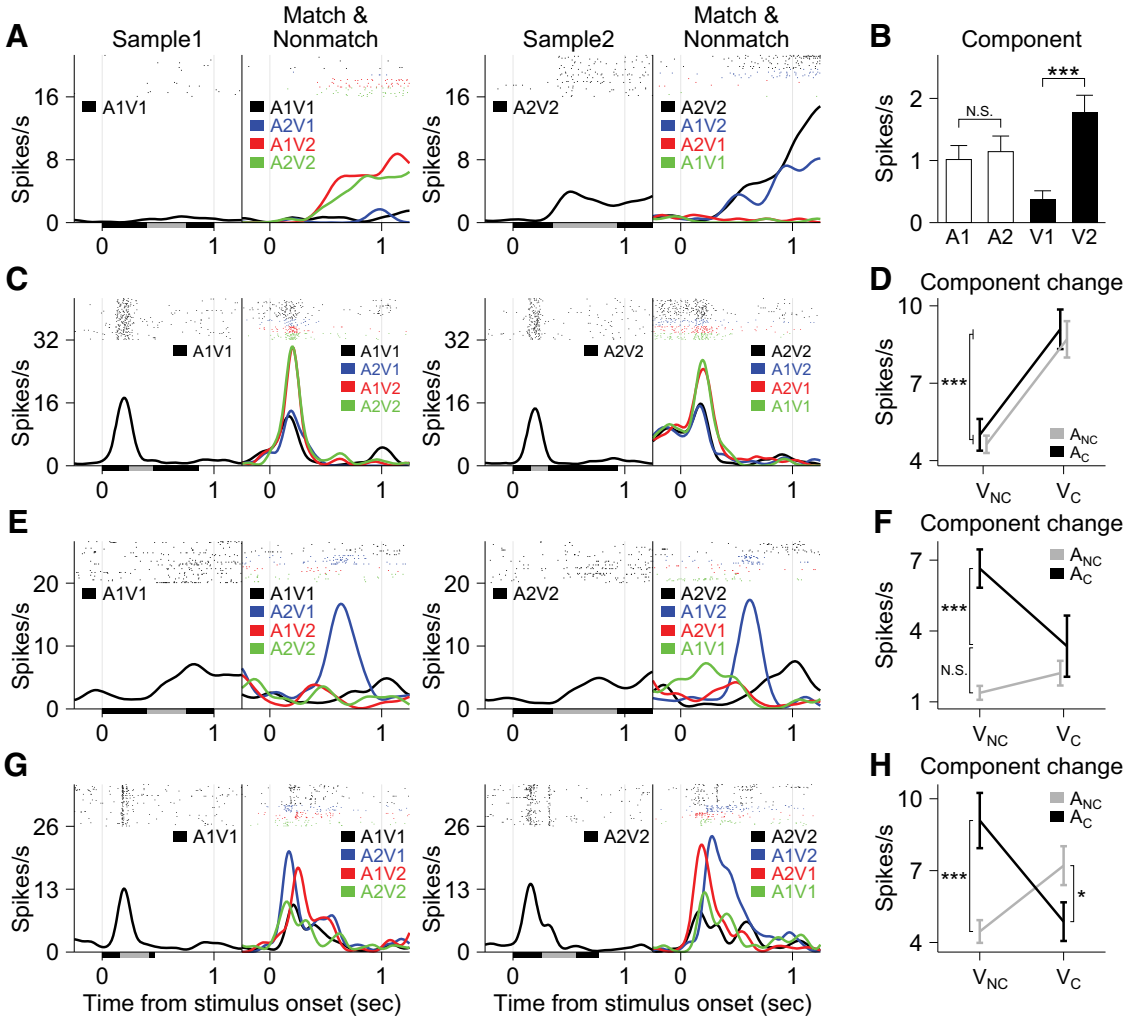

Figure 8. Four example neurons that are responsive to stimulus or component change(s) during the nonmatch periods. ***A***, ***C***, ***E***, ***G***, Spike-density functions and rasters are colored as follows: black, sample/match; blue, audio nonmatch; red, video nonmatch; green, AV nonmatch. Black and gray bars on the abscissae indicate the duration of the video and audio components of the sample stimuli, respectively. ***A***, ***B***, A neuron that responds to a particular visual component (V2) of the audiovisual stimuli. The activity during the match and nonmatch periods is sorted according to the auditory or visual component included in the stimuli and presented in ***B***, which shows the spike rate was greater for all stimuli containing V2. ***C***, ***D***, A neuron that had a significant change in firing for a unisensory (visual) component change between the samples and the nonmatches (V1→V2 or V2→V1). ***D***, The activity during the match and nonmatch periods is regrouped according to whether the auditory and visual components of the stimuli are repeated or changed. ***E–H***, Two neurons are shown which have responses that are dependent on the change status of both sensory components (multisensory component change neurons). ***E*** and ***F*** show a neuron with an increase in activity during the audio nonmatch where the auditory component changes but the visual component does not and ***G*** and ***H*** depict a neuron that responded significantly to both auditory and visual component changes but only when one component changes at a time. ***F*** and ***H*** are in the same format as ***D***. Error bars indicate SEM; ***$p < 0.001$; *$p < 0.05$. A_{NC}, audio no-change; A_C, audio change; V_{NC}, video no-change; V_C, video change.

(multisensory component change neuron). The neuron shown in Figure 8, *G* and *H*, is another multisensory component change neuron, which increased its activity when either the auditory or visual component changed alone (i.e., in the audio and video nonmatch conditions), but not when both or none of the components changed (i.e., in the AV nonmatch and match conditions).

The effect of the multiple component changes can be represented by either the main effects of both auditory and visual component changes (A_Δ and V_Δ, respectively) or their interaction ($A_\Delta V_\Delta$) in our model (Eq. 4). Of 32 nonmatch-type neurons, we found 8 (= 25.0%) multisensory component change neurons (1 neuron with the significant main effects of both A_Δ and V_Δ; 7 neurons with the interaction effect; Fig. 10). In addition, 14 (= 43.8%) were unisensory component change neurons and 10 neurons (= 31.3%) showed neither the effect of A_Δ nor V_Δ. Note that most of the unisensory component change neurons were sensitive to the visual component change (1 with the main effect of A_Δ alone; 13 with the main effect of V_Δ alone). We also noted that the neurons that were responsive to the component change(s), especially the multisensory component change neurons, were mostly located near the inferior prefrontal dimple (IPD) or lateral to IPD areas (Fig. 5), a location where previous studies have noted overlapping representation of face-, vocalization-, and multisensory-responsive neurons (Sugihara et al., 2006; Romanski and Averbeck, 2009).

Discussion

In this study, we investigated the activity of VLPFC neurons while nonhuman primates remembered and discriminated complex, natural audiovisual stimuli. Our task is a novel nonmatch-to-sample paradigm, which employs vocalizations and their accompanying facial gestures and requires using information from both auditory and visual stimuli simultaneously, as we do during face-to-face communication. VLPFC neurons were active during several phases of the audiovisual task, and more than half of the

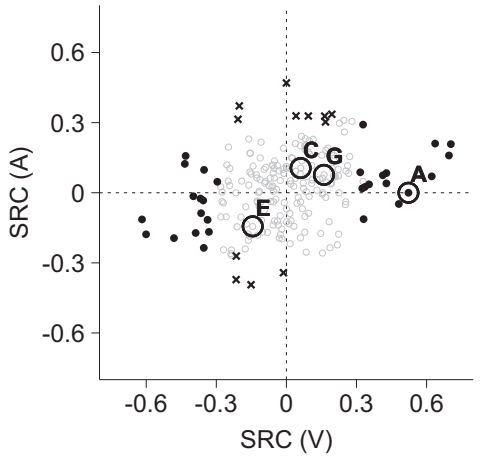

Figure 9. Effects of auditory and visual components. The standardized regression coefficients (SRCs) for the auditory components (A) in Equation 1 are plotted against the SRCs for the visual components (V). The positive SRC indicates that the neuron elicited a greater response to A2 (or V2) compared with A1 (or V1); the negative SRC conveys the opposite. Filled circles (black) and black crosses correspond to the neurons with a significant effect of the visual components or the auditory components, respectively. Open circles (gray) are the neurons with no significant effect of the auditory or visual components. A, C, E, and G indicate data points from the example neurons shown in Figure 8*A*, *C*, *E*, and *G*. Note that C, E, and G neurons did not show a significant effect of the auditory or visual components.

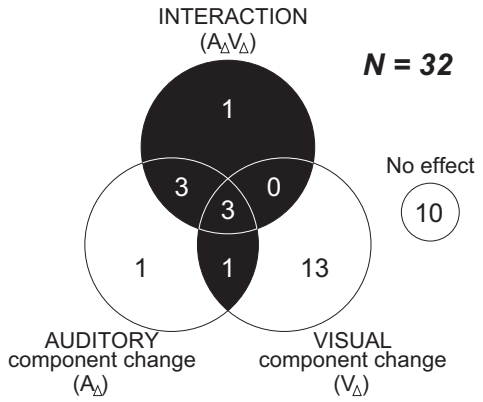

Figure 10. Summary of neuronal responses to the component change(s) during the nonmatch period. The neurons with an effect of the nonmatch type are grouped by the result of the regression analysis (Eq. 4). A_Δ indicates the change of the auditory component between the sample and the test stimulus (A1⟶A2 or A2⟶A1) and V_Δ indicates the change of the visual component (V1⟶V2 or V2⟶V1). $A_\Delta V_\Delta$ is the interaction. The black area indicates multisensory component change neurons that had significant main effects of both A_Δ and V_Δ, or a significant interaction effect.

neurons were responsive during the sample period and often maintained this activity during the delay period. In the nonmatch period, VLPFC neurons exhibited both stimulus and task-related activity with some neurons responding to the particular face or the vocalization presented, while other neurons showed evidence of WM for the nonmatch target. We also found that responses to the nonmatch stimuli were complex and multisensory, in the sense that the response to one auditory or visual component change was often dependent on a change in the other component. This suggests that information about auditory and visual stimuli maintained in WM is integrated across modalities in VLPFC and that VLPFC is an important site for audiovisual integration and memory.

Working memory and cross-modal information integration in VLPFC

Previous investigations have shown that the lateral prefrontal cortex is important for WM functions and that there may be areal segregation for spatial and nonspatial information processing (Wilson et al., 1993; Belger et al., 1998; Courtney et al., 1998; Hoshi et al., 2000; Rämä et al., 2004; Romanski, 2004; Nee et al., 2013). Evidence suggests that dorsolateral prefrontal cortex (DLPFC) may be specialized for processing visual and auditory spatial cues (Watanabe, 1981; Funahashi et al., 1989; Quintana and Fuster, 1992; Kikuchi-Yorioka and Sawaguchi, 2000; Constantinidis et al., 2001), whereas VLPFC is specialized for processing nonspatial features of stimuli (e.g., color and shape) and the identity of faces and vocalizations (Rosenkilde et al., 1981; Wilson et al., 1993; O Scalaidhe et al., 1997; Rämä and Courtney, 2005). Our results are in accordance with these findings in that VLPFC neurons showed task-related activity such as sustained delay activity and match/nonmatch preference in our WM task using face and vocalization stimuli and therefore support the idea that VLPFC has a role in object WM.

In addition, our findings indicate that VLPFC is involved in integrating information from different modalities during WM. During nonmatch detection, the activation of some VLPFC neurons was not dependent on whether the changed component was auditory or visual but whether the test stimulus was matching or nonmatching. This suggests that modality-specific information about the stimulus is combined and used to guide a decision based on the task rule. Moreover, the discovery of multisensory component change neurons, which attend to information from both auditory and visual channels and can encode whether the stimulus change occurs in both channels or in only one of them, provides direct evidence of cross-modal integration in VLPFC. While prefrontal neurons have previously been reported to integrate within-modal information in WM (Quintana and Fuster, 1993; Baker et al., 1996; Rao et al., 1997; D'Esposito et al., 1998; Rainer et al., 1998; Prabhakaran et al., 2000; Sala and Courtney, 2007), the role of the prefrontal cortex in cross-modal integration during WM has been investigated in only a few studies. Some showed that single prefrontal neurons represented behaviorally meaningful auditory and visual associations (Fuster et al., 2000; Zhang et al., 2014) and others demonstrated that neurons in DLPFC are engaged in parallel processing of auditory and visual spatial information during WM maintenance (Kikuchi-Yorioka and Sawaguchi, 2000; Artchakov et al., 2007). Our study is the first to demonstrate that VLPFC neurons perform nonspatial mnemonic functions across modalities.

The audiovisual task used in the current study to examine audiovisual WM is novel in its reliance on dynamic, naturalistic, and species-specific face-vocalization stimuli across all task epochs. Previous experiments on WM for communication stimuli presented face and vocal stimuli separately (Schumacher et al., 1996; Kamachi et al., 2003; Crottaz-Herbette et al., 2004) or let subjects choose which modality to remember based on saliency or by instruction (Parr, 2004; Rämä and Courtney, 2005). Our audiovisual paradigm allows us to examine how cross-modal integration of WM occurs using the same stimuli that comprise face-to-face communication and which VLPFC neurons have been shown to prefer (O Scalaidhe et al., 1997; Romanski and Goldman-Rakic, 2002; Tsao et al., 2008).

Auditory discrimination in nonhuman primates

Some recent studies, which showed that macaques had relatively poor performance in an auditory match-to-sample paradigm

(Scott et al., 2012, 2013), concluded that macaques have a "limited form" of auditory WM. This conflicts with our current results where we have noted sustained delay activity and match suppression/enhancement in VLPFC neurons when audiovisual comparisons including auditory discrimination were required. These are common neuronal correlates of WM (Miller et al., 1991, 1996; Plakke et al., 2013b) and we have shown for the first time that they are present with compound audiovisual stimuli. One difference from previous tasks is that we used a nonmatch-to-sample paradigm in which the intervening, repeated stimulus was identical to the sample and was therefore less susceptible to retroactive interference than memoranda in the match-to-sample task (Scott et al., 2012, 2013). This modification may account for some of the differences in neuronal activity and behavior.

We found that vocalization discrimination performance was inferior to face discrimination performance. Previous studies have shown that macaques are slow in learning auditory delayed match- or nonmatch-to-sample tasks and performance remains poor even after prolonged training (Wright et al., 1990; Fritz et al., 2005; Ng et al., 2009; Scott et al., 2012), compared with visual discrimination performance. This difference between auditory and visual discrimination may be related to behavior in their natural habitat, where most communication exchanges occur in close proximity, or face-to-face, since macaques are not arboreal like New World monkeys. Some researchers report that rhesus monkeys rely on vocalizations in social interaction only 5–32% of the time (Altmann, 1967; Partan, 2002), suggesting that facial cues may be more prominent in social interactions. Our analysis revealed that there are fewer VLPFC neurons responsive to the vocalization component of the audiovisual stimulus, or its change, compared with the number of neurons responsive to the face or the change of the face. These results parallel findings of greater numbers of visual unimodal responses compared with auditory unimodal responses in VLPFC (Sugihara et al., 2006; Diehl and Romanski, 2014), and in the monkey amygdala, another face-vocalization integration area, where many neurons are robustly responsive to faces (Gothard et al., 2007; Kuraoka and Nakamura, 2007; Mosher et al., 2010), but fewer neurons respond to corresponding vocalizations (Kuraoka and Nakamura, 2007). Moreover, even human subjects fare better at face recognition than voice recognition in many situations (Colavita, 1974; Hanley et al., 1998; Joassin et al., 2004; Calder and Young, 2005; Sinnett et al., 2007). Nonetheless, when we substituted an acoustically dissimilar noise stimulus as the nonmatch instead of a mismatching vocalization, discrimination performance was greatly improved, suggesting that vocal discrimination difficulties may therefore be due to feature similarity rather than mnemonic capacity (Scott et al., 2013).

Comparative similarities between VLPFC and human IFG

It has been suggested that VLPFC and the human IFG are homologous because they share similar cytoarchitectonic features (Petrides and Pandya, 2002) and may have some functional similarities as well (Romanski, 2012). The human IFG, including Broca's area, has long been linked with speech and language processes (Geschwind, 1970; Dronkers et al., 2007; Grodzinsky and Santi, 2008). Similarly, VLPFC has a discrete auditory region where neurons respond to complex stimuli including species-specific vocalizations (Romanski and Goldman-Rakic, 2002). These neurons tend to respond to multiple vocalizations on the basis of acoustic features (Romanski et al., 2005), which is consistent with a notion that VLPFC is part of the ventral auditory processing stream that analyzes the features of auditory objects (Belin et al., 2000; Binder et al., 2000; Scott et al., 2000; Zatorre et al., 2004). Moreover, macaque VLPFC is adjacent to vocal motor control regions (Petrides et al., 2005) and receives afferents from temporal lobe auditory regions (Petrides and Pandya, 1988; Barbas, 1992; Hackett et al., 1999; Romanski et al., 1999a, b; Saleem et al., 2014).

The human IFG is also involved in maintaining and integrating auditory and visual WM, which is necessary during communication. Human imaging studies have revealed that IFG is coactivated during auditory and visual verbal WM tasks or during delayed recognition tasks for unfamiliar faces and voices (Schumacher et al., 1996; Crottaz-Herbette et al., 2004; Rämä and Courtney, 2005), suggesting a role in nonspatial information processing independent of the stimulus modality, similar to the role of the neurons localized to the macaque VLPFC here. Therefore, our study provides support for the hypothesis that VLPFC and the human IFG share not only similar cytoarchitecture but also a similar function in the integration and mnemonic processing of cross-modal communication information.

References

Altmann SA (1967) The structure of primate social communication. In: Social communication among primates, pp 325–362. Chicago: University of Chicago Press.

Artchakov D, Tikhonravov D, Vuontela V, Linnankoski I, Korvenoja A, Carlson S (2007) Processing of auditory and visual location information in the monkey prefrontal cortex. Exp Brain Res 180:469–479. CrossRef Medline

Baker SC, Frith CD, Frackowiak RS, Dolan RJ (1996) Active representation of shape and spatial location in man. Cereb Cortex 6:612–619. CrossRef Medline

Barbas H (1992) Architecture and cortical connections of the prefrontal cortex in the rhesus monkey. Adv Neurol 57:91–115. Medline

Belger A, Puce A, Krystal JH, Gore JC, Goldman-Rakic P, McCarthy G (1998) Dissociation of mnemonic and perceptual processes during spatial and nonspatial working memory using fMRI. Hum Brain Mapp 6:14–32. CrossRef Medline

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000) Voice-selective areas in human auditory cortex. Nature 403:309–312. CrossRef Medline

Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET (2000) Human temporal lobe activation by speech and nonspeech sounds. Cereb Cortex 10:512–528. CrossRef Medline

Calder AJ, Young AW (2005) Understanding the recognition of facial identity and facial expression. Nat Rev Neurosci 6:641–651. CrossRef Medline

Calvert GA, Hansen PC, Iversen SD, Brammer MJ (2001) Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. Neuroimage 14:427–438. CrossRef Medline

Campanella S, Belin P (2007) Integrating face and voice in person perception. Trends Cogn Sci 11:535–543. CrossRef Medline

Chandrasekaran C, Ghazanfar AA (2009) Different neural frequency bands integrate faces and voices differently in the superior temporal sulcus. J Neurophysiol 101:773–788. CrossRef Medline

Colavita FB (1974) Human sensory dominance. Percept Psychophys 16:409–412. CrossRef

Constantinidis C, Franowicz MN, Goldman-Rakic PS (2001) The sensory nature of mnemonic representation in the primate prefrontal cortex. Nat Neurosci 4:311–316. CrossRef Medline

Courtney SM, Petit L, Maisog JM, Ungerleider LG, Haxby JV (1998) An area specialized for spatial working memory in human frontal cortex. Science 279:1347–1351. CrossRef Medline

Crottaz-Herbette S, Anagnoson RT, Menon V (2004) Modality effects in verbal working memory: differential prefrontal and parietal responses to auditory and visual stimuli. Neuroimage 21:340–351. CrossRef Medline

D'Esposito M, Aguirre GK, Zarahn E, Ballard D, Shin RK, Lease J (1998) Functional MRI studies of spatial and nonspatial working memory. Brain Res Cogn Brain Res 7:1–13. CrossRef Medline

Diehl MM, Romanski LM (2014) Responses of prefrontal multisensory

neurons to mismatching faces and vocalizations. J Neurosci 34:11233–11243. CrossRef Medline

Dronkers NF, Plaisant O, Iba-Zizen MT, Cabanis EA (2007) Paul Broca's historic cases: high resolution MR imaging of the brains of Leborgne and Lelong. Brain 130:1432–1441. CrossRef Medline

Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2001) Categorical representation of visual stimuli in the primate prefrontal cortex. Science 291:312–316. CrossRef Medline

Fritz J, Mishkin M, Saunders RC (2005) In search of an auditory engram. Proc Natl Acad Sci U S A 102:9359–9364. CrossRef Medline

Funahashi S, Bruce CJ, Goldman-Rakic PS (1989) Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. J Neurophysiol 61:331–349. Medline

Fuster JM (1973) Unit activity in prefrontal cortex during delayed-response performance: neuronal correlates of transient memory. J Neurophysiol 36:61–78. Medline

Fuster JM, Bodner M, Kroger JK (2000) Cross-modal and cross-temporal association in neurons of frontal cortex. Nature 405:347–351. CrossRef Medline

Geschwind N (1970) The organization of language and the brain. Science 170:940–944. 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

Ghazanfar AA, Chandrasekaran C, Morrill RJ (2010) Dynamic, rhythmic facial expressions and the superior temporal sulcus of macaque monkeys: implications for the evolution of audiovisual speech. Eur J Neurosci 31:1807–1817. CrossRef Medline

Goldman PS, Rosvold HE (1970) Localization of function within the dorsolateral prefrontal cortex of the rhesus monkey. Exp Neurol 27:291–304. CrossRef Medline

Gothard KM, Battaglia FP, Erickson CA, Spitler KM, Amaral DG (2007) Neural responses to facial expression and face identity in the monkey amygdala. J Neurophysiol 97:1671–1683. CrossRef Medline

Grodzinsky Y, Santi A (2008) The battle for Broca's region. Trends Cogn Sci 12:474–480. CrossRef Medline

Hackett TA, Stepniewska I, Kaas JH (1999) Prefrontal connections of the parabelt auditory cortex in macaque monkeys. Brain Res 817:45–58. CrossRef Medline

Hanley JR, Smith ST, Hadfield J (1998) I recognise you but I can't place you: an investigation of familiar-only experiences during tests of voice and face recognition. Q J Exp Psychol 51:179–195. CrossRef

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

Hoshi E, Shima K, Tanji J (2000) Neuronal activity in the primate prefrontal cortex in the process of motor selection based on two behavioral rules. J Neurophysiol 83:2355–2373. Medline

Joassin F, Maurage P, Bruyer R, Crommelinck M, Campanella S (2004) When audition alters vision: an event-related potential study of the cross-modal interactions between faces and voices. Neurosci Lett 369:132–137. 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

Kamachi M, Hill H, Lander K, Vatikiotis-Bateson E (2003) "Putting the face to the voice": matching identity across modality. Curr Biol 13:1709–1714. CrossRef Medline

Kikuchi-Yorioka Y, Sawaguchi T (2000) Parallel visuospatial and audiospatial working memory processes in the monkey dorsolateral prefrontal cortex. Nat Neurosci 3:1075–1076. CrossRef Medline

Kuraoka K, Nakamura K (2007) Responses of single neurons in monkey amygdala to facial and vocal emotions. J Neurophysiol 97:1379–1387. CrossRef Medline

Lee H, Noppeney U (2011) Physical and perceptual factors shape the neural mechanisms that integrate audiovisual signals in speech comprehension. J Neurosci 31:11338–11350. CrossRef Medline

McGurk H, MacDonald J (1976) Hearing lips and seeing voices. Nature 264:746–748. CrossRef Medline

Miller EK, Li L, Desimone R (1991) A neural mechanism for working and recognition memory in inferior temporal cortex. Science 254:1377–1379. CrossRef Medline

Miller EK, Erickson CA, Desimone R (1996) Neural mechanisms of visual working memory in prefrontal cortex of the macaque. J Neurosci 16:5154–5167. 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

Mosher CP, Zimmerman PE, Gothard KM (2010) Response characteristics of basolateral and centromedial neurons in the primate amygdala. J Neurosci 30:16197–16207. CrossRef Medline

Murray MM, Wallace MT (2012) The neural bases of multisensory processes. Boca Raton, FL: CRC.

Nee DE, Brown JW, Askren MK, Berman MG, Demiralp E, Krawitz A, Jonides J (2013) A meta-analysis of executive components of working memory. Cereb Cortex 23:264–282. CrossRef Medline

Ng CW, Plakke B, Poremba A (2009) Primate auditory recognition memory performance varies with sound type. Hear Res 256:64–74. CrossRef Medline

Noppeney U, Ostwald D, Werner S (2010) Perceptual decisions formed by accumulation of audiovisual evidence in prefrontal cortex. J Neurosci 30:7434–7446. 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

Parr LA (2004) Perceptual biases for multimodal cues in chimpanzee (Pan troglodytes) affect recognition. Anim Cogn 7:171–178. CrossRef Medline

Partan SR (2002) Single and multichannel signal composition: facial expressions and vocalizations of rhesus macaques (Macaca mulatta). Behaviour 139:993–1027. CrossRef

Paulesu E, Frith CD, Frackowiak RS (1993) The neural correlates of the verbal component of working memory. Nature 362:342–345. CrossRef Medline

Petrides M, Pandya DN (1988) Association fiber pathways to the frontal cortex from the superior temporal region in the rhesus monkey. J Comp Neurol 273:52–66. 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

Petrides M, Cadoret G, Mackey S (2005) Orofacial somatomotor responses in the macaque monkey homologue of Broca's area. Nature 435:1235–1238. CrossRef Medline

Plakke B, Romanski LM (2014) Auditory connections and functions of prefrontal cortex. Front Neurosci 8:199. CrossRef Medline

Plakke B, Diltz MD, Romanski LM (2013a) Coding of vocalizations by single neurons in ventrolateral prefrontal cortex. Hear Res 305:135–143. CrossRef Medline

Plakke B, Ng CW, Poremba A (2013b) Neural correlates of auditory recognition memory in primate lateral prefrontal cortex. Neuroscience 244:62–76. CrossRef Medline

Prabhakaran V, Narayanan K, Zhao Z, Gabrieli JD (2000) Integration of diverse information in working memory within the frontal lobe. Nat Neurosci 3:85–90. CrossRef Medline

Preuss TM, Goldman-Rakic PS (1991) Myelo- and cytoarchitecture of the granular frontal cortex and surrounding regions in the strepsirhine primate Galago and the anthropoid primate Macaca. J Comp Neurol 310:429–474. CrossRef Medline

Quintana J, Fuster JM (1992) Mnemonic and predictive functions of cortical neurons in a memory task. Neuroreport 3:721–724. CrossRef Medline

Quintana J, Fuster JM (1993) Spatial and temporal factors in the role of prefrontal and parietal cortex in visuomotor integration. Cereb Cortex 3:122–132. CrossRef Medline

Rainer G, Asaad WF, Miller EK (1998) Memory fields of neurons in the primate prefrontal cortex. Proc Natl Acad Sci U S A 95:15008–15013. CrossRef Medline

Rämä P, Courtney SM (2005) Functional topography of working memory for face or voice identity. Neuroimage 24:224–234. CrossRef Medline

Rämä P, Poremba A, Sala JB, Yee L, Malloy M, Mishkin M, Courtney SM

(2004) Dissociable functional cortical topographies for working memory maintenance of voice identity and location. Cereb Cortex 14:768–780. CrossRef Medline

Rao SC, Rainer G, Miller EK (1997) Integration of what and where in the primate prefrontal cortex. Science 276:821–824. CrossRef Medline

Romanski LM (2004) Domain specificity in the primate prefrontal cortex. Cogn Affect Behav Neurosci 4:421–429. 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]:10717–10724. CrossRef Medline

Romanski LM, Averbeck BB (2009) The primate cortical auditory system and neural representation of conspecific vocalizations. Annu Rev Neurosci 32:315–346. 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, Averbeck BB, Diltz M (2005) Neural representation of vocalizations in the primate ventrolateral prefrontal cortex. J Neurophysiol 93:734–747. CrossRef Medline

Rosenkilde CE, Bauer RH, Fuster JM (1981) Single cell activity in ventral prefrontal cortex of behaving monkeys. Brain Res 209:375–394. CrossRef Medline

Sala JB, Courtney SM (2007) Binding of what and where during working memory maintenance. Cortex 43:5–21. CrossRef Medline

Saleem KS, Miller B, Price JL (2014) Subdivisions and connectional networks of the lateral prefrontal cortex in the macaque monkey. J Comp Neurol 522:1641–1690. CrossRef Medline

Scalaidhe SP, Wilson FA, Goldman-Rakic PS (1999) Face-selective neurons during passive viewing and working memory performance of rhesus monkeys: evidence for intrinsic specialization of neuronal coding. Cereb Cortex 9:459–475. CrossRef Medline

Schumacher EH, Lauber E, Awh E, Jonides J, Smith EE, Koeppe RA (1996) PET evidence for an amodal verbal working memory system. Neuroimage 3:79–88. CrossRef Medline

Scott BH, Mishkin M, Yin P (2012) Monkeys have a limited form of short-term memory in audition. Proc Natl Acad Sci U S A 109:12237–12241. CrossRef Medline

Scott BH, Mishkin M, Yin P (2013) Effect of acoustic similarity on short-term auditory memory in the monkey. Hear Res 298:36–48. CrossRef Medline

Scott SK, Blank CC, Rosen S, Wise RJ (2000) Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123:2400–2406. CrossRef Medline

Sinnett S, Spence C, Soto-Faraco S (2007) Visual dominance and attention: the Colavita effect revisited. Percept Psychophys 69:673–686. 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

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

Watanabe M (1981) Prefrontal unit activity during delayed conditional discriminations in the monkey. Brain Res 225:51–65. 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

Wright AA, Shyan MR, Jitsumori M (1990) Auditorysame/different concept learning by monkeys. Anim Learn Behav 18:287–294. CrossRef

Zatorre RJ, Bouffard M, Belin P (2004) Sensitivity to auditory object features in human temporal neocortex. J Neurosci 24:3637–3642. CrossRef Medline

Zhang Y, Hu Y, Guan S, Hong X, Wang Z, Li X (2014) Neural substrate of initiation of cross-modal working memory retrieval. PLoS One 9:e103991. CrossRef Medline