Area Spt in the Human Planum Temporale Supports Sensory-Motor Integration for Speech Processing

Gregory Hickok,[1] **Kayoko Okada,**[1] **and John T. Serences**[2]

[1]*Department of Cognitive Sciences and Center for Cognitive Neuroscience, University of California, Irvine,* [2]*Department of Psychology, University of California, San Diego, California*

Hickok G, Okada K, Serences JT. Area Spt in the human planum temporale supports sensory-motor integration for speech processing. *J Neurophysiol* 101: 2725–2732, 2009. First published February 18, 2009; doi:10.1152/jn.91099.2008. Processing incoming sensory information and transforming this input into appropriate motor responses is a critical and ongoing aspect of our moment-to-moment interaction with the environment. While the neural mechanisms in the posterior parietal cortex (PPC) that support the transformation of sensory inputs into simple eye or limb movements has received a great deal of empirical attention—in part because these processes are easy to study in nonhuman primates—little work has been done on sensory-motor transformations in the domain of speech. Here we used functional magnetic resonance imaging and multivariate analysis techniques to demonstrate that a region of the planum temporale (Spt) shows distinct spatial activation patterns during sensory and motor aspects of a speech task. This result suggests that just as the PPC supports sensorimotor integration for eye and limb movements, area Spt forms part of a sensory-motor integration circuit for the vocal tract.

INTRODUCTION

Although most research on sensory-motor integration has focused on the neural mechanisms in posterior parietal cortex (PPC) that support interactions between vision and motor behaviors such as reaching and eye movements (Andersen 1997; Colby and Goldberg 1999; Cui and Andersen 2007), sensory-motor interactions are also critically important for speech-related behaviors including aspects of speech development (Doupe and Kuhl 1999; Hickok and Poeppel 2000, 2007; Hickok et al. 2000), sensory guidance of speech production (Guenther et al. 1998; Hickok 2000; Hickok and Poeppel 2007; Hickok et al. 2000; Warren et al. 2005), maintenance of parity between speech perception and production (Galantucci et al. 2006; Mattingly and Liberman 1988), and verbal working memory (Buchsbaum and D'Esposito 2008; Hickok et al. 2003; Jacquemot and Scott 2006). A region in the left posterior planum temporale (PT) in humans, area Spt, has been proposed as a critical node in a network that supports sensory-motor integration for speech and other vocal tract behaviors (Buchsbaum et al. 2001; Hickok and Poeppel 2000, 2004, 2007; Hickok et al. 2003; Okada and Hickok 2006; Pa and Hickok 2008). Area Spt exhibits properties parallel to those found in sensory-motor areas in the PPC of monkeys (Andersen 1997; Colby and Goldberg 1999; Cui and Andersen 2007). First, Spt has sensory-motor response properties, activating both during the auditory perception and covert production of speech (Buchsbaum et al. 2001, 2005b; Hickok et al. 2003; Okada and Hickok 2006). Second, Spt appears to be selective for the vocal tract effector system (Pa and Hickok 2008). Third, although Spt activates during speech functions, it is not speech selective: it responds to other vocal-tract related behaviors such as the perception and covert production (humming) of melodies (Hickok et al. 2003; Pa and Hickok 2008). Fourth, while Spt is responsive to speech stimulation, it is not critical for speech recognition (Hickok and Poeppel 2000, 2004, 2007), just as PPC sensory-motor areas are not critical for object recognition (Milner and Goodale 1995; Ungerleider and Mishkin 1982). Fifth, human cytoarchitectonic studies (Galaburda and Sanides 1980) and comparative studies in monkeys (Smiley et al. 2007) indicate that the posterior PT region is not part of unimodal auditory cortex. Finally, the posterior PT appears to be multi-sensory (Calvert and Campbell 2003; Calvert et al. 1997; Smiley et al. 2007). These similarities between PPC sensory-motor areas and Spt have led to the view that Spt supports some form of sensory-motor interaction that is critical to behaviors such as speech (Buchsbaum et al. 2001, 2005b; Hickok and Poeppel 2004; Hickok et al. 2003; Warren et al. 2005) and more precisely for motor behaviors that involve the vocal tract effector system (Hickok and Poeppel 2007; Pa and Hickok 2008).

Here we use functional magentic resonance imaging (fMRI) to test the hypothesis that subpopulations of neurons in Spt play specialized roles in sensory-motor integration, just as PPC sensory-motor areas contain subpopulations of cells with different sensory-motor weightings (e.g., ~33% were motor dominant, 26% visual dominant, and 37% sensory-motor in one study) (Sakata et al. 1995). Traditional fMRI analyses may miss such fine-grained distinctions because high-dimensional data sets are distilled into univariate estimates of the average response amplitude across all voxels within the region of interest (ROI). However, recently developed multivariate pattern classification methods exploit the fact that if some voxels contain more neurons of a particular persuasion (e.g., sensory vs. motor), they may exhibit a weak response preference. By pooling the output of many weakly selective voxels, it is possible to use machine-learning algorithms to distinguish characteristic voxel-by-voxel patterns of activation associated with different sensory-motor functions (Haxby et al. 2001; Kamitani and Tong 2005; Norman et al. 2006; Serences and Boynton 2007a,b). Establishing the existence of distinct sensory versus motor activation patterns would establish that

Address for reprint requests and other correspondence: G. Hickok, Center for Cognitive Neuroscience, Department of Cognitive Sciences, University of California, Irvine, CA 92697 (E-mail: greg.hickok@uci.edu) or to J. Serences, Department of Psychology, University of California, San Diego, CA 92093 (E-mail: jserences@ucsd.edu).

distinct subpopulations of neurons in Spt support sensory-motor integration for speech-related functions.

METHODS

Subjects

Twenty-two participants (10 females) between 18 and 35 yr of age were recruited from the University of California, Irvine (UCI), and received monetary compensation for their time. The volunteers were right-handed, native English speakers with normal or corrected-to-normal vision, no known history of neurological disease, and no other contraindications for MRI. Informed consent was obtained from each participant prior to participation in the study in accordance with UCI Institutional Review Board guidelines.

Materials and procedure

The data reported in this study were part of a larger experiment aimed at mapping responses to a range of sensory stimuli, including melodic sequences, noise bursts, auditory speech, and visual speech, each presented in blocks of 15 s. All stimulus types were randomly intermixed across the study and presented in equal ratios across the experiment and within each session (run).

Here we focus only on the auditory speech conditions. The stimuli were a set of "jabberwocky" sentences—sentences in which content words were replaced with nonsense words (e.g., "It is the glandor in my nedderop")—taken from a previous study (Hickok et al. 2003). There were three experimental conditions each presented in 15-s blocks (trials), "continuous speech": 15 s of listening to continuous speech (sets of sentences), "listen+rest": 3 s of listening to speech followed by 12 s of rest, "listen+rehearse": 3 s of listening to speech followed by 12 s of covert (subvocal) rehearsal of the heard stimuli (see Fig. 1). In addition, we included a null (rest) condition of the same 15-s duration. Each of six sessions (runs) contained two trials (15-s blocks) of each condition, including the null (rest) condition. A visual cue distinguished between the conditions: a fixation cross for the continuous listen condition that cued the subject to simply fixate and listen, a picture of an ear for the listen+rest condition that cued the

subject to listen and not rehearse after the offset of the auditory stimulus, and a picture of a mouth for the listen+rehearse condition that cued the subject to listen and then covertly rehearse the auditory stimulus until the end of the trial. These cues remained on the screen for the duration of the trial. The listen-rehearse paradigm has been used in previous experiments (Buchsbaum et al. 2001, 2005a,b; Hickok et al. 2003) and has been shown to drive activity in posterior planum temporale (Spt), which responds characteristically to both the listen and rehearse phases of the trial. This line of investigation assumes that covert rehearsal is a valid proxy for speech production. Evidence supporting this assumption comes from studies of speech production in more conventional naming tasks that demonstrate activity in this region (Graves et al. 2008; Levelt et al. 1998; Okada and Hickok 2006; Okada et al. 2003) and that show sensory (listening to words) and motor (naming pictures) overlap (Okada and Hickok 2006).

The experiment started with a short exposure session to familiarize subjects to all of the different experimental stimuli. Subjects were scanned during the exposure session to ensure they could comfortably hear the stimuli through the scanner noise and to acclimatize them to the fMRI environment. This was followed by five experimental sessions (runs). Each experimental session contained an equal number of trials (blocks) of each condition and a single scanning session was ~6 min long. Auditory stimuli were presented through an MR compatible headset, and stimulus delivery and timing were controlled using Cogent software (http://www.vislab.ucl.ac.uk/cogent_2000.php) implemented in Matlab 6 (Mathworks). To monitor subjects' attentiveness during the scans we presented occasional "oddball" stimuli (e.g., a speech stimulus presented in a female rather than male voice) to which subjects made a button press. These trials (3/session, 13%) were excluded from the subsequent analysis.

Scanning parameters

MRIs were obtained in a Philips Achieva 3T MR scanner (Philips Medical Systems, Andover, MA) fitted with an eight-channel RF receiver head coil, at the Research Imaging Center, UCI. We first collected a total of 620 EPI volumes over 5 sessions using Fast Echo EPI (sense reduction factor = 2.0, FOV = 220 × 180, matrix =
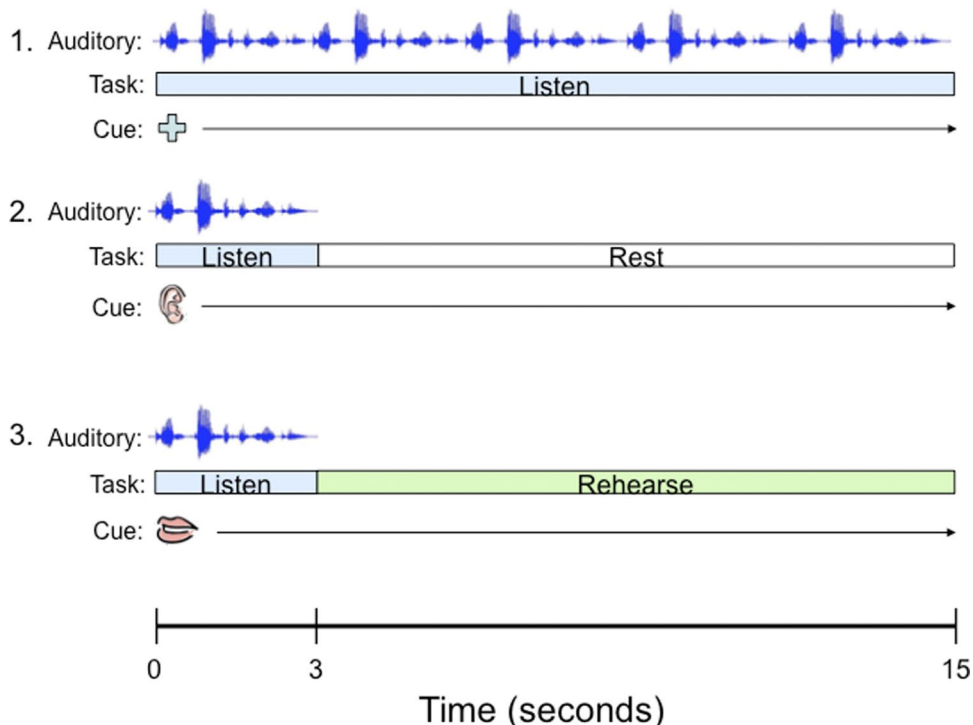

FIG. 1. Schematic diagram of the 3 conditions analyzed in the present study. *1*: the "continuous listen" condition. *2*: the "listen+rest" condition. And *3*: the "listen+rehearse condition.

112 × 112 mm , TR = 3.0s, TE = 25 ms, flip angle = 70°, size = 1.95 × 1.95 × 2 mm). After the functional scans, a high-resolution anatomical image was acquired with an MPRAGE pulse sequence in axial plane (matrix = 256 × 256 mm, TR = 8 ms, TE = 3.7 ms, flip angle = 8°, size = 1 × 1 × 1 mm).

Data analysis

The first three and the last images of each session were discarded prior to analysis. Preprocessing of the data were performed using Statistical Parametric Mapping (SPM5; Wellcome Department of Imaging Neuroscience, London, UK; www.fil.ion.ucl.ac.uk/spm) implemented in Matlab7 (Mathworks). First, motion correction was performed by creating a mean image from all of the volumes in the experiment and then realigning all volumes to that mean image using a six-parameter rigid-body transformation. Subsequent data analysis involved two stages. The first stage identified sensory-motor ROIs within the planum temporale (Spt) in individual subjects that were subsequently used in the second stage of analysis. The second stage used multivoxel pattern classification analysis to examine the spatial distribution of activation within the Spt ROI (see following text).

ROI identification analysis

First-level analysis was performed on each subject using AFNI software (Cox 1996). Due to the high anatomical variability of the posterior Sylvian region (Knaua et al. 2006), single-subject analysis was elected instead of group analysis. Images were smoothed with an isotropic 4-mm full-width half-maximum (FWHM) Gaussian kernel. Regression analysis was performed to find parameter estimates that best explained variability in the data. Each predictor variable representing the stimulus presentation time course for each event type was convolved with a standard hemodynamic response function (Boynton et al. 1996) and entered into the model along with six motion regressors.

As noted, our goal in this stage of the analysis was to identify area Spt in the posterior planum temporale. Spt is defined as a region within the left Sylvian fissure posterior to Heschl's gyrus that exhibits both auditory and motor-related response properties (Buchsbaum et al. 2005a,b; Hickok et al. 2003; Okada and Hickok 2006; Pa and Hickok 2008). This activation is strongly left dominant (Hickok et al. 2003). In the present study, sensory responsivity was measured using the contrast between the continuous speech condition and rest (the null rest blocks), while the motor-related response was measured using the contrast, speech+rehearse > speech+rest ($P < 0.001$, uncorrected). Thus in individual subjects, ROIs were defined by activations reflecting the conjunction of continuous speech > null rest blocks, and speech+rehearse > speech+rest that were located within the left planum temporale region (within the Sylvian fissure posterior to Heschl's gyrus), defined by coregistering each subject's activation maps with their own structural MRIs. Using these criteria, significant activations were found in 20 of the 22 subjects. ROIs in 14 of these 20 participants involved a sufficient number of voxels (≥10) to allow for pattern classification analysis (mean number of voxels ± SE: 70 ± 45). The analysis described in the following text was applied to this resulting dataset; however, a qualitatively (and statistically) similar pattern of results was obtained if all 20 subjects were included (see Supplemental Fig. S1[1]).

Multivoxel pattern analysis

This analysis focused on the continuous speech condition and the listen+rehearse condition. The rationale for using these two conditions is as follows. First, both of these conditions produce robust activations in the Spt ROI (see following text), and second, the trial structure is such that different time points within these conditions yield different predictions regarding processes that should be neurally distinguishable or not, providing a within-trial control for our classification analysis. Specifically, for the first 3 s following trial onset, the two conditions are similar in terms of their acoustic input properties. They diverge only during the last 12 s of the trial where one condition continues to involve sensory perception of speech and the other involves (covert) speech production without sensory speech stimulation (Fig. 1). Thus we expect increasing pattern discriminability as the trial progresses.

To test this hypothesis, we first normalized the time series from each voxel (using unsmoothed data) in each subject's Spt ROI on a run-by-run basis using a z-transform (to remove changes in mean signal intensity that occur between scanning runs). Then we segmented the normalized time series from each voxel within each Spt ROI into three separate temporal bins that each contained activation patterns measured on two successive TRs (0–3, 6–9, and 12–15 s post stimulus, where each number indicates the onset time of a volume acquisition). The pattern classification analysis was run separately on data from each temporal bin. Data from all but one run were extracted to form a "training" data set for the classification analysis; data from the remaining run were defined as a "test" set (note that we use the term "run" to refer to an entire 372-s data collection sequence so the training and test data sets were always independent). We then trained a support vector machine (Vapnik 1998) (the OSU-SVM implementation, downloaded from http://sourceforge.net/projects/svm/) based only on the training data and then used it to classify the task requirements (continuous speech versus listen+rehearse) on each trial in the test set (see e.g., Kamitani and Tong 2005; Serences and Boynton 2007a,b; Serences et al. 2009). Applying a SVM to the data set returns a weight for each voxel, such that the weighted sum of the values across all voxels is used to make the binary classification decision. Because zero is the decision boundary, positive sums indicate one choice, and negative sums the other. In the present experiment, positive weights were assigned to voxels that responded more during "listen" trials, and negative weights were assigned to voxels that responded more during "rehearse" trials; the most discriminating voxels of each class had the highest absolute values. To ensure that the classification algorithm was only using information about the spatially distributed voxel-by-voxel activation pattern—as opposed to amplitude differences between the conditions—we explicitly subtracted the mean activation level from each activation pattern before classification (this subtraction was carried out on a trial-by-trial basis, so the mean of the activation pattern associated with each trial was removed). This procedure was repeated using a 'hold-one-run-out' cross validation approach so that data from every scan were used as a test set in turn. Because each subject completed five runs/session, the overall classification accuracy for a subject was defined as the average classification accuracy across all five possible permutations of holding one run out as the test set and using the remaining runs as a training set. After computing classification accuracy within each of the three time bins separately for each of the subjects, we averaged the data across observers.

As a check to ensure that our analysis path was valid, we also verified that a data set composed of independent identically distributed (IID) noise did not yield above-chance classification accuracy when processed through the same analysis code (because there was no internally reliable signal). In addition, we repeated the analysis 10,000 times after randomly permuting the condition label assigned to each pattern to ensure that the code and/or the algorithm did not always produce above chance accuracy. (See Nichols and Holmes 2002 for a tutorial on permutation processes.) The dashed blue lines in Fig. 4 (and Supplemental Fig. S1) show classification accuracy on the upper and lower 5% of the 10,000 repetitions of the analysis using permuted trial labels, and all of our effects of interest survive these thresholds.

---

[1] The online version of this article contains supplemental data.

TABLE 1. *Talairach coordinates for the center of mass of the Spt ROI in each of 14 subjects that were entered into the pattern classification analysis*

	ROI. Area Spt		
S1	−47	−35	18
S2	−52	−40	22
S3	−52	−43	28
S4	−53	−50	15
S5	−49	−44	23
S6	−53	−39	22
S7	−51	−37	18
S8	−55	−47	22
S9	−49	−35	18
S10	−40	−41	14
S11	−61	−33	11
S12	−42	−32	16
S13	−40	−32	26
S14	−59	−47	17

ROI, region of interest; Spt, a region in the planum temporale.

RESULTS

Overall performance accuracy on the oddball detection task was 98%. This indicates that subjects were alert and attentive during the scanning procedure.

The Spt ROI localizer analysis (listen+rehearse > listen+rest *and* continuous listen > null rest blocks) detected reliable activation within the left posterior PT in 20 of the 22 subjects. Fourteen of these subjects had ROIs that contained a sufficient number of voxels ($\geq$10) to allow for pattern classification analysis. The results reported here focus on these 14 subjects. The location of sensory-motor ROIs were confirmed to be within the planum temporale (area Spt) in each subject based on coregistration of the functional images with a high-resolution MRI of each subject's own brain. The location of these activations in standardized space is presented in Table 1 and is displayed on a standardized brain image in Fig. 2 (*top row*). Activation foci in two individual subjects is shown in Fig. 2

(*bottom row*). The location of these ROIs is consistent with previous studies of area Spt (Buchsbaum et al. 2001, 2005b; Hickok et al. 2003).

The mean amplitude of the hemodynamic response for the listen+rehearse and the continuous listen conditions, collapsed across all voxels in the Spt ROIs, is presented in Fig. 3. An ANOVA carried out on these data confirmed that both conditions produced significant activations in the ROI [main effect of time point: $F(5,65) = 25.17$, $P = 0.001$] and that the activation levels for the two conditions differed during some time points [condition $\times$ time point interaction: $F(5,65) = 3.41$, $P = 0.008$]. During the first two time points, when the sensory-evoked response dominates the signal, the amplitudes are statistically identical ($P > 0.59$) for the two conditions. Beginning at time point 3, the curves begin to diverge with the listen+rehearse condition yielding greater amplitude [$t(13) = 2.49$, $P = 0.027$], presumably because the rehearsal process produces some additional activation in Spt which sums with the residual sensory response. The response in the listen+rehearse condition peaks and reaches its greatest difference from the continuous listen condition at time point 4 [$t(13) = 3.15$, $P = 0.008$] and then declines to equal the continuous listen condition in the last two time points [$t(13) = 1.19$, $P = 0.25$ and $t(13) = 0.08$, $P = 0.94$, respectively]. Because rehearsal continues for the duration of the block, this decline in amplitude presumably reflects the decay of the sensory component of the BOLD response. Note that our predictions regarding the evolution of pattern discriminability does not track the observed amplitude differences. Specifically we expect the greatest discriminability not at the peak of the amplitude difference between the two conditions, but at the end of the trial when the amplitude signals are not statistically different between the two conditions, i.e., when the signals are dominated by different functional sources (e.g., sensory vs. motor). It is this last time bin therefore that represents the strongest test of our hypothesis.
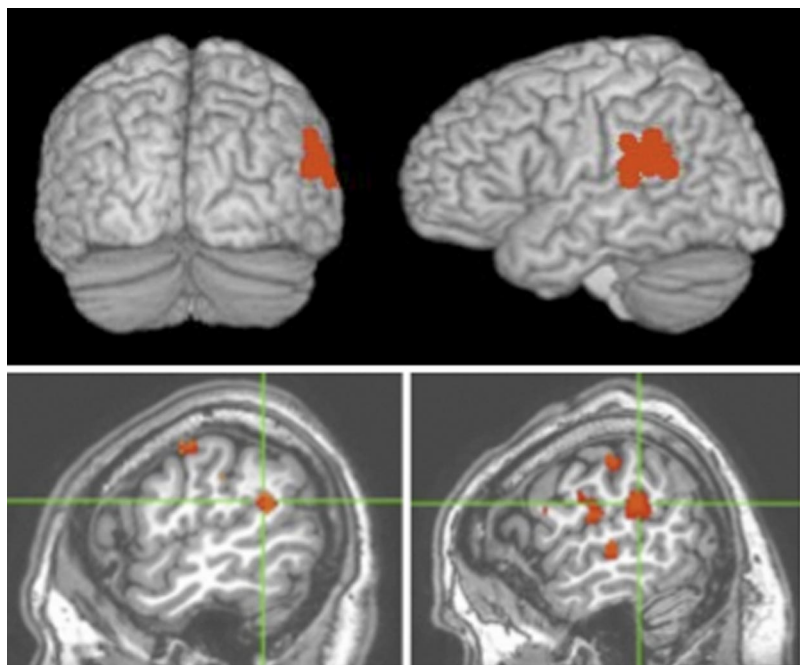

FIG. 2. *Top row*: location of the each subject's region of the planum temporale (Spt) region of interest (ROI) presented on a standardized 3-dimensensional (3D) brain image. Note that all activation foci were individually confirmed to be inside the posterior Sylvian fissure based on co-registration of the activation map onto each subject's own structural brain image. *Bottom row*: location of the Spt ROI (green crosshairs) in 2 subjects projected onto each subject's own anatomical magnetic resonance imaging (MRI) scan.
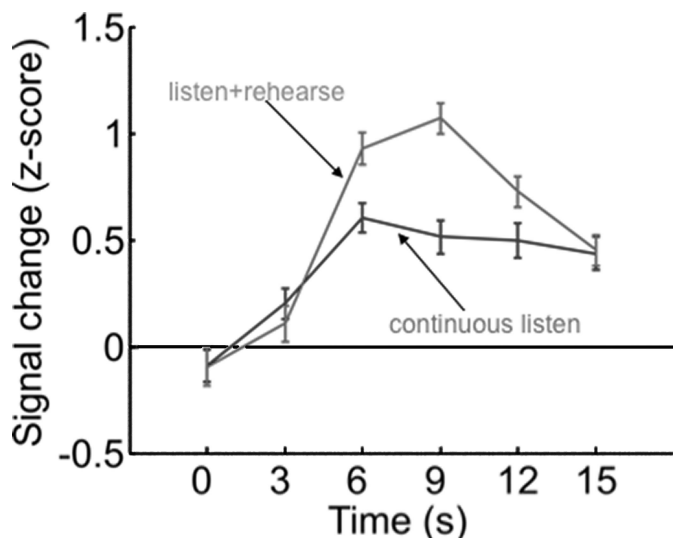
FIG. 3. Mean blood-oxygen-level-dependent (BOLD) signal amplitude time course (z-score) for the continuous listen and listen+rehearse conditions in the Spt ROI for 6 time points starting with trial onset. Speech stimuli were presented at time point 0, and trials ended at time point 12; time point 15 is the same as time point 0 in the next (random) trial and is included to show the continued evolution of the response function for these conditions.

To critically test the hypothesis that the Spt contains functionally distinct subdivisions of neurons, we next used a multivariate pattern classification analysis to determine if the voxel-by-voxel activation patterns within Spt discriminated speech perception (continuous listen condition) from speech production (listen+rehearse condition). In contrast to the time course of the mean BOLD amplitudes, which reconverged at the end of the trial, the activation patterns associated with each condition became more separable and pattern classification accuracy increased across the duration of the trial [1-way repeated-measures ANOVA, $F(2,26) = 6.7$, $P < 0.01$]. Specifically, classification accuracy was not significantly different from chance in the first time bin [0–3 s, $t(13) = 0.538$, $P = 0.60$] as expected. After this point, classification accuracy improves to above-chance levels [6–9 s, $t(13) = 3.58$, $P < 0.0034$] and reaches ~75% accuracy in the final time bin [$t(13) = 3.717$, $P < 0.0004$, Fig. 4; Supplemental Fig. S1].

Figure 5 shows a map of SVM weights within the area Spt ROI for two representative subjects; the relatively interdigitated distribution of positively and negatively weighted voxels implies that this region contains a mixture of neurons that are preferentially responsive to one cognitive operation over the other.

There are two potential problems with the above analysis. One is that the voxel selection procedure and the pattern classification analysis are not completely independent (Vul et al. 2009). The other is that visual cue for the listen+rehearse condition contained a namable object (mouth), whereas the continuous listen condition involved a simple fixation crosshair. It is conceivable that this visual cue difference is driving the pattern discrimination difference. To assess these possibilities, we used the even runs (sessions) to identify our ROI then ran the pattern classification analysis on data from only the odd runs (sessions). This was possible in 14 of our subjects using a slightly relaxed threshold of 0.005 (using only 2 runs to identify the ROIs necessarily reduces power). This split plot

approach removes any possible bias. In addition, not only did we (re)assess pattern classification accuracy on the continuous listen versus listen+rehearse, but we also assessed pattern classification accuracy on the continuous listen versus listen+rest condition that served as a control for the effects of the visual cue as this condition also contained a nameable cue (ear). We expected that continuous listen versus listen+rest would *not* result in above-chance classification.

The results of these analyses are presented in Fig. 6. Figure 6A shows the average hemodynamic response in the *odd* runs across subjects ($n = 14$) for three conditions, continuous listen, listen+rehearse, and listen+rest. Notice that in the first 6 s of the trial, which is dominated by the sensory response, signal amplitude is equivalent across conditions. After this point, the
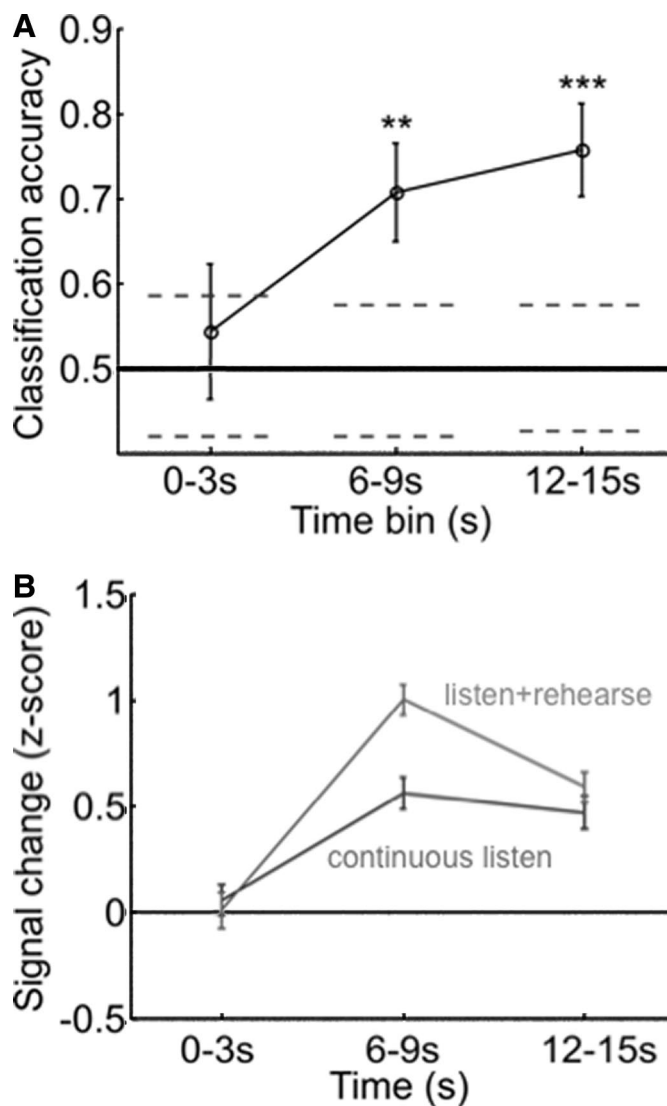

FIG. 4. *A*: pattern classification accuracy as a function of time within the trial. Data are organized into 3 bins corresponding to early, middle, and late portions of the trial. Chance classification accuracy = 0.50. Conditions are classified significantly better than chance in the 2nd and 3rd bins only, two asterisks, $P < 0.001$, three asterisks, $P < 0.0001$. Dashed lines indicate the upper and lower 5% of classification accuracies observed using a permutation test (see METHODS). *B*: BOLD amplitude for the 2 test conditions organized into the same time bins used in the pattern classification analysis for comparison. Note that classification accuracy does not track with amplitude differences between the 2 conditions.
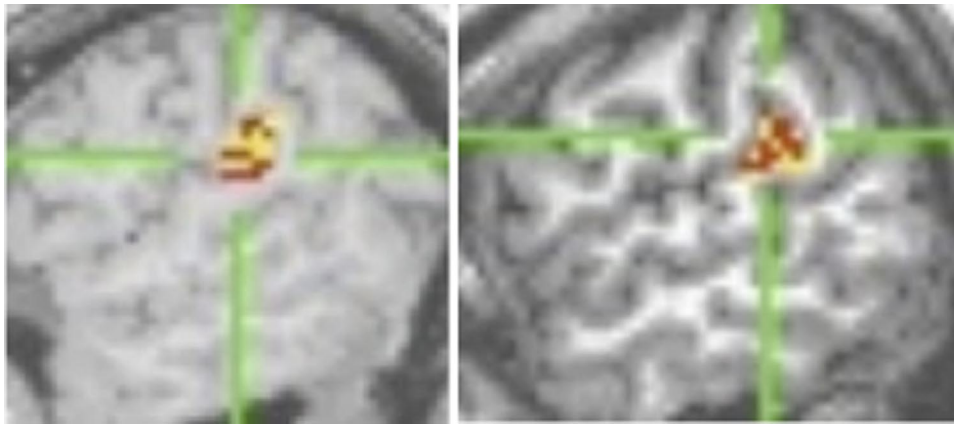
FIG. 5. Activity pattern within area Spt. Applying a SVM to the data set returns a weight for each voxel such that the weighted sum of the values across all voxels is used to make the binary classification decision. Because 0 is the decision boundary, positive sums indicate one choice, and negative sums the other. In the present experiment, positive weights were assigned to voxels that responded more during "listen" trials, and negative weights were assigned to voxels that responded more during "rehearse" trials; the most discriminating voxels of each class had the highest absolute values. Figure shows a map of SVM weights within the area Spt ROI for 2 representative subjects.

response in the three conditions diverge. In the continuous listen condition, the signal saturates and remains roughly the same for the balance of the trial. In the listen+rest condition, the signal drops off reflecting a return to baseline after speech stimulation is ended. In the listen+rehearse condition, the signal continues to increase, reaching its peak at 9 s presumably because the sensory response sums with the motor-related rehearsal response during this time window, then falls off during the final two time points, presumably reflecting the decay of the sensory response; the signal remains well above a resting baseline, however, due to the rehearsal. Again it is in this final phase of the trial when we expect to see maximal discriminability between the continuous listen and listen+rehearse conditions because the source of the activation, although equivalent in amplitude, derives from different underlying sources: sensory versus motor. Patten classification in the fully unbiased (odd run) dataset again confirmed this prediction

(Fig. 6B): classification accuracy was significantly above chance at the final time point [$t(13) = 3.017$, $P = 0.01$, 2-tailed, 0.03 after Bonferoni correction]. The same analysis carried out on the continuous listen versus listen+rest condition yielded no above-chance classifications (all P values > 0.19, 2-tailed, >0.50 after Bonferroni correction; Fig. 6C). We conclude from these analyses that pattern classification accuracy in the continuous listen versus listen+rehearse conditions is not an artifact of selection bias nor attributable to visual cue differences.

DISCUSSION

A region in the posterior planum temporale, area Spt, is activated both by the perception and production of speech (Buchsbaum et al. 2001, 2005a,b; Hickok et al. 2003; Okada and Hickok 2006). The present study showed that the pattern of
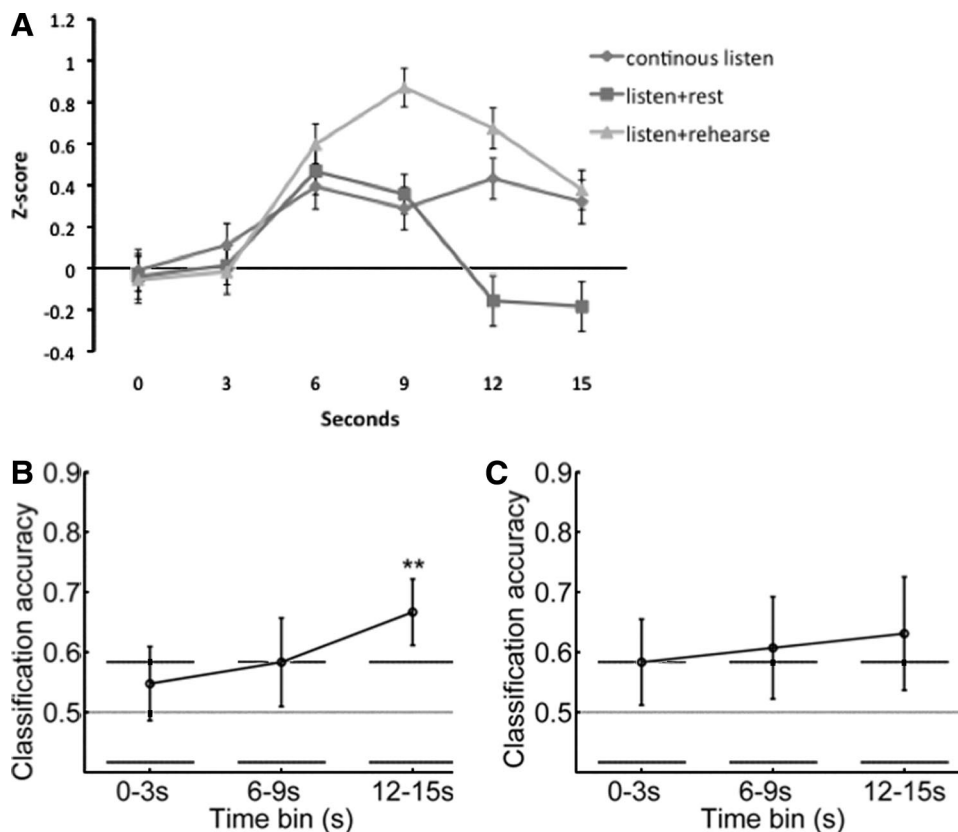

FIG. 6. BOLD amplitude and pattern classification accuracy in the split-plot analysis. A: BOLD amplitude in area Spt for the 3 conditions, continuous listen, listen+rehearse, and listen+rest, in odd runs only when Spt was defined using independent data from even runs. B: pattern classification accuracy for continuous listen vs. listen+rehearse as a function of time within the trial. C: pattern classification accuracy for continuous listen versus listen+rest as a function of time within the trial. Chance classification accuracy = 0.50, and horizontal lines indicate the upper and lower 5% of classification accuracies observed using a permutation test (see METHODS). Asterices, $P < 0.001$.

activity across voxels within Spt is different during speech perception compared with speech production-related processes (covert rehearsal in the present study). Pattern classification analysis revealed that the pattern of activity in Spt correctly predicted whether a trial involved speech perception or speech production at better than chance levels (~70% accuracy). This was particularly true in the final portion or the trial when the source of the signals was maximally different (sensory vs. motor-related). This classification cannot be attributed to general amplitude differences between the conditions because the amplitude was normalized across the conditions for the analysis and because even unnormalized amplitude was not significantly different during the final portion of the trial when discrimination accuracy was maximal. Instead the result demonstrates that the spatial pattern of activation across the ROI differs for speech perception and production, which may result from different spatial distributions of sensory and motor (and/or sensory-motor) responsive cell types across voxels in the ROI. The distinction in activation pattern for sensory versus motor conditions in Spt also argues against the possibility that motor response properties of Spt result from auditory imagery as well as the possibility that the sensory responses in Spt result from subvocal rehearsal: if sensory and motor activations result in distinguishable activation patterns, they are unlikely to be a consequence of the same process.

It is worth noting that the sensory-motor nature of the response of Spt appears to be reflected as well in the BOLD amplitude response, particularly in the fact that the listen+rehearse condition yields greater activation than the continuous speech condition in the middle time points of the trial, but then falls back to the level of the continuous speech condition by the end of the trial (Figs. 3, 4B, and 6A). If the response in Spt was purely sensory (e.g., if the "motor" response were simply auditory imagery), there is no explanation for why the signal is greater in the listen+rehearse condition where there is less auditory input. If, on the other hand, there are distinct populations of cells that are sensory- versus motor-weighted, then the increased amplitude for the listen+rehearse condition in the middle time bin can be explained as the summed hemodynamic response resulting from activity of these different cell types which would be evident in the mid-trial phase. By the end of the trial, the sensory contribution to the summed activity will have decayed in the listen+rehearse condition resulting in a drop in the overall activation level as observed. The listen+rest condition, which does not involve rehearsal, does not show this pattern but instead results in similar activity in the middle time points of the trial, and far less still in the final time points of the trial (Fig. 6A). Thus the pattern of BOLD amplitude activity is readily explainable on a sensory-motor account of Spt function consistent with our interpretation of the pattern classification results.

If our conclusion is correct that Spt contains both sensory-weighted and motor-weighted classes of cell types, it would indicate a further parallel between area Spt and sensory-motor integration areas in the posterior parietal lobe. Like parietal lobe sensory-motor integration areas, Spt exhibits both sensory and motor response properties (Buchsbaum et al. 2001; Dhankhar et al. 1997; Hickok et al. 2003), is relatively selective for motor modality (vocal tract) (Pa and Hickok 2008), appears to be multi-sensory (Dhankhar et al. 1997), is functionally connected to frontal motor areas (BA 44 in particular) (Buchsbaum et al. 2001), and appears to contain both sensory- and motor-weighted cell types (present study). This set of properties suggests that area Spt is a part of the collection of regions in the posterior parietal cortex that supports sensory-motor integration (Andersen 1997) with Spt tied to actions associated with the vocal tract (Hickok and Poeppel 2007; Pa and Hickok 2008).

Damage to tissue in the vicinity of area Spt has been associated with conduction aphasia (Damasio and Damasio 1983, 1980). Just as disruption of sensory-motor areas in the PPC result in disrupted motor function while sparing sensory recognition abilities (Milner and Goodale 1995; Ungerleider and Mishkin 1982), conduction aphasia is primarily a deficit of speech production while sparing speech recognition (Benson et al. 1973; A. R. Damasio 1991; Damasio and Damasio 198, 1980; Goodglass 1992, 1993; Goodglass and Kaplan 1983). Patients with conduction aphasia make frequent sound-based (phonemic) errors in their speech output and relatively few meaning-based errors, which are more prevalent in aphasic syndromes, such as Wernicke's aphasia, associated with more inferior temporal lobe damage (A. R. Damasio 1992; H. Damasio 1991; Hillis 2007). Conduction aphasics have difficulty in the verbatim repetition of speech, particularly with low-frequency utterances or pseudowords (Goodglass 1992). Verbatim repetition requires reference to a sensory-phonological trace for accurate reproduction—a requirement that is exaggerated with unfamiliar items or pseudowords—and therefore is a behavior that would be substantially impacted by damage to a speech-related sensory-motor integration system (Hickok 2000; Hickok et al. 2000). For these reasons, conduction aphasia has been interpreted as a syndrome that results from damage to the sensory-motor integration network in the posterior planum temporale, area Spt (Hickok 2001, 2000; Hickok and Poeppel 2004; Hickok et al. 2000, 2003). Thus the clinical effects of lesions involving area Spt are consistent with the proposed sensory-motor functions that region.

REFERENCES

Andersen R. Multimodal integration for the representation of space in the posterior parietal cortex. *Philos Trans R Soc Lond B Biol Sci* 352: 1421–1428, 1997.

Benson DF, Sheremata WA, Bouchard R, Segarra JM, Price D, Geschwind N. Conduction aphasia: a clincopathological study. *Arch Neurol* 28: 339–346, 1973.

Boynton GM, Engel SA, Glover GH, Heeger DJ. Linear systems analysis of functional magnetic resonance imaging in human V1. *J Neurosci* 16: 4207–4221, 1996.

Buchsbaum B, Hickok G, Humphries C. Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cogn Sci* 25: 663–678, 2001.

Buchsbaum BR, D'Esposito M. The search for the phonological store: from loop to convolution. *J Cogn Neurosci* 20: 762–778, 2008.

Buchsbaum BR, Olsen RK, Koch P, Berman KF. Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. *Neuron* 48: 687–697, 2005a.

Buchsbaum BR, Olsen RK, Koch PF, Kohn P, Kippenhan JS, Berman KF. Reading, hearing, and the planum temporale. *Neuroimage* 24: 444–454, 2005b.

Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SCR, McGuire PK, Woodruff PWR, Iversen SD, David AS. Activation of auditory cortex during silent lipreading. *Science* 276: 593–596, 1997.

Calvert GA, Campbell R. Reading speech from still and moving faces: The neural substrates of visible speech. *J Cogn Neurosci* 15: 57–70, 2003.

Colby CL, Goldberg ME. Space and attention in parietal cortex. *Annu Rev Neurosci* 22: 319–349, 1999.

Cox RW, AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* 29: 162–173, 1996.

Cui H, Andersen RA. Posterior parietal cortex encodes autonomously selected motor plans. *Neuron* 56: 552–559, 2007.

Damasio AR. Signs of aphasia. In: *Acquired Aphasia*, edited by Sarno MT. San Diego, CA: Academic, 1991, p. 27–43.

Damasio AR. *Aphasia. N Engl J Med* 326: 531–539, 1992.

Damasio H. Neuroanatomical correlates of the aphasias. In: *Acquired aphasia*, edited by Sarno M. San Diego, CA: Academic, 1991, p. 45–71.

Damasio H, Damasio AR. The anatomical basis of conduction aphasia. *Brain* 103: 337–350, 1980.

Damasio H, Damasio AR. Localization of lesions in conduction aphasia. In: *Localization in Neuropsychology*, edited by Kertesz A. San Diego, CA: Academic, 1983, p. 231–243.

Dhankhar A, Wexler BE, Fulbright RK, Halwes T, Blamire AM, Shulman RG. Functional magnetic resonance imaging assessment of the human brain auditory cortex response to increasing word presentation rates. *J Neurophysiol* 77: 476–483, 1997.

Doupe AJ, Kuhl PK. Birdsong and human speech: common themes and mechanisms. *Annu Rev Neurosci* 22: 567–631, 1999.

Galaburda A, Sanides F. Cytoarchitectonic organization of the human auditory cortex. *J Comp Neurol* 190: 597–610, 1980.

Galantucci B, Fowler CA, Turvey MT. The motor theory of speech perception reviewed. *Psychon Bull Rev* 13: 361–377, 2006.

Goodglass H. Diagnosis of conduction aphasia. In: *Conduction Aphasia*, edited by Kohn SE. Hillsdale, NJ: Erlbaum, 1992, p. 39–49.

Goodglass H. *Understanding Aphasia.* San Diego, CA: Academic, 1993.

Goodglass H, Kaplan E. *The assesasment of apahasia and Related disordears.* Philadelphia: Lea and Febiger, 1983.

Graves WW, Grabowski TJ, Mehta S, Gupta P. The left posterior superior temporal gyrus participates specifically in accessing lexical phonology. *J Cogn Neurosci* 20: 1698–1710, 2008.

Guenther FH, Hampson M, Johnson D. A theoretical investigation of reference frames for the planning of speech movements. *Psychol Rev* 105: 611–633, 1998.

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293: 2425–2430, 2001.

Hickok G. Speech perception, conduction aphasia, and the functional neuroanatomy of language. In: *Language and the Brain*, edited by Grodzinsky Y, Shapiro L, Swinney D. San Diego, CA: Academic, 2000, p. 87–104.

Hickok G. Functional anatomy of speech perception and speech production: psycholinguistic implications. *J Psycholinguistic Res* 30: 225–234, 2001.

Hickok G, Buchsbaum B, Humphries C, Muftuler T. Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. *J Cogn Neurosci* 15: 673–682, 2003.

Hickok G, Erhard P, Kassubek J, Helms-Tillery AK, Naeve-Velguth S, Strupp JP, Strick PL, Ugurbil K. A functional magnetic resonance imaging study of the role of left posterior superior temporal gyrus in speech production: implications for the explanation of conduction aphasia. *Neurosci Lett* 287: 156–160, 2000.

Hickok G, Poeppel D. Towards a functional neuroanatomy of speech perception. *Trends Cognit Sci* 4: 131–138, 2000.

Hickok G, Poeppel D. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* 92: 67–99, 2004.

Hickok G, Poeppel D. The cortical organization of speech processing. *Nat Rev Neurosci* 8: 393–402, 2007.

Hillis AE. Aphasia: progress in the last quarter of a century. *Neurology* 69: 200–213, 2007.

Jacquemot C, Scott SK. What is the relationship between phonological short-term memory and speech processing? *Trends Cogn Sci* 10: 480–486, 2006.

Kamitani Y, Tong F. Decoding the visual and subjective contents of the human brain. *Nat Neurosci* 8: 679–685, 2005.

Knaua TA, Bollich AM, Corey DM, Lemen LC, Foundas AL. Variability in perisylvian brain anatomy in healthy adults. *Brain Language* 97: 219–232, 2006.

Levelt WJM, Praamstra P, Meyer AS, Helenius P, Salmelin R. An MEG study of picture naming. *J Cogn Neurosci* 10: 553–567, 1998.

Mattingly IG, Liberman AM. Specialized perceiving systems for speech an other biologically significant sounds. In: *Auditory Function: Neurological Bases of Hearing*, edited by Edelman GMG, Gall WE, Cowan WM. New York: Wiley, 1988, p. 775–793.

Milner AD, Goodale MA. *The Visual Brain in Action.* Oxford, UK: Oxford Univ. Press, 1995.

Nichols TE, Holmes AP. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum Brain Mapp* 15: 1–25, 2002.

Norman KA, Polyn SM, Detre GJ, Haxby JV. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn Sci* 10: 424–430, 2006.

Okada K, Hickok G. Left posterior auditory-related cortices participate both in speech perception and speech production: neural overlap revealed by fMRI. *Brain Lang* 98: 112–117, 2006.

Okada K, Smith KR, Humphries C, Hickok G. Word length modulates neural activity in auditory cortex during covert object naming. *Neuroreport* 14: 2323–2326, 2003.

Pa J, Hickok G. A parietal-temporal sensory-motor integration area for the human vocal tract: evidence from an fMRI study of skilled musicians. *Neuropsychologia* 46: 362–368, 2008.

Sakata H, Taira M, Murata A, Mine S. Neural mechanisms of visual guidance of hand action in the parietal cortex of the monkey. *Cereb Cortex* 5: 429–438, 1995.

Serences JT, Boynton GM. Feature-based attentional modulations in the absence of direct visual stimulation. *Neuron* 55: 301–312, 2007a.

Serences JT, Boynton GM. The representation of behavioral choice for motion in human visual cortex. *J Neurosci* 27: 12893–12899, 2007b.

Serences JT, Ester E, Vogel E, Awh E. Stimulus-specific delay activity in human primary visual cortex. *Psychological Science* 20: 201–214, 2009.

Smiley JF, Hackett TA, Ulbert I, Karmas G, Lakatos P, Javitt DC, Schroeder CE. Multisensory convergence in auditory cortex. I. Cortical connections of the caudal superior temporal plane in macaque monkeys. *J Comp Neurol* 502: 894–923, 2007.

Ungerleider LG, Mishkin M. Two cortical visual systems. In: *Analysis of Visual Behavior*, edited by Ingle DJ, Goodale MA, Mansfield RJW. Cambridge, MA: MIT Press, 1982, p. 549–586.

Vapnik VN. *Statistical Learning Theory.* New York: Wiley, 1998.

Vul E, Harris C, Winkilman P, and Pashler H. Voodoo correlations in social neuroscience. *Perspect Psychol Sci* In press.

Warren JE, Wise RJ, Warren JD. Sounds do-able: auditory-motor transformations and the posterior temporal plane. *Trends Neurosci* 28: 636–643, 2005.