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

Gregory Hickok¹*, Kayoko Okada¹, & John T. Serences²*

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

Keywords: Planum temporale, speech, area Spt, sensory-motor integration, auditory dorsal stream, fMRI

*Corresponding author

Greg Hickok
Center for Cognitive Neuroscience
Department of Cognitive Sciences
University of California
Irvine, CA 92697
greg.hickok@uci.edu

Or,

John Serences
Department of Psychology
University of California, San Diego
La Jolla, CA 92093-0109
jsерences@ucsd.edu
Abstract

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 non-human primates – little work has been done on sensory-motor transformations in the domain of speech. Here, we used fMRI 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, the Spt forms a sensory-motor integration circuit for the vocal tract.
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 et al. 2000; Hickok and Poeppel 2007; 2000), sensory guidance of speech production (Guenther et al. 1998; Hickok 2000; Hickok et al. 2000; Hickok and Poeppel 2007; 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 et al. 2003; Hickok and Poeppel 2004; 2007; 2000; 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; Buchsbaum et al. 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 2004; 2007; 2000), 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
multisensory (Calvert et al. 1997; Calvert and Campbell 2003; 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; Buchsbaum et al. 2005b; Hickok et al. 2003; Hickok and
Poeppel 2004; 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 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.,
approximately 33% are motor dominant, 26% visual dominant, & 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 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 years 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 & 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 seconds. 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 second blocks (trials), (i) “continuous speech”: 15s of listening to continuous speech (sets of sentences), (ii) “listen+rest”: 3s of listening to speech followed by 12s of rest, (iii) “listen+rehearse”: 3s of listening to speech followed by 12s of covert (subvocal) rehearsal of the heard stimuli (see Figure 1). In addition, we included a null (rest) condition of the same 15s duration. Each of six sessions (runs) contained two trials (15s blocks) of each condition, including the null (rest) condition. A visual cue distinction between the conditions: a fixation cross for the continuous listen condition which cued the subject to simply fixate and listen, a picture of an ear for the listen+rest condition which 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 which cued the subject to listen and then 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; Buchsbaum et al. 2005a; Buchsbaum et al. 2005b; 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 which 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 5 experimental sessions (runs). Each experimental session contained an equal number of trials (blocks) of each condition and a single scanning session was approximately 6 minutes 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, Inc, USA). 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**

MR images were obtained in a Philips Achieva 3T MR scanner (Philips Medical Systems, Andover, MA) fitted with an 8 channel RF receiver head coil, at the Research Imaging Center, University of California, Irvine. We first collected a total of 620 EPI volumes over 5 sessions using Fast Echo EPI (sense reduction factor=2.0, FOV = 220x180, matrix=112x112mm , TR=3.0s, TE=25ms, flip angle=70°, size=1.95x1.95x2mm). After the functional scans, a high-resolution anatomical image was acquired with an MPRAGE pulse sequence in axial plane (matrix=256x256mm, TR=8ms, TE=3.7ms, flip angle=8°, size=1x1x1mm).

**Data Analysis**
The first three and the last image of each session was discarded prior to analysis. Preprocessing of the data was performed using Statistical Parametric Mapping (SPM5; Wellcome Department of Imaging Neuroscience, London, UK; www.fil.ion.ucl.ac.uk/spm) implemented in Matlab7 (Mathworks, Inc, USA). 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 6-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 below).

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; Buchsbaum et al. 2005b; Hickok et al. 2003; Okada and Hickok 2006; Pa and Hickok 2008). This activation is strongly left dominant (Hickok et al. 2003; Hickok and Poeppel 2007). 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 < .001, uncorrected). Thus, in individual subjects, ROIs were defined by (i) activations reflecting the conjunction of continuous speech > null rest blocks, and speech+rehearse > speech+rest that (ii) 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 or more) to allow for pattern classification analysis (mean number of voxels, ±S.E.M.: 70±45). The analysis described below 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 Figure S1).

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 below), 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 seconds following trial onset the two conditions are similar in terms of their acoustic input properties. They diverge only during the last 12 seconds of the trial where one condition continues to involve sensory perception of speech and the other involves (covert) speech production without sensory speech stimulation (Figure 1). Thus, we expect increasing pattern discriminability as the trial progresses.

To test this hypothesis, we first normalized the timeseries 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 timeseries from each voxel within each Spt ROI into three separate temporal bins that each contained activation patterns measured on two successive TRs (0s-3s, 6s-9s, and 12s-15s 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 372s 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 vs. listen+rehearse) on each trial in the test set (see e.g. (Kamitani and Tong 2005; Serences and Boynton 2007a; b; Serences et al. in press)). 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. Since 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. Since each subject completed 5 runs/session, the overall classification accuracy for a subject was defined as the average classification accuracy across all 5 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 comprised 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 Figure 4 (and Supplemental Figure 1) 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.

**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 (10 or more) 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 Figure 2 (top row). Activation foci in two individual subjects is shown in Figure 2 (bottom row). The location of these ROIs is consistent with previous studies of area Spt (Buchsbaum et al. 2001; Buchsbaum et al. 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 Figure 3. An analysis of variance 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 = .001), and that the activation levels for the two conditions differed during some time points (condition x time point interaction: F(5, 65) = 3.41, p = .008). During the first two time points, when the sensory-evoked response dominates the signal, the amplitudes are statistically identical (p > .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 = .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 = .008) and then declines to equal the continuous listen condition in the last two time points (t(13) = 1.19, p = .25 and t(13) = .08, p = .94, respectively). Since 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 do 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 at this last time bin, therefore, that represents the strongest test of our hypothesis.

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 the Spt discriminated speech.
perception (continuous listen condition) from speech production (listen+rehearse condition). In contrast to the timecourse 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 (one-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 (0s-3s, $t(13)=0.538$, $p=.60$), as expected. After this point, classification accuracy improves to above-chance levels (6s-9s, $t(13)=3.58$, $p<0.0034$), and reaches approximately 75% accuracy in the final time bin ($t(13)=3.717$, $p<0.0004$, Figure 4; Supplemental Figure 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 is not completely independent (Vul et al. in press). 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 from only the odd runs (sessions). This was possible in 14 of our subjects using a slightly relaxed threshold of 0.005 (using only two 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 which 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 Figure 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 six
seconds of the trial, which is dominated by the sensory response, signal amplitude is
equivalent across conditions. After this point, the 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 9s, 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. Pattern classification in the fully unbiased (odd
run) dataset again confirmed this prediction (Figure 6B): classification accuracy was
significantly above chance at the final time point (t(13) = 3.017, p = .01, two-tailed, .03 after Bonferroni correction). The same analysis carried out on the continuous listen versus listen+rest condition yielded no above-chance classifications (all p-values > .19, two-tailed, >.50 after Bonferroni correction) (Figure 6C). We conclude from these analyses that pattern classification accuracy in the continuous listen vs. 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; Buchsbaum et al. 2005a; Buchsbaum et al. 2005b; Hickok et al. 2003; Okada and Hickok 2006). The present study showed that the pattern of activity across voxels within Spt is different during speech perception compared to 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 were maximally different (sensory vs. motor-related). This classification cannot be attributed to general amplitude differences between the conditions (i) because the amplitude was normalized across the conditions for the analysis, and (ii) because even unnormalized amplitude was not significantly different during the final portion of the trial with 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 (Figure 3, 4b, 6A). If the response in Spt were 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 (Figure 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 (i) exhibits both sensory and motor response properties (Buchsbaum et al. 2001; Dhankhar et al. 1997; Hickok et al. 2003), (ii) is relatively selective for motor modality (vocal tract) (Pa and Hickok 2008), (iii) appears to be multisensory (Dhankhar et al. 1997), (iv) is functionally connected to frontal motor areas (BA 44 in particular)(Buchsbaum et al. 2001), and (v) 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; Damasio 1991a; Damasio and Damasio 1983; 1980; Goodglass 1992; Goodglass 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 (Damasio 1992; Damasio 1991b; 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 et al. 2003; Hickok et al. 2000; Hickok and Poeppel 2004). Thus, the clinical effects of lesions involving area Spt are consistent with the proposed sensory-motor functions that region.
References


Table Captions

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.
Figure Captions

Figure 1. Schematic diagram of the three conditions analyzed in the present study. 1. The “continuous listen” condition. 2. The “listen+rest” condition. 3. The “listen+rehearse” condition. The pattern classification analysis used condition 1 vs. 3.

Figure 2. Top row: Location of the each subject’s Spt ROI presented on a standardized 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 two subjects projected onto each subject’s own anatomical MRI scan.

Figure 3. Mean BOLD signal amplitude timecourse (z-score) for the continuous listen and listen+rehearse conditions in the Spt ROI for six 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.

Figure 4. A. Pattern classification accuracy as a function of time within the trial. Data is organized into three 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 second and third bins only, ***p<0.001. Dashed blue lines indicate the upper and lower 5% of classification accuracies observed using a permutation test (see
Methods). B. BOLD amplitude for the two 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 two conditions.

**Figure 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. Since 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. Figure shows a map of SVM weights within the area Spt ROI for two representative subjects.

**Figure 6.** BOLD amplitude and pattern classification accuracy in the split-plot analysis.

A. BOLD amplitude in area Spt for the three 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 vs. listen+rest as a function of time within the trial. Chance classification accuracy = 0.50, and dashed blue lines indicate the upper and lower 5% of classification accuracies observed using a permutation test (see Methods). **p<0.001.
Figure S1. Pattern classification accuracy as a function of time within the trial using all 20 subjects who showed activation within area Spt. Data is organized into three 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 second and third bins only, **p<.005, ***p<0.001. Dashed blue lines indicate the upper and lower 5% of classification accuracies observed using a permutation test (see Methods).
Figure 3

[Graph showing signal change (z-score) over time (s) for 'listen+rehearse' and 'continuous listen'.]