Neural Organization of Linguistic Short-term Memory is Sensory Modality-dependent: Evidence from Signed and Spoken Language

Judy Pa[1], Stephen M. Wilson[1], Herbert Pickell[2], Ursula Bellugi[2], and Gregory Hickok[1]

Abstract

■ Despite decades of research, there is still disagreement regarding the nature of the information that is maintained in linguistic short-term memory (STM). Some authors argue for abstract phonological codes, whereas others argue for more general sensory traces. We assess these possibilities by investigating linguistic STM in two distinct sensory–motor modalities, spoken and signed language. Hearing bilingual participants (native in English and American Sign Language) performed equivalent STM tasks in both languages during functional magnetic resonance imaging scanning. Distinct, sensory-specific activations were seen during the maintenance phase of the task for spoken versus signed language. These regions have been previously shown to respond to non-linguistic sensory stimulation, suggesting that linguistic STM tasks recruit sensory-specific networks. However, maintenance-phase activations common to the two languages were also observed, implying some form of common process. We conclude that linguistic STM involves sensory-dependent neural networks, but suggest that sensory-independent neural networks may also exist. ■

INTRODUCTION

Despite decades of research, the nature of the information that is maintained in linguistic[1] short-term memory (STM) is not fully understood. For example, if one is asked to maintain a list of auditorily presented words, this information could be represented and maintained as a set of acoustic traces, a sequence of actions sufficient to reproduce the words, a sequence of abstract representations, such as phonological forms, or some combination of these. The range of existing accounts of linguistic STM, based primarily on behavioral data, encompasses these possibilities (Baddeley & Larsen, 2007a, 2007b; Jones, Hughes, & Macken, 2007; Boutla, Supalla, Newport, & Bavelier, 2004; Jones, Macken, & Nicholls, 2004; Wilson, 2001; Jones & Tremblay, 2000; Neath, 2000; Jones & Macken, 1996; Baddeley, 1992). For example, Wilson (2001) argues explicitly for sensory–motor codes underlying linguistic STM, whereas Baddeley and Larsen (2007b) argue for a specifically phonologically based storage system. Although one could imagine that phonological codes may or may not be a form of sensory (or sensory–motor) representation, a common view is that the phonological representations underlying linguistic STM are postsensory, amodal codes. Jones, Hughes,

and Macken (2006) summarize this view, "... the representations with which [Baddeley's phonological] store deals ... can neither be acoustic nor articulatory but must, rather, be post-categorical, 'central' representations that are functionally remote from more peripheral perceptual or motoric systems. Indeed, the use of the term phonological seems to have been deliberately adopted in favor of the terms acoustic or articulatory (see, e.g., Baddeley, 1992) to indicate the abstract nature of the phonological store's unit of currency" (p. 266).

Recent neurobiological work in several domains of working memory has provided evidence for sensory system involvement in the temporary storage of items held in STM, and this has led to the view that STM involves, at least in part, the active maintenance of sensory traces (see Postle, 2006; Pasternak & Greenlee, 2005; Ruchkin, Grafman, Cameron, & Berndt, 2003 for reviews). Although this view is often generalized to include linguistic STM, it is difficult to demonstrate unequivocally the involvement of sensory systems in the linguistic domain. For example, studies presenting auditory information typically report superior temporal sulcus (STS) and posterior planum temporale (Spt) activation during maintenance of speech information (Stevens, 2004; Hickok, Buchsbaum, Humphries, & Muftuler, 2003; Buchsbaum, Hickok, & Humphries, 2001). Although both of these regions are often associated with the auditory system, both have also been shown to be involved in processing

[1]University of California, Irvine, [2]Salk Institute for Biological Studies, La Jolla, CA

information from other sensory (and motor) modalities (Beauchamp, Argall, Bodurka, Duyn, & Martin, 2004; Hickok et al., 2003; Griffiths & Warren, 2002). This observation raises the possibility of amodal (perhaps phonological) codes underlying linguistic STM.

One approach to this question is to study linguistic STM using different input modalities, such as auditory versus visual speech. In this case, the sensory encoding is varied while perhaps keeping phonological information constant. If STM for auditory versus visual word forms activated distinct brain networks, one might argue for sensory-specific codes. A recent functional magnetic resonance imaging (fMRI) study (Buchsbaum, Olsen, Koch, & Berman, 2005) assessed this possibility, and found delay activity in the planum temporale region to be insensitive to input modality, but found a preference for the auditory modality in the STS. This relative preference for the maintenance of information in the auditory modality in the STS is suggestive of a sensory-specific effect. However, there are complications with using written word stimuli in that visual word forms can be decoded (and therefore presumably represented) in multiple ways, including the use of phonological mechanisms as well as in terms of visual word forms. It is possible, therefore, that linguistic STM relies primarily on amodal phonological forms, and the differences in activation in the STS reflects the strength of activation of this network: Auditory word forms may activate this network strongly, whereas visual word forms may activate it less strongly because alternate (i.e., visual) coding is also possible.

The goal of the present study was to shed light on these questions from a unique perspective, comparing the neural organization of linguistic STM in two language modalities, spoken versus signed. Comparison of spoken versus signed language provides a unique perspective on linguistic STM because, although the two systems are radically different in terms of sensory–motor modalities, they are quite similar at an abstract linguistic level, including the involvement of abstract phonological forms (see Emmorey, 2002 for a review). If distinct, sensory modality-specific activations are found to support STM for spoken versus signed language, this would constitute strong evidence for sensory-based codes underlying linguistic STM. On the other hand, similarities in the neural systems supporting linguistic STM in the two language systems could indicate modality neutral processes, although there are other possibilities (see below).

Research on STM for sign language has revealed impressive similarities to the core behavioral effects found in studies on STM for spoken language (Wilson & Emmorey, 2003; Wilson, 2001; Wilson & Emmorey, 1998). These effects include the phonological similarity effect (worse performance for lists of similar sounding items), the articulatory suppression effect (worse performance when articulatory rehearsal is disrupted), the irrelevant speech effect (worse performance when to-be-ignored auditory stimuli are presented), and the word length effect (worse performance for longer items) (Baddeley, 1992). All of these effects hold in their sign analog forms, suggesting that the organization of linguistic STM is highly similar in the two language modalities (Wilson, 2001; Wilson & Emmorey, 1997, 1998). This does not necessarily imply, however, a modality-independent linguistic STM circuit. In fact, evidence from the irrelevant speech and irrelevant sign effects suggest a modality-*dependent* storage mechanism, because irrelevant stimuli within sensory modality yield the greatest disruption on STM performance (Wilson & Emmorey, 2003). Thus, although the basic organization appears to be similar across modalities, the evidence suggests that linguistic information is stored, at least partly, in modality-dependent systems.

Functional neuroimaging has the potential to test the possibility that linguistic STM storage systems are modality-dependent. A recent positron emission tomography study compared STM for speech (Swedish) and sign (Swedish Sign Language) in hearing/speaking subjects who also acquired sign language at a young age (Ronnberg, Rudner, & Ingvar, 2004). This study found extensive differences in the neural activation patterns during STM for speech versus sign. STM for sign was greater than for speech in visual-related areas bilaterally (ventral occipital and occipital–temporal cortex), and in the posterior parietal cortex bilaterally, whereas STM for speech yielded greater activation in auditory-related areas in the superior temporal lobe bilaterally, as well as in some frontal regions. Taken at face value, this study would seem to make a strong case for modality-dependent storage systems in linguistic STM. However, because the encoding and storage phases of the task were not experimentally dissociated in the design—activation reflected both components—it is very likely that the bulk of the sensory-specific activations resulted simply from the sensory processing of the stimuli. If additional, storage-related activity was present, their design could not easily detect it.

Another study examined STM for sign language in deaf native signers using fMRI (Buchsbaum, Pickell, et al., 2005) and compared its findings to a similar previously published study involving speech and hearing nonsigner subjects (Hickok et al., 2003). This study used a design with a several second delay period between encoding and recall, which allowed for the measurement of storage-related activity. The pattern of activation during the retention phase was substantially different from what had been found in hearing participants performing a similar task with speech. Short-term maintenance of sign language stimuli produced prominent activations in the posterior parietal lobe, which were not found in the speech study. This parietal activation was interpreted as a reflection of visual–motor integration processes. Posterior parietal regions have been implicated in visual–motor integration in both human and non-

human primates (Andersen, 1997; Milner & Goodale, 1995), and studies of gesture-imitation in hearing subjects report parietal activation (Chaminade, Meltzoff, & Decety, 2005; Peigneux et al., 2004). It seems likely, therefore, that parietal activation in an STM task for sign language does not reflect activity of a sensory store, but instead results from sensory–motor processes underlying the interaction of storage and manual articulatory rehearsal (Buchsbaum, Pickell, et al., 2005). Additional maintenance activity was found in the posterior superior temporal lobe (left Spt and posterior STS [pSTS] bilaterally), as well as in posterior frontal regions, all of which have been shown to activate during maintenance of speech information, suggestive of some form of common process. However, because cross-modality comparisons could only be made between subjects and studies, it is difficult to make solid inferences about patterns of overlap and dissociation. No maintenance activity was found in visual-related areas in that study, such as the ventral temporal–occipital regions that were so strongly activated during sign perception (Buchsbaum, Pickell, et al., 2005; MacSweeney et al., 2002; Petitto et al., 2000). Activation in these regions would provide more convincing support for sensory-dependent STM storage systems.

In sum, the current imaging evidence does not provide unequivocal support for sensory modality-specific storage of linguistic material in STM, although there is evidence for nonidentical sensory–motor integration networks, and evidence for significant cross-modality overlap in STM networks in frontal, as well as posterior superior temporal lobe regions. The goal of the present study was to examine the neural networks supporting linguistic STM using a within-subject design involving hearing native bilingual participants (American Sign Language [ASL] and English), and using a delay phase between encoding and retrieval that allows us to identify regions active specifically during the maintenance of linguistic material in STM.

METHODS

Participants

Fifteen healthy participants (4 men; ages 19–40 years; 27.33 ± 5.45) participated in this study. One participant was excluded from the study due to excessive head movement during scanning. All participants were right-handed by self-report. Due to the nature of the task conditions, participants who were hearing, native English speakers, and ASL signers were used in this paradigm. Participants were recruited through a local distribution list-serve of CODAs (Children of Deaf Adults) and personal communication of participating researchers and participants. Prior to scanning, participants were trained on the task via a practice session to ensure their understanding of the task. All participants gave informed consent, and the University of California, Irvine Institutional Review Board approved this study.

Experimental Design

There were two types of trials: speech and sign. Each trial began with a 3-sec stimulus containing a set of either three nonsense words (pseudowords) or nonsense signs (pseudosigns). The 3-sec stimulus was followed by 15 sec of covert rehearsal. This was followed by another 3-sec stimulus, which contained the same three words or signs, in either the same or different order. Participants then responded via a button press whether the order was the same or different. Each trial was followed by a 15-sec rest period, which allowed the hemodynamic response to return to baseline.

The pseudowords were a mixture of two-, three-, and four-syllable items such as "plinkit, gazendin, bingalopof." Individual words were chosen for inclusion in a given three-item list such that the combined length was as close to 3 sec as possible. The lists were then digitally edited to 3 sec in duration. We did not explicitly control such factors as phonotactic frequency and lexical neighborhood density because these factors were internally controlled: The same items appeared in both the pre-rehearse and pre-rest sensory stimulation phases.

Pseudosigns were bimanual gestures that did not carry meaning but conformed to the phonotactic rules of ASL. Pseudosigns with a high degree of similarity to real signs were avoided to minimize semantic coding. Pseudosigns were used (1) for consistency with the previously published study using spoken pseudowords, and (2) to minimize semantic-related processing. Previous work has indicated that pseudosigns, like pseudowords, are processed phonologically as opposed to being processed as nonstructured manual gestures (Emmorey, 1995). Sign stimuli were generated and digitally video-recorded by a team of native ASL signers (Buchsbaum, Pickell, et al., 2005). Nonsense stimuli (pseudowords, pseudosigns) were used to minimize the activation of semantic regions and cross-modal processing.

Stimuli were presented using the Matlab toolbox Cogent 2000 (www.vislab.ucl.ac.uk/Cogent). There were two functional runs, and speech and sign trials were intermixed in random order within each run. Each subject completed 25 trials per condition. A black crosshair was displayed as a fixation point on the center of the gray screen throughout the runs to help participants maintain focus, except during the visual presentation of signs. The stimuli were presented through VisuaStim XGA MRI-compatible head-mounted goggles and earphones (Resonance Technology, Northridge, CA).

fMRI Procedures

Data were collected at the University of California, Irvine in a Phillips-Picker 1.5-T scanner interfaced with

a Phillips-Eclipse console for the pulse sequence generation and data acquisition. A high-resolution anatomical image was acquired (axial plane) with a 3-D SPGR pulse sequence for each subject (FOV = 250 mm, TR = 13 msec, flip angle = 20°, size = 1 mm × 1 mm × 1 mm). A series of echo-planar imaging acquisitions were then collected. fMRI data were acquired using gradient-echo, echo-planar imaging (FOV = 250 mm, TR = 2000 msec, TE = 40 msec, flip angle = 80°). A total of 910 volumes were collected (455 volumes per run) consisting of 22 axial, 5 mm slices per volume covering all of the cerebrum and the majority of the cerebellum.

Data Analysis

The fMRI data were preprocessed using tools from FSL (Smith et al., 2004). Skull stripping was performed with BET (Smith, 2002), motion correction was carried out with MCFLIRT (Jenkinson & Smith, 2001), and the program IP was used to smooth the data with a Gaussian kernel (8 mm full width at half maximum [FWHM]) and to normalize mean signal intensity across subjects.

Functional images were aligned using FLIRT to high-resolution anatomical images via an affine transformation with 6 degrees of freedom. High-resolution anatomical images were then aligned to the standard Montreal Neurological Institute (MNI) average of 152 brains using an affine transformation with 12 degrees of freedom.

A general linear model was fit to the data from each voxel in each subject using the FMRISTAT toolbox (Worsley et al., 2002) in MATLAB (Mathworks, Natick, MA). Explanatory variables were included which modeled perception of speech, rehearsal of speech, perception of sign, and rehearsal of sign. Perception events were 3 sec long, and rehearsal events were 15 sec long, and these were immediately adjacent to one another in the order: perception–rehearsal–perception–rest. Each of these variables was convolved with a hemodynamic response function modeled as a gamma function with 5.4 sec time to peak and 5.2 sec FWHM. Because of the delay of the hemodynamic response, it is difficult to properly attribute BOLD responses to temporally adjacent events, therefore additional statistical calculations were performed based on time courses of important regions of interest (ROIs), as described below.

Although there are cognitive differences between the first and second perception events (i.e., the first involves encoding, whereas the second involves a process of comparison), we did not attempt to model them separately because differences between them would inevitably be confounded with the button press as well as the rehearsal event which follows the former but precedes the latter, and cannot be entirely factored out due to its invariant temporal proximity.

Temporal drift was removed by adding a cubic spline in the frame times to the design matrix (one covariate per 2 min of scan time), and spatial drift was removed by adding a covariate in the whole volume average. Six motion parameters (three each for translation and rotation) were also included as confounds of no interest. Autocorrelation parameters were estimated at each voxel and used to whiten the data and design matrix. The two runs within each subject were combined using a fixed effects model, and then the resulting statistical images were registered to MNI space.

Group analysis was performed with FMRISTAT using a mixed effects linear model (Worsley et al., 2002). Standard deviations from individual-subject analyses were passed up to the group level. The resulting t statistic images were thresholded at $t > 2.95$ ($df = 15$, $p < .005$, uncorrected) at the voxel level, with a minimum cluster size then applied so that only clusters significant at $p < .05$ (corrected) according to Gaussian Random Field (GRF) theory were reported. Conjunction analyses were performed by taking the minimums of two t statistic images, thus voxels shown as activated in conjunction analyses were independently active in each of the two conditions.

Significantly activated regions were overlaid on a high-resolution single-subject T1 image (Holmes et al., 1998) using a custom MATLAB program. In the tables of regions showing significant signal increases or decreases, anatomical labels were determined manually by inspecting significant regions in relation to the anatomical data averaged across the subjects, with reference to an atlas of neuroanatomy (Duvernoy, 1999). In cases of large activated areas spanning more than one region, prominent local maxima were identified and tabulated separately.

Plots of time courses from ROIs in the group analysis were based on a Gaussian (8 mm FWHM) around the peak voxel of the relevant activation for the contrast of interest (Figures 1 and 2). A time course for a left pSTS region showing maintenance effects for both speech and sign was also constructed based on data from individual subjects. For each subject, the peak voxel in the left pSTS was identified in a conjunction image where the value of each voxel was the minimum of the perception and maintenance t statistic images. The time courses from each subject were then averaged together (Figure 4).

RESULTS
Behavioral Results

The behavioral data acquired while scanning were analyzed for accuracy to ensure that participants were performing the task and to determine whether there was an inherent difference in difficulty levels of each task. 6.5% of trials had no response and were excluded from the behavioral analysis. Performance levels across the two conditions (speech, sign) were high. The mean percent correct for speech was 87.6 ± 12.3%, whereas
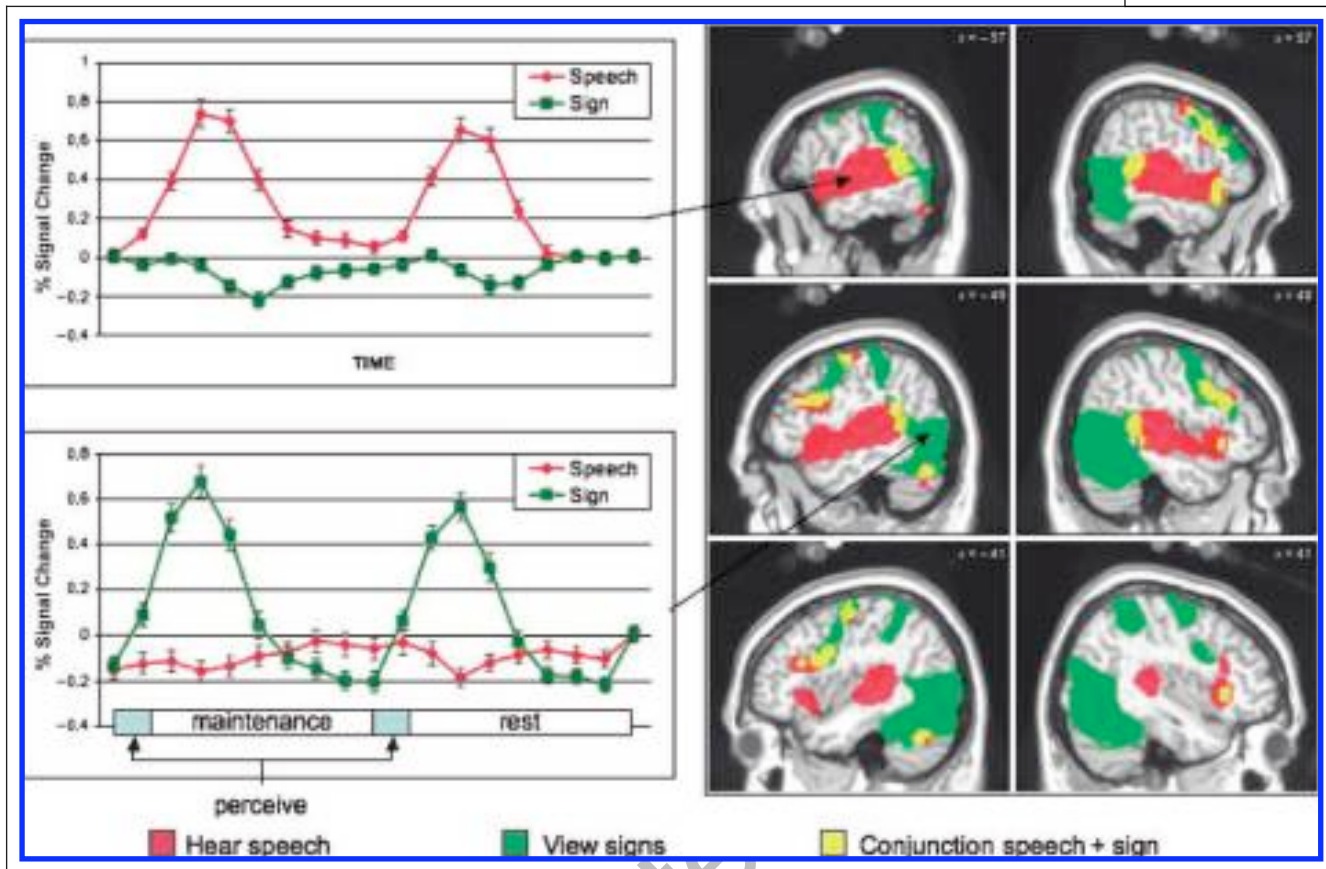
Figure 1. Activation maps and selected time courses associated with the perception/encoding of speech or sign.

for sign was 89.5 ± 10.9%. A paired t test showed that the levels of accuracy were not statistically different $[t(14) = -0.49, p = .63]$, suggesting that there was no difference in difficulty between the two tasks.

fMRI Results

Activations associated with the perception of speech or sign stimuli were largely bilateral. Predictably, perception of nonsigns produced activation in occipital and ventral occipital–temporal regions, whereas the perception of spoken nonwords activated the superior temporal lobe. Sign language perception also activated parietal regions, whereas speech did not. These same regions were also significantly differentially activated when perception of speech and perception of sign were compared directly (data not shown). Several regions showed overlap in activation for speech and sign, including the posterior superior temporal lobe, the inferior frontal gyrus, the premotor cortex, and the cerebellum, all bilateral (see Figure 1, Table 1).

Regions active during the maintenance phase of the task were more left-dominant, and showed areas of coactivation for speech and sign, as well as prominent regions of modality-specific activation (Figures 2 and 3,

Table 1). Frontal lobe areas showed extensive activation for both speech and sign maintenance, predominantly in the left hemisphere. The inferior posterior parietal lobe also showed coactivation for speech and sign maintenance, as well as a small cluster in the left cerebellum. Several modality-specific activation foci were identified. Specifically, sign tended to activate more anterior regions (in the frontal cortex), whereas speech tended to activate the more posterior areas (with overlapping regions in between). As previously documented, sign-specific activation was also noted in the parietal lobe, primarily in the left hemisphere. Finally, and particularly relevant, modality-specific activations were found in sensory-related areas. Speech maintenance activated the left middle and pSTS, whereas sign maintenance activated the left posterior ventral occipital–temporal cortex. A region of overlap was observed during the maintenance of both speech and sign in the pSTS (Figures 2 and 3). These modality-specific activations were also found when maintenance of speech and maintenance of sign were directly compared, with the exception of the anterior–posterior distinction in the left frontal cortex, which was not significant in the direct comparison.

To insure that the region of overlap in the pSTS for the maintenance of speech and sign was not an artifact
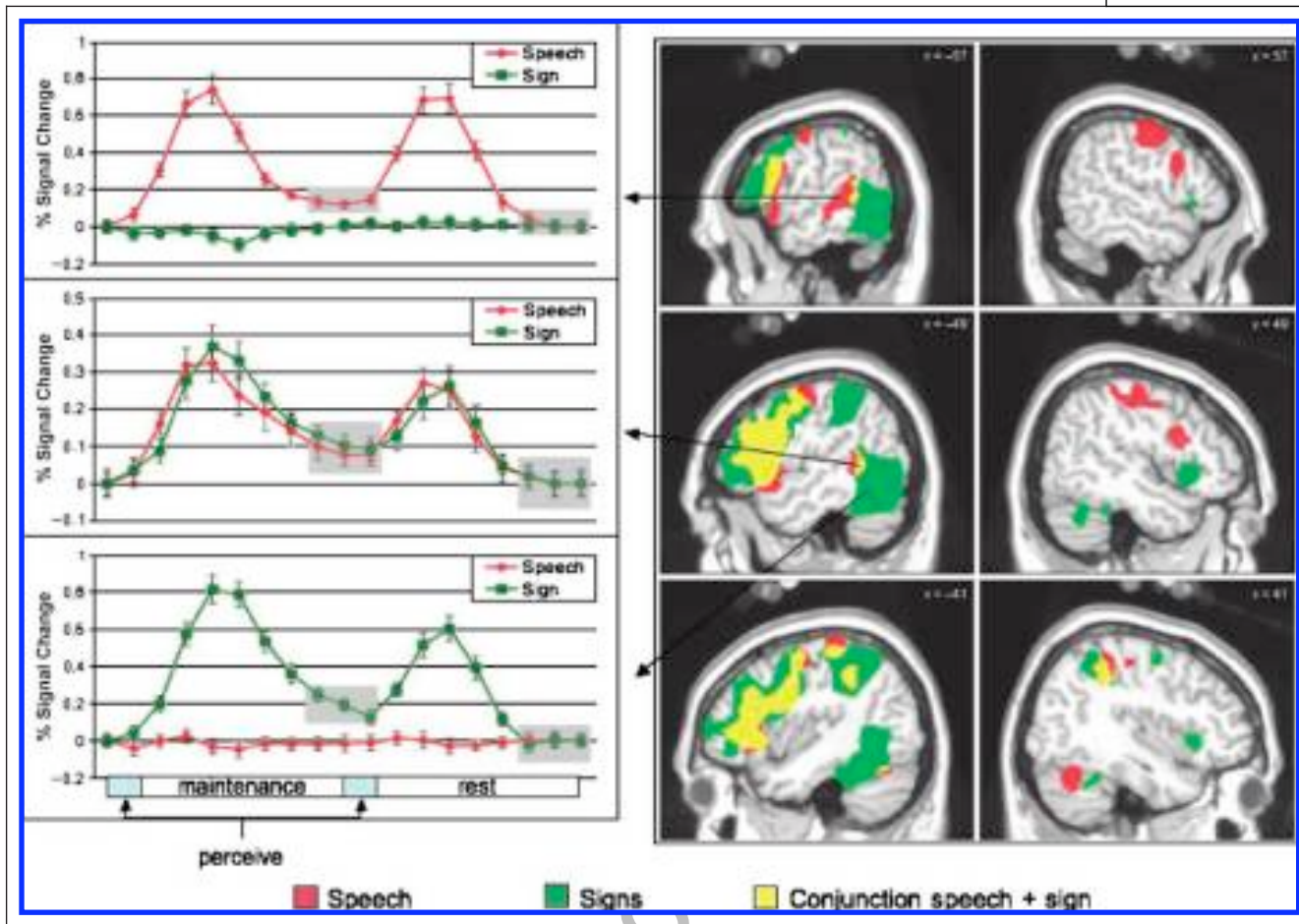
Figure 2. Activation maps and time courses associated with the maintenance portion of the task in three ROIs, a speech-specific region in the pSTS/STG (top time-course panel), a region jointly activated by the maintenance of speech and sign (middle time-course panel), and a sign-specific region in the ventral temporal–occipital cortex (bottom time-course panel). Time-course data show that these regions are also highly responsive to sensory stimulation. Note that the maintenance signal—the difference between first and second time-course regions highlighted in gray—is substantially smaller than the sensory response. However, it is nonetheless highly reliable (see Figure 3), and similar in amplitude to frontal maintenance responses, as well to previous reports of LOC maintenance activity in a visual-object STM task. Gray boxes indicate that portion of the time course that was used to generate bar graphs in Figure 3.

of averaging or smearing from our group analysis (due to individual variability of speech-specific and sign-specific activations), we carried out an individual-subject analysis. The pSTS ROI was identified in each subject by looking for the conjunction of sign and speech activations during the maintenance phase of the task. Statistically significant activations were found in each participant in the pSTS (mean MNI coordinates: −56, −48, 11). An average time course associated with activity in this area across subjects is shown in Figure 4. Note that the amplitude of the response in this individual-subject analysis is roughly twice that found in the group-based analysis, suggesting that the latter may underestimate amplitude measurements because of across-subject variability in the location of the peak response.

Only a subset of areas involved in sensory processing is involved in active maintenance of linguistic material. Figure 5 contains activation maps showing regions that are responsive to the sensory phase of the task, the maintenance phase, and the conjunction of the sensory and maintenance phases. For speech, the (nonfrontal) region showing both sensory and maintenance activity is in the left posterior superior temporal gyrus (STG). For sign, the conjunction of the sensory and maintenance stage revealed activations more posterior and ventral, involving visual-related regions. There was some overlap in the sensory + maintenance conjunctions for speech and sign in the left pSTS (see also Figure 4).

DISCUSSION

The primary goal of this study was to assess whether modality-specific activations could be identified in *sensory-responsive cortices* during the active maintenance of speech versus sign stimuli. Such activations were found. Active maintenance of speech stimuli recruited middle and posterior portions of the STS/STG of the left

Table 1. Corresponding MNI Coordinates, *t* Statistics, and *p* Values of Activated Regions in the Conditions of Hear/Rehearse Speech, See/Rehearse Sign, and the Conjunctions of Hear Speech/See Sign and Rehearse Speech/Sign

Region	MNI Coordinates			Volume (mm³)	Max t	p
	x	y	z			
Hear Speech						
Right temporal and frontal regions, and bilateral subcortical structures				87,744		<.0001
Right superior temporal gyrus and sulcus	62	−26	−4		14.54	
Right precentral gyrus	56	−4	52		5.26	
Right inferior frontal gyrus, pars opercularis and triangularis	48	22	20		5.20	
Right anterior insula	36	24	−4		6.41	
Left putamen/globus pallidus	−16	6	−4		9.16	
Right putamen/globus pallidus	20	2	2		5.09	
Left medial geniculate nucleus	−8	22	0		6.63	
Right medial geniculate nucleus	10	−18	0		6.37	
Left superior colliculus	−4	−36	−6		5.14	
Right superior colliculus	8	−36	−6		5.29	
Left inferior colliculus	−12	−28	−14		4.96	
Right inferior colliculus	12	−30	−16		5.72	
Left superior temporal gyrus and sulcus	−56	−20	−4	46,632	14.51	<.0001
Left cerebellum and occipital cortex				27,456		<.0001
Occipital cortex	−2	−90	−10		5.11	
Left cerebellum	−36	−64	−30		6.07	
Right cerebellum	32	−68	−28	10,608	5.44	<.0001
Bilateral medial prefrontal cortex	−4	16	40	10,376	6.15	<.0001
Left inferior frontal gyrus, pars opercularis and triangularis	−46	12	22	5440	5.88	.0014
Left central sulcus/precentral gyrus	−46	−12	56	2728	5.22	.039
Rehearse Speech						
Left fronto-parietal network				45,824		<.0001
Left postcentral gyrus	−38	−34	60		7.00	
Left intraparietal sulcus	−30	−70	46		6.70	
Left precentral gyrus	−46	−6	56		5.27	
Left inferior frontal gyrus, pars opercularis	−52	10	16		7.88	
Left inferior frontal gyrus, pars triangularis	−42	30	6		6.84	
Supplementary motor area and medial prefrontal cortex	−4	16	46	14,336	5.63	<.0001
Right precentral gyrus, central sulcus, and inferior parietal lobule	54	−14	52	9864	5.89	<.0001
Right cerebellum	32	−66	−26	7864	6.48	.0002
Left superior temporal sulcus and middle temporal gyrus	−62	−34	2	5800	5.78	.001
Left cerebellum	−16	−70	−24	4376	4.99	.0045
Right inferior frontal gyrus, pars opercularis	52	8	28	4176	6.54	.0057

Table 1. (continued)

Region	MNI Coordinates			Volume (mm³)	Max t	p
	x	y	z			
See Sign						
Occipital and parietal regions				255,712		<.0001
Left intraparietal sulcus	−28	−62	48		10.34	
Right intraparietal sulcus	28	−64	56		10.01	
Left posterior superior temporal sulcus	−60	−48	10		6.38	
Right posterior superior temporal sulcus	62	−38	8		10.52	
Left visual motion area MT	−42	−76	−16		24.51	
Right visual motion area MT	52	−72	−14		17.74	
Primary visual cortex	4	−82	−4		14.86	
Right frontal regions				17,672		<.0001
Right precentral gyrus	46	−8	62		8.06	
Right inferior precentral sulcus and IFG, pars opercularis	48	4	32		7.84	
Left frontal regions				11,144		<.0001
Left precentral gyrus	−46	−10	60		6.75	
Left inferior frontal gyrus, pars opercularis	−44	14	24		5.21	
Right lateral geniculate nucleus	12	−18	4	6752	6.61	.0004
Left lateral geniculate nucleus	−18	−32	−6	5688	7.65	.0011
Right Sylvian fissure/anterior superior temporal gyrus	54	16	−24	5208	6.68	.0018
Supplementary motor area and medial prefrontal cortex	−2	−2	52	3280	5.63	.018
Pontine tegmentum	4	−34	−36	2968	4.76	.027
Rehearse Sign						
Left fronto-parietal network				111,280		<.0001
Supplementary motor area and medial prefrontal cortex	−4	−4	54		7.27	
Left intraparietal sulcus	−28	−72	50		10.15	
Left precentral gyrus	−42	0	48		8.51	
Left middle frontal gyrus	−46	34	30		8.42	
Left inferior frontal gyrus, pars opercularis	−48	14	16		7.77	
Left occipito-temporal cortex	−58	−68	−8	17,328	7.22	<.0001
Right inferior frontal gyrus, pars opercularis	52	20	20	3336	5.24	.016
Right precentral sulcus/superior frontal sulcus	32	−6	56	2848	4.60	.032
Right intraparietal sulcus	40	−44	46	2632	3.92	.045
Conjunction of Hear Speech and See Sign						
Occipital cortex	4	−92	−2	7304	6.40	.0003
Right posterior superior temporal sulcus	62	−38	6	7096	7.69	.0003
Supplementary motor area	−2	−4	52	6400	6.18	.0006
Left superior temporal sulcus	−58	−50	6	5896	5.65	.001

Table 1. (*continued*)

Region	MNI Coordinates			Volume (*mm³*)	Max *t*	*p*
	x	*y*	*z*			
Right inferior frontal gyrus, pars opercularis	52	10	36	5648	5.17	.001
Left inferior frontal gyrus, pars opercularis	−42	10	24	3712	5.43	.01
Left cerebellum	−42	−64	−30	3600	4.89	.012
Anterior superior temporal gyrus	54	16	−10	3296	4.99	.017
Conjunction of Rehearse Speech and Rehearse Sign						
Left inferior frontal cortex				31,168		<.0001
Left inferior frontal gyrus, pars triangularis	−40	32	6		9.51	
Left inferior frontal gyrus, pars opercularis	−50	12	14		7.56	
Left precentral gyrus	−42	−2	26		4.81	
Supplementary motor area	−4	−10	66	11,864	5.95	<.0001
Left intraparietal sulcus	−30	−70	50	6376	5.80	<.0001

hemisphere—regions that were also active during the perception of speech stimuli. Conversely, active maintenance of sign stimuli recruited ventral occipital–temporal cortices in the left hemisphere—regions that were also active during the perception of sign stimuli. This result provides strong evidence for modality-specific codes in linguistic STM (Wilson, 2001). However, a left pSTS region was active during the maintenance of linguistic
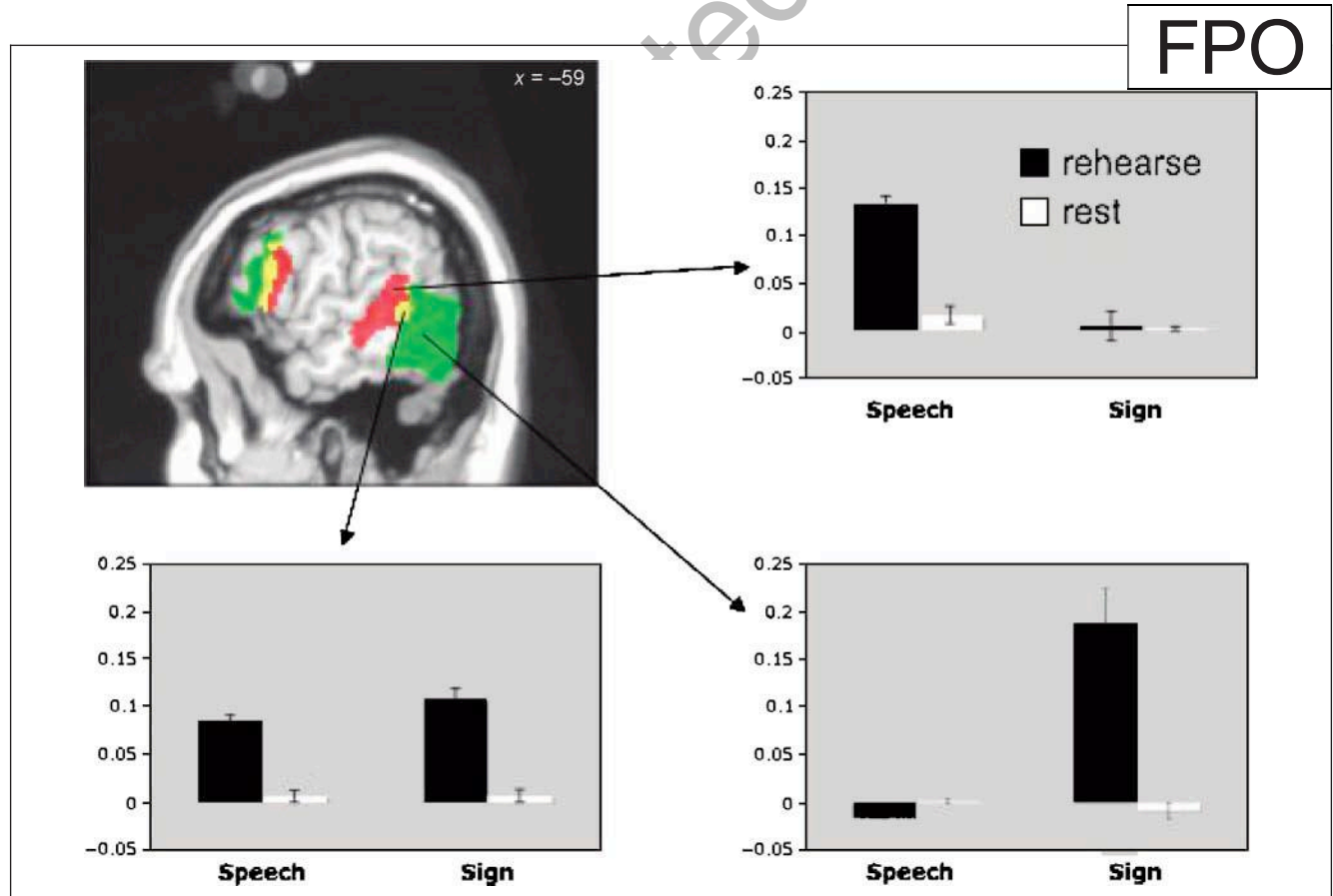

Figure 3. Activation map from a representative slice showing speech (red) versus sign (green) maintenance activations, as well as their conjunction (yellow) in the left hemisphere. Bar graphs show amplitude of the response with standard error bars during maintenance versus rest for the sign and speech conditions.

Figure 4. Time course generated by an individual-subject analysis of the conjunction of speech and sign activations during the maintenance phase of the task. Note that the amplitude of the response is roughly twice that found in time courses derived from the group analysis.
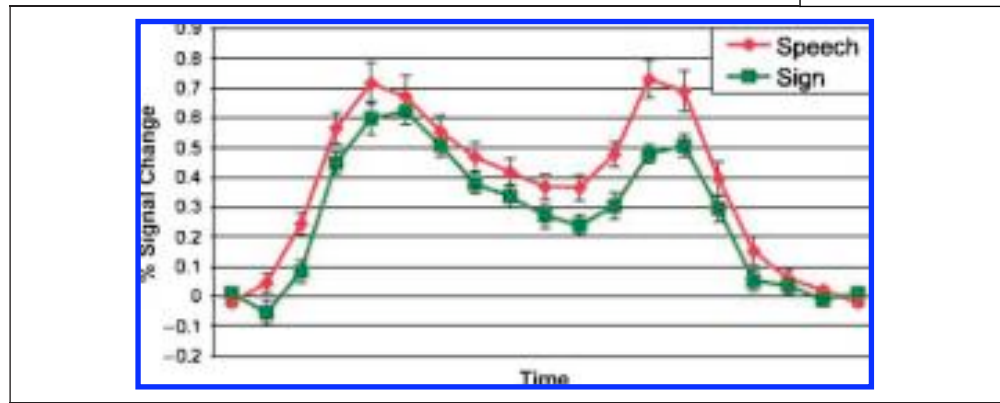
information independent of modality, raising the possibility of a modality-neutral code or process in linguistic STM (see below for further discussion).

The visual areas activated during both the perception and maintenance of sign information encompass the lateral occipital complex (LOC) and, likely, area MT. The LOC is thought to be involved in object recognition and has been shown to be most responsive to visual stimuli with clear, three-dimensional shape (Grill-Spector, Kourtzi, & Kanwisher, 2001), such as the passive viewing/ encoding of our dynamic sign stimuli in the present study. The ventral occipital–temporal cortex activation to sign perception is also consistent with previous research that found lateral occipito-temporal cortex involvement during the perception of human body parts (Downing, Jiang, Shuman, & Kanwisher, 2001). A number of studies have also implicated LOC in (nonlinguistic) visual STM (Pessoa, Gutierrez, Bandettini, & Ungerleider, 2002; Courtney, Ungerleider, Keil, & Haxby, 1997), including both encoding and maintenance (Xu & Chun, 2006). As the observed sign maintenance activations appear to involve distinct visual areas, we suggest that there may be multiple modality-specific circuits, corresponding to different sensory features (e.g.,

motion, form) that can be actively maintained in STM (Postle, 2006; Pasternak & Greenlee, 2005).

Speech-specific responses for the perception and active maintenance of speech were found in the posterior half of the left STS and STG extending dorsally into the posterior planum (Spt), consistent with previous studies (Buchsbaum, Olsen, et al., 2005; Hickok et al., 2003; Buchsbaum et al., 2001). The STS bilaterally has been implicated in aspects of phonological processing (Hickok & Poeppel, 2007), whereas the posterior planum has been implicated in auditory–motor interaction (Hickok & Poeppel, 2004, 2007; Buchsbaum, Olsen, et al., 2005; Hickok et al., 2003; Buchsbaum et al., 2001). The present findings are consistent with previous claims that STM for speech is supported by sensory/ phonological systems in the pSTS interacting with frontal motor articulatory systems via an interface region (Spt) in the posterior planum temporale (Hickok & Poeppel, 2000, 2004, 2007; Jacquemot & Scott, 2006; Hickok et al., 2003).

We also found suggestive evidence for recruitment of modality-independent networks involved in the active short-term maintenance of linguistic information. Consistent with previous studies (Buchsbaum, Pickell, et al., 2005; Ronnberg et al., 2004), we found extensive overlap

Figure 5. Activation maps associated with the perception, maintenance, and conjunction of perception and maintenance of speech versus sign stimuli.
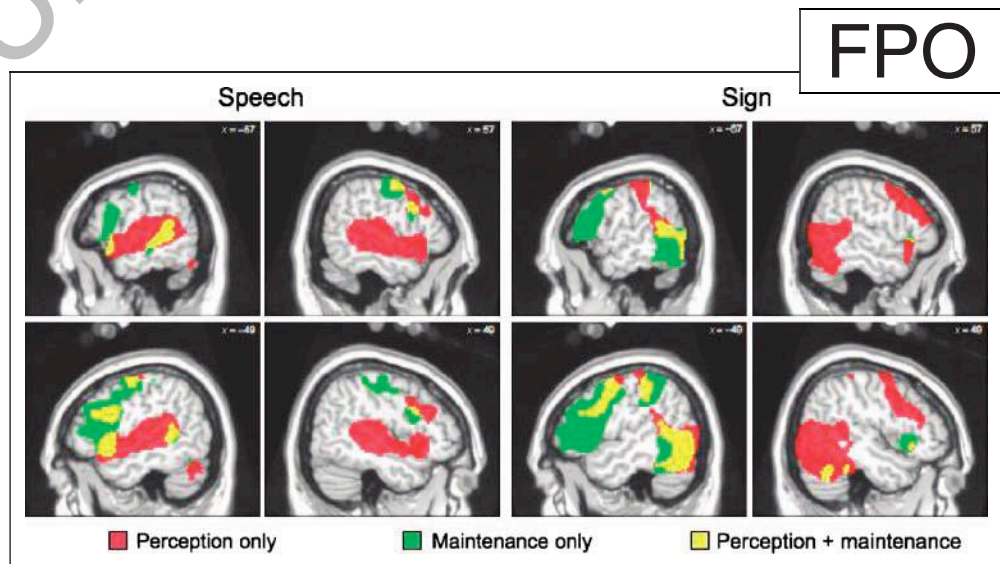
between speech and sign maintenance activations in the frontal lobe, including the left inferior frontal gyrus and left premotor regions. The overlap between speech and sign maintenance in the left frontal lobe is consistent with recent observations of overlap in this region between speech and sign in an object-naming task (Emmorey, Mehta, & Grabowski, 2007). It is unclear what functions may be supported by these regions of overlap in the frontal lobe, but candidates implicated in previous studies include articulatory rehearsal (Smith & Jonides, 1999), cognitive control functions (Thompson-Schill, Bedny, & Goldberg, 2005), and selecting or inhibiting action plans (Koechlin & Jubault, 2006).

An additional region showing coactivation for speech and sign maintenance was found in the posterior superior temporal lobe (Figure 4). This region also appeared to be sensitive to the sensory perception of both speech and sign, as evidenced by robust responses to sensory stimulation of either type, which was roughly twice the amplitude of the maintenance response. One interpretation of this finding is that this region is a polysensory field that also participates in the short-term maintenance of linguistic information. Such an interpretation is consistent with the demonstration of multisensory organization of the pSTS by Beauchamp et al. (2004) where high spatial resolution fMRI revealed small (subcentimeter) intermixed patches of the cortex that are sensitive to auditory stimulation alone, visual stimulation alone, and both auditory and visual stimulation. Using more typical, lower-resolution imaging such as that employed in the present study, this region of the STS appears uniformly multisensory (Beauchamp et al., 2004), consistent with our findings for joint speech/sign activation. If this analysis is correct, our study adds to the work on the multisensory organization of the STS by demonstrating its mnemonic involvement. It is unclear what kind of information might be coded in this region, however. As both speech and sign stimuli involve phonological representations, one might be tempted to conclude that this modality-independent area is coding information at that level. However, as noted above, sensory interference with short-term maintenance of linguistic information appears to be modality specific (Wilson & Emmorey, 2003). Further, as noted above, there is no a priori reason why phonological representations need be modality-independent. This pSTS region has been found to activate during the perception of biological motion (Grossman et al., 2000) and to the perception of faces (Beauchamp et al., 2004), two domains that may link speech and sign perception. Another possibility is that the posterior superior temporal region is coding some form of sequence information for events (Jones & Macken, 1996). Additional data on the response properties of this area are needed to understand its functional role in speech/sign processing.

As discussed in the Introduction, previous studies of STM for sign language failed to find maintenance activity in visual-related regions. This is somewhat surprising, particularly in the case of the Buchsbaum, Pickell, et al. (2005) study, which used the same stimuli and a similar design. There are several possible explanations for the discrepancy. First, the present study, unlike that of Buchsbaum, Pickell, et al., required an overt button-press response indicating whether the list maintained in STM was in the same order as a probe list. This may have induced increased task vigilance and, therefore, amplified maintenance-related activity in visual areas. Another possibility is the different subject groups that participated in the two studies. Buchsbaum, Pickell, et al. studied native deaf signers, whereas we studied native hearing signers. Sensory deprivation is known to affect brain organization for perceptual functions (Corina et al., 2007; Finney, Clementz, Hickok, & Dobkins, 2003) and so this must be considered as a possible explanation. Nonetheless, even if this is the explanation for the discrepancy, it does not detract from the significance of our findings. In fact, one could argue that demonstrating sensory-specific STM effects in hearing, bilingual participants makes a stronger case for sensory coding of linguistic STM because one cannot explain our findings by appealing to plastic reorganization of STM systems resulting from sensory deprivation.

In summary, the present experiment provides strong support for models of STM that posit modality-specific sensory storage systems (Postle, 2006; Wilson, 2001; Fuster, 1995). These systems appear to be a subset of those involved in the sensory processing of the information that is to be remembered. We suggest further that these modality-specific STM systems involve multiple, feature-specific circuits (e.g., for motion vs. form in the vision domain; Postle, 2006; Pasternak & Greenlee, 2005). However, the present study, as well as previous work comparing linguistic STM for speech and sign, also provides evidence for some form of modality-independent processes. Thus, STM for linguistic information is supported by a complex network of circuits, perhaps coding different aspects of the stimuli, rather than a single dedicated STM circuit.

Note

1. The common term "verbal" short-term memory is often interpreted as specifically tied to speech-related language systems. We use the term "linguistic" to refer to language-related short-term memory generally.

REFERENCES

Andersen, R. (1997). Multimodal integration for the representation of space in the posterior parietal cortex. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 352,* 1421–1428.

Baddeley, A. D. (1992). Working memory. *Science, 255,* 556–559.

Baddeley, A. D., & Larsen, J. D. (2007a). The phonological loop unmasked? A comment on the evidence for a ''perceptual–gestural'' alternative. *Quarterly Journal of Experimental Psychology (Colchester), 60,* 497–504.

Baddeley, A. D., & Larsen, J. D. (2007b). The phonological loop: Some answers and some questions. *Quarterly Journal of Experimental Psychology (Colchester), 60,* 512–518.

Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H., & Martin, A. (2004). Unraveling multisensory integration: Patchy organization within human STS multisensory cortex. *Nature Neuroscience, 7,* 1190–1192.

Boutla, M., Supalla, T., Newport, E. L., & Bavelier, D. (2004). Short-term memory span: Insights from sign language. *Nature Neuroscience, 7,* 997–1002.

Buchsbaum, B., Hickok, G., & Humphries, C. (2001). Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cognitive Science, 25,* 663–678.

Buchsbaum, B., Pickell, B., Love, T., Hatrak, M., Bellugi, U., & Hickok, G. (2005). Neural substrates for verbal working memory in deaf signers: fMRI study and lesion case report. *Brain and Language, 95,* 265–272.

Buchsbaum, B. R., Olsen, R. K., Koch, P., & Berman, K. F. (2005). Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. *Neuron, 48,* 687–697.

Chaminade, T., Meltzoff, A. N., & Decety, J. (2005). An fMRI study of imitation: Action representation and body schema. *Neuropsychologia, 43,* 115–127.

Corina, D., Chiu, Y. S., Knapp, H., Greenwald, R., San Jose-Robertson, L., & Braun, A. (2007). Neural correlates of human action observation in hearing and deaf subjects. *Brain Research, 1152,* 111–129.

Courtney, S. M., Ungerleider, L. G., Keil, K., & Haxby, J. V. (1997). Transient and sustained activity in a distributed neural system for human working memory. *Nature, 386,* 608–611.

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science, 293,* 2470–2473.

Duvernoy, H. M. (1999). *The human brain: Surface, three-dimensional sectional anatomy with MRI, and blood supply*. New York: Springer-Verlag Wein.

Emmorey, K. (1995). Processing the dynamic visual–spatial morphology of signed languages. In L. B. Feldman (Ed.), *Morphological aspects of language processing: Crosslinguistic perspectives* (pp. 29–54). Hillsdale, NJ: Erlbaum.

Emmorey, K. (2002). *Language, cognition, and the brain: Insights from sign language research*. Mahwah, NJ: Erlbaum.

Emmorey, K., Mehta, S., & Grabowski, T. J. (2007). The neural correlates of sign versus word production. *Neuroimage, 36,* 202–208.

Finney, E. M., Clementz, B. A., Hickok, G., & Dobkins, K. R. (2003). Visual stimuli activate auditory cortex in deaf subjects: Evidence from MEG. *NeuroReport, 14,* 1425–1427.

Fuster, J. M. (1995). *Memory in the cerebral cortex*. Cambridge: MIT Press.

Griffiths, T. D., & Warren, J. D. (2002). The planum temporale as a computational hub. *Trends in Neurosciences, 25,* 348–353.

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research, 41,* 1409–1422.

Grossman, E., Donnelly, M., Price, R., Pickens, D., Morgan, V., Neighbor, G., et al. (2000). Brain areas involved in perception of biological motion. *Journal of Cognitive Neuroscience, 12,* 711–720.

Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory–motor interaction revealed by fMRI: Speech, music, and working memory in area Spt. *Journal of Cognitive Neuroscience, 15,* 673–682.

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences, 4,* 131–138.

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

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8,* 393–402.

Holmes, C. J., Hoge, R., Collins, L., Woods, R., Toga, A. W., & Evans, A. C. (1998). Enhancement of MR images using registration for signal averaging. *Journal of Computer Assisted Tomography, 22,* 324–333.

Jacquemot, C., & Scott, S. K. (2006). What is the relationship between phonological short-term memory and speech processing? *Trends in Cognitive Sciences, 10,* 480–486.

Jenkinson, M., & Smith, S. M. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis, 5,* 143–156.

Jones, D. M., Hughes, R. W., & Macken, W. J. (2006). Perceptual organization masquerading as phonological storage: Further support for a perceptual–gestural view of short-term memory. *Journal of Memory and Language, 54,* 265–281.

Jones, D. M., Hughes, R. W., & Macken, W. J. (2007). The phonological store abandoned. *Quarterly Journal of Experimental Psychology (Colchester), 60,* 505–511.

Jones, D. M., & Macken, W. J. (1996). Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 369–381.

Jones, D. M., Macken, W. J., & Nicholls, A. P. (2004). The phonological store of working memory: Is it phonological and is it a store? *Journal of Experimental Psychology. Learning, Memory, and Cognition, 30,* 656–674.

Jones, D. M., & Tremblay, S. (2000). Interference in memory by process or content? A reply to Neath. *Psychonomic Bulletin & Review, 7,* 550–558.

Koechlin, E., & Jubault, T. (2006). Broca's area and the hierarchical organization of human behavior. *Neuron, 50,* 963–974.

MacSweeney, M., Woll, B., Campbell, R., McGuire, P. K., David, A. S., Williams, S. C., et al. (2002). Neural systems underlying British Sign Language and audio-visual English processing in native users. *Brain, 125,* 1583–1593.

Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action*. Oxford: Oxford University Press.

Neath, I. (2000). Modeling the effects of irrelevant speech on memory. *Psychonomic Bulletin & Review, 7,* 403–423.

Pasternak, T., & Greenlee, M. W. (2005). Working memory in primate sensory systems. *Nature Reviews Neuroscience, 6,* 97–107.

Peigneux, P., Van der Linden, M., Garraux, G., Laureys, S., Degueldre, C., Aerts, J., et al. (2004). Imaging a cognitive model of apraxia: The neural substrate of gesture-specific cognitive processes. *Human Brain Mapping, 21,* 119–142.

Pessoa, L., Gutierrez, E., Bandettini, P., & Ungerleider, L. (2002). Neural correlates of visual working memory: fMRI amplitude predicts task performance. *Neuron, 35,* 975–987.

Petitto, L. A., Zatorre, R. J., Gauna, K., Nikelski, E. J., Dostie, D., & Evans, A. C. (2000). Speech-like cerebral activity in profoundly deaf people processing signed languages: Implications for the neural basis of human language. *Proceedings of the National Academy of Sciences, U.S.A., 97,* 13961–13966.

Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience, 139,* 23–38.

Ronnberg, J., Rudner, M., & Ingvar, M. (2004). Neural correlates of working memory for sign language. *Brain Research, Cognitive Brain Research, 20,* 165–182.

Ruchkin, D. S., Grafman, J., Cameron, K., & Berndt, R. S. (2003). Working memory retention systems: A state of activated long-term memory. *Behavioral and Brain Sciences, 26,* 709–777.

Smith, E. E., & Jonides, J. (1999). Storage and executive processing the frontal lobes. *Science, 283,* 1657–1661.

Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping, 17,* 143–155.

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., et al. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage, 23(Suppl. 1),* S208–S219.

Stevens, A. A. (2004). Dissociating the cortical basis of memory for voices, words and tones. *Cognitive Brain Research, 18,* 162–171.

Thompson-Schill, S. L., Bedny, M., & Goldberg, R. F. (2005). The frontal lobes and the regulation of mental activity. *Current Opinion in Neurobiology, 15,* 219–224.

Wilson, M. (2001). The case for sensorimotor coding in working memory. *Psychonomic Bulletin & Review, 8,* 44–57.

Wilson, M., & Emmorey, K. (1997). A visuospatial ''phonological loop'' in working memory: Evidence from American Sign Language. *Memory & Cognition, 25,* 313–320.

Wilson, M., & Emmorey, K. (1998). A ''word length effect'' for sign language: Further evidence for the role of language in structuring working memory. *Memory & Cognition, 26,* 584–590.

Wilson, M., & Emmorey, K. (2003). The effect of irrelevant visual input on working memory for sign language. *Journal of Deaf Studies and Deaf Education, 8,* 97–103.

Worsley, K. J., Liao, C., Aston, J., Petre, V., Duncan, G. H., Morales, F., et al. (2002). A general statistical analysis for fMRI data. *Neuroimage, 15,* 1–15.

Xu, Y., & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature, 440,* 91–95.