Auditory–Motor Interaction Revealed by fMRI: Speech, Music, and Working Memory in Area Spt

Gregory Hickok, Bradley Buchsbaum, Colin Humphries, and Tugan Muftuler

Abstract

■ The concept of auditory–motor interaction pervades speech science research, yet the cortical systems supporting this interface have not been elucidated. Drawing on experimental designs used in recent work in sensory–motor integration in the cortical visual system, we used fMRI in an effort to identify human auditory regions with both sensory and motor response properties, analogous to single-unit responses in known visuomotor integration areas. The sensory phase of the task involved listening to speech (nonsense sentences) or music (novel piano melodies); the "motor" phase of the task involved covert rehearsal/humming of the auditory stimuli. A small set of areas in the superior temporal and temporal–parietal cortex responded both during the listening phase and the rehearsal/humming phase. A left lateralized region in the posterior Sylvian fissure at the parietal–temporal boundary, area Spt, showed particularly robust responses to both phases of the task. Frontal areas also showed combined auditory + rehearsal responsivity consistent with the claim that the posterior activations are part of a larger auditory–motor integration circuit. We hypothesize that this circuit plays an important role in speech development as part of the network that enables acoustic–phonetic input to guide the acquisition of language-specific articulatory-phonetic gestures; this circuit may play a role in analogous musical abilities. In the adult, this system continues to support aspects of speech production, and, we suggest, supports verbal working memory. ■

INTRODUCTION

Researchers in several fields have postulated a link between auditory and motor representations of speech: Neurological tradition dictates that there is a connection between the left posterior auditory fields (Wernicke's area) and the left frontal articulatory systems (Broca's area), which support aspects of speech production including the ability to repeat heard speech (Benson et al., 1973). In speech perception research, the idea of articulatory-based representations supporting speech perception (the Motor Theory; Liberman & Mattingly, 1985) has a long history and has garnered recent attention with the discovery of mirror neurons (Rizzolatti & Arbib, 1998) ("motor" cells that respond to the perception of action). In the area of verbal working memory, scientists have argued that articulatory circuits can be used to refresh the contents of a sensory storage system (Wilson, 2001; Baddeley, 1992). In addition, workers in speech development have argued that young children must compare the speech sounds they hear in their environment with their own speech output attempts as a means of tuning their articulatory system (Doupe & Kuhl, 1999). Despite the widespread agreement that some form of auditory–motor interface system must exist, little progress has been made in mapping the neural basis of this network. In fact, Doupe and Kuhl (1999), in a discussion of this point in the context of both speech and birdsong, have stated, "Despite its clear importance, the link between perception and production is surprisingly ill understood in both speech and song systems . . ." (p. 606).

At the same time, work in the visual domain has been quite successful in identifying visuomotor interface systems in the dorsal (parietal) processing stream. Several parietal regions in monkey have been identified which appear to be optimized for interfacing visual input with various motor effector systems (Andersen, 1997; Rizzolatti, Fogassi, & Gallese, 1997). For example, area AIP contains a class of neurons which have visuomotor response properties: They respond both to the visual presentation of a 3-D object, and during grasping of that object (even when grasping is carried out in the dark) (Murata, Gallese, Kaseda, & Sakata, 1996). AIP is reciprocally connected to frontal area F5 that also contains neurons responsive during grasping or manipulation with the hand, and deactivation of either AIP or F5 produces grasping deficits (Gallese, Fadiga, Fogassi, Luppino, & Murata, 1997). AIP, then, appears to be part of a visuomotor integration circuit that relates perceptual codes for object shape/orientation to motor codes for grasping/manipulation with the hand.

University of California, Irvine

Journal of Cognitive Neuroscience 15:5, pp. 673–682

Guided by the recent success in mapping visuomotor integration systems, we have recently hypothesized that an auditory–motor interface system is located in the inferior parietal lobe in humans, as part of the auditory dorsal stream (Hickok & Poeppel, 2000). Consistent with this proposal is the observation that the inferior parietal lobe is activated during verbal working memory tasks (Jonides et al., 1998), tasks that can be viewed as involving a kind of auditory–motor integration (Wilson, 2001). Although most authors attribute the parietal lobe activation to the operations of a verbal "storage" component (the "phonological store"), we have suggested instead that it reflects an auditory–motor interface system which translates between auditory representations of speech in the superior temporal lobe and motor representations of speech in the frontal lobe (Hickok & Poeppel, 2000). This interpretation predicts (i) that inferior parietal areas (auditory–motor integration) as well as portions of the superior temporal lobe (sensory coding) should be active during verbal working memory tasks, and (ii) that these sites should show both auditory and motor response properties (analogous to visuomotor response properties of many AIP neurons, for example). We have reported a preliminary fMRI study which supports these predictions (Buchsbaum, Hickok, & Humphries, 2001). In that study, subjects listened to sets of three multisyllabic nonsense words (the sensory phase), and then silently rehearsed them (the motor phase)[1] for several seconds. Regions that responded to both phases of the task were found in two posterior sites in every subject: a lateral site in the posterior superior temporal sulcus (STS), and a more dorsal site in the left posterior Sylvian fissure at the parietal–temporal boundary (area Spt).

Another recent report has identified the left posterior dorsal STG (our Spt) as a site active during the motor act of speech (Wise et al., 2001). In this PET study, this region was active when subjects repeated a phrase out loud and mouthed the phrase silently, but not when subjects were asked to "think" of the phrase repeatedly. If subjects were silently rehearsing in the "think" condition, we would have expected Spt to be active, yet it was not. Thus, the evidence to date has not provided unequivocal support for the hypothesis that Spt is a region with auditory–motor response properties analogous to single-unit responses in the dorsal visual stream.

The present fMRI experiment had two goals: first, to determine whether or not Spt demonstrates auditory–motor response properties, and second, to determine the stimulus specificity of the response in area Spt by contrasting speech stimuli with melodic tonal stimuli. If this temporal–parietal site functions as a phonological store, as has been proposed, it should be less responsive in tasks that involve nonphonemic stimuli. Alternatively, if this region subserves auditory–motor integration more generally, as we have proposed, it may be equally responsive in phonemic and nonphonemic tasks.

Five subjects listened to, and then covertly rehearsed, either nonsense ("jabberwocky") sentences, or melodic tonal sequences (subjects covertly hummed the tonal sequences) while the hemodynamic response was monitored using fMRI. Using multiple regression analysis, three classes of responses where identified: "auditory," in which the MR signal increased in response to acoustic stimulation but not during the rehearsal phase; "rehearsal," in which the signal increased during the rehearsal phase but not during auditory stimulation; and, "auditory + rehearsal," in which signal increased both during the auditory and rehearsal phases of the task. Again, articulatory rehearsal was carried out subvocally, so activation during the rehearsal phase cannot be a result of hearing one's own voice. A "listen-only" condition was also included (carried out in separate runs) in which participants simply listened to the same set of materials without rehearsing. Comparing activation between the listen-only condition and the rehearse condition provided another means to identify regions with auditory + rehearsal responses.

RESULTS

Auditory responses were found bilaterally in the superior temporal lobe, but also in small foci of activation in the right frontal cortex (lateral premotor) for both stimulus classes. Rehearsal responses were found predominantly in the posterior frontal lobes (inferior and middle frontal gyri) and anterior insula, although there were foci of rehearsal activity both in the parietal (e.g., supramarginal gyrus) and in the superior temporal lobes (posterior STS) for both music and speech.

Auditory + rehearsal responses, our primary focus, were found in the left Spt in every subject for both the speech and music conditions. In the group analysis, Spt activations were more extensive in the left hemisphere for speech, as well as for music, and were centered at Talairach coordinates $x = -51$, $y = -46$, $z = 16$, and $x = -54$, $y = -39$, $z = 20$, respectively (see Figure 1). The group analysis also identified auditory + rehearsal responses more ventrally within the STS. Two STS foci were noted for speech, only in the left hemisphere, one in the posterior sector of STS ($x = -45$, $y = -54$, $z = 4$), the other in the middle sector ($x = -59$, $y = -30$, $z = 0$). Two STS activation foci were also found in the left hemisphere in similar posterior and middle locations ($x = -45$, $y = -55$, $z = 4$ and $x = -59$, $y = -32$, $z = 4$, respectively), and an additional focus for music was observed in the right hemisphere ($x = 53$, $y = -47$, $z = 4$), but see below for further discussion of hemisphere differences in the STS. Auditory and rehearsal foci were also noted in the posterior frontal cortex including the lateral premotor and inferior frontal gyrus, consistent with our previous report (Buchsbaum et al., 2001) (see Figure 1). Because the focus of the present study is on posterior temporal activations, these frontal activations will not be discussed further. The pattern of activity in the posterior temporal lobes seen in the group analysis held up in individual subject data: We performed an analysis of variance (ANOVA) with Hemisphere (left, right), Area (Spt, STS), and Condition (speech, music) entered as factors, and number of activated voxels calculated separately for each subject entered as the dependent variable. The main effects of Hemisphere, Area, and Condition failed to reach significance ($p > .10$), but there was a significant Hemisphere × Area interaction, $F(1,4) = 9.34$, $p = .038$, which reflected the fact that there was much greater activation in the left Spt (for both conditions) than in the right Spt, left STS or right STS, all of which responded roughly equally (Figure 2). The three-way interaction also reached significance, $F(1,4) = 20.41$, $p = .01$, reflecting the fact that responses across the two conditions and regions in the right hemisphere were similar whereas the responses in the left hemisphere varied by condition and region: in particular, the difference between Spt and STS was greater for music than for speech in the left

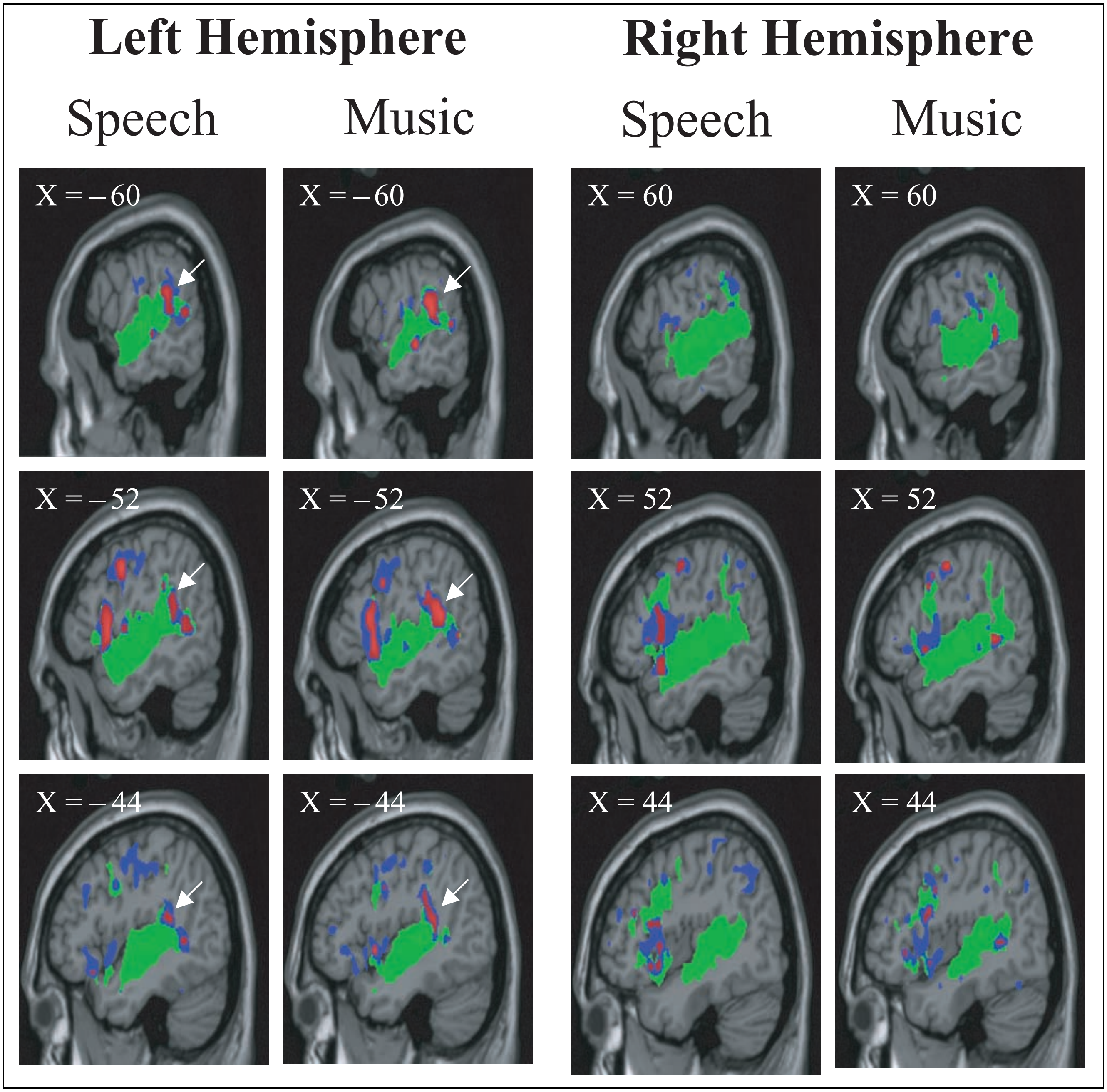

Figure 1. Activation maps in the lateral left and right hemispheres for three types of the response patterns shown separately in the speech and music trials. ''Auditory'' responses are shown in green; ''rehearsal'' responses are shown in blue; and ''auditory + rehearsal'' responses are shown in red. Arrows indicate the location of area Spt. See text for details.

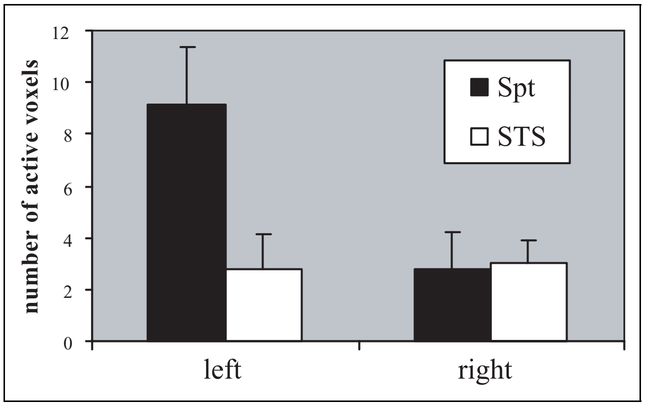

Figure 2. Average number of voxels passing threshold for auditory + rehearsal responses across 5 subjects in area Spt versus STS and in the left versus the right hemisphere, collapsed across stimulus conditions.

hemisphere (mean activated voxels in the left hemisphere for music: Spt = 11.2, STS = 1.8, and for speech: Spt = 7.0, STS = 4.0; in the right hemisphere for music: Spt = 3.2, STS = 3.8, and for speech: Spt = 2.4, STS = 2.2). Finally, the group-averaged activation maps show what appears to be a Stimulus × Hemisphere interaction in the STS: Speech appears to activate the left STS more than music, whereas music appears to activate the right STS more than speech. However, when active voxels in the STS were tabulated individually in each subject and submitted to an ANOVA, the difference was not statistically significant. This suggests similar degrees of activation in the STS on both sides by music and speech (unlike Spt activation, which is clearly larger on the left for both conditions).

Both stimulus types yielded robust auditory + rehearsal responses in left Spt. Are there magnitude differences that might suggest some degree of specialization for speech/phonemic stimuli? To test this possibility directly, we first identified the spatial extent of Spt in each subject individually during speech trials. The response during speech trials was then compared with the response during the music trials in these same speech-defined voxels. Thus, we are testing whether the region that has the best auditory + rehearsal response for speech is speech specific or whether it also demonstrates an auditory + rehearsal response for music. The timecourse of the two stimuli in this speech-defined region was nearly identical (Figure 3A, red traces), confirming that speech-defined Spt activates robustly to nonspeech stimuli. We also, however, identified the spatial extent of Spt in each subject individually during the music trials. The response during music trials was then compared with the response during the speech trials in these same music-defined voxels. (The set of voxels picked out by the speech versus music trials overlapped but were nonidentical as careful inspection of Figure 1 demonstrates.) Thus, we are testing whether the region that has the best auditory + rehearsal response for music is music specific or whether it also demonstrates an auditory + rehearsal response for speech. The timecourse of the two stimuli in this music-defined region was not identical. Although both stimulus classes showed an auditory + rehearsal response, it was weaker for the speech stimuli in this music-defined region (Figure 3B, red traces). This suggests a modest degree of stimulus specificity in music-defined Spt.

To ensure that area Spt truly does show auditory and rehearsal responses (rather than, for example, just a slower-than-normal return to baseline levels after auditory stimulation), we examined the response pattern when subjects were asked to rehearse during a portion of the trial compared with the response pattern when subjects were asked to simply listen during the whole trial. If Spt is indeed responsive both during the auditory and rehearsal phases of the listen + rehearse trials, we should see substantially less activation in the listen-only trials during that period when subjects were rehearsing on listen + rehearse trials. In fact, in the listen-only condition, we would expect the activation to drop back down to the same baseline levels found during the "rest" period. This is indeed what we found. Figure 3 shows the timecourse of activation in the listen-only trials (black traces) in the same voxels that showed the best auditory + rehearsal responses for speech (Figure 3A) and music (Figure 3B). The observed differences between listen + rehearse and listen-only during the "rehearse" phase of the trial proved highly reliable for both the music and speech trials (repeated-measures ANOVA on data collected between 16 and 20 sec from trial onset, $p = .006$ and $p = .01$, respectively).

To provide one further confirmation of our finding that the left Spt shows auditory + rehearsal responses, the experiment was replicated (except for the listen-only trials) on a high-field (3 T) MRI system with four additional participants, and regression analyses performed as with the 1.5-T experiment. Again auditory + rehearsal responses were found in each participant in the left Spt (average coordinate $x = -48$, $y = -44$, $z = 19$). And again, analyses looking at the timecourse of activation in Spt (defined using speech trials) showed nearly identical responses for both conditions (Figure 4), confirming that speech-defined Spt is not speech specific.

DISCUSSION

This study was aimed at addressing the question of whether an auditory–motor interface system exists in the posterior Sylvian cortex (area Spt). We used two types of auditory stimuli, music and speech, and used covert rehearsal as the motor–articulatory task. The present finding of auditory–motor responsivity in the left area Spt (i.e., responses during both the auditory and rehearsal phases of the trial) parallels well-documented visual–motor response properties in the dorsal visual stream (Andersen, 1997; Rizzolatti et al., 1997), and therefore supports (i) the hypothesis that this region is part of the auditory dorsal stream, and (ii) the claim that one function of this stream is auditory–motor integration (Hickok & Poeppel, 2000). The fact that both speech and melodic stimuli activate Spt suggests that this system is not speech-dedicated, and argues against the possibility that activations in this region in verbal working memory tasks reflect the operation of a specifically phonological store. The fact that there were some voxels in Spt that responded better to music than to speech may indicate some degree of specificity for tonal stimuli within portions of Spt. Additional work will need to be performed to test this possibility thoroughly.

Damage to the left posterior Sylvian region is associated with conduction aphasia, a syndrome in which

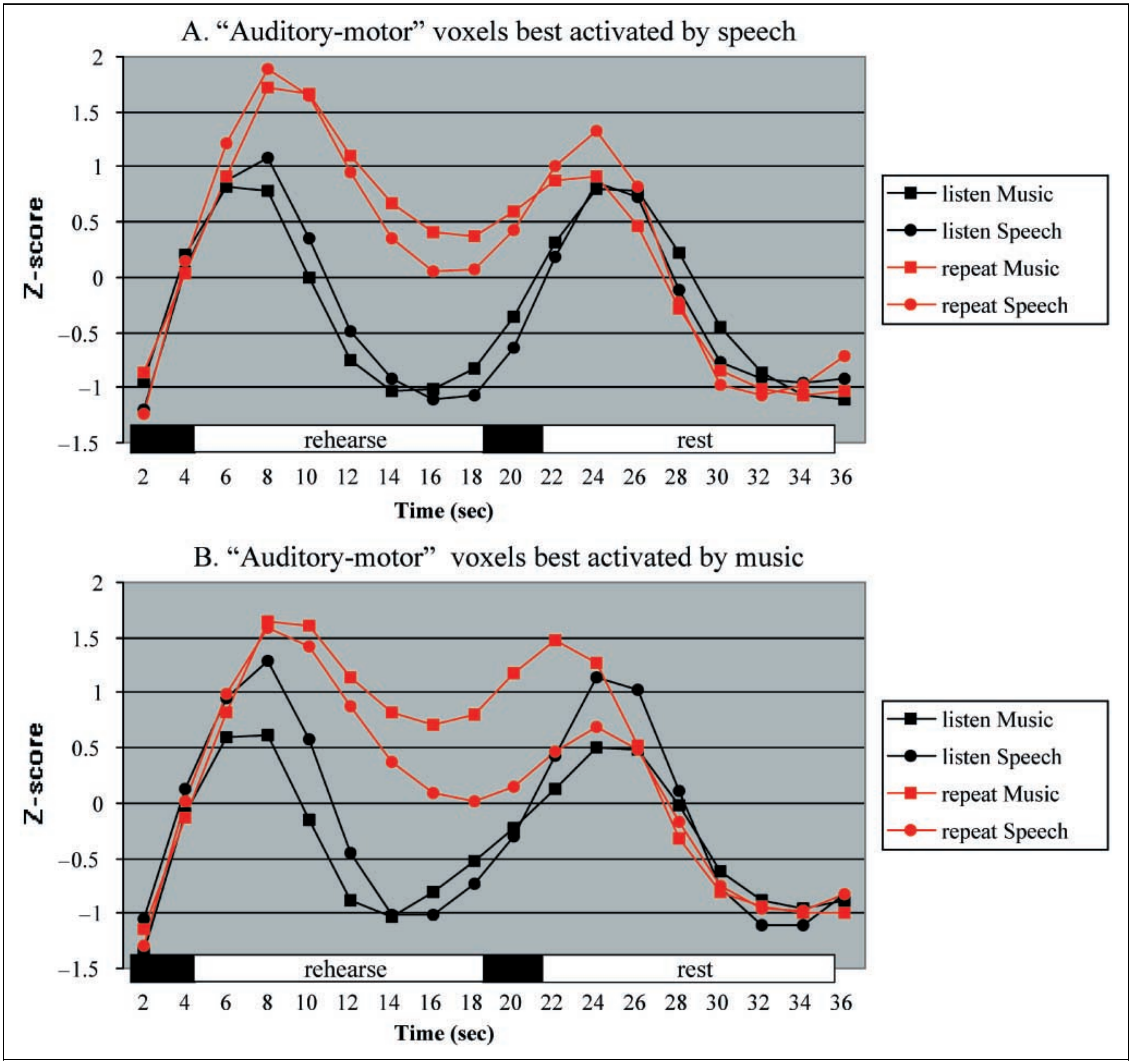

Figure 3. (A) Averaged timecourse of activation over the course of a trial in area Spt (functionally defined using the speech condition) for speech and music (24 trials per condition, 5 subjects). Timeline at the bottom shows the structure of each trial; black bars indicate auditory stimulus presentation. Red traces indicate activation during listen + rehearse runs, black traces indicate activation during listen-only runs in which subjects did not rehearse stimuli at all. Note the auditory response is reflected in the two peaks of activation associated with auditory stimulation, and the rehearsal response is reflected both by the difference in activation level between the rehearsal period and the rest period in the red traces, as well as the difference in activation between the red traces (rehearse) and the black traces (listen only). (B) Averaged timecourse of activation over the course of a trial in area Spt (functionally defined using the music condition) for speech and music (24 trials per condition, 5 subjects).

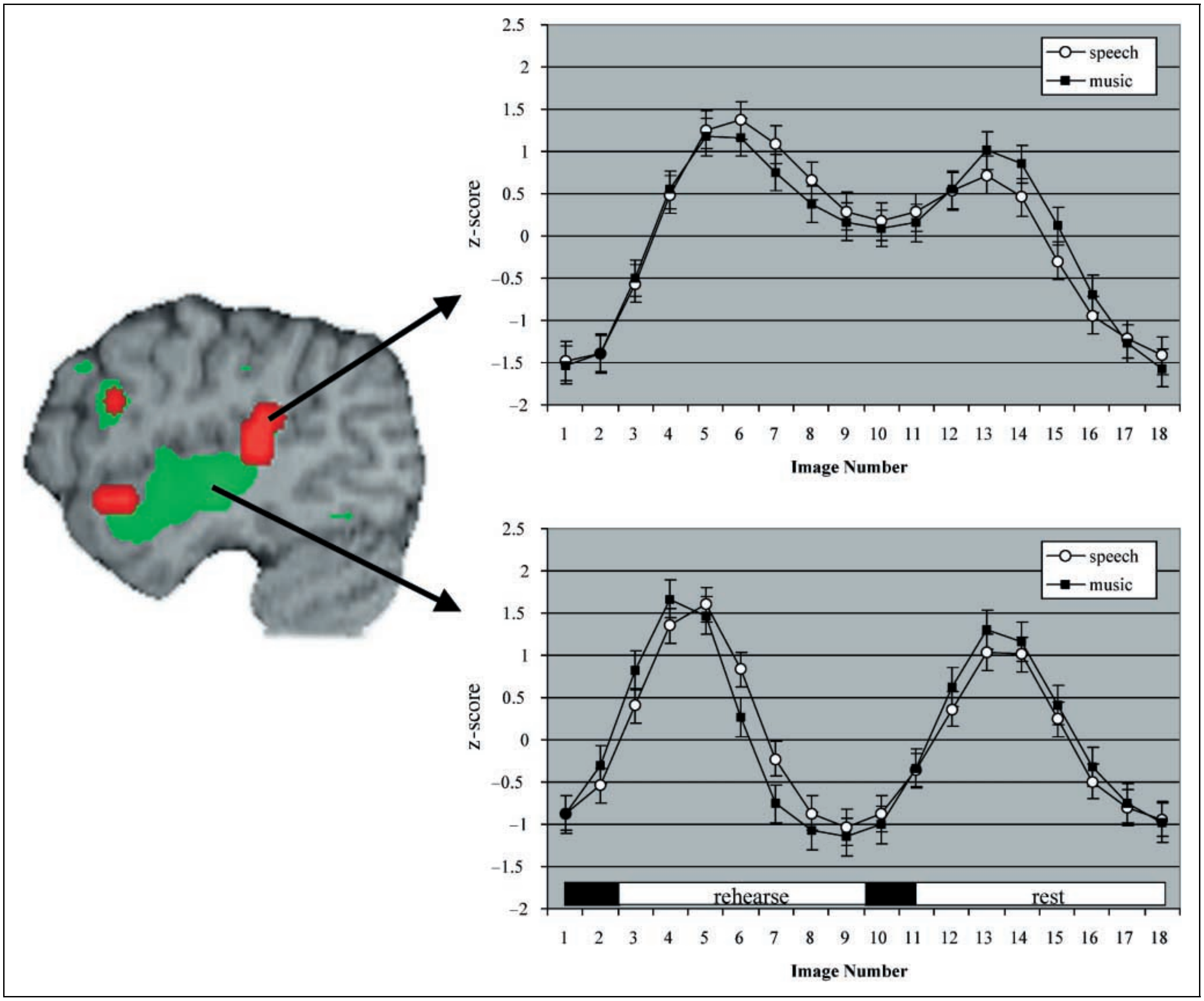

Figure 4. High-field fMRI replication. Brain image shows the activation map on a single left hemisphere slice in a representative subject (green = ''auditory'' responses; red = ''auditory + rehearsal'' responses). Timecourse data are averaged over the 4 subjects (24 trials per condition, per subject). Top panel shows ''auditory + rehearsal'' activation in area Spt, bottom panel shows ''auditory'' activation in the middle STG regions on the left.

phonemic errors in production and poor immediate verbal recall (repetition) are dominant symptoms (Goodglass, 1992). These symptoms are exacerbated with increased phonological load, as for instance, if the patient is asked to repeat multisyllabic words or pseudowords (Goodglass, 1992). Although classically considered a disconnection syndrome resulting from a white matter lesion isolating posterior language areas from anterior language areas (Geschwind, 1965), there is now good evidence supporting the view that conduction aphasia results from cortical dysfunction in posterior Sylvian regions (Hickok et al., 2000; Anderson et al., 1999). The deficit and lesion pattern associated with conduction aphasia is consistent with a disorder involving an auditory–motor integration network (Hickok, 2000). The repetition disorder can be explained straightforwardly: Immediate verbal recall requires that phonological strings be perceived and maintained in the auditory cortex and then mapped onto motor articulatory systems for production. Damage to the auditory-to-motor mapping system will interrupt this process. But why should phonemic errors occur in spontaneous speech production? The answer lies in the observation that left auditory-related cortices participate not only in speech perception but also in phonological aspects of speech production. Evidence for this idea comes from a range of studies including lesion work (for reviews, see Hickok, 2000, 2001), cortical stimulation (Anderson et al., 1999), fMRI and PET (Hickok et al., 2000; Indefrey & Levelt, 2000), and MEG (Levelt, Praamstra, Meyer, Helenius, & Salmelin, 1998). This hypothesis is also consistent with the recent claim coming from computational work that speech gestures are planned in auditory representational space and then mapped onto motor–articulatory coordinates for production (Guenther, Hampson, & Johnson, 1998). We suggest that phonological representations in auditory cortices, which are accessed during speech production, influence motor speech planning via area Spt. Therefore, conduction aphasics present with phonological errors in production for the same reason they have trouble with immediate verbal recall, namely, there is a disruption in the normal pathway by which auditory representations of speech are mapped onto motor representations of speech.

Speech-defined area Spt is also robustly activated by melodic stimuli. Do conduction aphasics have difficulty with immediate recall of novel tonal sequences? This ability is not typically tested, but present data suggest that they do. Indeed, we are aware of two reports which examined immediate auditory recall of speech and tonal stimuli—one a case study (Strub & Gardner, 1974), and one a group study of 28 aphasics including 5 conduction aphasics (Gordon, 1983)—and both report impairment for speech as well as tonal stimuli in conduction aphasia. In the group study, the conduction aphasics actually performed worse in the immediate recall of a binary sequence of tones (e.g., high–low–low–high–high) than in recalling a binary sequence of digits (e.g., 1–2–2–1–1): The means were 3.55 and 4.40, respectively. This is suggestive of a possible co-occurrence of deficits in these two abilities. This is clearly an area that deserves further investigation, particularly from a neuropsychological perspective.

If area Spt is defined using melodic stimuli, and the response to speech stimuli in that region is examined, we find a weaker motor (rehearsal phase) response with speech than with musical stimuli. The observation that there are regions of the left posterior Sylvian cortex which respond better in this way to music than to speech, and more generally, the observation that the peak auditory–motor activation for speech versus music differed in that region, suggests some degree of stimulus-dependent organization to area Spt. Thus, area Spt might be best characterized as a region that supports auditory–motor integration generally, but which is comprised of subfields that are weighted more towards one stimulus type or another.

It might be argued that rehearsal-driven activation in auditory-responsive cortical fields is simply a reflection of auditory imagery (Halpern & Zatorre, 1999), and not as we are suggesting, the operation of an auditory–motor integration network. We would agree that our temporal lobe activations are indeed reflecting a form of auditory imagery, but this interpretation is not at odds with our auditory–motor integration claim. We would argue instead that auditory imagery is supported, at least in part, by an auditory–motor integration network. That is, auditory images are reactivated via circuits that interface the auditory systems with frontal–motor (and perhaps higher-order) systems. In support of this hypothesis, we point out that previous studies of auditory imagery (e.g., see Halpern & Zatorre, 1999) find activation not only in auditory-related cortices, but also in frontal regions.

Recent functional imaging studies of verbal working memory have identified some of the same regions that showed auditory–motor responses in the present study (for a review, see Smith & Jonides, 1997). This is perhaps not surprising given that one part of the present task involved articulatory rehearsal, a subcomponent of working memory. What the present study adds to existing research on verbal working memory is the following. First, it has shown that some but not all of the cortical fields that show retention-period activity are responsive to acoustic input. If one is interested in localizing the storage component of verbal working memory, this is an important observation because the aural presentation of irrelevant speech interferes with items being retained (the "irrelevant speech effect") (Salamé & Baddeley, 1982). This effect has been used to argue that acoustic information has obligatory access to the storage component of verbal working memory, a claim that predicts that the neural substrate supporting the storage component should respond both to auditory input and during a retention interval (Becker, MacAndrew, & Fiez, 1999). We note that at least two parietal areas have been implicated in the storage component of verbal working memory, a dorsal parietal region (roughly Brodmann's area 7), and a more inferior parietal region (roughly area 40) (Becker et al., 1999). We found evidence for auditory + rehearsal activity at the temporal–parietal boundary (Spt), which may correspond to the area 40 site found previously, but found no evidence for auditory + rehearsal activity in the superior parietal cortex (although activations in this vicinity were noted during the rehearsal phase alone, particularly in the right hemisphere), suggesting that this region is functioning in a different capacity. This is consistent with Becker et al.'s (1999) claim that the dorsal parietal site may be involved in a higher-order operation such as some form of attention.

A second contribution of the present study to the working memory literature is the clear replication of our previous finding (Buchsbaum et al., 2001) of auditory + rehearsal activations in the temporal lobe, in particular, in the posterior STS. This site has not been among the regions typically activated in previous functional imaging studies of verbal working memory, but its involvement in a verbal working memory circuit is not wholly unexpected. For example, from a single-unit physiology perspective, it has long been known that in addition to the working memory properties of prefrontal units (Fuster, 1989), many sensory areas also show sustained activity during a retention phase in a delayed response paradigm, and temporary "lesions" (via cooling) to these areas can produce deficits on delayed response tasks, just as is the case with the prefrontal cortex (for review, see Fuster, 1995). Thus, sensory cortical areas appear to be a component of working memory circuits. Given that sensory cortices involved in acoustic–phonetic perception are located in the superior temporal lobe (Hickok & Poeppel, 2000), one might reason, by analogy, that maintenance of such information would involve activating these superior temporal fields (Hickok & Poeppel, 2000). This line of thinking views working memory as a subsystem (or an active state) of existing sensory and motor systems, rather than an autonomous buffer system dedicated solely to the temporary storage of information (Fuster, 1995).

A third contribution of this study is the demonstration that at least one component of the neural systems supporting verbal working memory (Spt) is not speech specific. This might, at first, seem at odds with the behaviorally demonstrated irrelevant speech effect, which has been claimed to hold for auditory speech but not for noise stimuli (Salamé & Baddeley, 1982). However, other behavioral studies have shown that the irrelevant speech effect is not specific to speech: Pitch-varying tonal stimuli can also interfere with serial recall (Jones & Macken, 1996). Thus, the lack of stimulus specificity in Spt is consistent with existing behavioral data.

In summary, we hypothesize that area Spt is part of a network in the auditory dorsal stream, which performs a coordinate transformation between auditory and motor representations, analogous to visual–motor transformation systems in the dorsal visual stream (Hickok & Poeppel, 2000). Area Spt is left dominant, although a right homologue did show similar but weaker responses. Other components of this network appear to involve portions of the STS (bilaterally), and some premotor areas in the frontal lobe (bilaterally, and including portions of Broca's area as well as more dorsal sites). This system, including Spt, is not exclusive to speech. We did find weak evidence, however, suggesting some degree of stimulus dependence to the organization of Spt. We suggest that this auditory–motor integration circuit plays a role in normal speech production, as evidenced by the effect of lesions to this region, and supports auditory- and speech-based forms of working memory. In addition, we hypothesize that this network plays a critical role in speech development, particularly in the process of tuning articulatory gestures to match the acoustic patterns found in the speech of the child's linguistic environment (Hickok & Poeppel, 2000; Doupe & Kuhl, 1999).

METHODS

A total of 9 right-handed subjects (5 in the 1.5-T experiment, 4 in the 3-T replication) with a mean age of 27, including 6 men and 3 women, participated after giving written informed consent.

Twenty-one short unfamiliar piano melodies, each of 3 sec duration, were recorded using a midi synthesizer. Melodies were composed of white key pitches on the piano. Each melody outlined a common major or minor chord in the system of tonal Western harmony. Durations in the sequence were chosen to sound relatively rhythmic according to typical Western rhythmic patterns. There were from 5 to 17 pitches in each melodic sequence (mean = 8 pitches/melody). Pitch durations ranged from 106 to 1067 msec (mean = 369 msec, *SD* = 208 msec). An equal number of "jabberwocky" sentences (nouns and verbs were replaced with pseudowords), also approximately 3 sec in length, were recorded and digitized using studio quality audio equipment. Stimuli were delivered through air-conduction tubes by a Macintosh Powerbook (Cupertino, CA) sitting in a room adjoining the MRI scanner.

Each trial in this event-related design consisted of an initial stimulus presentation (music or sentence; 3 sec duration), followed by a 15-sec covert rehearsal (or rest, see below) period, a re-presentation of the stimulus (signaling the subject to stop rehearsing), and ended with 15 sec of rest. The next trial was then initiated by a novel stimulus item. In another set of experimental runs, trials were formatted as described above, except that subjects were instructed to rest in between all stimulus presentations. This latter paradigm amounted to a passive listening condition. The two conditions alternated by scanning run and were counterbalanced across subjects. In both tasks (listen-only and listen-and-rehearse), a single run consisted of 7 pseudorandomized trials (music or sentence), where each full trial extended over a period of 36 sec. A minimum of 24 trials (12 music, 12 sentences) were collected for each condition in each subject. Subjects were pretrained on sample stimuli prior to fMRI testing: They heard both speech and music stimuli and were asked to rehearse the stimulus out loud. All subjects were able to perform this task efficiently. During debriefing, all subjects indicated that they performed the task as instructed and without difficulty.

Scanning was conducted at 1.5 T on a Siemens Vision scanner using a gradient-echo EPI sequence with the following parameters: TE = 40, TR = 2 sec, FOV = 256 mm, matrix = 64 by 64, voxel dimensions = 6 × 4 × 4 mm. Sixteen (6 mm thick) contiguous axial slices were acquired so as to easily cover brain regions relevant to the study. In addition, all subjects received a high-resolution (180 sagittal slices, voxel dimensions = 1 × 1 × 1 mm) MPRAGE scan for use as an anatomically detailed background for statistical activation maps.

The 3-T replication was conducted on a whole-body magnet (Magnex Scientific, Concord, CA), with a Marconi Medical EDGE console, modified to operate at 128 MHz (3 T field). fMRI data were acquired using single-shot EPI with 128 by 128 matrix size and 240 mm FOV. Slice thickness was 5 mm. Multiple 2-D slices were acquired from each subject to cover the entire brain volume. The TR was 2 sec and TE was 31 msec.

For each subject, the entire sequence of EPI image volumes were realigned to the first of the session using a

6-parameter rigid-body 3-D motion correction algorithm (Woods, Grafton, Holmes, Cherry, & Mazziotta, 1998). Time-series were then band-pass filtered (0.2–0.015 Hz) and mean-centered. Finally, after co-registering the EPI images with the structural MRI, both sets of images were then transformed into standardized Talaraich space using a 169-parameter nonlinear warping algorithm (Woods et al., 1998). No spatial smoothing was applied to minimize the possibility that we might see auditory–motor responses simply by averaging an auditory area with a motor area.

Analyses were first carried out on single-subject data using multiple regression to assess the separate contribution of task components overlapping in time. Hemodynamic responses to the auditory phase (music or sentence) and to the motor phase were modeled using gaussian-shaped (mean 4.8 sec, *SD* 1.8 sec) impulse response functions. Indicator variables were then applied to the model in order to specify the trial type (e.g., music or sentence). Thus, for each trial type, regressors were created for the two trial phases (auditory vs. motor) (see Ollinger, Shulman, & Corbetta, 2001 for a description of fMRI designs with compound events). Significance probability maps (SPM) were computed for each trial phase (music/auditory, sentence/auditory, music/motor, sentence/motor) representing (in z-transformed units) the contribution to the overall model of each individual regressor. Intersection maps could then be created (using a threshold of $p < .0001$) to reveal areas showing significant activation to two or more of the four trial components. Of specific interest in this study were the two intersection maps representing respective within-class (e.g., sentence or music) auditory + rehearse responses. A group analysis was also performed using the same procedure except for the addition of subject as a random effects variable (Pinheiro, 2000).

A possible problem with the analysis described above is that because the modeled hemodynamic responses to each phase of the trial (for either sentence or music conditions) overlap in time, any deviation in the actual hemodynamic response from the modeled response may cause a bias in the attribution of variance towards one of the two trial phases. However, if auditory + rehearse activations designated in the regression analysis were the result of this statistical artifact, a direct comparison between the trials that required rehearsal (listen and rehearse) and those that did not (listen-only) would not reveal any differences. Specifically, if variance being attributed to the "motor" regressor was in fact merely the effect of lingering activation from the auditory stimulation, one would not expect a difference in the amplitude of the hemodynamic response during the period approximately ranging from 16 to 20 sec after the onset of the first auditory stimulus. Such a time window spans that portion of the trial during which two conditions are met: (1) 16 sec represents an outside estimate of the time it would take for activity resulting from a 3-sec auditory stimulation to again reach baseline and (2) the range occurs roughly in the middle of rehearsal phase for one condition (listen-and-rehease) and during period of rest for the other (listen-only). Because a robust difference was found during this time period between listen-and-rehearse and listen-only trials, we can rule out this potential confound.

Acknowledgments

We thank Susan Anderson for her help in generating the music stimuli. This work was supported by NIH R01 DC03681.

Reprint requests should be sent to Gregory Hickok, Department of Cognitive Sciences, University of California, Irvine, CA 92697, USA, or via e-mail: gshickok@uci.edu.

The data reported in this experiment have been deposited in the fMRI Data Center (http://www.fmridc.org). The accession number is 2-2003-113H8.

Note

1. We use the term "motor" here rather loosely, and are fully aware that the response during this phase may not be exclusively motor in a strict sense.

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.

Anderson, J. M., Gilmore, R., Roper, S., Crosson, B., Bauer, R. M., Nadeau, S., Beversdorf, D. Q., Cibula, J., Rogish , M., III, Kortencamp, S., Hughes, J. D., Gonzalez Rothi, L. J., & Heilman, K. M. (1999). Conduction aphasia and the arcuate fasciculus: A reexamination of the Wernicke–Geschwind model. *Brain and Language, 70,* 1–12.

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

Becker, J. T., MacAndrew, D. K., & Fiez, J. A. (1999). A comment on the functional localization of the phonological storage subsystem of working memory. *Brain and Cognition, 41,* 27–38.

Benson, D. F., Sheremata, W. A., Bouchard, R., Segarra, J. M., Price, D., & Geschwind, N. (1973). Conduction aphasia: A clincopathological study. *Archives of Neurology, 28,* 339–346.

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.

Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: Common themes and mechanisms. *Annual Review of Neuroscience, 22,* 567–631.

Fuster, J. M. (1989). *The prefrontal cortex: Anatomy, physiology, and neuropsychology of the frontal lobe* (2nd ed.). New York: Raven Press.

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

Gallese, V., Fadiga, L., Fogassi, L., Luppino, G., & Murata, A. (1997). A parietal–frontal circuit for hand and grasping movements in the monkey: Evidence from reversible inactivation experiments. In P. Thier & H.-O. Karnath (Eds.), *Parietal lobe contributions to orientation in 3D space* (pp. 255–270). Heidelberg: Springer-Verlag.

Geschwind, N. (1965). Disconnexion syndromes in animals and man. *Brain, 88,* 237–294, 585–644.

Goodglass, H. (1992). Diagnosis of conduction aphasia. In S. E. Kohn (Ed.), *Conduction aphasia* (pp. 39–49). Hillsdale, NJ: Erlbaum.

Gordon, W. P. (1983). Memory disorders in aphasia: I. Auditory immediate recall. *Neuropsychologia, 21,* 325–339.

Guenther, F. H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review, 105,* 611–633.

Halpern, A. R., & Zatorre, R. J. (1999). When that tune runs through your head: A PET investigation of auditory imagery for familiar melodies. *Cerebral Cortex, 9,* 697–704.

Hickok, G. (2000). Speech perception, conduction aphasia, and the functional neuroanatomy of language. In Y. Grodzinsky, L. Shapiro, & D. Swinney (Eds.), *Language and the brain* (pp. 87–104). San Diego: Academic Press.

Hickok, G. (2001). Functional anatomy of speech perception and speech production: Psycholinguistic implications. *Journal of Psycholinguistic Research, 30,* 225–234.

Hickok, G., Erhard, P., Kassubek, J., Helms-Tillery, A. K., Naeve-Velguth, S., Strupp, J. P., Strick, P. L., & Ugurbil, K. (2000). 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. *Neuroscience Letters, 287,* 156–160.

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

Indefrey, P., & Levelt, W. J. M. (2000). The neural correlates of language production. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 845–865). Cambridge: MIT Press.

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.

Jonides, J., Schumacher, E. H., Smith, E. E., Koeppe, R. A., Awh, E., Reuter-Lorenz, P. A., Marshuetz, C., & Willis, C. R. (1998). The role of parietal cortex in verbal working memory. *Journal of Neuroscience, 18,* 5026–5034.

Levelt, W. J. M., Praamstra, P., Meyer, A. S., Helenius, P., & Salmelin, R. (1998). An MEG study of picture naming. *Journal of Cognitive Neuroscience, 10,* 553–567.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21,* 1–36.

Murata, A., Gallese, V., Kaseda, M., & Sakata, H. (1996). Parietal neurons related to memory-guided hand manipulation. *Journal of Neurophysiology, 75,* 2180–2186.

Ollinger, J. M., Shulman, G. L., & Corbetta, M. (2001). Separating processes within a trial in event-related functional MRI. *Neuroimage, 13,* 210–217.

Pinheiro, J. C. (2000). Mixed-effects models in S and S-PLUS. In J. C. Pinheiro & D. M. Bates (Eds.), *Statistics and computing.* New York: Springer.

Rizzolatti, G., & Arbib, M. (1998). Language within our grasp. *Trends in Neurosciences, 21,* 188–194.

Rizzolatti, G., Fogassi, L., & Gallese, V. (1997). Parietal cortex: From sight to action. *Current Opinion in Neurobiology, 7,* 562–567.

Salamé, P., & Baddeley, A. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior, 21,* 150–164.

Smith, E. E., & Jonides, J. (1997). Working memory: A view from neuroimaging. *Cognitive Psychology, 33,* 5–42.

Strub, R. L., & Gardner, H. (1974). The repetition defect in conduction aphasia: Mnestic or linguistic? *Brain and Language, 1,* 241–255.

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

Wise, R. J. S., Scott, S. K., Blank, S. C., Mummery, C. J., Murphy, K., & Warburton, E. A. (2001). Separate neural sub-systems within "Wernicke's area". *Brain, 124,* 83–95.

Woods, R. P., Grafton, S. T., Holmes, C. J., Cherry, S. R., & Mazziotta, J. C. (1998). Automated image registration: I. General methods and intrasubject, intramodality validation. *Journal of Computer Assisted Tomography, 22,* 141–154.