Attentional and linguistic interactions in speech perception

Merav Sabri,* Jeffrey R. Binder, Rutvik Desai, David A. Medler,
Michael D. Leitl, and Einat Liebenthal

Medical College of Wisconsin, Department of Neurology, MEB 4550, 8701 Watertown Plank Rd., Milwaukee, WI 53226, USA

The role of attention in speech comprehension is not well understood. We used fMRI to study the neural correlates of auditory word, pseudoword, and nonspeech (spectrally rotated speech) perception during a bimodal (auditory, visual) selective attention task. In three conditions, Attend Auditory (ignore visual), Ignore Auditory (attend visual), and Visual (no auditory stimulation), 28 subjects performed a one-back matching task in the assigned attended modality. The visual task, attending to rapidly presented Japanese characters, was designed to be highly demanding in order to prevent attention to the simultaneously presented auditory stimuli. Regardless of stimulus type, attention to the auditory channel enhanced activation by the auditory stimuli (Attend Auditory > Ignore Auditory) in bilateral posterior superior temporal regions and left inferior frontal cortex. Across attentional conditions, there were main effects of speech processing (word + pseudoword > rotated speech) in left orbitofrontal cortex and several posterior right hemisphere regions, though these areas also showed strong interactions with attention (larger speech effects in the Attend Auditory than in the Ignore Auditory condition) and no significant speech effects in the Ignore Auditory condition. Several other regions, including the postcentral gyri, left supramarginal gyrus, and temporal lobes bilaterally, showed similar interactions due to the presence of speech effects only in the Attend Auditory condition. Main effects of lexicality (word > pseudoword) were isolated to a small region of the left lateral prefrontal cortex. Examination of this region showed significant word > pseudoword activation only in the Attend Auditory condition. Several other brain regions, including left ventromedial frontal lobe, left dorsal prefrontal cortex, and left middle temporal gyrus, showed Attention × Lexicality interactions due to the presence of lexical activation only in the Attend Auditory condition. These results support a model in which neutral speech presented in an unattended sensory channel undergoes relatively little processing beyond the early perceptual level. Specifically, processing of phonetic and lexical–semantic information appears to be very limited in such circumstances, consistent with prior behavioral studies.
© 2007 Elsevier Inc. All rights reserved.

Introduction

The extent to which unattended linguistic information is automatically (implicitly) processed has been the focus of numerous studies. There is ample behavioral evidence for lexical processing following visual presentation of task-irrelevant meaningful linguistic stimuli. In the well-known Stroop task, subjects name the color of incongruent color words (e.g., the word RED printed in blue ink) more slowly than the color of squares (MacLeod, 1991; Stroop, 1935). Similarly, in implicit semantic priming tasks, subjects' classification of a visible target word is affected by the immediately preceding undetectable masked prime word (Abrams et al., 2002; Marcel, 1983).

Corroborating evidence from neuroimaging studies is also available. In a functional magnetic resonance imaging (fMRI) study of the Stroop effect, Banich and colleagues (2000) found activation in the left precuneus and left inferior and superior parietal lobe when comparing the color–word with the color–object task, both of which required color naming, suggesting additional processing of irrelevant words in the former (see also Melcher and Gruber, 2006). Similar findings were reported by Peterson and colleagues (1999) using the color–word task. In this study, results from a factor analysis revealed a "phonological encoding" factor with regional activations in the left posterior middle temporal gyrus (MTG), the precuneus, posterior superior temporal gyrus (STG), and angular and supramarginal gyri (SMG). A recent Stroop-like fMRI study reported obligatory processing of kanji and kana words in literate Japanese subjects, with similar word-related activations for size judgments, when words were task irrelevant, and for lexical decision tasks when words were relevant (Thuy et al., 2004). In a positron emission tomography (PET) study, Price and colleagues (1996) found activations associated with task-irrelevant visual word processing in the left posterior temporal lobe, the left inferior parietal lobe, the cuneus, and the left inferior frontal gyrus (IFG) when subjects were engaged in a nonlinguistic visual feature detection task (i.e., detection of one or more ascenders within word and nonword stimuli). Finally, event-related potential studies of the mismatch negativity (MMN) component provide evidence for detection of lexical or phonemic changes without directed attention

* Corresponding author. Fax: +1 414 456 6562.
  E-mail address: msabri@mcw.edu (M. Sabri).

to the auditory stream (Endrass et al., 2004; Naatanen, 2001; Pettigrew et al., 2004; Pulvermuller and Shtyrov, 2006; Pulvermuller et al., 2004; Shtyrov et al., 2004).

In the majority of these imaging studies, however, task-relevant and task-irrelevant information was delivered within the same sensory channel, such that the irrelevant information constituted a feature or set of features within the attended stimulus. Thus, whether task irrelevant lexical–semantic information is processed when truly unattended (i.e., presented in an unattended channel) cannot be answered based on the current literature. In the auditory modality, the separation of relevant and irrelevant channels is usually accomplished by employing a dichotic listening paradigm in which the relevant information is presented to one ear and the irrelevant information to the other. The extent to which information delivered to the irrelevant ear is processed was the focus of Cherry's (1953) seminal shadowing experiments. Cherry employed a dichotic listening task in which two different spoken "messages" were presented simultaneously, one to each ear. Participants were asked to pay attention to just one of the messages and to "shadow" it (i.e., repeat the message aloud). Later, when asked to describe the contents of the ignored message, participants appeared not to have noticed a change in language (from English to German), a change to speech played backward, or specific words or phrases. They did notice, however, gross physical changes in the stimuli, such as when the unattended speaker switched gender, or when unattended speech was replaced with a 400-Hz tone, suggesting that at least low-level sensory information can be perceived implicitly without directed attention. There are surprisingly few studies that employed similar selective-listening protocols using words. A recent dichotic priming study reported priming effects to be dependent on the saliency of lexical primes presented in the unattended channel (Dupoux et al., 2003). The vast majority of dichotic listening studies employing words were focused on laterality effects, however, and thus used this paradigm for studying perceptual asymmetry when attention was directed simultaneously to both channels (i.e., ears) (Vingerhoets and Luppens, 2001; Voyer et al., 2005).

Channel separation can also be achieved using a bimodal selective attention paradigm in which the relevant information is presented in one modality (e.g., visual) and the irrelevant information in another (e.g., auditory). This approach produces larger selection effects compared with the within-modality selective attention paradigm (i.e., dichotic listening) (Duncan et al., 1997; Rees et al., 2001; Treisman, 1973). Functional imaging studies using this paradigm showed increased activation in brain regions implicated in processing of the attended modality (Johnson and Zatorre, 2005; O'Leary et al., 1997; Woodruff et al., 1996). Petkov and colleagues (2004) examined BOLD activations in the auditory cortex to unattended tones during a primary visual task and by the same tones when attended. Interestingly, lateral auditory association cortex was recruited when tones were at the focus of attention, whereas more medial areas near primary auditory cortex were recruited when the tones were ignored. The question of implicit linguistic processing in the absence of directed attention remains unanswered, however, since none of these studies employed auditory words or word-like stimuli.

Our aim in the current study was to investigate implicit processing in brain regions responsive to auditory words, pseudowords, and acoustically complex nonspeech (spectrally rotated speech, Blesser, 1972). Unlike previous imaging studies in which the unattended words were presented in the same sensory modality

as the attended stimuli, in the current study the separation between attended and unattended information was maximized by presentation of the irrelevant words and rotated speech in a different sensory modality. The primary task, a one-back matching task that required a button-press response for identical consecutive Japanese characters, was presented in the visual modality, whereas the simultaneous unattended words and rotated speech were presented in the auditory modality. The visual task was intentionally very demanding (compared to the relatively easy distractor tasks used in most prior studies) to minimize allocation of attentional resources to the simultaneously presented auditory sounds. Findings of stronger activations for unattended words compared to pseudowords would support the claim that unattended auditory words are processed obligatorily at a lexical–semantic level. Alternatively, findings of similar activations for both words and pseudowords compared to nonphonemic auditory control sounds (rotated speech) would suggest obligatory processing of speech sounds only at a prelexical (phonemic) level.

Methods

Subjects

Participants were 28 healthy adults (13 women, mean age=26.5 years, SD=6.9) with no history of neurological or hearing impairments and normal or corrected-to-normal visual acuity. The participants were native English speakers, and all were right-handed according to the Edinburgh Handedness Inventory (Oldfield, 1971). Informed consent was obtained from each participant prior to the experiment, in accordance with the Medical College of Wisconsin Institutional Review Board.

Task design and procedure

The study employed an event-related design blocked by condition. There were seven fMRI runs, each of which lasted 8.93 min and included 12 blocks with both auditory and visual stimuli and six blocks with only visual stimuli. For the 12 blocks containing both auditory and visual stimuli, attention was directed either toward the sounds (Attend Auditory, 6 blocks) or toward the visual display (Ignore Auditory, 6 blocks). For the six blocks containing only visual stimuli, attention was directed toward the visual display (Visual). The task was signaled by a 4-s instruction cue presented at fixation prior to each 24-s block: "Attend Auditory" for the Attend Auditory condition, "Attend Visual" for the Ignore Auditory condition, and "Attend Visual*" for the Visual condition, with the asterisk (*) sign used in the last case to warn subjects that no auditory sounds would be presented. The order of blocks was randomized within each run. Each auditory block (Attend Auditory or Ignore Auditory) included six auditory events, for a total of 72 auditory events per run, distributed equally among three types of stimuli. Across the entire experiment, there were 84 auditory events per attentional condition per stimulus type. Subjects performed a one-back matching task, requiring them to signal any instance of an identical stimulus occurring on two consecutive trials in the attended modality. Subjects were instructed to respond as quickly and as accurately as possible, pressing a key using their index finger. In the Attend Auditory condition, subjects were instructed to keep their eyes focused on the screen while ignoring the visual characters.

Training was provided prior to the experiment, incorporating three blocks of each condition. Performance was monitored online, and additional training was provided when necessary.

Stimuli

Three types of auditory stimuli – words (Word), pseudowords (Pseudo), and spectrally rotated speech (Rotated) – were used in the experiment. Words were 140 concrete monosyllabic nouns, 3 to 6 letters long, with a minimum concreteness rating of 500 in the MRC database (Wilson, 1988) (www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm) and minimum frequency of 25 per million in the CELEX database (Baayen et al., 1995). The 140 pseudowords were chosen from the Irvine Phonotactic Online Dictionary (www.iphod.com). The pseudowords and words were matched on number of phonemes, phonological neighborhood size (total log frequency of phonological neighbors), and average biphone log frequency (average log frequency of each phoneme pair in the word). All stimuli were recorded from a male native English speaker during a single recording session.

Spectrally rotated speech (Blesser, 1972) consisted of 70 rotated words and 70 rotated pseudowords chosen randomly from the lists described above. Rotated speech is produced by a filtering process that inverts the signal in the frequency domain, preserving the acoustic complexity of the original speech but leaving phonemes largely unrecognizable. Mean durations of the words, pseudowords, and rotated speech were 653 ms, 679 ms, and 672 ms, respectively. These durations were not statistically different ($p > 0.05$). All sounds were normalized to an equivalent overall power level. Stimulus onset asynchrony (SOA) for the auditory stimuli varied randomly between 2, 4, and 6 s (mean SOA = 4 s) to better estimate the single trial BOLD response to each stimulus (Dale, 1999). Auditory stimuli were repeated only once on 17% of the trials within each condition; these repeats served as targets for the one-back task.

Visual stimuli consisted of eight randomly alternating single black Japanese characters with SOA of 500 ms, centered in the middle of the screen (stimulus duration was 100 ms and font size 72 points). Visual stimuli were repeats of the previous stimulus on 17% of the trials within each condition.

Auditory stimulation was delivered binaurally through plastic tubing attached to foam earplugs using a commercial system (Avotec, Jensen Beach, FL). Visual stimuli were rear-projected onto a screen at the subject's feet with a high-luminance LCD projector and were viewed through a mirror mounted on the MR head coil. Stimulus delivery was controlled by a personal computer running E-prime software (Psychology Software Tools, Inc., Pittsburgh, PA).

Image acquisition and analysis

Images were acquired on a 1.5-T GE Signa scanner (GE Medical Systems, Milwaukee, WI). Functional data consisted of T2*-weighted, gradient-echo, echo-planar images (TR = 2000 ms, TE = 40 ms, flip angle = 90°, NEX = 1). The 1876 image volumes were each reconstructed from 21 axially oriented contiguous slices with 3.75 × 3.75 × 5.50 mm voxel dimensions. High-resolution anatomical images of the entire brain were obtained using a 3-D spoiled gradient-echo sequence (SPGR) with 0.94 × 0.94 × 1.2 mm voxel dimensions. Head movement was minimized by foam padding.

Image analysis was conducted using the AFNI software package (Cox, 1996). Within-subject analysis consisted of spatial co-registration (Cox and Jesmanowicz, 1999) to minimize motion artifacts and voxel-wise multiple linear regression with reference functions representing the stimulus conditions compared to the baseline. We defined the baseline as the Visual condition. Six separate regressors coded the occurrence of attended and un-attended trials of each of the three auditory stimulus types. Another regressor was added to code the "resting" state (i.e., the inter-stimulus interval) between auditory stimuli in the Attend Auditory condition, as there is considerable evidence for activation of a "default" conceptual system during the awake resting state (Binder et al., 1999; McKiernan et al., 2006; Raichle et al., 2001). Translation and rotation movement parameters estimated during image registration were included in the regression model to remove residual variance associated with motion-related changes in BOLD signal, as well as regressors coding the presence of instructions and the motor responses. A Gaussian kernel of 6 mm FWHM was used for spatial smoothing. The time course shape and magnitude of the hemodynamic response was estimated from the stimulus time series using the AFNI program 3dDeconvolve. General linear tests were conducted to obtain contrasts of interest between conditions collapsed across time lags 2 and 3 (the response peak). The individual statistical maps and the anatomical scans were projected into standard stereotaxic space (Talairach and Tournoux, 1988) by linear re-sampling, and group maps were created in a random-effects analysis. The group maps were thresholded at voxel-wise $p < 0.01$ and corrected for multiple comparisons by removing clusters smaller than 19 voxels (1463 μl) to a map-wise two-tailed $p < 0.05$. This cluster threshold was determined through Monte Carlo simulations that provide the chance probability of spatially contiguous voxels exceeding the threshold. Only voxels inside the brain were used in the analyses, so that fewer comparisons are performed and a smaller volume correction is required.

Results

Behavioral performance

The d' measure of perceptual sensitivity (Macmillan and Creelman, 1991) was calculated for each condition. Overall d' in the Attend Auditory condition was 3.20 (SD = 0.43). Performance varied by auditory stimulus type ($F_{2,54} = 5.045$, $p < 0.01$), with d' for words = 3.35 (SD = 0.63), pseudowords = 3.31 (SD = 0.62), and rotated = 2.95 (SD = 0.59). Newman–Keuls post hoc (0.05 criterion level) testing indicated that the d' was significantly smaller for rotated speech compared with words and pseudowords, suggesting greater task difficulty in the former. The overall d' in the visual task was 2.70 (SD = 0.40), confirming that the visual task was indeed demanding. Performance did not differ between the Ignore Auditory ($d' = 2.72$, SD = 0.41) and Visual ($d' = 2.69$, SD = 0.40) conditions, suggesting that subjects successfully ignored the auditory sounds and focused on the visual stimuli during the Ignore Auditory condition.

fMRI

The imaging results are presented in the form of group maps of the contrasts between conditions of interest, created with a random-effects analysis and thresholded at corrected $p < 0.05$. Table 1 provides the locations of local extrema for each contrast.

Table 1

Contrast	x	y	z	Z-score	Side	Anatomical location
Attend Auditory>Ignore Auditory	43	−27	−3	6.19	R	STS
	56	−42	−5	6.06	R	Posterior MTG
	39	22	35	5.72	R	MFG
	37	47	−1	5.46	R	MFG
	26	51	10	4.48	R	SFG
	25	12	47	4.78	R	SFG/MFG
	4	35	−5	5.16	R	Inferior rostral g.
	26	−40	61	4.70	R	Postcentral g.
	41	−65	41	5.50	R	Angular g.
	56	−53	21	4.85	R	Angular g.
	6	−56	47	5.62	R	Precuneus
	16	−60	−6	6.69	R	Lingual g.
	5	−61	3	6.46	R	Lingual g.
	17	−38	−5	5.31	R	Parahippocampus
	−63	1	4	5.42	L	Anterior STG
	−49	3	−1	4.69	L	Anterior STG
	−50	−41	4	4.38	L	MTG
	−58	−32	−3	4.65	L	MTG
	−54	27	16	5.43	L	IFG
	−58	−5	11	5.14	L	IFG
	−42	53	6	4.94	L	MFG
	−40	20	38	4.70	L	MFG
	−20	30	−10	5.65	L	OFC
	−64	−25	25	4.03	L	SMG
	−49	−51	36	4.00	L	SMG
	−5	−10	41	5.07	L	Cingulate g.
	−3	−81	14	7.36	L	Cuneus
	−13	−81	33	5.54	L	Cuneus
	−16	−92	17	6.00	L	Cuneus
	−11	−63	−3	6.76	L	Lingual g.
	−10	−85	5	6.06	L	Lingual g.
Speech > Rotated	18	−82	27	4.23	R	Superior occipital g.
	47	−72	5	3.75	R	Middle occipital g.
	16	−54	56	4.23	R	Postcentral s.
	−21	37	−8	4.60	L	OFC
Rotated > Speech	57	−22	14	6.54	R	STG/PT
	24	40	11	4.37	R	MFG
	37	−58	49	4.48	R	SMG/IPS
	27	11	1	5.00	R	Putamen
	27	−57	−24	3.85	R	Cerebellum
	0	−26	26	4.37	R/L	Posterior CC
	−48	−28	9	5.38	L	STG/PT
	−10	15	45	5.31	L	SFG
	−31	14	0	4.72	L	Anterior insula
	−37	−56	39	4.84	L	SMG/IPS
	−17	−69	−34	4.44	L	Cerebellum
Attention × Speech	48	4	−24	4.18	R	Anterior MTG
	23	−14	−14	4.45	R	Hippocampus
	42	−13	−17	4.54	R	Fusiform g.
	38	−39	−16	4.26	R	Fusiform g.
	28	−25	−18	4.56	R	Collateral s.
	25	−46	−7	4.39	R	Parahippocampus
	29	−67	16	4.78	R	Middle occipital g.
	43	−73	5	5.52	R	Middle occipital g.
	30	−85	17	4.44	R	Middle occipital g.
	24	−78	29	4.27	R	Superior occipital g.
	9	−9	57	4.78	R	SMA
	5	−11	39	4.27	R	Cingulate g.
	22	−38	50	4.37	R	Postcentral g.
	25	−3	−1	3.92	R	Putamen
	60	−21	3	−4.94	R	STG
	57	−34	6	−4.42	R	STG/PT

(continued on next page)

Table 1 (*continued*)

Contrast	x	y	z	Z-score	Side	Anatomical location
Attention × Speech	35	2	52	−4.43	R	MFG
	40	5	33	−4.70	R	MFG/IFG
	42	10	21	−4.49	R	IFG
	45	19	28	−4.49	R	IFG
	1	20	47	−3.82	R	SFG
	45	−55	50	−4.80	R	SMG/IPS
	35	−56	37	−4.15	R	Angular g.
	−51	−8	−15	4.24	L	Anterior MTG
	−23	−10	−18	4.71	L	Hippocampus
	−42	−42	−18	4.15	L	Fusiform g.
	−26	−72	18	4.09	L	Middle occipital g.
	−48	−73	−8	5.29	L	Middle occipital g.
	−8	27	−16	3.91	L	OFC
	−43	30	−11	4.11	L	IFG
	−6	53	15	4.40	L	SFG
	−21	−14	52	4.25	L	SFG
	−12	61	17	4.09	L	SFG
	−19	35	−1	4.24	L	Anterior cingulate g.
	−22	−38	53	4.86	L	Postcentral g.
	−17	−56	49	4.66	L	Postcentral g.
	−15	−58	20	4.51	L	Posterior cingulate g.
	−39	−26	21	4.43	L	Parietal operculum
	−26	6	−3	4.17	L	Putamen
	−9	−55	−19	4.78	L	Cerebellum
	−45	14	26	−3.90	L	IFG
	−45	40	2	−3.70	L	IFG/MFG
	−38	−47	35	−3.94	L	SMG/IPS
Word > Pseudo	−36	15	36	3.79	L	MFG
Pseudo > Word	−60	−21	13	3.68	L	STG
Attention × Lexicality	−31	28	43	4.32	L	MFG
	−14	36	−7	3.58	L	OFC
	−61	−30	−11	3.27	L	MTG

Abbreviations: R=right; L=left; s=sulcus; g=gyrus; CC=corpus callosum; IFG=inferior frontal gyrus; IPS=intraparietal sulcus; ITG=inferior temporal gyrus; MFG=middle frontal gyrus; MTG=middle temporal gyrus; OFC=orbital frontal cortex; PT=planum temporale; SFG=superior frontal gyrus; SMA=supplementary motor area; SMG=supramarginal gyrus; STG=superior temporal gyrus; STS=superior temporal sulcus; TTG=transverse temporal gyrus.

Attention effects

To examine the main effect of attention on auditory processing, we contrasted all auditory events (words, pseudowords, and rotated speech) presented under the Attend Auditory condition with all auditory events presented under the Ignore Auditory condition. This contrast, Attend Auditory–Ignore Auditory, is presented in Fig. 1A and Table 1. Regardless of stimulus type, greater activation under the Attend Auditory condition (orange-yellow areas in Fig. 1A) was observed in bilateral posterior temporal regions, including STG, superior temporal sulcus (STS), and MTG. Additional activation involved a large region of left lateral prefrontal cortex, including the inferior and middle frontal gyri (IFG and MFG). Extensive activations also involved the medial occipital lobes (cuneus and lingual gyrus) bilaterally, extending anteriorly into the precuneus and posterior parahippocampal gyrus. These medial occipital activations, which likely reflect shifts of attention from foveal to peripheral visual locations (Brefczynski and DeYoe, 1999; Kastner et al., 1999; Tootell et al., 1998), are not relevant to the aims of this study and will not be discussed further. Additional activations involved the left frontoparietal operculum and supramarginal gyrus, right dorsal prefrontal cortex (MFG and superior frontal sulcus), bilateral suborbital sulcus and gyrus rectus, left orbitofrontal cortex (OFC), left mid-cingulate gyrus, and right angular gyrus. The reverse contrast (Ignore Auditory–Attend Auditory, blue areas in Fig. 1A) showed activations in bilateral superior, middle, and inferior occipital gyri, in the intraparietal sulcus (IPS), mainly on the right side, and in the frontal eye field bilaterally. As these activations are time-locked to presentation of the ignored auditory events, we speculate that they may reflect transient refocusing or strengthening of visual attention on the visual stimulus to avoid interference from the auditory stimuli. Separate comparisons between attended and ignored words, attended and ignored pseudowords, and attended and ignored rotated speech revealed activations in very similar areas to those observed in the overall Attend Auditory–Ignore Auditory contrast.

Speech effects

To test for the differential processing of speech sounds, we contrasted the combined word and pseudoword events with the combined rotated speech events across attentional conditions (main effect of speech). This contrast, Speech-Rotated, is presented in Fig. 1B and Table 1. Regardless of attention, greater activation for speech was observed in left OFC, right superior and middle occipital gyri, and right postcentral gyrus. Greater activation for rotated speech was observed in extensive bilateral, right-dominant dorsal temporal regions, including the transverse temporal gyrus (TTG), STG, and posterior STS. Other regions showing stronger activation for rotated speech included the anterior insula and adjacent frontal operculum bilaterally, the medial aspect of SFG and posterior corpus callosum, anterior intraparietal sulcus and

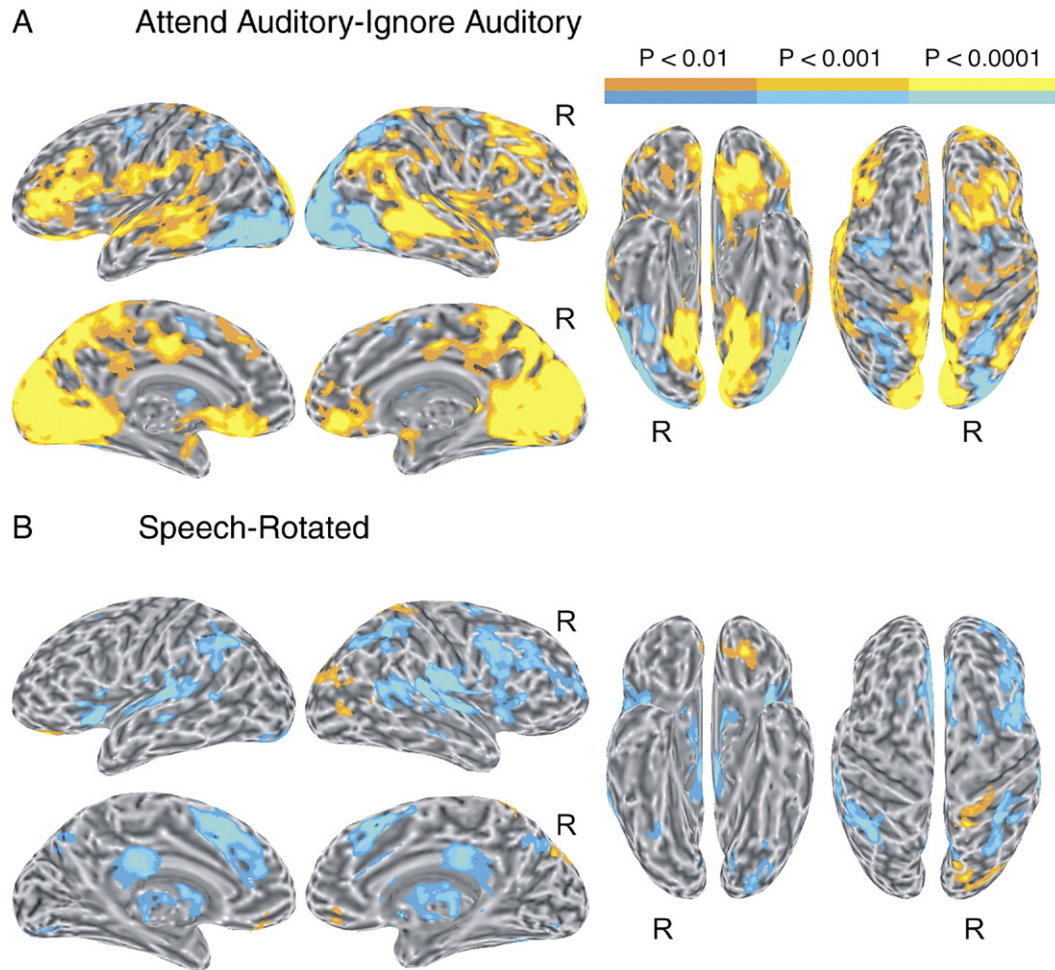## A Attend Auditory-Ignore Auditory

B Speech-Rotated

Fig. 1. (A) Brain activation for the contrast between attended and ignored auditory events (main effect of attention). (B) Brain activation for the contrast between speech (i.e., words + pseudowords) and rotated speech (main effect of speech). In all figures, left and right lateral and medial views of the inflated brain surface are shown in the left half of the figure, ventral and dorsal views in the right half. The color scale indicates voxel-wise probability values.

adjacent SMG bilaterally, mid-cingulate gyrus bilaterally, and right MFG.

Attention–speech interactions

The interaction between attention and speech ((Attend Speech–Attend Rotated)–Ignore Speech–Ignore Rotated)) is presented in Fig. 2A and Table 1. To examine the nature of the interactions, graphs showing response profiles (collapsed across time lags 2 and 3) for a sample of functionally defined representative regions in the four relevant conditions compared to baseline are plotted in Fig. 2B. Three types of interaction phenomena were observed. First, many of the regions with positive interaction values (yellow-orange colors in Fig. 2A) showed stronger activation by speech (relative to rotated speech) but only in the Attend condition. These included all of the regions that had shown main effects of speech>rotated, including left OFC and adjacent ventromedial frontal lobe (graph 1), right postcentral gyrus (graph 2), and right superior/middle occipital gyri (graph 3). Other regions showing this pattern included the left postcentral gyrus (graph 4), supplementary motor area (SMA) and adjacent cingulate gyrus, left and right anterior temporal lobe (graph 5), left hippocampus, left posterior cingulate gyrus, and left and right ventral temporal lobe (including portions of parahippocampus,

fusiform gyrus, ITG, and MTG; graph 6). Post hoc *t*-tests for dependent samples confirmed significant speech>rotated activation in the Attend condition for each of these areas, but in none of these regions was there significant speech>rotated activation in the Ignore condition. Furthermore, *t*-tests showed no significant speech> rotated activation in the Ignore condition in any of the other regions in Fig. 2A. This negative result was further confirmed by a voxel-wise contrast between ignored speech and ignored rotated speech (Supplementary Fig. 1), which showed no speech>rotated activation surviving a whole-brain corrected $p < 0.05$ threshold. Thus, we observed no evidence of speech-related activation in the Ignore condition.

A second phenomenon contributing to the interaction in many of these regions was a significant activation for rotated speech over speech only in the Ignore condition. This activation favoring rotated speech in the Ignore condition usually occurred in conjunction with activation favoring speech in the Attend condition, creating a crossover interaction. This pattern was seen in left OFC (graph 1), left postcentral gyrus (graph 4), left SMA/cingulate, left and right anterior temporal lobe (graph 5), left hippocampus, and left posterior cingulate gyrus. In a few regions, including the left parietal operculum (graph 7), left parahippocampus, left putamen, right
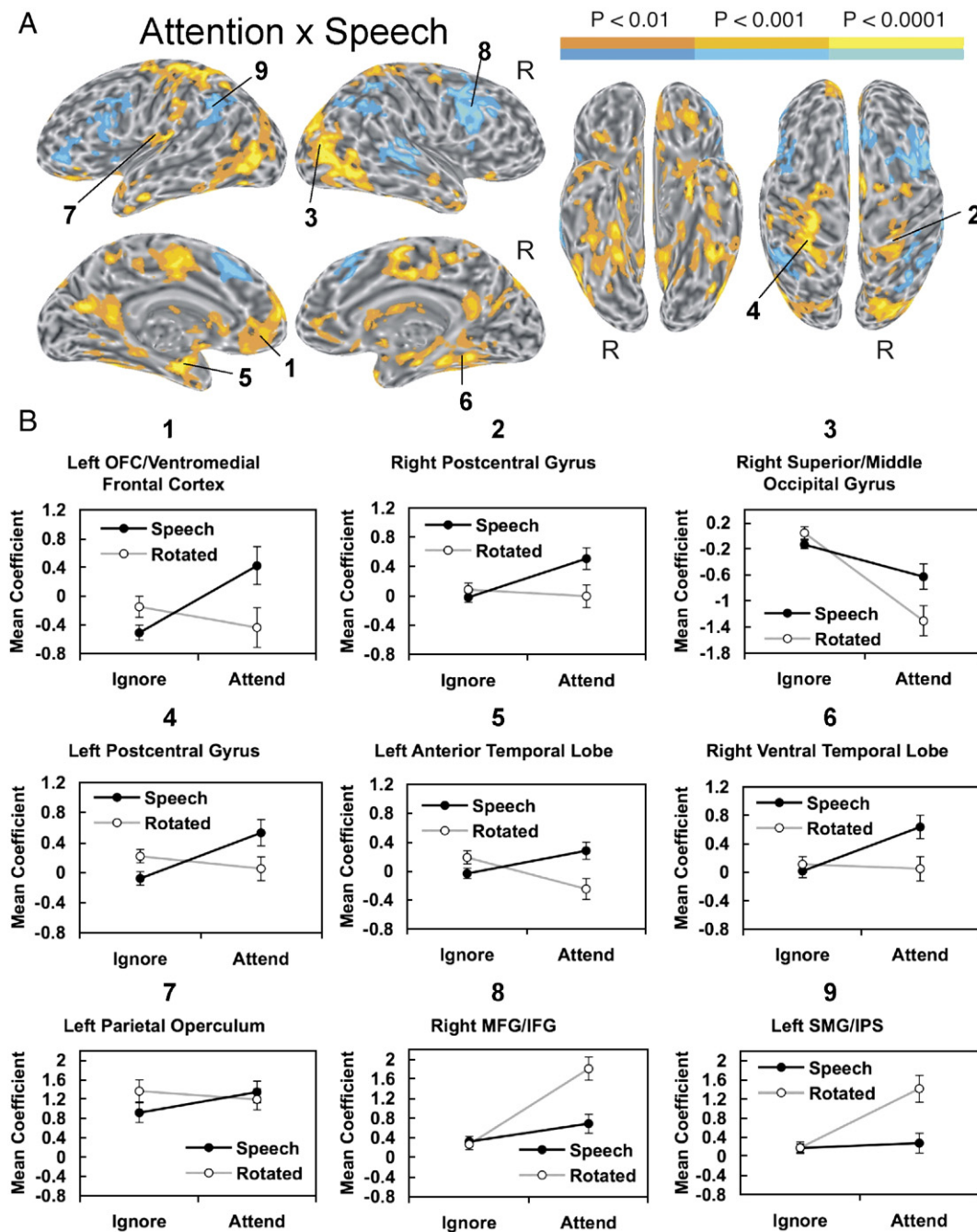
Fig. 2. (A) Brain activation for the interaction between attention and speech. (B) Response profiles (collapsed across time lags 2 and 3) with standard error bars (SEM) across subjects in representative regions across the four relevant conditions compared to baseline.

caudate, and left and right superior cerebellum, there was rotated > speech activation in the Ignore condition without any accompanying crossover (speech > rotated) activation in the Attend condition. Though unexpected, this widespread pattern of activation favoring rotated speech in the Ignore condition was striking and further supports the contention that measurable phonemic processing of speech sounds occurred only in the Attend condition.

A third phenomenon producing interaction effects in some regions was rotated > speech activation only in the Attend condition. This pattern (blue areas in Fig. 2A) was observed in left and right

MFG/IFG (graph 8), anterior inferior frontal sulcus, medial SFG, right STG and adjacent STS, and left and right SMG/IPS (graph 9). These areas largely overlapped those that had shown a main effect of rotated speech > speech (blue areas in Fig. 1B). The interaction analysis demonstrates that the activation favoring rotated speech in these areas was largely restricted to the Attend condition. Exceptions to this rule include the primary auditory regions bilaterally and the anterior insula bilaterally, all of which showed a main effect favoring rotated speech but no evidence of an Attention × Speech interaction.

Lexicality effects

To examine the hypothesized effects of lexicality, we contrasted the word with the pseudoword conditions, collapsed across attentional tasks (main effect of lexicality). This contrast, Word-Pseudo, is presented in Fig. 3A. Greater activation for words was observed in left prefrontal cortex at the junction of MFG and IFG (local maximum at −36, 15, 36). Pseudowords produced greater activation in left dorsal temporal areas, including TTG and anterior STG.

Attention–lexicality interactions

Examination of the interaction between attention and lexicality ((Attend Word–Attend Pseudo)–(Ignore Word–Ignore Pseudo)), presented in Fig. 3B and Table 1, showed significant interactions in left OFC and adjacent ventromedial frontal cortex, left MFG, and

left MTG. Graphs showing response profiles in these regions in the four relevant conditions compared to baseline are plotted in Fig. 3C. T-tests for dependent samples revealed word>pseudoword activation in all of these regions that was present only in the Attend condition and not in the Ignore condition. A significant pseudoword>word activation was present in left MTG in the Ignore condition (graph 1).

The left MFG region identified in the interaction analysis, although in close proximity, did not overlap anatomically with the slightly more posterior left MFG region that had shown a lexicality main effect (Fig. 3A). Given that our main aim in the current study is to assess the degree to which language processing occurs for unattended stimuli, we were curious to know whether the more posterior MFG region showing a main effect of lexicality was
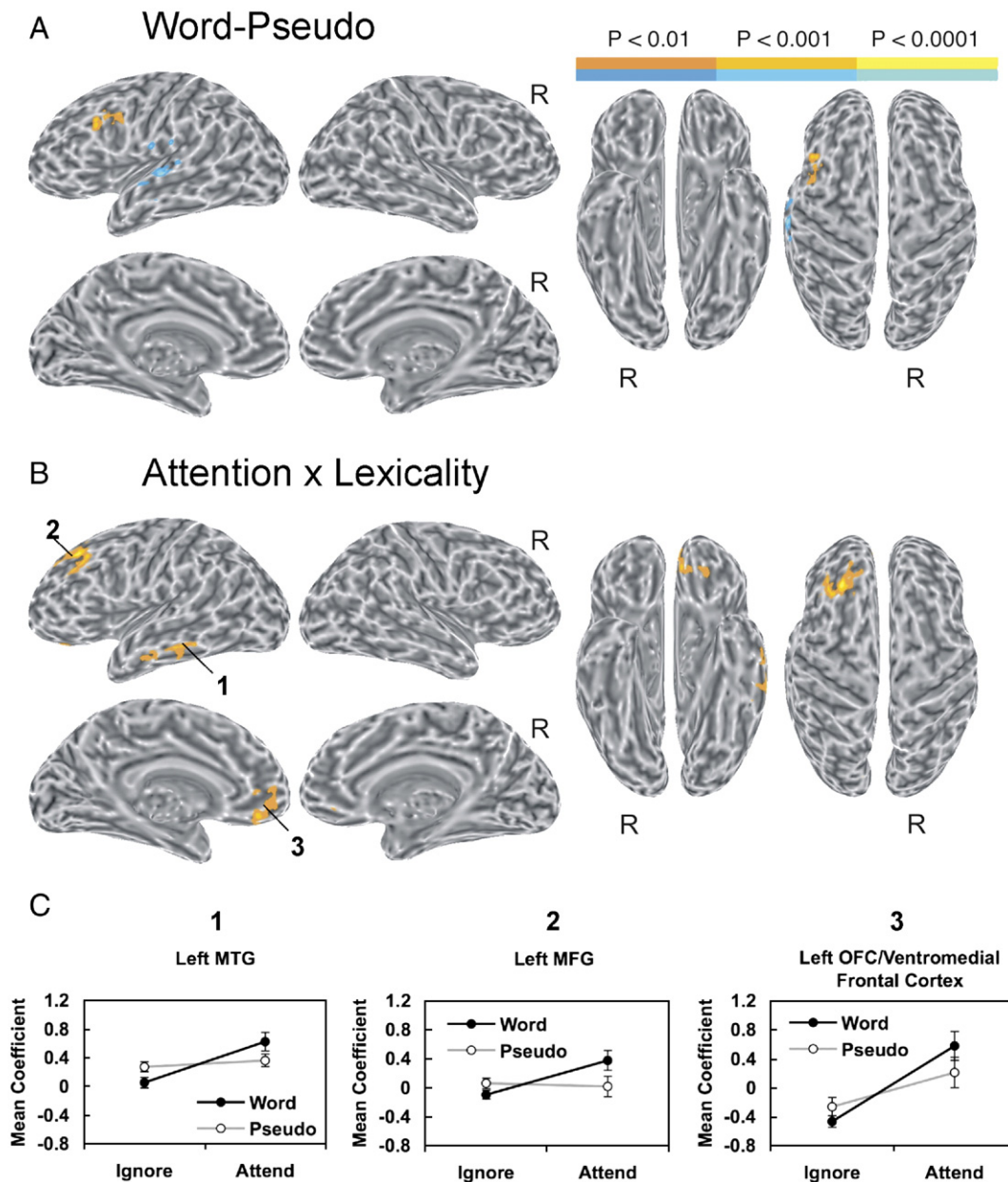

Fig. 3. (A) Brain activation for the contrast between words and pseudowords (main effect of lexicality). (B) Brain activation for the interaction between attention and lexicality. (C) Response profiles (collapsed across time lags 2 and 3) with standard error bars (SEM) across subjects in selected regions of interest across the four relevant conditions compared to baseline.
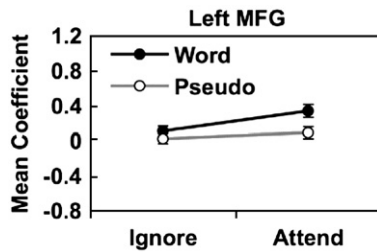
Fig. 4. Response profile in the left posterior frontal region showing a main effect of lexicality (word > pseudoword) in Fig. 3A.

actually activated by words over pseudowords in the Ignore condition. Activation values for this ROI, shown in Fig. 4, reveal that the lexicality effect is driven mainly by the Attend condition. *T*-tests for dependent samples revealed word > pseudoword activation in the Attend condition ($p < 0.001$) but not in the Ignore condition ($p = 0.24$). These results were further confirmed by voxel-wise contrasts between words and pseudowords under each of the attentional conditions (Supplementary Figs. 2 and 3), which showed significant word > pseudoword activation in both of the left MFG regions shown in Fig. 3 when these stimuli were attended, but no word > pseudoword activation when stimuli were ignored. Thus, we observed no evidence of significant word > pseudoword activation in the Ignore condition.

Discussion

Directing attention to auditory stimuli (whether words, pseudowords, or spectrally rotated speech) resulted in greater activation in bilateral auditory areas (STG, STS, and MTG), left inferior parietal cortex (SMG and frontoparietal operculum), and a variety of prefrontal areas (IFG, MFG, left OFC) than when these stimuli were ignored during a demanding primary visual task. This general attentional effect is consistent with numerous auditory ERP studies (Hillyard et al., 1973; Woldorff et al., 1993; Woldorff and Hillyard, 1991) as well as with prior imaging studies using various speech (Grady et al., 1997; Hashimoto et al., 2000; Hugdahl et al., 2000, 2003; Jancke et al., 2001; Jancke et al., 1999; O'Leary et al., 1997; Pugh et al., 1996; Vingerhoets and Luppens, 2001) and nonspeech (Hall et al., 2000; Jancke et al., 2003) auditory stimuli. Contrasts emphasizing speech sound processing (Speech-Rotated) and lexical processing (Word-Pseudo) showed not only main effects of these stimulus variables but also strong interactions between attention and stimulus type. In the following, we discuss these stimulus and attentional effects in more detail, focusing on the question of whether and to what degree processing of auditory stimuli occurs when these stimuli are presented in an unattended sensory channel.

Processing of speech compared to unfamiliar rotated speech sounds

Brain areas involved in the phonemic perception of speech sounds were identified by contrasting the combined word and pseudoword trials, both of which involve presentation of phonemic speech sounds, with the rotated-speech trials. Main effects favoring the speech stimuli were observed in the left orbitofrontal region, right postcentral gyrus, and right lateral occipital cortex (middle and

superior occipital gyri). Interpretation of these main effects, however, is qualified by the presence of significant Attention × Speech interactions in all three of these regions (Fig. 2). In each case, the positive speech effect was present in the Attend Auditory condition but not in the Ignore Auditory condition. Moreover, no brain regions met significance criteria in a voxel-wise speech > rotated contrast under the Ignore condition (Supplementary Fig. 1). These results suggest that unattended neutral speech sounds (i.e., presented during active attention to the visual channel) undergo no more processing than unfamiliar, largely nonphonemic sounds, i.e., that phoneme recognition is mostly absent under such conditions. This result is consistent with prior behavior studies, which show that subjects fail to detect even obvious phonetic manipulations of an unattended speech signal, such as temporal reversal of the sounds or switching to a foreign language (e.g., Cherry, 1953; Wood and Cowan, 1995a).

On the other hand, this conclusion seems to contradict findings from the ERP literature, specifically the MMN component elicited "pre-attentively" (i.e., without directed attention) in response to a phonemic change in an otherwise repetitive auditory stream (Endrass et al., 2004; Naatanen, 2001; Pettigrew et al., 2004; Pulvermuller and Shtyrov, 2006; Pulvermuller et al., 2004; Shtyrov et al., 2004). However, it is also known that the MMN can be susceptible to attentional manipulations, as its amplitude is attenuated when participants are engaged in a demanding primary task (e.g., Sabri et al., 2006; Woldorff et al., 1991). Tasks commonly employed in speech-MMN studies, such as passive listening, reading, or silent film watching, do not prevent the subject from directing attention toward the irrelevant sounds. In the absence of a demanding primary task, it is likely that participants still listen to the irrelevant sounds, and consequently any conclusion regarding pre-attentive processing is then uncertain.

Although the available data thus suggest very limited cortical processing of ignored speech, there is evidence that emotionally or intrinsically significant words are an exception to this rule. For example, a substantial minority (~30–35%) of subjects can detect their own name when presented in an ignored auditory channel (e.g., Cherry, 1953; Wood and Cowan, 1995b), and such sounds are commonly followed by an involuntary orienting response to the ignored ear (Cherry, 1953; Buchner et al., 2004; Wood and Cowan, 1995b). Unlike the neutral stimuli used in the present experiment, one's own name has a specific and intrinsic significance, the entire purpose of which is to elicit an orienting response. Because the name refers to oneself, it is also of intrinsic emotional significance. One's name is also one of the first words acquired during development, and so may be processed much more automatically and wholistically than other words. Thus, personal names may be a special case of a highly familiar auditory object that elicits an over-learned response without intermediary phonemic processing. In fact, many nonhuman animals learn to respond to their name as a wholistic token without evidence of phonemic analysis. One's own name can even elicit a specific event-related potential response within sleep (e.g., Portas et al., 2000).

A number of other regions also showed strong Attention × Speech interactions due to a speech > rotated effect in the Attend but not the Ignore condition. These included the left postcentral gyrus, the anterior temporal lobes bilaterally, and the ventral temporal lobes bilaterally. Many of these ventral and anterior temporal regions have been linked previously with lexical–semantic processes during word and sentence comprehension (e.g., Binder et al., 2003; Humphries et al., 2006; Scott et al., 2000; Spitsyna et al.,

2006; Vandenberghe et al., 2002). Their activation in the speech > rotated contrast probably reflects greater activation of auditory word form, semantic memory, and episodic memory encoding systems by the words and pseudowords compared to the much less meaningful rotated speech sounds. Again, this language-related activation appeared only in the Attend condition and was absent in the Ignore condition.

In contrast to this dependence of speech processing on attention, greater activation for the rotated-speech trials was observed in several regions regardless of attentional state. For example, main effects favoring rotated speech over speech were observed in the superior temporal lobes bilaterally, particularly the TTG and superior temporal plane, including the planum temporale. With the exception of the right posterior STG, most of these regions showed no Attention × Speech interaction and were also activated in the rotated > speech contrast under the Ignore Auditory condition (Supplementary Fig. 1). This finding was unexpected and to our knowledge has not been reported previously. One possible explanation relates to spectral differences between the speech and rotated speech. The process of spectral inversion increases the average energy at higher frequencies and decreases energy at lower frequencies relative to normal speech, thus some of the observed effects might represent stronger activation of high-frequency areas in tonotopic cortex by the rotated speech. This account, however, would predict accompanying low-frequency regions activated preferentially by the speech sounds, which were not observed. An alternative explanation, which we consider more likely, is that the novelty of the rotated speech sounds produced a momentary shift in attention toward the auditory stream even when subjects were attending to the visual task (Escera et al., 1998).

Several other regions showed activation for the rotated-speech trials that appeared to interact with attentional condition, probably reflecting the greater difficulty of the one-back task for the rotated speech items. Performance on this task was significantly worse for the rotated speech, suggesting greater difficulty encoding and storing these sounds in short-term memory. Rotated speech contains acoustic features that are similar to normal speech but is nevertheless difficult for listeners to understand (Blesser, 1972). Words and pseudowords, in contrast, can be encoded at higher linguistic (phonemic, lexical, and semantic) levels, with resultant facilitation of short-term memory. The fMRI data are consistent with this model, in that attended rotated speech produced stronger activations, compared to attended speech, in bilateral MFG and IFG, bilateral IPS, and medial frontal regions (SFG), all of which were previously shown to be activated by increasing task difficulty (Adler et al., 2001; Binder et al., 2005; Braver et al., 2001, 1997; Desai et al., 2006; Honey et al., 2000; Sabri et al., 2006; Ullsperger and von Cramon, 2001).

A somewhat unexpected finding was the absence of activation in the posterior STS for speech over rotated speech. Several recent studies comparing phonetic sounds to acoustically matched non-phonetic sounds (Dehaene-Lambertz et al., 2005; Liebenthal et al., 2005; Mottonen et al., 2006) or to noise (Binder et al., 2000; Rimol et al., 2005) have shown activation specifically in this brain region. Two factors might explain these discordant findings. First, although rotated speech is difficult to decode at a lexical level, some phonemes are still recognizable (Blesser, 1972), in contrast to the nonphonetic control conditions used in several of the studies just mentioned. Second, the greater difficulty of maintaining a memory trace for the rotated speech sounds in the current study may have

obscured any differences in phoneme-level activation in the posterior STS. That is, although the word and pseudoword stimuli contained more phonemic information, the small amount of phonemic information available in the rotated speech may have received more "top-down" activation as a result of the task demands. Our results are consistent with Scott et al.'s (2000) PET study, which also did not show activation in the posterior STS when comparing intelligible with rotated-speech sentences, but not with Narain et al.'s (2003) fMRI study using a similar comparison. There are several methodological differences between these three studies in terms of imaging modality, experimental design, and analyses, which could be contributing to the discrepancy in the results. For example, the current study employed random effects whole-brain analysis, which is statistically more conservative than the fixed effects model used by Narain and colleagues.

Processing of words compared to pseudowords: effects of lexicality

Collapsed across attentional conditions, words produced stronger activation than pseudowords in a small region of left lateral prefrontal cortex near the posterior border between MFG and IFG. Examination of the activation levels in these voxels showed that this effect was driven primarily by a lexicality effect in the Attend condition, with no significant effect in the Ignore condition using either an ROI or voxel-wise analysis (Fig. 4, Supplementary Figs. 2 and 3). In contrast, three regions in the left hemisphere – a dorsal prefrontal region at the junction between MFG and SFG, a ventral–medial frontal region including orbital frontal and medial prefrontal cortex, and the left MTG – showed Attention × Lexicality interactions due to the presence of a significant word > pseudoword effect only in the Attend condition. Finally, no brain regions met significance criteria in the voxel-wise word > pseudoword contrast under the Ignore condition (Supplementary Fig. 3). These results suggest that unattended auditory words (i.e., presented during active attention to the visual channel) undergo no more processing than unattended pseudowords, i.e., that lexical–semantic processing does not occur under such conditions. This result is consistent with prior behavior studies, which show that subjects fail to detect even obvious linguistic manipulations of an unattended speech signal and are generally unable to report the presence or absence of specific words or phrases in the unattended materials (Cherry, 1953; Wood and Cowan, 1995a).

The three regions that showed lexical (word > pseudoword) responses only in the Attend condition are among those frequently identified in imaging studies of semantic processing (e.g., Binder et al., 1999, 2003, 2005; Demonet et al., 1992; Mummery et al., 1998; Price et al., 1997; Roskies et al., 2001; Scott et al., 2003; Spitsyna et al., 2006). Other areas often identified in these studies include the anterior and ventral temporal lobes (particularly the fusiform and parahippocampal gyri), angular gyrus, and posterior cingulate gyrus. Several of these regions (anterior and ventral temporal lobes, posterior cingulate gyrus) appeared in the Speech × Attention interaction analysis and showed speech > rotated activation in the Attend condition but did not show lexicality main effects or Lexicality × Attention interactions. These latter areas thus appear to have been activated as much by pseudowords as by words, calling into question our experimental distinction between "speech" and "lexical" levels of processing. There is, in fact, independent evidence that auditory pseudowords produce at least transient activation of the lexical–semantic system. For example, auditory lexical decisions for both pseudowords and words are slower when the

item shares its initial phonemes with a large number of words (i.e., has a large "phonological cohort") compared to when the phonological cohort is small (Luce and Pisoni, 1998). Luce and Lyons (1999) have explained this effect as due to transient activation of an initially large cohort of possible lexical items, which is then gradually narrowed down over time as more and more of the auditory stimulus is perceived. Thus, our results are consistent with previous reports of a widely distributed lexical–semantic system located in the temporal lobes, posterior cingulate gyrus, dorsal prefrontal cortex, and ventral–medial frontal lobe, though many of these regions appear to have been activated by auditory pseudowords as well as by words.

The left posterior STG showed a main effect of lexicality, but in the opposite direction, with greater activation for pseudowords compared to words. Though this region did not show a significant Lexicality × Attention interaction in a voxel-wise analysis, an ROI analysis across the four conditions showed a significant pseudoword > word activation only in the Ignore condition, a pattern that was corroborated by the voxel-wise simple effects contrasts of word vs. pseudoword separately for the Attend and Ignore conditions (Supplementary Figs. 2 and 3). Thus, the lexicality main effect (pseudoword > word) in this region was driven mainly by a difference in processing during the Ignore condition. This region of secondary auditory cortex also showed a main effect of speech, with greater activation by rotated speech compared to speech. Analogous to the earlier discussion, the most likely explanation for the pseudoword > word response in this region is that the more novel pseudowords produced a larger momentary shift in attention toward the auditory channel during the visual task than did the words. The lack of such an effect in the Attend condition can then be explained by the fact that both pseudowords and words were fully attended, and minor variation in attention due to novelty had relatively little additional effect.

Conclusion

This study provides the first neuroimaging evidence regarding linguistic processing of unattended speech stimuli. Taken together, the findings of the current study support a model in which attentional resources are necessary for both phonemic and lexical–semantic processing of neutral auditory words. Exceptions can include processing of emotionally loaded or conditionally significant stimuli, such as one's own name. When subjects are engaged in a demanding primary task, higher level processing of speech presented in an unattended channel is minimal. Directing attention to the auditory stream allows speech to be processed at higher levels, though the extent of this linguistic processing probably depends on task demands. On the other hand, novel sounds like rotated speech, and to some extent pseudowords, appear to momentarily capture perceptual and attentional resources even when attention is directed toward a difficult task in another sensory channel.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2007.09.052.

References

Abrams, R.L., Klinger, M.R., Greenwald, A.G., 2002. Subliminal words activate semantic categories (not automated motor responses). Psychon. Bull. Rev. 9, 100–106.

Adler, C.M., Sax, K.W., Holland, S.K., Schmithorst, V., Rosenberg, L., Strakowski, S.M., 2001. Changes in neuronal activation with increasing attention demand in healthy volunteers: an fMRI study. Synapse 42, 266–272.

Baayen, R.H., Piepenbrock, R., Gulikers, L., 1995. The CELEX lexical database [CD ROM]. Philadelphia.

Banich, M.T., Milham, M.P., Atchley, R., Cohen, N.J., Webb, A., Wszalek, T., et al., 2000. FMRI studies of Stroop tasks reveal unique roles of anterior and posterior brain systems in attentional selection. J. Cogn. Neurosci. 12, 988–1000.

Binder, J.R., Frost, J.A., Hammeke, T.A., Bellgowan, P.S., Rao, S.M., Cox, R.W., 1999. Conceptual processing during the conscious resting state. A functional MRI study. J. Cogn. Neurosci. 11, 80–95.

Binder, J.R., Frost, J.A., Hammeke, T.A., Bellgowan, P.S., Springer, J.A., Kaufman, J.N., et al., 2000. Human temporal lobe activation by speech and nonspeech sounds. Cereb. Cortex 10, 512–528.

Binder, J.R., McKiernan, K.A., Parsons, M.E., Westbury, C.F., Possing, E.T., Kaufman, J.N., et al., 2003. Neural correlates of lexical access during visual word recognition. J. Cogn. Neurosci. 15, 372–393.

Binder, J.R., Medler, D.A., Desai, R., Conant, L.L., Liebenthal, E., 2005. Some neurophysiological constraints on models of word naming. NeuroImage 27, 677–693.

Blesser, B., 1972. Speech perception under conditions of spectral transformation. I. Phonetic characteristics. J. Speech Hear. Res. 15, 5–41.

Braver, T.S., Cohen, J.D., Nystrom, L.E., Jonides, J., Smith, E.E., Noll, D.C., 1997. A parametric study of prefrontal cortex involvement in human working memory. NeuroImage 5, 49–62.

Braver, T.S., Barch, D.M., Gray, J.R., Molfese, D.L., Snyder, A., 2001. Anterior cingulate cortex and response conflict: effects of frequency, inhibition and errors. Cereb. Cortex 11, 825–836.

Brefczynski, J.A., DeYoe, E.A., 1999. A physiological correlate of the 'spotlight' of visual attention. Nat. Neurosci. 2, 370–374.

Buchner, A., Rothermund, K., Wentura, D., Mehl, B., 2004. Valence of distractor words increases the effects of irrelevant speech on serial recall. Mem. Cogn. 32, 722–731.

Cherry, E.C., 1953. Some experiments on the recognition of speech, with one and with two ears. J. Acoust. Soc. Am. 25, 975–979.

Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29, 162–173.

Cox, R.W., Jesmanowicz, A., 1999. Real-time 3D image registration of functional MRI. Magn. Reson. Med. 42, 1014–1018.

Dale, A.M., 1999. Optimal experimental design for event-related FMRI. Hum. Brain Mapp. 8, 109–114.

Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A., Dehaene, S., 2005. Neural correlates of switching from auditory to speech perception. NeuroImage 24, 21–33.

Demonet, J.F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J.L., Wise, R., et al., 1992. The anatomy of phonological and semantic processing in normal subjects. Brain 115, 1753–1768.

Desai, R., Conant, L.L., Waldron, E., Binder, J.R., 2006. FMRI of past tense processing: the effects of phonological complexity and task difficulty. J. Cogn. Neurosci. 18, 278–297.

Duncan, J., Martens, S., Ward, R., 1997. Restricted attentional capacity within but not between sensory modalities. Nature 387, 808–810.

Dupoux, E., Kouider, S., Mehler, J., 2003. Lexical access without attention?

Explorations using dichotic priming. J. Exp. Psychol. Hum. Percept. Perform. 29, 172–184.

Endrass, T., Mohr, B., Pulvermuller, F., 2004. Enhanced mismatch negativity brain response after binaural word presentation. Eur. J. Neurosci. 19, 1653–1660.

Escera, C., Alho, K., Winkler, I., Naatanen, R., 1998. Neural mechanisms of involuntary attention to acoustic novelty and change. J. Cogn. Neurosci. 10, 590–604.

Grady, C.L., Van Meter, J.W., Maisog, J.M., Pietrini, P., Krasuski, J., Rauschecker, J.P., 1997. Attention-related modulation of activity in primary and secondary auditory cortex. NeuroReport 8, 2511–2516.

Hall, D.A., Haggard, M.P., Akeroyd, M.A., Summerfield, A.Q., Palmer, A.R., Elliott, M.R., et al., 2000. Modulation and task effects in auditory processing measured using fMRI. Hum. Brain Mapp. 10, 107–119.

Hashimoto, R., Homae, F., Nakajima, K., Miyashita, Y., Sakai, K.L., 2000. Functional differentiation in the human auditory and language areas revealed by a dichotic listening task. NeuroImage 12, 147–158.

Hillyard, S.A., Hink, R.F., Schwent, V.L., Picton, T.W., 1973. Electrical signs of selective attention in the human brain. Science 182, 177–180.

Honey, G.D., Bullmore, E.T., Sharma, T., 2000. Prolonged reaction time to a verbal working memory task predicts increased power of posterior parietal cortical activation. NeuroImage 12, 495–503.

Hugdahl, K., Law, I., Kyllingsbaek, S., Bronnick, K., Gade, A., Paulson, O.B., 2000. Effects of attention on dichotic listening: an 15O-PET study. Hum. Brain Mapp. 10, 87–97.

Hugdahl, K., Thomsen, T., Ersland, L., Rimol, L.M., Niemi, J., 2003. The effects of attention on speech perception: an fMRI study. Brain Lang. 85, 37–48.

Humphries, C., Binder, J.R., Medler, D.A., Liebenthal, E., 2006. Syntactic and semantic modulation of neural activity during auditory sentence comprehension. J. Cogn. Neurosci. 18, 665–679.

Jancke, L., Mirzazade, S., Shah, N.J., 1999. Attention modulates activity in the primary and the secondary auditory cortex: a functional magnetic resonance imaging study in human subjects. Neurosci. Lett. 266, 125–128.

Jancke, L., Buchanan, T.W., Lutz, K., Shah, N.J., 2001. Focused and nonfocused attention in verbal and emotional dichotic listening: an fMRI study. Brain Lang. 78, 349–363.

Jancke, L., Specht, K., Shah, J.N., Hugdahl, K., 2003. Focused attention in a simple dichotic listening task: an fMRI experiment. Brain Res.: Cogn. Brain Res. 16, 257–266.

Johnson, J.A., Zatorre, R.J., 2005. Attention to simultaneous unrelated auditory and visual events: behavioral and neural correlates. Cereb. Cortex 15, 1609–1620.

Kastner, S., Pinsk, M.A., De Weerd, P., Desimone, R., Ungerleider, L.G., 1999. Increased activity in human visual cortex during directed attention in the absence of visual stimulation. Neuron 22, 751–761.

Liebenthal, E., Binder, J.R., Spitzer, S.M., Possing, E.T., Medler, D.A., 2005. Neural substrates of phonemic perception. Cereb. Cortex 15, 1621–1631.

Luce, P.A., Lyons, E.A., 1999. Processing lexically embedded spoken words. J. Exp. Psychol. Hum. Percept. Perform. 25, 174–183.

Luce, P.A., Pisoni, D.B., 1998. Recognizing spoken words: the neighborhood activation model. Ear Hear. 19, 1–36.

MacLeod, C.M., 1991. Half a century of research on the Stroop effect: an integrative review. Psychol. Bull. 109, 163–203.

Macmillan, N.A., Creelman, C.D., 1991. Detection theory: a user's guide: (1991). xv, 407 pp. New York, NY, US: Cambridge University Press.

Marcel, A.J., 1983. Conscious and unconscious perception: experiments on visual masking and word recognition. Cogn. Psychol. 15, 197–237.

McKiernan, K.A., D'Angelo, B.R., Kaufman, J.N., Binder, J.R., 2006. Interrupting the "stream of consciousness": an fMRI investigation. NeuroImage 29, 1185–1191.

Melcher, T., Gruber, O., 2006. Oddball and incongruity effects during Stroop task performance: a comparative fMRI study on selective attention. Brain Res. 1121, 136–149.

Mottonen, R., Calvert, G.A., Jaaskelainen, I.P., Matthews, P.M., Thesen, T.,

Tuomainen, J., et al., 2006. Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. NeuroImage 30, 563–569.

Mummery, C.J., Patterson, K., Hodges, J.R., Price, C.J., 1998. Functional neuroanatomy of the semantic system: divisible by what? J. Cogn. Neurosci. 10, 766–777.

Naatanen, R., 2001. The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). Psychophysiology 38, 1–21.

Narain, C., Scott, S.K., Wise, R.J., Rosen, S., Leff, A., Iversen, S.D., et al., 2003. Defining a left-lateralized response specific to intelligible speech using fMRI. Cereb. Cortex 13, 1362–1368.

O'Leary, D.S., Andreasen, N.C., Hurtig, R.R., Torres, I.J., Flashman, L.A., Kesler, M.L., et al., 1997. Auditory and visual attention assessed with PET. Hum. Brain Mapp. 5, 422–436.

Oldfield, R.C., 1971. The assessment and analysis of handedness: The Edinburgh Inventory. Neuropsychologia 9, 97–113.

Peterson, B.S., Skudlarski, P., Gatenby, J.C., Zhang, H., Anderson, A.W., Gore, J.C., 1999. An fMRI study of Stroop word-color interference: evidence for cingulate subregions subserving multiple distributed attentional systems. Biol. Psychiatry 45, 1237–1258.

Petkov, C.I., Kang, X., Alho, K., Bertrand, O., Yund, E.W., Woods, D.L., 2004. Attentional modulation of human auditory cortex. Nat. Neurosci. 7, 658–663.

Pettigrew, C.M., Murdoch, B.E., Ponton, C.W., Finnigan, S., Alku, P., Kei, J., et al., 2004. Automatic auditory processing of English words as indexed by the mismatch negativity, using a multiple deviant paradigm. Ear Hear. 25, 284–301.

Portas, C.M., Krakow, K., Allen, P., Josephs, O., Armony, J.L., Frith, C.D., 2000. Auditory processing across the sleep–wake cycle: simultaneous EEG and fMRI monitoring in humans. Neuron 28, 991–999.

Price, C.J., Wise, R.J., Frackowiak, R.S., 1996. Demonstrating the implicit processing of visually presented words and pseudowords. Cereb. Cortex 6, 62–70.

Price, C.J., Moore, C.J., Humphreys, G.W., Wise, R.J.S., 1997. Segregating semantic from phonological processes during reading. J. Cogn. Neurosci. 9, 727–733.

Pugh, K.R., Shaywitz, B.A., Shaywitz, S.E., Fulbright, R.K., Byrd, D., Skudlarski, P., et al., 1996. Auditory selective attention: an fMRI investigation. NeuroImage 4, 159–173.

Pulvermuller, F., Shtyrov, Y., 2006. Language outside the focus of attention: the mismatch negativity as a tool for studying higher cognitive processes. Prog. Neurobiol. 79, 49–71.

Pulvermuller, F., Shtyrov, Y., Kujala, T., Naatanen, R., 2004. Word-specific cortical activity as revealed by the mismatch negativity. Psychophysiology 41, 106–112.

Raichle, M.E., MacLeod, A.M., Snyder, A.Z., Powers, W.J., Gusnard, D.A., Shulman, G.L., 2001. A default mode of brain function. Proc. Natl. Acad. Sci. U. S. A. 98, 676–682.

Rees, G., Frith, C., Lavie, N., 2001. Processing of irrelevant visual motion during performance of an auditory attention task. Neuropsychologia 39, 937–949.

Rimol, L.M., Specht, K., Weis, S., Savoy, R., Hugdahl, K., 2005. Processing of sub-syllabic speech units in the posterior temporal lobe: an fMRI study. NeuroImage 26, 1059–1067.

Roskies, A.L., Fiez, J.A., Balota, D.A., Raichle, M.E., Petersen, S.E., 2001. Task-dependent modulation of regions in the left inferior frontal cortex during semantic processing. J. Cogn. Neurosci. 13, 829–843.

Sabri, M., Liebenthal, E., Waldron, E.J., Medler, D.A., Binder, J.R., 2006. Attentional modulation in the detection of irrelevant deviance: a simultaneous ERP/fMRI study. J. Cogn. Neurosci. 18, 689–700.

Scott, S.K., Blank, C.C., Rosen, S., Wise, R.J.S., 2000. Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123, 2400–2406.

Scott, S.K., Leff, A.P., Wise, R.J., 2003. Going beyond the information given: a neural system supporting semantic interpretation. NeuroImage 19, 870–876.

Shtyrov, Y., Hauk, O., Pulvermuller, F., 2004. Distributed neuronal networks for encoding category-specific semantic information: the mismatch negativity to action words. Eur. J. Neurosci. 19, 1083–1092.

Spitsyna, G., Warren, J.E., Scott, S.K., Turkheimer, F.E., Wise, R.J., 2006. Converging language streams in the human temporal lobe. J. Neurosci. 26, 7328–7336.

Stroop, J.R., 1935. Studies of interference in serial verbal reactions. J. Exp. Psychol. 18, 643–662.

Talairach, J., Tournoux, P., 1988. Co-planar Stereotaxic Atlas of the Human Brain. Thieme Medical, New York.

Thuy, D.H., Matsuo, K., Nakamura, K., Toma, K., Oga, T., Nakai, T., et al., 2004. Implicit and explicit processing of kanji and kana words and non-words studied with fMRI. NeuroImage 23, 878–889.

Tootell, R.B.H., Hadjikhani, N.K., Mendola, J.D., Marrett, S., Dale, A.M., 1998. From retinotopy to recognition: fMRI in human visual cortex. Trends Cogn. Sci. 2, 174–183.

Treisman, M., 1973. Relation between signal detectability theory and the traditional procedures for measuring thresholds. An addendum. Psychol. Bull. 79, 45–47.

Ullsperger, M., von Cramon, D.Y., 2001. Subprocesses of performance monitoring: a dissociation of error processing and response competition revealed by event-related fMRI and ERPs. NeuroImage 14, 1387–1401.

Vandenberghe, R., Nobre, A.C., Price, C.J., 2002. The response of left temporal cortex to sentences. J. Cogn. Neurosci. 14, 550–560.

Vingerhoets, G., Luppens, E., 2001. Cerebral blood flow velocity changes during dichotic listening with directed or divided attention: a transcranial Doppler ultrasonography study. Neuropsychologia 39, 1105–1111.

Voyer, D., Szeligo, F., Russell, N.L., 2005. Dichotic target detection with words: a modified procedure. J. Clin. Exp. Neuropsychol. 27, 400–411.

Wilson, M., 1988. The MRC psycholinguistic database: machine readable dictionary, version 2. Behav. Res. Methods Instrum. Comput. 20, 6–11.

Woldorff, M.G., Hillyard, S.A., 1991. Modulation of early auditory processing during selective listening to rapidly presented tones. Electroencephalogr. Clin. Neurophysiol. 79, 170–191.

Woldorff, M.G., Hackley, S.A., Hillyard, S.A., 1991. The effects of channel-selective attention on the mismatch negativity wave elicited by deviant tones. Psychophysiology 28, 30–42.

Woldorff, M.G., Gallen, C.C., Hampson, S.A., Hillyard, S.A., Pantev, C., Sobel, D., et al., 1993. Modulation of early sensory processing in human auditory cortex during auditory selective attention. Proc. Natl. Acad. Sci. U. S. A. 90, 8722–8726.

Wood, N.L., Cowan, N., 1995a. The cocktail party phenomenon revisited: attention and memory in the classic selective listening procedure of Cherry (1953). J. Exp. Psychol. Gen. 124, 243–262.

Wood, N., Cowan, N., 1995b. The cocktail party phenomenon revisited: how frequent are attention shifts to one's name in an irrelevant auditory channel? J. Exp. Psychol. Learn. Mem. Cogn. 21, 255–260.

Woodruff, P.W., Benson, R.R., Bandettini, P.A., Kwong, K.K., Howard, R.J., Talavage, T., et al., 1996. Modulation of auditory and visual cortex by selective attention is modality-dependent. NeuroReport 7, 1909–1913.