Behavioral/Systems/Cognitive

Effortful Listening: The Processing of Degraded Speech Depends Critically on Attention

Conor J. Wild,[1] Afiqah Yusuf,[2] Daryl E. Wilson,[2] Jonathan E. Peelle,[3,4] Matthew H. Davis,[4] and Ingrid S. Johnsrude[2,5]

[1]Centre for Neuroscience Studies and [2]Department of Psychology, Queen's University, Kingston, Ontario, Canada, K7L 3N6, [3]Center for Cognitive Neuroscience and Department of Neurology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, [4]MRC Cognition and Brain Sciences Unit, Cambridge CB2 7EF, United Kingdom, [5]Linnaeus Centre for Hearing and Deafness (HEAD), Department of Behavioural Sciences and Learning, Linköping University, SE-581 83 Linköping, Sweden

The conditions of everyday life are such that people often hear speech that has been degraded (e.g., by background noise or electronic transmission) or when they are distracted by other tasks. However, it remains unclear what role attention plays in processing speech that is difficult to understand. In the current study, we used functional magnetic resonance imaging to assess the degree to which spoken sentences were processed under distraction, and whether this depended on the acoustic quality (intelligibility) of the speech. On every trial, adult human participants attended to one of three simultaneously presented stimuli: a sentence (at one of four acoustic clarity levels), an auditory distracter, or a visual distracter. A postscan recognition test showed that clear speech was processed even when not attended, but that attention greatly enhanced the processing of degraded speech. Furthermore, speech-sensitive cortex could be parcellated according to how speech-evoked responses were modulated by attention. Responses in auditory cortex and areas along the superior temporal sulcus (STS) took the same form regardless of attention, although responses to distorted speech in portions of both posterior and anterior STS were enhanced under directed attention. In contrast, frontal regions, including left inferior frontal gyrus, were only engaged when listeners were attending to speech and these regions exhibited elevated responses to degraded, compared with clear, speech. We suggest this response is a neural marker of effortful listening. Together, our results suggest that attention enhances the processing of degraded speech by engaging higher-order mechanisms that modulate perceptual auditory processing.

Introduction

Conversations in everyday life are often made more challenging by poor listening conditions that degrade speech (e.g., electronic transmission, background noise) or by tasks that distract us from our conversational partner. Research exploring how we perceive degraded speech typically considers situations in which speech is the sole (or target) signal, and not how distraction may influence speech processing (Miller et al., 1951; Kalikow et al., 1977; Pichora-Fuller et al., 1995; Davis et al., 2005). Attention may play a critical role in processing speech that is difficult to understand.

It has been hypothesized that perceiving degraded speech consumes more attentional resources than does clear speech (Rabbitt, 1968, 1990). This "effortful listening" hypothesis is usually tested indirectly by showing that attending to degraded (compared with clear) speech interferes with downstream cognitive processing, such as encoding words into memory (Rabbitt, 1990; Murphy et al., 2000; Stewart and Wingfield, 2009). Here, we examine how distraction (compared with full attention) affects the processing of spoken sentences: if processing degraded speech requires more attentional resources than clear speech, then distraction should interfere more with the processing of degraded speech. Using functional magnetic resonance imaging (fMRI), we tested this hypothesis by directly comparing neural responses to degraded and clear sentences when these stimuli are attended or unattended.

Under directed attention, spoken sentence comprehension activates a distributed network of brain areas involving left frontal and bilateral temporal cortex (Davis and Johnsrude, 2003; Davis et al., 2007; Hickok and Poeppel, 2007; Obleser et al., 2007). This speech-sensitive cortex is arranged in a functional hierarchy: cortically early regions (e.g., primary auditory cortex) are sensitive to the acoustic form of speech, whereas activity in higher-order temporal and frontal regions correlates with speech intelligibility regardless of acoustic characteristics (Davis and Johnsrude, 2003), suggesting that these areas contribute to the processing of more abstract linguistic information. Frontal and periauditory regions, which respond more actively to degraded, compared with clear, speech, have been proposed to compensate for distortion (Davis and Johnsrude, 2003, 2007; Shahin et al., 2009; Wild et al., 2012). We expected that attention would selectively modulate speech-evoked responses in these higher order areas, because lower-level periauditory responses to speech do not seem to depend on attentional state (Heinrich et al., 2011).

Received March 28, 2012; revised July 18, 2012; accepted Aug. 17, 2012.

Author contributions: C.J.W., D.W., M.H.D., and I.S.J. designed research; C.J.W. and A.Y. performed research; C.J.W., J.E.P., M.H.D., and I.S.J. contributed unpublished reagents/analytic tools; C.J.W. analyzed data; C.J.W. wrote the paper.

This work was supported by Natural Science and Engineering Research Council of Canada and Canadian Institutes of Health Research.

The authors declare no financial conflicts of interest.

Correspondence should be addressed to Conor Wild, The Brain and Mind Institute, Western University, London, Ontario, Canada N6A 5B7. E-mail: conorwild@gmail.com.

DOI:10.1523/JNEUROSCI.1528-12.2012

 0270-6474/12/3214010-12$15.00/0

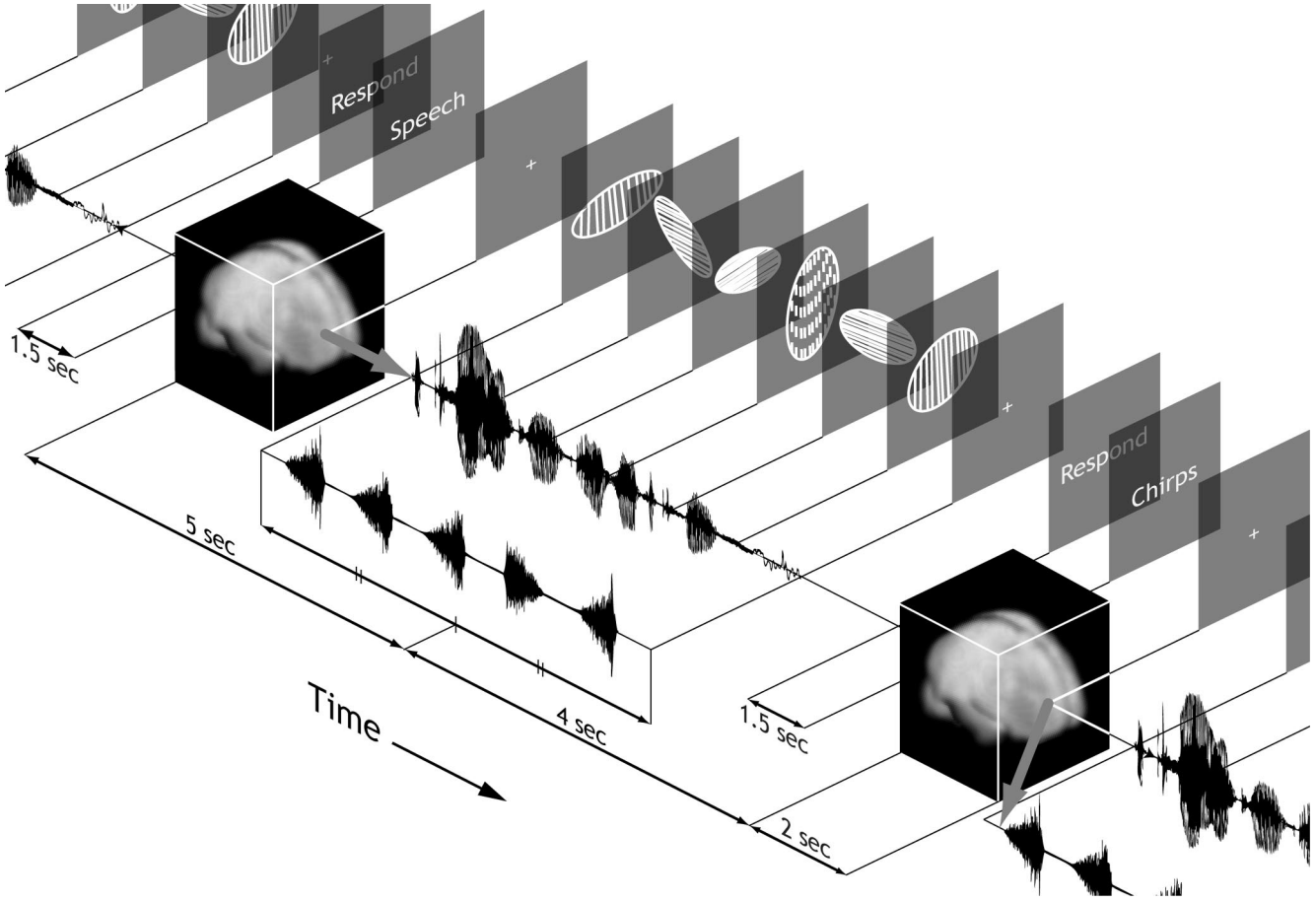

Figure 1. A schematic representation of an experimental trial (TR = 9000 ms). Image acquisition (2000 ms) was clustered at the end each trial, leaving 7000 ms of silence during which stimuli were presented. The midpoint of the stimulus (three-source composite) was positioned exactly 4000 ms before the scan. Subjects' attention was directed to one of the three sources (the gray arrow pointed toward the speech signal in this instance) by a cue during the preceding scan.

In the present study, we use fMRI to compare how sentences of varying acoustic clarity—and hence, intelligibility—are processed when attended, or ignored in favor of engaging distracter tasks. We also use a recognition memory posttest to assess how well sentences from the scanning session are processed as a function of stimulus clarity and attentional state. This factorial design, with intelligibility and attentional task as main effects, allows us to identify regions that are not only sensitive to differences in speech intelligibility or attentional focus, but, critically, areas where the processing of speech depends on attention (i.e., the interaction). Elevated responses to degraded speech that occur only when attention is directed toward speech would suggest a neural signature of effortful listening.

Materials and Methods

Participants. We tested 21 undergraduate students (13 females) between 19 and 27 years of age (mean, 21 years; SD, 3.0 years) from Queen's University (Ontario, Canada). All participants were recruited through poster advertisement and the Queen's Psychology 100 Subject Pool. A separate group of 13 undergraduate students (11 females, 18–35 years of age) were tested to pilot the materials and the procedure. They underwent the same experimental protocol as the other participants, including the presentation of all three stimulus sources, in an isolated soundbooth.

All subjects were right-handed native speakers of English, with normal self-reported hearing, normal or corrected-to-normal vision, and no known attentional or language processing impairments. Participants reported no history of seizures or psychiatric or neurological disorders, and no current use of any psychoactive medications. Participants also complied with magnetic resonance imaging safety standards: they reported no prior surgeries involving metallic implants, devices, or objects. This study was cleared by the Health Sciences and Affiliated Teaching Hospitals Research Ethics Board (Kingston, ON, Canada), and written informed consent was received from all subjects.

Experimental design. To avoid acoustic confounds associated with continuous echoplanar imaging, fMRI scanning was conducted using a sparse imaging design (Edmister et al., 1999; Hall et al., 1999) in which stimuli were presented in the 7 s silent gap between successive 2 s volume acquisitions. On every trial, subjects were cued to attend to one of three simultaneously presented stimuli (Fig. 1)—a sentence [speech stimulus (SP)], an auditory distracter (AD), or a visual distracter (VD)—and performed a decision task associated with the attended stimulus. The speech stimulus on every trial was presented at one of four levels of clarity (for details of stimulus creation, see Speech stimuli, below). Together, these yielded a factorial design with 12 conditions (4 speech intelligibility levels × 3 attention conditions). A silent baseline condition was also included: participants simply viewed a fixation cross, and no other stimuli were presented.

Speech stimuli. Sentence stimuli consisted of 216 meaningful English sentences (e.g., "His handwriting was very difficult to read") recorded by a female native speaker of North American English in a soundproof chamber using an AKG C1000S microphone with an RME Fireface 400 audio interface (sampling at 16 bits, 44.1 khz). We manipulated speech clarity, and hence intelligibility, using a noise-vocoding technique (Shannon et al., 1995) that preserves the temporal information in the speech envelope but reduces the amount of spectral clarity. Noise-vocoded (NV) stimuli were created by filtering each audio recording into contiguous (approximately) logarithmically spaced frequency bands (selected

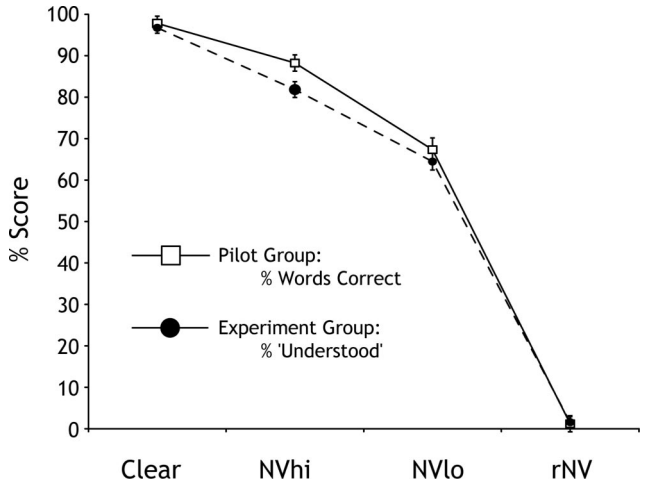

Figure 2. Intelligibility of speech as a function of stimulus clarity for pilot and experimental subjects. The pilot group (white squares) performed a word report task; intelligibility was scored as the percentage of words reported correctly for each sentence. The group tested in the scanner (black circles) made a binary response indicating whether or not they understood the gist of the sentence. Error bars represent SEM, adjusted for repeated-measures data (Loftus and Masson, 1994).

to be approximately equally spaced along the basilar membrane) (Greenwood, 1990). Filtering was performed using finite impulse response Hann bandpass filters with a window length of 801 samples. The amplitude envelope from each frequency band was extracted by full-wave rectifying the band-limited signal and applying a low-pass filter (30 Hz cutoff, using a fourth-order Butterworth filter). Each envelope was then applied to bandpass-filtered noise of the same frequency range, and all bands were recombined to produce the final NV utterance.

With this process, we created four levels of speech varying in acoustic clarity (Fig. 2): clear speech, which is easily understood and highly intelligible; six-band NV stimuli (NV-hi), which is spectrally degraded but still quite intelligible; compressed six-band NV stimuli (NV-lo), which is more difficult to understand than regular six-band NV stimuli; and spectrally rotated NV (rNV) stimuli, which is acoustically very similar to NV stimuli, but impossible to understand. NV-lo stimuli differ from NV-hi items in that their channel envelopes were amplitude-compressed (by taking the square root) to reduce dynamic range before applying them to the noise carriers. To create rNV stimuli items, the envelope from the lowest frequency band was applied to the highest frequency noise band (and vice versa), the envelope from the second lowest band was applied to the second highest band (and vice versa), and envelopes from the inner two bands were swapped. Spectrally rotated speech is completely unintelligible but retains the same overall temporal profile and spectral complexity as the nonrotated version, and hence serves as a closely matched control. After processing, all stimuli (864 audio files) were normalized to have the same total root mean square power.

Twelve sets of 18 sentences were constructed from the corpus of 216 items. The sets were statistically matched for number of words (mean = 9.0, SD = 2.2), number of syllables (mean = 20.1, SD = 7.3), length in milliseconds (mean = 2499, SD = 602.8), and the logarithm of the sum word frequency (Thorndike and Lorge written frequency, mean = 5.5, SD = 0.2). Each set of sentences was assigned to one of the 12 experimental conditions, such that sets and conditions were counterbalanced across subjects to eliminate item-specific effects.

The pilot participants, when instructed to attend to the speech stimulus, repeated back as much of the sentence as they could, which was scored to give a percentage of words correct measure of attended speech intelligibility. A repeate-measures ANOVA on the average proportion of words reported correctly showed a significant main effect of speech type ($F_{(3,36)} = 451.21$, $p < 0.001$; Fig. 2), and *post hoc* tests (Bonferroni corrected for multiple comparisons) showed that that clear speech was reported more accurately than NV-hi ($t_{(12)} = 5.38$, $p < 0.001$), which was reported more accurately than NV-lo ($t_{(12)} = 5.55$, $p < 0.001$), which was reported more accurately than rNV speech ($t_{(12)} = 20.38$, $p < 0.001$).

Auditory distracters. Auditory distracters were sequences of 400 ms narrow-band ramped noise bursts separated by a variable amount of silence (220–380 ms). The number of sounds in each sequence was selected so that the durations of the auditory distracter and the sentence stimulus were approximately equal (Fig. 1). Each noise burst was created by passing 400 ms of broadband white noise through a filter with a fixed bandwidth of 1000 Hz and a center frequency that was randomly selected to be between 4500 and 5500 Hz. The noise bursts were amplitude-modulated to create linear onsets of 380 ms and sharp linear offsets of 20 ms. Target sounds in this stream possessed a sharp onset (20 ms) and a long offset (380 ms) (Fig. 1). Half of all experimental trials were selected to contain a single target sound, which never occurred first in the sequence of noise bursts.

Auditory stimuli (distracter sequences and speech stimuli) were presented diotically over MR-compatible high-fidelity electrostatic earphones, placed in ear defenders that attenuated the background sound of the scanner by ~30 dB (NordicNeurolab AudioSystem).

Data from the auditory distracter task were analyzed using signal detection theory by comparing the *z*-score of the proportion of hits to the *z*-score of the proportion of false alarms, yielding a d' score for each participant. For the pilot group, the average d' score was 2.30 (SD = 0.88), which was significantly greater than chance ($d' > 0$; $t_{(12)} = 9.51$, $p < 0.001$), indicating that participants were able to perform the target detection task.

Visual distracters. The visual distracters were series of cross-hatched white ellipses presented on a black background (Fig. 1), length-matched to the duration of the speech stimulus on every trial. These visual stimuli have been shown to be effective distracters in other experiments manipulating focus of attention (Carlyon et al., 2003). Every 200 ms, a new ellipse, which randomly varied in terms of horizontal and vertical scaling factors and could be reflected in the vertical or horizontal axis, was presented (Fig. 1). Half of all trials (balanced across experimental conditions) were selected to contain a visual target: an ellipse with dashed, instead of solid, lines. If present in a trial, the visual target would appear within ±1 s of the midpoint of the speech stimulus. Visual stimuli were displayed by a video projector on to a rear-projection screen viewed by participants through an angled mirror mounted on the head coil.

Again, data were analyzed with signal detection theory to give a d' score for each participant. For the pilot group, the average d' score was 3.89 (SD = 0.35), which was significantly greater than chance levels (i.e., $d' > 0$; $t_{(12)} = 39.91$, $p < 0.001$), indicating that participants were able to perform the target detection task.

Procedure. On each trial, participants were cued to attend to a single stimulus stream with a visual prompt presented during the scan of the previous trial (Fig. 1). The cue word "Speech" instructed participants to attend to the speech stimulus, "Chirps" cued the participants to the auditory distracter, and "Footballs" cued the visual sequence.

When cued to attend to the speech stimulus, participants listened and indicated at the end of the trial whether or not they understood the gist of the sentence (with a two-alternative, yes/no keypress), providing a measure of the intelligibility of the attended speech. When cued to attend to the visual or auditory distracter, participants monitored the stream for a single target stimulus and indicated at the end of the trial whether or not the target was present (with a two-alternative, yes/no keypress). Subjects were instructed to press either button at the end of each silent trial. A response window of 1.5 s (prompted by the word "Respond") occurred before the onset of the image acquisition period (Fig. 1). Participants made their responses with a button wand held in their right hand, using the index finger button for "yes" and the middle finger button for "no."

Participants experienced 18 trials of each of the 12 experimental conditions (4 speech types × 3 attention tasks) and 10 trials of the silent baseline (226 trials total). The 226 trials were divided into four blocks of 56 or 57 trials, each with approximately the same number of trials from each condition. Two extra images were added to the start of each block to allow the magnetization to reach a steady state; these dummy images were discarded from all preprocessing and analysis steps. We implemented an event-related design such that participants did not know

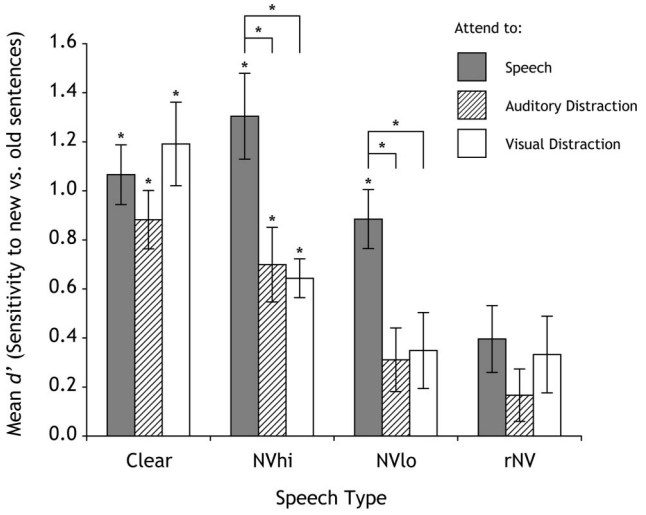

Figure 3. Results of the postscan old/new discrimination task. Bar height indicates average d' across participants and error bars represent SEM adjusted for repeated-measures data. Asterisks above a brace indicate a significant difference between conditions, asterisks above a bar indicate that the average d' was significantly different from zero (i.e., chance); significance levels are adjusted for multiple comparisons using a Bonferroni ($N = 12$) correction.

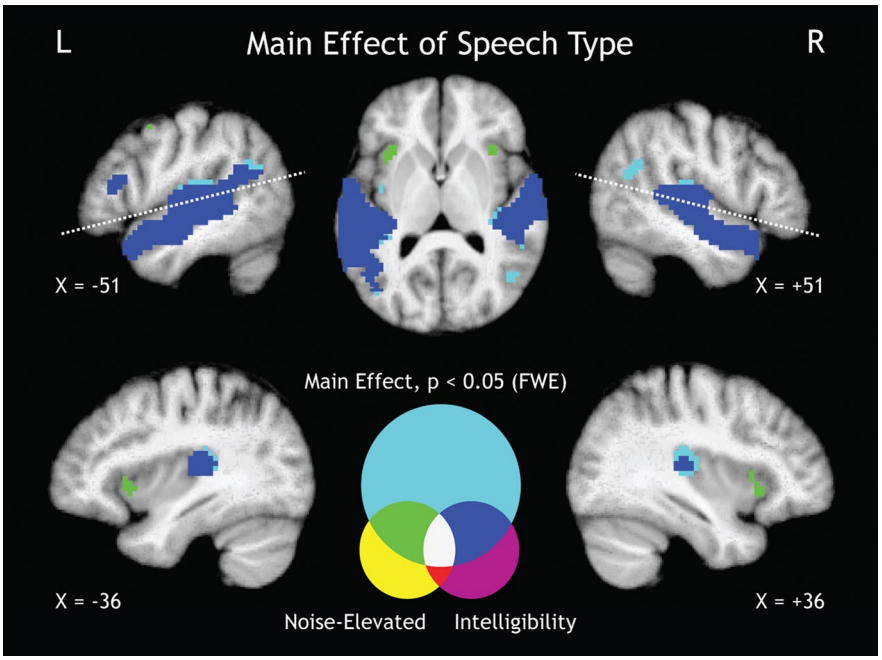

Figure 4. The *F* contrast for the main effect of speech type (cyan) is logically combined with two simple-effects contrasts: voxels where BOLD signal correlates with sentence intelligibility scores (obtained from pilot subjects) (magenta); and voxels that show a noise-elevated response (yellow). Thus, dark blue voxels demonstrate a significant overall main effect of speech type and a correlation with intelligibility, whereas green voxels exhibit a significant main effect that takes the form of a noise-elevated response. The Venn diagram depicts an overlap between these two simple effects because they are not orthogonal contrasts; however, no voxels demonstrate a significant response to both these contrasts. There are no pure yellow or magenta voxels in these panels, because there were no voxels that demonstrated a significant simple effect in the absence of a significant main effect. Conversely, cyan voxels show a significant main effect, but of a form that is not captured by our *t* contrasts. Dotted lines indicate the location and angle of the top middle axial slice. L, Left; R, right.

which task they would perform on any given trial until a cue appeared. However, we reduced task switching to make the experiment easier on participants by constraining the number of consecutive trials with the same task. The distribution was approximately Poisson shaped, such that it was most likely for there to be at least two trials in a row with the same task, but never more than six in a row. Despite the pseudorandomized distribution of tasks, the speech stimulus on every trial was fully randomized and silent trials were fully interspersed throughout the experiment.

All participants (including pilot subjects who generated the performance data reported in Participants, above) underwent extensive training on all three tasks—on each task individually and with task switching. Because the intelligibility of NV speech depends on experience, all participants were also trained with six-band NV stimuli before the study to ensure that intelligibility of the NV-hi and NV-lo stimuli were asymptotic before beginning the actual experiment (Davis et al., 2005).

Behavioral posttest. Immediately after the scanning session, subjects performed a surprise recognition task. This posttest measured participants' memory for a subset (50%) of sentences they heard during the experiment (half of the sentences from each condition randomly selected for each participant). An additional 55 new foil sentences (recorded by the same female speaker) were intermixed with the 108 target sentences. Subjects made an old/new discrimination for each stimulus, responding via button press. Sensitivity (d') for each condition was determined by comparing the *z*-score of the proportion of hits in each condition (out of a maximum of 9) to the *z*-score of the proportion of false alarms for all foils. These scores were analyzed using two-factor (speech type × attention task) repeated-measures ANOVA.

fMRI protocol and data acquisition. Imaging was performed on the 3.0 tesla Siemens Trio MRI system in the Queens Centre for Neuroscience Studies MR Facility (Kingston, Ontario, Canada). T2*-weighted functional images were acquired using GE-EPI sequences (field of view, 211 mm × 211 mm; in-plane resolution, 3.3 mm × 3.3 mm; slice thickness, 3.3 mm with a 25% gap; TA, 2000 ms per volume; TR, 9000 ms; TE, 30 ms; flip angle, 78°). Acquisition was transverse oblique, angled away from the eyes, and in most cases covered the whole brain (in a very few cases, slice positioning excluded the top of the superior parietal lobule). Each stimulus sequence was positioned in the silent interval such that the middle of the sequence occurred 4 s before the onset of the next scan (Fig. 1). In addition to functional data, a whole-brain 3D T1-weighted anatomical image (voxel resolution, 1.0 mm^3) was acquired for each participant at the start of the session.

fMRI data preprocessing. fMRI data were processed and analyzed using Statistical Parametric Mapping (SPM8; Wellcome Centre for Neuroimaging, London, UK). Data preprocessing steps for each subject included: (1) rigid realignment of each EPI volume to the first of the session; (2) coregistration of the structural image to the mean EPI; (3) normalization of the structural image to common subject space (with a subsequent affine registration to MNI space) using the group-wise DARTEL registration method included with SPM8 (Ashburner, 2007); and (4) warping of all functional volumes using deformation flow fields generated from normalization step, which simultaneously resampled the images to isotropic 3 mm voxels and spatially smoothed them with a three-dimensional Gaussian kernel with a full-width at half-maximum of 8 mm. Application of this smoothing kernel resulted in an estimated smoothness of ~15 mm in the group analyses.

fMRI analysis. Analysis of each participant's data was conducted using a general linear model in which each scan was coded as belonging to one of 13 conditions. The four runs were modeled as one session within the design matrix, and four regressors were used to remove the mean signal from each of the runs. Six realignment parameters were included to account for movement-related effects (i.e., three degrees of freedom for translational movement in the *x*, *y*, and *z* directions, and three degrees of freedom for rotational motion: yaw, pitch, and roll). Two additional regressors coded the presence of a target in the visual and

auditory distracter streams. Button presses were not modeled because a button was pressed on every trial. Due to the long TR of this sparse-imaging paradigm, no correction for serial autocorrelation was necessary. A high-pass filter with a cutoff of 216 s was modeled to eliminate low-frequency signal confounds such as scanner drift. These models were then fitted using a least-mean-squares method to each individual's data, and parameter estimates were obtained. Contrast images for each of the 12 experimental conditions were generated by comparing each of the condition parameter estimates (i.e., 12 betas) to the silent baseline condition. These images were primarily used to obtain plots of estimated signal within voxels for each condition.

The group-level analysis was conducted using a 4 (Speech Type) × 3 (Attentional Task) factorial partitioned-error repeated-measures ANOVA, in which separate models were constructed for each main effect and for the interaction of the two factors (Henson and Penny, 2003). For whole-brain analyses of the main effects and their interaction, we used a voxelwise threshold of $p < 0.05$, corrected for multiple comparisons over the whole brain using a nonparametric permutation test as implemented in SnPM (www.sph.umich.edu/ni-stat/SnPM) (Nichols and Holmes, 2002). This test has been shown to have strong control over experiment-wise type I error (Holmes et al., 1996).

A significant main effect or interaction in an ANOVA can be driven by many possible simple effects. In our study, for example, a main effect of speech type might mean that activity correlates with intelligibility (i.e., high activity for clear speech, intermediate activity for degraded speech, and low activity for unintelligible speech), that activity is increased for degraded compared with clear speech, or that there is some other difference in activity between the four levels of the speech type factor. Therefore, the thresholded *F*-statistic images showing overall main effects (and interaction) were parsed into simple effects by inclusively masking with specific *t*-contrast images (i.e., simple effects) that were thresholded at $p < 0.001$, uncorrected. The *t*-contrasts were combined to determine logical intersections of the simple effects; in this way, significant voxels revealed by *F*-contrasts were labeled as being driven by one or more simple effects. Peaks were localized using the LONI probabilistic brain atlas (LPBA40) (Shattuck et al., 2008) and confirmed by visual inspection of the average structural image. Results of the fMRI analysis are shown on the average normalized T1-weighted structural image.

Results

Due to technical difficulties with the stimulus-delivery and response-collection computer program, behavioral and fMRI data were unavailable for two subjects. Analyses of fMRI data, and behavioral data obtained during scanning, are based on the remaining 19 subjects. Posttest data were unavailable for one subject, and so the results of this test are based on data from 20 subjects.

Behavioral results

Speech task

When attending to speech stimuli, participants indicated on every trial whether or not they understood the gist of the sentence. A one-way repeated-measures ANOVA of the proportion of sentences understood, treating speech type as a four-level within-subjects factor, demonstrated a significant main effect of speech type ($F_{(3,54)} = 275.34$, $p < 0.001$; Fig. 2). *Post hoc* pairwise comparisons, corrected for multiple comparisons (Sidak, 1967), indicated that these subjective reports of intelligibility did not reliably differ between clear speech and NV-hi items. NV-hi sentences were reported as understood significantly more often than NV-lo ($t_{(18)} = 5.90$, $p < 0.001$), which were reported as understood significantly more often than rNV sentences ($t_{(18)} = 13.70$, $p < 0.001$). These results closely matched the intelligibility data collected from the pilot group (Fig. 2).

Table 1. Results of the group-level ANOVA; peaks that demonstrate a significant main effect of speech type ($p < 0.05$, corrected family-wise for multiple comparisons)

F contrast	Coordinates (mm) *x*	*y*	*z*	*F*	Voxels in cluster	Location	Simple effect(s)
Main effect:	**−63**	**−9**	**−9**	**125.57**	**1437**	**L A STG**	**Intell**
Speech Type	*−57*	*−15*	*3*	*100.19*		*L STG*	*Intell*
	−36	*−30*	*9*	*71.31*		*L Heschl's G*	*Intell*
	−54	*−36*	*6*	*70.20*		*L P STG/STS*	*Intell*
	−48	*−24*	*6*	*68.29*		*L Heschl's G*	*Intell*
	−51	*−51*	*21*	*26.13*		*L Angular G*	*Intell*
	−42	*−66*	*24*	*18.11*		*L Angular G*	*Intell*
	60	**−3**	**−6**	**74.74**	**962**	**R A STG**	**Intell**
	57	*−18*	*3*	*74.08*		*R P STG*	*Intell*
	57	*6*	*−15*	*69.16*		*R A STG*	*Intell*
	51	*−6*	*−15*	*51.82*		*R A STS*	*Intell*
	45	*−36*	*6*	*18.39*		*R P STS*	*Intell*
	−54	**24**	**15**	**23.75**	**52**	**LIFG**	**Intell**
	−30	**−81**	**15**	**20.24**	**28**	**L M Occipital G**	**Other**
	48	**−60**	**21**	**18.09**	**47**	**R Angular G**	**Other**
	0	**21**	**39**	**18.01**	**22**	**A Cingulate**	**NoiseElev**
	−15	**−69**	**36**	**17.92**	**21**	**L Precuneus**	**Other**
	−39	**−3**	**3**	**16.80**	**3**	**L Insula**	**Other**
	36	**18**	**−3**	**16.52**	**19**	**R A Insula**	**NoiseElev**
	−33	**15**	**−6**	**16.09**	**31**	**L A Insula**	**NoiseElev**
	−48	**−78**	**−3**	**15.95**	**18**	**L M Occipital G**	**Other**
	63	**−48**	**0**	**13.89**	**3**	**R M Temporal G**	**Other**
	−30	**−54**	**−6**	**13.72**	**1**	**L Fusiform G**	**Other**
	−51	**3**	**48**	**13.72**	**1**	**L Premotor C**	**NoiseElev**

Bold entries represent the most significant peak in the cluster; italics indicate significant subpeaks within the cluster. L, Left; R, right; P, posterior; A, anterior; I, inferior; S, superior; G, gyrus. Simple effects were determined with *t* contrasts: Intell, intelligibility-related response, NoiseElev, moise-elevated response; Other, demonstrates the main effect but neither of the tested simple effects.

Distracter tasks

Mean sensitivities (i.e., d') for the auditory and visual distracter tasks were 2.15 (SD = 1.30) and 3.17 (SD = 0.55), respectively. Both scores were significantly greater than chance levels ($t_{(18)} = 7.22$, $p < 0.001$; $t_{(18)} = 25.02$, $p < 0.001$), suggesting that participants in the scanner were attending to the correct stimulus stream. A pairwise comparison showed that the auditory distraction task was significantly more challenging than the visual task ($t_{(18)} = 7.22$, $p < 0.005$).

The d' scores were also broken down by speech type to see whether the unattended speech stimulus affected performance in the distracter tasks. ANOVAs showed no significant effect of the unattended speech type on target detection performance for the auditory or visual distracter tasks.

Old/new discrimination posttest

Results of the posttest are shown in Figure 3. There was a significant main effect of Speech Type ($F_{(3,57)} = 29.12$, $p < 0.001$), with a pattern similar to the pilot subjects, where recognition scores for Clear and NV-hi sentences were not reliably different, but recognition for NV-hi items was significantly better than for NV-lo ($t_{(19)} = 4.48$, $p < 0.001$), which was significantly greater than recognition of rNV sentences ($t_{(19)} = 3.50$, $p < 0.005$). There was also a significant main effect of Attention ($F_{(2,38)} = 11.25$, $p < 0.001$), such that d' values were significantly higher for attended sentences compared with those presented when attention was elsewhere (SP > AD: $t_{(19)} = 4.23$, $p < 0.001$; SP > VD: $t_{(19)} = 3.74$, $p < 0.001$). We note that memory for sentences presented during the distraction tasks did not differ significantly. Importantly, there was a significant Speech Type × Attention interaction ($F_{(6,114)} = 3.61$, $p < 0.005$), where pairwise (Sidak-corrected) comparisons showed that recognition of degraded

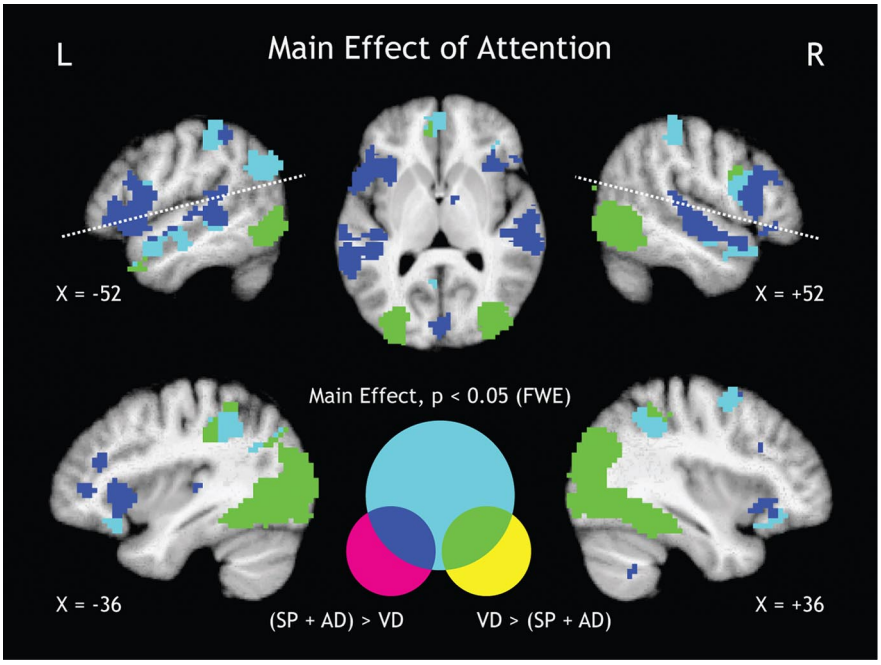

Figure 5. The *F* contrast for the main effect of attention (cyan) is logically combined with two simple-effects contrasts: SP + AD > 2VD, which compares all attended speech and auditory conditions to all attended visual conditions (magenta); and the reverse (yellow). These two contrasts are orthogonal, and hence do not overlap in the Venn diagram. Cyan voxels indicate those voxels that demonstrate a significant main effect of attention, but without this being attributable to the tested simple-effects contrasts. There are no yellow or magenta voxels in these panels because there were no voxels that demonstrated a significant simple effect in the absence of a significant main effect. Dotted lines indicate the location and angle of the top middle axial slice. L, Left; R, Right; FWE, Family-wise error.

sentences (i.e., NV-hi and NV-lo) was significantly enhanced by attention to the speech stimulus (NV-hi SP > AD: $t_{(19)} = 3.27$, $p < 0.005$; NV-hi SP > VD: $t_{(19)} = 3.93$, $p < 0.001$; NV-lo SP > AD: $t_{(19)} = 3.46$, $p < 0.005$; NV-lo SP > VD: $t_{(19)} = 3.64$, $p < 0.005$), whereas attention had no effect on the recognition of clear speech or rotated NV speech items (Fig. 3*A*).

The Speech Type × Attention interaction can also be explained by comparing how recognition of (potentially intelligible) noise-vocoded speech items compares to clear speech across attentional tasks. For attended speech, recognition of clear speech sentences did not differ from NV-hi or NV-lo. However, when attention was directed toward the auditory distracter, clear sentences were remembered significantly better than NV-lo ($t_{(19)} = 4.76$, $p < 0.001$), but not NV-hi, and when attention was directed toward the visual distracter, recognition of clear sentences was better than both NV-hi ($t_{(19)} = 3.63$, $p < 0.01$) and NV-lo ($t_{(19)} = 4.15$, $p < 0.005$).

One-sample *t* tests were conducted on d' scores for each condition (12 per group) to determine whether recognition of sentences presented in those conditions was greater than chance (i.e., $d' > 0$). Performance was significantly better than chance ($d' > 0$; $p < 0.05$, Bonferroni-corrected for 12 comparisons) for all clear and high-intelligibility NV speech conditions and for attended NV-lo items. Recognition of unattended NV-lo items did not differ from chance. The unintelligible rNV stimuli were never recognized above chance levels.

fMRI results

Main effect of speech type

The contrast assessing the main effect of speech type revealed activation of left inferior frontal gyrus (LIFG) and large bilateral activations of the temporal cortex, ranging along the full length of the superior temporal gyrus (STG), superior temporal sulcus (STS), and the superior temporal plane (Fig. 4; Table 1). There are many ways in which four speech-type conditions can differ from each other, but we were interested in two specific patterns of difference, which we tested with specific *t* contrasts. First, we searched for areas where blood oxygen level-dependent (BOLD) signal correlated with speech intelligibility scores (i.e., an intelligibility-related response). Intelligibility scores collected from the pilot subjects were used to construct this contrast because they provided a more objective and continuous measure than the binary subjective response made by participants in the scanner (Davis and Johnsrude, 2003). The pilot and in-MR measures were highly correlated with each other (Fig. 2). Second, a noise-elevated response, which was assessed with the contrast (NV-hi + NV-lo)/2 > Clear was used to identify regions that were more responsive to degraded than clear speech, and therefore might be involved in compensating for acoustic degradation. The unintelligible rNV stimuli were not included in this contrast (i.e., weighted with a zero), because it is not clear whether listeners would try very hard to understand them or just give up. Responses within bilateral temporal and inferior frontal regions demonstrated a significant correlation with intelligibility (Fig. 4, dark blue voxels), largely consistent with a previous correlational intelligibility analysis (Davis and Johnsrude, 2003). Noise-elevated responses were found in left premotor (Fig. 4, top left) and bilateral insular cortex. These did not overlap with any regions that demonstrated a correlation with intelligibility. Interestingly, a noise-elevated response was not observed in left inferior frontal cortex as might have been expected from previous findings (Davis et al., 2003; Giraud et al., 2004; Shahin et al., 2009). The lack of a noise-elevated response in LIFG collapsed across attention conditions is due to a strong interaction with attention, as we discuss below.

Main effect of attention task

The contrast assessing the main effect of attention condition revealed widespread activity (Fig. 5; Table 2). We tested for two simple effects: regions where attention to an auditory stimulus resulted in enhanced responses compared with attention to the visual stimulus [(SP + AD) > 2VD] and areas that demonstrated the opposite pattern [2VD > (SP + AD)]. In accordance with previous research, we observed that attention modulated activity in sensory cortices, such that responses in the sensory cortex corresponding to the modality of the attended stimulus were enhanced (Heinze et al., 1994; Petkov et al., 2004; Johnson and Zatorre, 2005, 2006; Heinrich et al., 2011). This confirmed that our attentional manipulation was effective.

Interaction (Speech Type × Attention Task)

Most interesting were areas that demonstrated an interaction between Speech Type and Attention; that is, areas in which the relationship between acoustic quality of sentences and BOLD signal depended on the listeners' attentional state. Several clusters

Table 2. Results of the group-level ANOVA; peaks that demonstrate a significant main effect of attention task ($p < 0.05$, corrected family-wise for multiple comparisons)

Contrast	Coordinates (mm) x	y	z	F	Voxels in cluster	Location	Simple effect
Main effect: Attention	**33**	**−81**	**12**	**101.23**	**1888**	**R M Occipital G**	**VD > (SP + AD)**
	27	*−54*	*51*	*86.30*		*R Intraparietal S*	*VD > (SP + AD)*
	48	*−60*	*−9*	*84.95*		*R I Temporal G*	*VD > (SP + AD)*
	30	*−72*	*30*	*73.07*		*R Intraparietal S*	*VD > (SP + AD)*
	45	*−54*	*3*	*59.37*		*R P M Temporal G*	*VD > (SP + AD)*
	33	*−45*	*−15*	*46.87*		*R Fusiform G*	*VD > (SP + AD)*
	45	*−33*	*48*	*33.81*		*R Intraparietal S*	*Other*
	48	*−81*	*21*	*28.52*		*R M Occipital G*	*VD > (SP + AD)*
	57	*−24*	*48*	*23.71*		*R Supramarginal G*	*Other*
	−27	**−93**	**12**	**99.31**	**3247**	**L M Occipital G**	**VD > (SP + AD)**
	−42	*−69*	*−3*	*95.69*		*L M Occipital G*	*VD > (SP + AD)*
	−33	*−60*	*−6*	*68.95*		*L Fusiform G*	*VD > (SP + AD)*
	−54	*9*	*3*	*65.27*		*LIFG*	*VD > (SP + AD)*
	−30	*−51*	*54*	*59.83*		*L S Parietal G*	*VD > (SP + AD)*
	−24	*−72*	*30*	*57.69*		*L S Occipital G*	*VD > (SP + AD)*
	−60	*−39*	*15*	*55.81*		*L ST/Planum Temporale*	*(AP + AD) > VD*
	−33	*−84*	*−3*	*55.18*		*L M Occipital G*	*VD > (SP + AD)*
	−39	*−39*	*39*	*52.26*		*L Intraparietal S*	*Other*
	−51	*−36*	*0*	*47.38*		*L P STS*	*(SP + AD) > VD*
	−60	*−6*	*−9*	*44.91*		*L A STG/STS*	*Other*
	−42	*−75*	*39*	*40.44*		*L Angular G*	*VD > (SP + AD)*
	−45	*18*	*21*	*39.10*		*LIFG*	*(SP + AD) > VD*
	−45	*−63*	*27*	*38.34*		*L Angular G*	*Other*
	−36	*42*	*3*	*37.90*		*LIFG*	*(SP + AD) > VD*
	−51	*−36*	*48*	*36.09*		*L Supramarginal G*	*(SP + AD) > VD*
	−36	*24*	*3*	*33.68*		*LIFG*	*(SP + AD) > VD*
	−42	*−24*	*0*	*33.42*		*L Heschl's G*	*(SP + AD) > VD*
	−66	*−27*	*27*	*29.35*		*L Supramarginal G*	*(SP + AD) > VD*
	−60	*−27*	*45*	*28.70*		*L Intraparietal S*	*Other*
	−3	**15**	**54**	**83.98**	**292**	**L SMA**	*(SP + AD) > VD*
	−9	*3*	*63*	*27.03*		*L SMA*	*(SP + AD) > VD*
	6	*33*	*39*	*25.94*		*R SMA*	*(SP + AD) > VD*
	54	**−15**	**−3**	**57.15**	**482**	**R STG/STS**	*(SP + AD) > VD*
	63	*−30*	*3*	*36.97*		*R P STG/STS*	*(SP + AD) > VD*
	54	*−3*	*−9*	*36.46*		*R A STG*	*(SP + AD) > VD*
	51	*6*	*−18*	*36.04*		*R A STS/STG*	*(SP + AD) > VD*
	36	*30*	*−12*	*26.10*		*R I Orbitofrontal G*	*(SP + AD) > VD*
	−42	**33**	**21**	**52.63**	**40**	**L M Frontal G**	*(SP + AD) > VD*
	54	**15**	**18**	**52.41**	**301**	**R I Frontal G**	*(SP + AD) > VD*
	45	*33*	*27*	*37.47*		*R M Frontal G*	*(SP + AD) > VD*
	60	**−21**	**33**	**39.59**	**25**	**R Supramarginal G**	**VD > (SP + AD)**
	−3	**48**	**−18**	**38.44**	**308**	**L S Frontal G**	**VD > (SP + AD)**
	−6	*57*	*6*	*34.98*		*L S Frontal G*	*VD > (SP + AD)*
	0	*33*	*−18*	*31.87*		*R S Frontal G*	*VD > (SP + AD)*
	6	*63*	*30*	*30.14*		*R S Frontal G*	*Other*
	18	*57*	*30*	*22.58*		*R M Frontal G*	*Other*
	33	**−54**	**−45**	**31.05**	**6**	**Cerebellum**	**(SP + AD) > VD**
	0	**−87**	**21**	**31.05**	**133**	**L Cuneus**	**(SP + AD) > VD**
	6	*−75*	*6*	*29.78*		*R Lingual G*	*(SP + AD) > VD*
	−15	**60**	**30**	**29.65**	**14**	**L S Frontal G**	**Other**
	−30	**3**	**51**	**29.49**	**41**	**L Premotor C**	**(SP + AD) > VD**
	−3	**−39**	**39**	**28.52**	**96**	**L P Cingulate G**	**VD > (SP + AD)**
	−3	*−54*	*18*	*26.82*		*L Precuneus*	*VD > (SP + AD)*
	39	**0**	**57**	**28.25**	**57**	**R Premotor C**	**Other**
	−54	**9**	**−30**	**25.49**	**15**	**L A M Temporal G**	**VD > (SP + AD)**
	30	**15**	**−18**	**24.80**	**8**	**R Insula**	**Other**

Bold entries represent the most significant peak in the cluster; italics indicate significant subpeaks within the cluster. L, Left; R, right; P, posterior; A, anterior; I, inferior; S, superior; G, gyrus. Simple effects were determined with *t* contrasts.

of voxels demonstrated a significant interaction (Fig. 6, bright blue voxels; Table 3). These included bilateral posterior STS/STG, left anterior STS/STG, the LIFG (specifically partes triangularis and opercularis), bilateral angular gyri, bilateral anterior insulae, left supplementary motor area (SMA), and the caudate. Interestingly, areas of the STG corresponding to primary auditory cortex and much of the superior temporal plane showed no evidence of an interaction, even at a threshold of $p < 0.05$, uncorrected for multiple comparisons (Fig. 6, red voxels).

Attention influences speech processing differently in frontal and temporal cortex

It is possible that speech-evoked responses in these areas are modulated by attention to different extents or in different ways. Such a difference would manifest as a three-way (Region × Speech Type × Attention) interaction. To quantitatively compare the interaction patterns in temporal and frontal cortices, we conducted three-way repeated-measures ANOVAs on the parameter estimates extracted from four areas: left anterior STS, left posterior STS, right posterior STS, and LIFG. A single LIFG response was created by averaging the parameter estimates from the two LIFG peaks listed in Table 3 (they were within 15 mm of each other, which, given the effective smoothness of the data, is an unresolvable difference). With all four regions entered into the model, a significant three-way interaction ($F_{(18,324)} = 2.13$, $p < 0.01$) indicated that Speech Type × Attention interaction patterns truly differed among these regions. Follow-up comparisons were performed using three-way ANOVAs (Region × Speech Type × Attention) on two regions at a time. Interactions from the three temporal peaks (left anterior and bilateral posterior STS) were not reliably different, but they all differed significantly from the LIFG (left anterior STS vs LIFG: $F_{(6,108)} = 2.52$, $p < 0.05$, left posterior STS vs LIFG: $F_{(6,108)} = 4.33$, $p < 0.001$; right posterior STS vs LIFG: $F_{(6,108)} = 2.83$, $p < 0.05$). Given the lack of difference among them, the three temporal peaks were averaged to create a single STS response, which differed significantly from the LIFG interaction pattern ($F_{(6,108)} = 4.15$, $p < 0.005$). The distinct interaction profiles in LIFG and in temporal cortex are illustrated in Figure 6, *b* and *c*.

Characterizing the influence of attention on speech processing in frontal and temporal cortex

It can be seen that, for both LIFG and STS, the two-way (Speech Type × Attention) interaction is due at least in part to elevated signal for degraded speech when it is attended, compared with when it is not (Fig. 6*b*,*c*). This was confirmed by pairwise comparisons that showed that NV-hi sentences evoked significantly greater activity when they were attended, than when they were ignored (Table 4; Fig. 6*b*,*c*). Also, in both regions, rNV stimuli elicited greater activity when attention was directed toward it, or the auditory distracter, than when attention was directed toward the visual stimulus (Table 4; Fig. 6*b*,*c*).

Despite this common enhancement of activity for degraded speech that was attended, overall speech-evoked responses differed in the LIFG and STS (as evidenced by the significant three-way interaction). To quantify these differences, we compared these responses between areas with two-way ANOVAs (Region × Speech Type). Attended speech elicited a significantly different pattern of activity in LIFG than in the STS ($F_{(3,54)} = 6.12$, $p < 0.001$). A *post hoc* contrast that compared degraded speech (NV-hi and NV-lo) against clear speech revealed a significant noise-elevated response in the LIFG ($F_{(1,18)} = 15.51$, $p < 0.001$; Fig. 6*b*) but not in the STS (absent in Fig. 6*c*). Responses to unattended speech also differed significantly between these areas, as demonstrated by significant Region × Speech Type (2 × 4) interactions at both levels of distraction (Auditory Distracter: $F_{(3,54)} = 29.61$, $p < 0.001$; Visual Distracter: $F_{(3,54)} = 17.84$, $p < 0.001$). Linear interaction contrasts showed that unattended speech produced a steeper linear response (i.e., decreasing activity with decreasing intelligibility) in the STS than in the LIFG (Auditory Distracter: $F_{(1,18)} = 74.95$, $p < 0.001$; Visual Distracter:

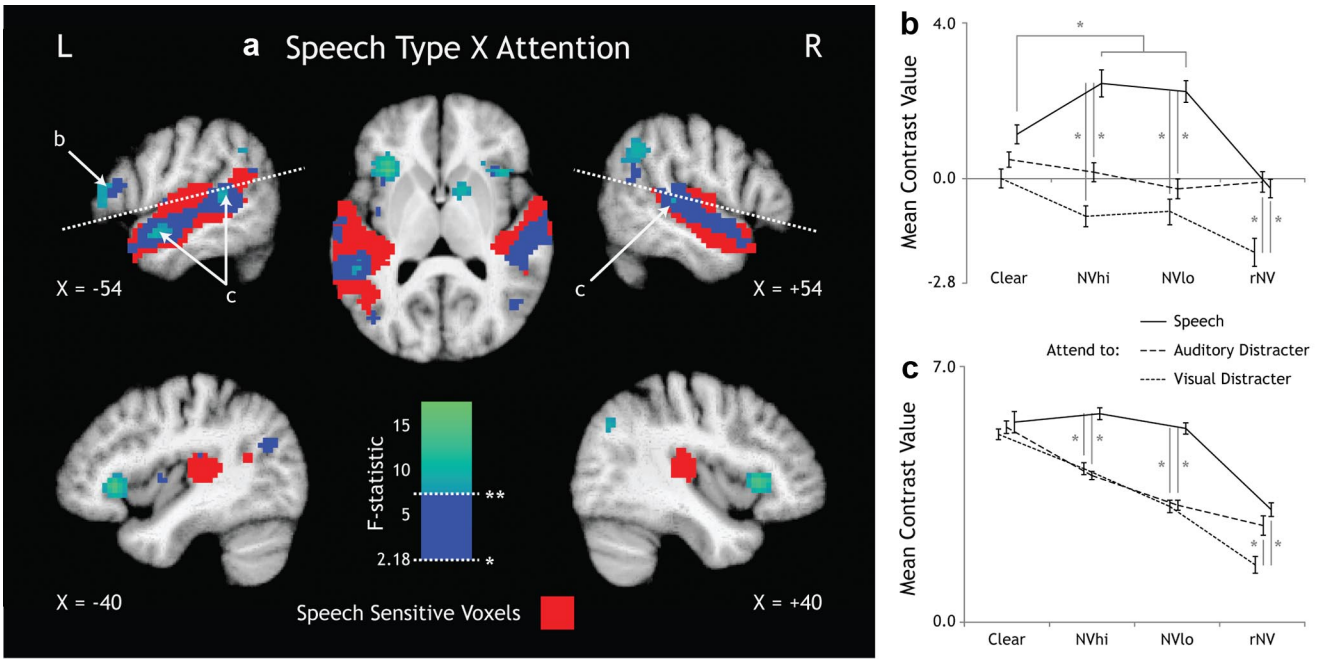

Figure 6. ***a***, The Speech Type × Attention interaction *F* contrast (dark blue color) is thresholded at $p < 0.05$, uncorrected (*F* value of 2.18, indicated with *). The critical F-value determined by nonparametric permutation testing, representing the cutoff for $p < 0.05$ corrected family-wise for multiple comparisons (FWE), is indicated with **. Thus, lighter blue voxels demonstrated a significant interaction at the whole-brain level. Red voxels indicate those that are sensitive to the different types of speech (i.e., demonstrate a significant main effect of speech type—all voxels in Fig. 4), yet show no evidence for an interaction with Attention ($p > 0.05$, uncorrected). ***b***, ***c***, Contrast values (i.e., estimated signal relative to baseline; arbitrary units) are plotted for LIFG (***b***) and anterior and posterior STS peaks (***c***). Only one STS response is plotted (i.e., the average of the left anterior and bilateral posterior STS responses) because the interaction patterns were not significantly different. Error bars represent the SEM suitable for repeated-measures data (Loftus and Masson, 1994). Vertical lines with asterisks indicate significant pairwise comparisons ($p < 0.05$, Bonferroni-corrected for 12 comparisons).

Table 3. Results of the group-level ANOVA of whole-brain data; peaks that demonstrate a significant speech type by attention interaction ($p < 0.05$, corrected family-wise for multiple comparisons)

Contrast	Coordinates (mm)			*F*	Voxels in cluster	Location
	x	*y*	*z*			
Interaction: Speech	**−33**	**24**	**−3**	**17.25**	**197**	**L A Insula**
Type × Attention	*−45*	*21*	*9*	*8.85*		*L I Frontal G*
	−54	*30*	*6*	*8.79*		*L I Frontal G*
	39	**21**	**3**	**14.46**	**115**	**R A Insula**
	−51	**−39**	**9**	**10.71**	**23**	**L P STG / STS**
	51	**−51**	**36**	**10.40**	**56**	**R Angular G**
	−6	**9**	**57**	**10.29**	**29**	**L SMA**
	12	**9**	**0**	**10.26**	**40**	**R Caudate**
	42	**−69**	**36**	**9.81**	**31**	**R Angular G**
	−54	**0**	**−15**	**9.67**	**28**	**L A STG/STS**
	−60	*−6*	*−9*	*8.93*		*L A STG/STS*
	−6	**21**	**39**	**9.37**	**6**	**L S Frontal G**
	−63	**−33**	**−12**	**8.90**	**5**	**L MTG**
	54	**−30**	**6**	**8.34**	**2**	**R STG/STS**
	−57	**−54**	**36**	**7.80**	**5**	**L Angular G**
	−51	**0**	**48**	**5.83***		**L Premotor Cortex**

Asterisks indicate marginal significance. Bold entries represent the most significant peak in the cluster; italics indicate significant subpeaks within the cluster. L, Left; R, right; P, posterior; A, anterior; I, inferior; S, superior; G, gyrus.

Table 4. Results of statistical pairwise comparisons of parameter estimates in two areas: the STS (average response) and LIFG

Speech type	Attention comparison	STS		LIFG	
		$t_{(18)}$	*p*	$t_{(18)}$	*p*
Clear	SP > AD	1.07	0.30	2.32	0.032
	SP > VD	1.11	0.28	3.04	0.01
	AD > VD	−0.44	0.66	1.36	0.19
NV-hi	SP > AD	7.35	<0.001*	6.15	<0.001*
	SP > VD	7.90	<0.001*	6.94	<0.001*
	AD > VD	−2.60	0.02	2.51	0.02
NV-lo	SP > AD	8.49	<0.001*	6.71	<0.001*
	SP > VD	8.70	<0.001*	6.94	<0.001*
	AD > VD	−1.3	0.21	1.18	0.25
rNV	SP > AD	1.90	0.07	−0.48	0.95
	SP > VD	5.56	<0.001*	4.16	<0.001*
	AD > VD	5.41	<0.001*	3.86	<0.001*

These statistics correspond to the results displayed in Figure 6, *b* and *c*. Asterisks indicate significant differences, Bonferonni corrected for 12 comparisons in each region.

$F_{(1,18)} = 39.84$, $p < 0.001$). These results can be observed in Figure 6, *b* and *c*: in the STS, activity elicited by unattended speech decreases with intelligibility, whereas this pattern is less apparent in the LIFG. Although STS regions are significantly active (relative to rest) regardless of attention condition, responses in the LIFG region are above baseline only when attention is on the speech stimulus. Furthermore, the interaction pattern in LIFG explains why this area did not show a noise-elevated response for the main effect of speech type (Fig. 4): this response was present only for speech that was attended.

It is interesting that the noise-elevated response only for attended speech can be qualitatively observed in other brain regions that demonstrated a significant interaction. Figure 7 depicts the interaction patterns from left SMA (Fig. 7*a*), left insula (Fig. 7*b*), right caudate (Fig. 7*c*), and right angular gyrus (Fig. 7*d*). Again, attended speech elicited a noise-elevated response that was absent when attention was focused elsewhere.

Attention-dependent speech processing occurs on the upper bank of the STS

Finally, we wished to improve our localization of the temporal lobe activations revealed in the interaction analysis. It has been

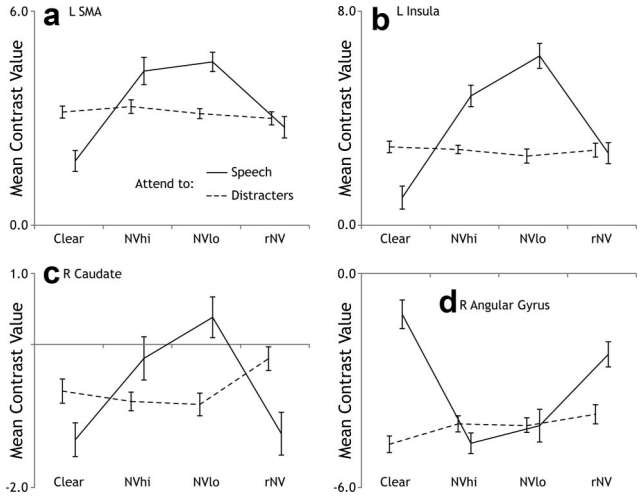

Figure 7. Average contrast values (i.e., estimated signal relative to baseline; arbitrary units) for peaks demonstrating a significant Speech Type × Attention interaction in the whole-brain analysis: ***a***, left supplementary motor area; ***b***, left insula; ***c***, right caudate; ***d***, right angular gyrus. Solid lines indicate conditions in which attention was directed toward the speech signal; dashed lines indicate conditions in which attention was directed toward the distracter stimuli (responses are collapsed across auditory and visual distracter conditions). Error bars indicate SEM.

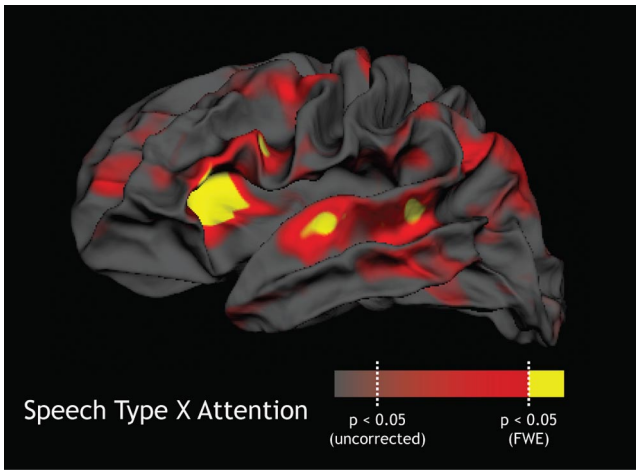

Figure 8. Results of the surface-based analysis of the Speech Type × Attention interaction. Red areas indicate regions demonstrating an interaction at $p < 0.05$, uncorrected for multiple comparisons; yellow areas represent regions in which the *F* statistic was greater than the critical *F* obtained from permutation testing (i.e., $p < 0.05$, FWE).

observed from anatomical studies of rhesus monkeys that the STS is a large and heterogeneous area of cortex, containing several distinct areas that can be parcellated according to their cytoarchitectonic and myeloarchitectonic properties, as well as their afferent and efferent connectivity (Seltzer and Pandya, 1978, 1989, 1991; Padberg et al., 2003). These include unisensory regions—auditory area TAa, along the upper bank and lip of the STS, and visual areas TEa and TEm, along the lower bank of the sulcus—and polymodal regions TPO and PGa, which lie along the upper bank and in the depth of the sulcus (Seltzer and Pandya, 1978). Area TPO itself is composed of three distinct subdivisions (TPOc, TPOi, and TPOr) which receive inputs of varying strength from frontal (ventral and prearcuate cortex) and parietal regions (Padberg et al., 2003). Therefore, precise localization of the STS peaks, which could provide important functional information (e.g., auditory vs visual vs multisensory processing), is confounded by volumetric smoothing: BOLD signal on one side of the sulcus is smoothed across to the physically close, yet cortically distant, bank of the opposite side. Smoothing in two dimensions (along the cortical sheet) overcomes this issue. Accordingly, we performed a surface-based analysis (with the Freesurfer image analysis suite: http://surfer.nmr.mgh.harvard.edu/) of the Speech Type × Attention interaction model simply for visualization purposes, so that we could more accurately locate the area exhibiting interaction within the STS region. Contrast images were inflated and smoothed along the cortical sheet, then submitted to a group-level analysis. This visualization suggests that the rostral STS peak lies on the upper bank of the STS, and so may correspond to the auditory area TAa, whereas the more caudal peak lies more in the depth of the sulcus, but still on the upper bank, and may therefore correspond to multisensory TPO cortex (Fig. 8).

Discussion

This study demonstrates that the comprehension of speech that varies in intelligibility and the engagement of brain areas that support speech processing depend on the degree to which listeners attend to speech. Our behavioral and fMRI results suggest that, in our paradigm at least, clear speech is generally processed and remembered regardless of whether listeners are instructed to attend to it, but speech that is perceptually degraded yet highly intelligible is processed quite differently when listeners are distracted.

The postscan recognition data show that unattended clear speech was encoded into memory, suggesting successful comprehension; listeners were able to remember clear sentences with similar accuracy whether attended or unattended. There was no difference in recognition accuracy for sentences presented in the distraction tasks, despite the difference in task difficulty, which suggests that it is not solely attentional load that determines whether unattended speech is processed. Nonetheless, we cannot discount the possibility that more challenging tasks could disrupt the processing of unattended clear speech, and future work will address this by manipulating load with a broader range of task difficulties. Our conclusion agrees with other studies that demonstrate effects of unattended clear speech on behavioral measures (Salamé and Baddeley, 1982; Hanley and Broadbent, 1987; Kouider and Dupoux, 2005; Rivenez et al., 2006) and electrophysiological responses (Shtyrov, 2010; Shtyrov et al., 2010), but conflicts with findings that listeners usually cannot remember unattended speech when listening to another talker (Cherry, 1953; Wood and Cowan, 1995). Speech signals are acoustically very similar, and attention is likely needed to segregate the target stream from interfering talkers, thereby reducing the resources available for processing unattended speech. This may be similar to our observation that attention is required to process degraded speech: significant Speech Type × Attention interactions in our behavioral and fMRI data indicate that processing of to-be-ignored (degraded) speech is significantly disrupted.

The combination of neural and behavioral interactions provides the first demonstration that the processing of degraded speech depends critically on attention. Degraded speech was highly intelligible when it was attended, but cortical processing and subsequent memory for those sentences was greatly diminished (to chance levels for NV-lo sentences) when attention was focused elsewhere. The recognition data strongly suggest that distraction impaired perception of degraded speech more than clear speech, consistent with our on-line BOLD measures of speech processing during distraction. In both the STS and LIFG,

the processing of degraded (but not clear) speech was significantly enhanced by attention.

Previous fMRI studies of speech perception have either failed to observe similar interactions or have not assessed the degree to which unattended speech is processed at a behavioral level. For instance, Heinrich et al. (2011) showed that low-level auditory processes contributing to the continuity illusion for vowels remain operational during distraction, and thus low-level, speech-specific responses in posterior STG remain intact. Sabri et al. (2008) demonstrated that speech-evoked fMRI responses are attenuated and that lexical effects are absent, or reversed, during distraction. However, without behavioral evidence, it is hard to conclude (as proposed by Sabri et al., 2008) that processing is significantly diminished when speech is ignored. Furthermore, the noise associated with continuous fMRI scanning would have created a challenging listening situation that (according to our findings) might be equally responsible for the absence of neural responses to unattended speech.

Our fMRI results demonstrate that the distributed hierarchy of brain areas underlying sentence comprehension (Davis and Johnsrude, 2003; Davis et al., 2007; Hickok and Poeppel, 2007; Obleser et al., 2007) can be parcellated by the degree to which patterns of speech-related activity depend on attention. It is only brain regions more distant from auditory cortex, probably supporting higher-level processes, that show attentional modulation. Response patterns in primary auditory regions did not depend on attention (i.e., there was no interaction), despite a reliable main effect consistent with other studies of auditory attention (Alho et al., 2003; Hugdahl et al., 2003; Petkov et al., 2004; Fritz et al., 2007). This suggest that early auditory cortical processing of speech is largely automatic and independent of attention, but can be enhanced (or suppressed) by attention.

In contrast, areas of left frontal and bilateral temporal cortex exhibited robust changes in patterns of speech-evoked activity due to changes in attentional state. In both regions, this dependence manifested primarily as an increase in activity for degraded speech when it was attended compared with when in was ignored. However, the dissimilarity of patterns in these regions (i.e., the significant three-way interaction) provides evidence that attention influences speech processing differently in these areas. When speech was attended, LIFG activity for degraded speech was greater than for clear speech (i.e., a noise-elevated response), whereas in the STS, activity for degraded speech was enhanced to the level of clear speech. During distraction conditions, LIFG activity did not depend on speech type, but activity in the STS was correlated with intelligibility. Together, these results suggest that the LIFG only responds to degraded speech when listeners are attending to it, whereas the STS responds to speech intelligibility, regardless of attention or how that intelligibility is achieved. Increased activity for attended degraded speech may reflect the improvement in intelligibility afforded by explicit, effortful processing, or by additional cognitive processes (such as perceptual learning) that are engaged under directed attention (Davis et al., 2005; Eisner et al., 2010). A recent behavioral study demonstrated that perceptual learning of NV stimuli is significantly impaired by distraction under conditions similar to those studied here (Huyck and Johnsrude, 2012).

These fMRI results are consistent with the proposal that speech comprehension in challenging listening situations is facilitated by top-down influences on early auditory processing (Davis and Johnsrude, 2007; Sohoglu et al., 2012; Wild et al., 2012). Due to their anatomical connectivity, regions of prefrontal cortex—including LIFG and premotor cortex—are able to modulate activity within early auditory belt and parabelt cortex (Hackett et al., 1999; Romanski et al., 1999) and intermediate stages of processing on the dorsal bank of the STS either directly (Seltzer and Pandya, 1989, 1991; Petrides and Pandya, 2002a,b) or indirectly through parietal cortex (Petrides and Pandya, 1984, 2009; Rozzi et al., 2006). LIFG has been shown to contribute to the processes involved in accessing and combining word meanings (Thompson-Schill et al., 1997; Wagner et al., 2001; Rodd et al., 2005), and this information could be used to recover words and meaning from an impoverished speech signal. Somatomotor representations may also be helpful for parsing difficult-to-understand speech (Davis and Johnsrude, 2007), including NV stimuli (Wild et al., 2012; Hervais-Adelman et al., 2012). We note that many of the fMRI and transcranial magnetic stimulation studies that implicate left premotor regions in speech processing have similarly used degraded speech or other stimuli that are difficult for listeners to understand (Fadiga et al., 2002; Watkins et al., 2003; Watkins and Paus, 2004; Wilson et al., 2004; Wilson and Iacoboni, 2006; Osnes et al., 2011). We also observed significant interactions in bilateral insular cortex and in the left caudate nucleus. These areas connect with primary auditory cortex, prefrontal cortex, (supplementary) motor regions, and temporoparietal regions (Alexander et al., 1986; Middleton and Strick, 1994, 1996; Yeterian and Pandya, 1998; Clower et al., 2005) and have been shown to be involved in phonological processing (Abdullaev and Melnichuk, 1997; Bamiou et al., 2003; Tettamanti et al., 2005; Booth et al., 2007; Christensen et al., 2008). The interactions observed in these areas are consistent with the idea that motoric representations are engaged during effortful speech perception.

In light of our results, we propose that the interaction pattern observed in higher-order speech-sensitive cortex is a neural signature of effortful listening. Effort has recently become a topic of great interest to applied hearing researchers and is typically assessed through indirect measures; for example, autonomic arousal (Zekveld et al., 2010; Zekveld et al., 2011; Mackersie and Cones, 2011), the degree to which participants are able to perform a secondary task (Howard et al., 2010), or, as in our study, the ability of listeners to remember what they had heard (Rabbitt, 1968, 1990; Stewart and Wingfield, 2009; Tun et al., 2009). We propose that fMRI can be used to operationalize listening effort more directly by comparing the effortful BOLD response between attended and unattended speech conditions in the network of frontal areas we have observed. To validate this proposal, future work will attempt to relate individual differences in this BOLD response to listener attributes, such as the ability to divide attention, experience with degraded speech, and other cognitive factors. Neural measures of effortful listening might provide a novel means of assessing the efficacy and comfort of hearing prostheses, and help researchers to optimize the benefit obtained from these devices.

Our findings unequivocally demonstrate that the extent to which degraded speech is processed depends on the listener's attentional state. Whereas clear speech can be processed even when ignored, comprehension of degraded speech appears to require focused attention. Our fMRI results are consistent with the idea that enhanced processing of degraded speech is accomplished by engaging higher-order language-related processes that modulate earlier perceptual auditory processing.

References

Abdullaev YG, Melnichuk KV (1997) Cognitive operations in the human caudate nucleus. Neurosci Lett 234:151–155. CrossRef Medline

Alexander GE, DeLong MR, Strick PL (1986) Parallel organization of func-

tionally segregated circuits linking basal ganglia and cortex. Annual review of neuroscience 9:357–381. CrossRef Medline

Alho K, Vorobyev VA, Medvedev SV, Pakhomov SV, Roudas MS, Tervaniemi M, van Zuijen T, Näätänen R (2003) Hemispheric lateralization of cerebral blood-flow changes during selective listening to dichotically presented continuous speech. Brain Res Cogn Brain Res 17:201–211. CrossRef Medline

Ashburner J (2007) A fast diffeomorphic image registration algorithm. Neuroimage 38:95–113. CrossRef Medline

Bamiou DE, Musiek FE, Luxon LM (2003) The insula (Island of Reil) and its role in auditory processing: literature review. Brain Res Rev 42:143–154. CrossRef Medline

Booth JR, Wood L, Lu D, Houk JC, Bitan T (2007) The role of the basal ganglia and cerebellum in language processing. Brain Res 1133:136–144. CrossRef Medline

Carlyon RP, Plack CJ, Fantini DA, Cusack R (2003) Cross-modal and non-sensory influences on auditory streaming. Perception 32:1393–1402. CrossRef Medline

Cherry EC (1953) Some experiments on the recognition of speech, with one and with two ears. J Acoust Soc Am 25:975–979. CrossRef

Christensen TA, Antonucci SM, Lockwood JL, Kittleson M, Plante E (2008) Cortical and subcortical contributions to the attentive processing of speech. Neuroreport 19:1101–1105. CrossRef Medline

Clower DM, Dum RP, Strick PL (2005) Basal ganglia and cerebellar inputs to "AIP". Cereb Cortex 15:913–920. CrossRef Medline

Davis MH, Johnsrude IS (2003) Hierarchical processing in spoken language comprehension. J Neurosci 23:3423–3431. Medline

Davis MH, Johnsrude IS (2007) Hearing speech sounds: top-down influences on the interface between audition and speech perception. Hear Res 229:132–147. CrossRef Medline

Davis MH, Johnsrude IS, Hervais-Adelman A, Taylor K, McGettigan C (2005) Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. J Exp Psychol Gen 134:222–241. CrossRef Medline

Davis MH, Coleman MR, Absalom AR, Rodd JM, Johnsrude IS, Matta BF, Owen AM, Menon DK (2007) Dissociating speech perception and comprehension at reduced levels of awareness. Proc Natl Acad Sci U S A 104:16032–16037. CrossRef Medline

Edmister WB, Talavage TM, Ledden PJ, Weisskoff RM (1999) Improved auditory cortex imaging using clustered volume acquisitions. Hum Brain Mapp 7:89–97. CrossRef Medline

Eisner F, McGettigan C, Faulkner A, Rosen S, Scott SK (2010) Inferior frontal gyrus activation predicts individual differences in perceptual learning of cochlear-implant simulations. J Neurosci 30:7179–7186. CrossRef Medline

Fadiga L, Craighero L, Buccino G, Rizzolatti G (2002) Short communication: speech listening specifically modulates the excitability of tongue muscles: a TMS study. Eur J Neurosci 15:399–402. CrossRef Medline

Fritz JB, Elhilali M, David SV, Shamma SA (2007) Auditory attention: focusing the searchlight on sound. Curr Opin Neurobiol 17:437–455. CrossRef Medline

Giraud AL, Kell C, Thierfelder C, Sterzer P, Russ MO, Preibisch C, Kleinschmidt A (2004) Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. Cereb Cortex 14:247–255. CrossRef Medline

Greenwood DD (1990) A cochlear frequency-position function for several species: 29 years later. J Acoust Soc Am 87:2592–2605. CrossRef Medline

Hackett TA, Stepniewska I, Kaas JH (1999) Prefrontal connections of the parabelt auditory cortex in macaque monkeys. Brain Res 817:45–58. CrossRef Medline

Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW (1999) "sparse" temporal sampling in auditory fMRI. Hum Brain Mapp 7:213–223. CrossRef Medline

Hanley JR, Broadbent C (1987) The effect of unattended speech on serial recall following auditory presentation. Br J Psychol 78:287–297. CrossRef

Heinrich A, Carlyon RP, Davis MH, Johnsrude IS (2011) The continuity illusion does not depend on attentional state: fMRI evidence from illusory vowels. J Cogn Neurosci 23:2675–2689. CrossRef Medline

Heinze HJ, Mangun GR, Burchert W, Hinrichs H, Scholz M, Münte TF, Gös A, Scherg M, Johannes S, Hundeshagen H, Gazzaniga MS, Hillyard SA (1994) Combined spatial and temporal imaging of brain activity during visual selective attention in humans. Nature 372:543–546. CrossRef Medline

Henson RN, Penny WD (2003) ANOVAs and SPM. Technical report. London: Wellcome Department of Imaging Neuroscience.

Hervais-Adelman AG, Carlyon RP, Johnsrude IS, Davis MH (2012) Brain regions recruited for the effortful comprehension of noise-vocoded words. Lang Cogn Process 27:1145–1166. CrossRef

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

Holmes AP, Blair RC, Watson JD, Ford I (1996) Nonparametric Analysis of Statistic Images from Functional Mapping Experiments. J Cereb Blood Flow Metab 16:7–22. Medline

Howard CS, Munro KJ, Plack CJ (2010) Listening effort at signal-to-noise ratios that are typical of the school classroom. Int J Audiol 49:928–932. CrossRef Medline

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

Huyck JJ, Johnsrude IS (2012) Rapid perceptual learning of noise-vocoded speech requires attention. J Acoust Soc Am 131:EL236–EL242. CrossRef Medline

Johnson JA, Zatorre RJ (2005) Attention to simultaneous unrelated auditory and visual events: behavioral and neural correlates. Cereb Cortex 15:1609–1620. CrossRef Medline

Johnson JA, Zatorre RJ (2006) Neural substrates for dividing and focusing attention between simultaneous auditory and visual events. Neuroimage 31:1673–1681. CrossRef Medline

Kalikow DN, Stevens KN, Elliott LL (1977) Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. J Acoust Soc Am 61:1337–1351. CrossRef Medline

Kouider S, Dupoux E (2005) Subliminal speech priming. Psychol Sci 16:617–625. CrossRef Medline

Loftus G, Masson M (1994) Using confidence-intervals in within-subject designs. Psychonom Bull Rev 1:476–490. CrossRef

Mackersie CL, Cones H (2011) Subjective and psychophysiological indices of listening effort in a competing-talker task. J Am Acad Audiol 22:113–122. CrossRef Medline

Middleton FA, Strick PL (1994) Anatomical evidence for cerebellar and basal ganglia involvement in higher cognitive function. Science 266:458–461. CrossRef Medline

Middleton FA, Strick PL (1996) The temporal lobe is a target of output from the basal ganglia. Proc Natl Acad Sci U S A 93:8683–8687. CrossRef Medline

Miller GA, Heise GA, Lichten W (1951) The intelligibility of speech as a function of the context of the test materials. J Exp Psychol 41:329–335. CrossRef Medline

Murphy DR, Craik FI, Li KZ, Schneider BA (2000) Comparing the effects of aging and background noise of short-term memory performance. Psychol Aging 15:323–334. CrossRef Medline

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

Obleser J, Wise RJ, Alex Dresner M, Scott SK (2007) Functional integration across brain regions improves speech perception under adverse listening conditions. J Neurosci 27:2283–2289. CrossRef Medline

Osnes B, Hugdahl K, Specht K (2011) Effective connectivity analysis demonstrates involvement of premotor cortex during speech perception. Neuroimage 54:2437–2445. CrossRef Medline

Padberg J, Seltzer B, Cusick CG (2003) Architectonics and cortical connections of the upper bank of the superior temporal sulcus in the rhesus monkey: an analysis in the tangential plane. J Comp Neurol 467:418–434. CrossRef Medline

Petkov CI, Kang X, Alho K, Bertrand O, Yund EW, Woods DL (2004) Attentional modulation of human auditory cortex. Nat Neurosci 7:658–663. CrossRef Medline

Petrides M, Pandya DN (1984) Projections to the frontal cortex from the posterior parietal region in the rhesus monkey. J Comp Neurol 228:105–116. CrossRef Medline

Petrides M, Pandya DN (2002a) Comparative cytoarchitectonic analysis of the human and the macaque ventrolateral prefrontal cortex and corticocortical connection patterns in the monkey. Eur J Neurosci 16:291–310. CrossRef Medline

Petrides M, Pandya DN (2002b) Association pathways of the prefrontal cortex and functional observations. In: Principles of frontal lobe function, 1st ed. (Stuss DT, Knight RT), pp 31–50. New York: Oxford UP.
Petrides M, Pandya DN (2009) Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. PLoS Biol 7:e1000170. CrossRef Medline
Pichora-Fuller MK, Schneider BA, Daneman M (1995) How young and old adults listen to and remember speech in noise. J Acoust Soc Am 97:593–608. CrossRef Medline
Rabbitt PM (1968) Channel-capacity, intelligibility and immediate memory. Q J Exp Psychol 20:241–248. CrossRef Medline
Rabbitt P (1990) Mild hearing loss can cause apparent memory failures which increase with age and reduce with IQ. Acta Otolaryngol Suppl 476:167–175; discussion 176. Medline
Rivenez M, Darwin CJ, Guillaume A (2006) Processing unattended speech. J Acoust Soc Am 119:4027–4040. CrossRef Medline
Rodd JM, Davis MH, Johnsrude IS (2005) The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. Cereb Cortex 15:1261–1269. Medline
Romanski LM, Bates JF, Goldman-Rakic PS (1999) Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. J Comp Neurol 403:141–157. CrossRef Medline
Rozzi S, Calzavara R, Belmalih A, Borra E, Gregoriou GG, Matelli M, Luppino G (2006) Cortical connections of the inferior parietal cortical convexity of the Macaque monkey. Cereb Cortex 16:1389–1417. Medline
Sabri M, Binder JR, Desai R, Medler DA, Leitl MD, Liebenthal E (2008) Attentional and linguistic interactions in speech perception. Neuroimage 39:1444–1456. CrossRef Medline
Salamé P, Baddeley A (1982) Disruption of short-term memory by unattended speech: implications for the structure of working memory. J Verb Learn Verb Behav 21:150–164. CrossRef
Seltzer B, Pandya DN (1978) Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. Brain Res 149:1–24. CrossRef Medline
Seltzer B, Pandya DN (1989) Frontal lobe connections of the superior temporal sulcus in the rhesus monkey. J Comp Neurol 281:97–113. CrossRef Medline
Seltzer B, Pandya DN (1991) Post-rolandic cortical projections of the superior temporal sulcus in the rhesus monkey. J Comp Neurol 312:625–640. CrossRef Medline
Shahin AJ, Bishop CW, Miller LM (2009) Neural mechanisms for illusory filling-in of degraded speech. Neuroimage 44:1133–1143. CrossRef Medline
Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. Science 270:303–304. CrossRef Medline
Shattuck DW, Mirza M, Adisetiyo V, Hojatkashani C, Salamon G, Narr KL, Poldrack RA, Bilder RM, Toga AW (2008) Construction of a 3D probabilistic atlas of human cortical structures. Neuroimage 39:1064–1080. CrossRef Medline
Shtyrov Y (2010) Automaticity and attentional control in spoken language processing: neurophysiological evidence. Mental Lexicon 5:255–276. CrossRef
Shtyrov Y, Kujala T, Pulvermüller F (2010) Interactions between Language and Attention Systems: Early Automatic Lexical Processing? J Cogn Neurosci 22:1465–1478. Medline
Sidak Z (1967) Rectangular confidence regions for the means of multivariate normal distributions. J Am Stat Assoc 62:626–633. CrossRef
Sohoglu E, Peelle JE, Carlyon RP, Davis MH (2012) Predictive top-down integration of prior knowledge during speech perception. J Neurosci 32:8443–8453. CrossRef Medline
Stewart R, Wingfield A (2009) Hearing loss and cognitive effort in older adults' report accuracy for verbal materials. J Am Acad Audiol 20:147–154. CrossRef Medline
Tettamanti M, Moro A, Messa C, Moresco RM, Rizzo G, Carpinelli A, Matarrese M, Fazio F, Perani D (2005) Basal ganglia and language: phonology modulates dopaminergic release. Neuroreport 16:397–401. CrossRef Medline
Thompson-Schill SL, D'Esposito M, Aguirre GK, Farah MJ (1997) Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. Proc Natl Acad Sci U S A 94:14792–14797. CrossRef Medline
Tun PA, McCoy S, Wingfield A (2009) Aging, hearing acuity, and the attentional costs of effortful listening. Psychol Aging 24:761–766. CrossRef Medline
Wagner AD, Paré-Blagoev EJ, Clark J, Poldrack RA (2001) Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. Neuron 31:329–338. CrossRef Medline
Watkins K, Paus T (2004) Modulation of motor excitability during speech perception: the role of Broca's area. J Cogn Neurosci 16:978–987. CrossRef Medline
Watkins KE, Strafella AP, Paus T (2003) Seeing and hearing speech excites the motor system involved in speech production. Neuropsychologia 41:989–994. CrossRef Medline
Wild CJ, Davis MH, Johnsrude IS (2012) Human auditory cortex is sensitive to the perceived clarity of speech. Neuroimage 60:1490–1502. CrossRef Medline
Wilson SM, Iacoboni M (2006) Neural responses to non-native phonemes varying in producibility: evidence for the sensorimotor nature of speech perception. Neuroimage 33:316–325. CrossRef Medline
Wilson SM, Saygin AP, Sereno MI, Iacoboni M (2004) Listening to speech activates motor areas involved in speech production. Nat Neurosci 7:701–702. CrossRef Medline
Wood NL, Cowan N (1995) The cocktail party phenomenon revisited: Attention and memory in the classic selective listening procedure of Cherry (1953). J Exp Psychol Gen 124:243–262. CrossRef Medline
Yeterian EH, Pandya DN (1998) Corticostriatal connections of the superior temporal region in rhesus monkeys. J Comp Neurol 399:384–402. CrossRef Medline
Zekveld AA, Kramer SE, Festen JM (2010) Pupil response as an indication of effortful listening: the influence of sentence intelligibility. Ear Hear 31:480–490. CrossRef Medline
Zekveld AA, Kramer SE, Festen JM (2011) Pupil dilation uncovers extra listening effort in the presence of a single-talker masker. Ear Hear 32:498–510. CrossRef Medline