
Cortical Mechanisms of Speech Perception in Noise

Patrick C. M. Wong
Ajith K. Uppunda
Todd B. Parrish
Sumitrajit Dhar

Northwestern University, Evanston, IL

Purpose: The present study examines the brain basis of listening to spoken words in noise, which is a ubiquitous characteristic of communication, with the focus on the dorsal auditory pathway.

Method: English-speaking young adults identified single words in 3 listening conditions while their hemodynamic response was measured using fMRI: speech in quiet, speech in moderately loud noise (signal-to-noise ratio [SNR] 20 dB), and in loud noise (SNR -5 dB).

Results: Behaviorally, participants' performance (both accuracy and reaction time) did not differ between the quiet and SNR 20 dB condition, whereas they were less accurate and responded slower in the SNR -5 dB condition compared with the other 2 conditions. In the superior temporal gyrus (STG), both left and right auditory cortex showed increased activation in the noise conditions relative to quiet, including the middle portion of STG (mSTG). Although the right posterior STG (pSTG) showed similar activation for the 2 noise conditions, the left pSTG showed increased activation in the SNR -5 dB condition relative to the SNR 20 dB condition.

Conclusion: We found cortical task-independent and noise-dependent effects concerning speech perception in noise involving bilateral mSTG and left pSTG. These results likely reflect demands in acoustic analysis, auditory-motor integration, and phonological memory, as well as auditory attention.

KEY WORDS: auditory cortex, speech perception, auditory perception, noise

Verbal communication in the modern world often occurs in the presence of interfering background noise. As such, listeners must develop sensory, cognitive, and neural resources for handling noise to achieve successful speech (spoken words) processing and communication. The current study investigates the neural/cortical correlates associated with listening to speech in noise by younger adults with normal auditory functions. Much research investigating the neural bases of speech perception focuses on problems such as the perception of speech and nonspeech sounds (Binder, Liebenthal, Possing, et al., 2004), the perception of consonants (Burton, Small, & Blumstein, 2000), and the perception of native and nonnative speech sounds (Wong, Parson, Martinez, & Diehl, 2004b) to identify neural networks that are associated with different aspects of speech. While important, these studies do not focus on how perceiving speech in noise is accomplished, which is arguably a hallmark of our speech processing system that enables communication in a variety of natural but unfavorable listening situations. Reduced ability for understanding speech in noise is a primary symptom of (central) auditory processing disorders (Bamiou, Musiek, & Luxon, 2001), which is estimated to affect between 2% and 3% of children (Chermak & Musiek, 1997) and 22.6% of adults over 60 years of age (Cooper & Gates, 1991). It is also a characteristic of many individuals with learning disabilities

(King, Lombardino, Crandell, & Leonard, 2003; Hugdahl et al., 1998), specific language impairment (Wright et al., 1997), and attention-deficit/hyperactivity disorder (Chermak, Somers, & Seikel, 1998). These disorders are often attributed to neuroanatomic and neurophysiologic anomalies (Hugdahl et al., 1998), although no evidence has been provided with regard to how such anomalies affect an individual's ability to listen to speech in noise.

It has been found that normal-hearing individuals take advantage of spectral and temporal “dips” to understand speech in noise (Peters, Moore, & Baer, 1998). The term *spectral dips* refers to differences in spectral characteristics between the target and background in which the spectrum of the target becomes more salient. The term *temporal dips* refers to momentary low-energy segments of the background noise where the target signal is more audible. Individuals with peripheral hearing loss commonly complain of difficulty understanding speech in the presence of background noise and have demonstrated difficulties with using spectral and temporal dips (Peters et al., 1998). Others with known and probable deficits in the central nervous system have also shown difficulties using these dips, including high-functioning individuals with autism or Asperger's syndrome (Alcantara, Weisblatt, Moore, & Bolton, 2004) as well as patients with multiple sclerosis who had confirmed demyelination in the central auditory system but normal peripheral function (Rappaport et al., 1994). Furthermore, it has been found that many children with a learning disability show marked behavioral deficits with listening to speech in noise (Bradlow, Kraus, & Hayes, 2003) concomitant with neurophysiologic deficits at the brainstem and cortical levels as revealed by abnormal electrophysiological responses to speech sounds presented in noise (King, Warrier, Hayes, & Kraus, 2002).

The importance of the corticofugal pathways (from the auditory cortex to the cochlea) in hearing in noisy environments and perception of speech in noise—although less understood—has been examined, especially in the animal literature (Rajan, 2000). It has been suggested that the efferent auditory pathway enhances the encoding of signals in noise. These enhancements include improvement in threshold detection, intensity discrimination of tones in noise, and perception of speech in noise (Giraud, Garnier, Micheyl et al., 1997; Kawase, Delgutte, & Liberman, 1993; Zeng, Martino, Linthicum, & Soli, 2000). Furthermore, it has been reported that transient evoked otoacoustic emissions are higher in level and are suppressed to a lesser degree by contralateral stimulation in children with auditory processing disorders in comparison with an age-matched control group (Muchnik et al., 2004). This has led to speculation that the efferent auditory pathways are dysfunctional in at

least some children with auditory processing disorders. Although a majority of this evidence pertains to the brainstem efferent auditory pathways, the corticofugal pathways have been shown to impact directly on the excitation and inhibition of the inferior colliculus—the pathways thought to be functionally associated with frequency analysis and acoustic orientation (Sun, Jen, Sun, & Zhang, 1989), arguably important processes for listening in noisy environments. A potentially contributing region in the central auditory system, which marks the beginning stage of the corticofugal pathway, would certainly be the lateral auditory cortex (superior temporal gyrus [STG]), which has been found to be important not only in processing speech (Liebenthal, Binder, Spitzer, Possing, & Medler, 2005; Scott & Wise, 2004; Wong, Nusbaum, & Small, 2004a) but also spectrally complex nonspeech signals (Hall et al., 2002) and signals with different amplitude characteristics (Jäncke, Shah, Posse, Grosse-Ryukun, & Müller-Gärtner, 1998). Although the middle portion of STG (mSTG) bilaterally might be important for performing acoustic analyses that are pertinent to separating the signal from noise, including the extraction of phonetically relevant acoustic information (Hickok & Poeppel, 2004), it is possible that other connected cortical regions are involved. Specifically, we hypothesize that increased phonological working memory might be needed to more successfully perform the task of higher phoneme-based demands. Baddeley (1992) postulates that such memory involves storage (posterior temporal regions) and rehearsal (frontal regions) components. Such a model is cited to support more recent neuroimaging models advocating a dorsal auditory–motor integration network involving posterior STG (pSTG) and prefrontal cortex (PFC), possibly the dorsal lateral aspect of PFC (Hickok & Poeppel, 2004; Scott & Wise, 2004). Support for this model also comes from an extensive review of neuroimaging studies showing pSTG to be especially relevant in phonological code retrieval (Indefrey & Levelt, 2000; cf. Cowan, 1995; Chein, Ravizza, & Fiez, 2003). Despite the fact that the presence of background noise is a ubiquitous property of communication (including the requirement to process spoken words in noise), only a handful of studies have been conducted examining cortical responses of speech perception in noise by younger adults (e.g., Salvi et al., 2002; Scott, Rosen, Wickham, & Wise, 2004; Zekveld, Heslenfeld, Festen, & Schoonhoven, 2006), although several others have used auditory stimuli embedded in noise without directly examining speech perception in noise as a mechanism that entails the use of different neural resources (e.g., Davis & Johnsrude, 2003; Wong et al., 2004b). Studies on the cortical electrophysiology of speech perception in noise have been conducted (e.g., Martin, Kurtzberg, & Stapells, 1999; Shtyrov et al., 1998, 1999); however, these studies are limited by the

spatial resolution of the methodology and cannot reveal the precise anatomical location of the neural structures involved.

Two positron emission tomography (PET) neuroimaging studies have been conducted to examine cortical mechanisms related to the perception of spoken words in noise (Salvi et al., 2002; Scott et al., 2004). Results from these studies are ambiguous about the role of the superior temporal region, leaving room for attributing speech perception in noise to cognitive and attentional (auditory independent), rather than sensory, processes. Salvi et al. (2002) asked younger participants with normal hearing to perceive speech in quiet and in monaurally presented noise. Half of the participants heard the stimuli from the left ear, half from the right ear. In addition, some participants were left-handed. Compared with listening to speech in quiet, listening to speech in noise showed additional activation in the medial frontal and cerebellar areas (but not the auditory cortex), implicating the involvement of attentional and motor planning networks. Because of the heterogeneity of the participant population and because not all participants received the same experimental condition, it is difficult to conclude whether those areas were the only ones involved in listening to speech in noise. It is worth noting that greater left STG activation was found in the speech in noise condition when the stimuli were presented to the right ear relative to the left ear; however, this might not be indicative of additional STG involvement during speech perception in noise relative to quiet, especially when the same pattern of results (left > right activation) was found in the perception of simple tones presented monaurally regardless of whether stimuli was targeted to the left or right ear (Devlin et al., 2003).

In another PET study, Scott et al. (2004) examined brain activation corresponding to “passively” listening to speech in two types of noise by 7 right-handed native English-speaking adult men with a mean age of 42 years. The noise types, namely broadband noise and sentences produced by a single talker, were used to simulate energetic and informational masking, respectively. Relative to the broadband noise condition, listening to speech in single-talker noise activated the left prefrontal cortex and the right posterior parietal lobe, implicating the engagement of attentional and cognitive networks. Listening in the broadband noise condition, relative to the single-talker noise condition, activated the superior temporal gyrus. Understanding the role of the STG activation in the previously mentioned results is not straightforward. Because language comprehension is associated with the STG activation, listeners might still be able to comprehend the single talker noise (e.g., Hickok & Poeppel, 2004). Furthermore, increasing the level of the single talker noise did not lead to an increase in STG activation, making any link between them even more

tenuous. It is also worth noting that neither of these comparisons addressed differences between listening to speech in quiet and in noise (more specifically, no condition required participants to listen to speech in quiet).

In the present study, we seek to more comprehensively investigate the cortical mechanisms (especially the lateral auditory cortex) in speech perception in noise. Three experimental conditions were included: listening to speech in quiet and listening to speech in two levels of multitalker babble noise. Participants in all conditions were asked to match single words with pictures. The three experimental conditions allowed for the examination of neural differences between listening in quiet and listening in noise, as well as listening in different levels of noise. Multitalker babble noise is often used in clinical testing; it represents many ecologically relevant listening situations where patients complain of listening difficulties (e.g., in a crowded restaurant), and it eliminates potential semantic effects, as it is unintelligible. Upon extensive behavioral testing in pilot experiments, we identified two signal-to-noise ratios (SNRs); one resulted in identical behavioral performance compared with listening in quiet, whereas the other one significantly impaired participants' behavioral performance. Thus, our design allowed us to examine noise-dependent brain responses with or without behavioral differences. Our design involving the collection of behavioral data in the scanner also provides an objective way for ascertaining participants' active participation during scanning.

Although PET generally has lower spatial resolution than fMRI, PET is chosen in many auditory experiments because there is minimal background noise generated by the scanner, unlike the MR environment. Recently, more auditory fMRI experiments have been conducted using a technique called *sparse sampling*, which uses long image repetition time (TR) so there is no imaging noise during stimulus presentation (e.g., Belin, Zattore, Hoge, Evans, & Pike, 1999; Gaab & Schlaug, 2003; Hall et al., 1999; Margulis, Mlsna, Uppunda, Parrish, & Wong, in press; Wong, Perrachione, & Parrish, 2007). The sparse sampling method samples the hemodynamic response once per TR at specific time points relative to stimulus presentation. The current study takes advantage of this method. We hypothesized that even when no behavioral differences are present (between the quiet and SNR 20 conditions, as discussed in subsequent paragraphs), listening to speech in noise will involve the auditory cortex for performing acoustic analysis that includes extracting phonetically relevant information from the noise-embedded speech signal (especially with the involvement of mSTG bilaterally). As the noise level and task difficulty increase, the more dorsal pathway (left pSTG) will show increased activation, signaling the

involvement of phonological memory, auditory–motor integration, and auditory attention.

Materials and Method

Participants

Participants were 11 young adult native speakers of American English with the mean age of 24.3 years (range = 20–34 years; 7 females, 4 males) who reported no audiologic or neurologic deficits. The participants were right-handed as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971). All participants passed a pure-tone screening at 25 dB at 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz.

Stimuli and Experimental Procedures

There were three experimental conditions; all included a set of single words. These single words, taken from Raymer et al. (1990), occurred at a low frequency rate in American English (see Appendix A for the word list). These words were produced by a native male speaker of American English in a sound-attenuated chamber and were recorded via a SHURE SM58 microphone onto a Pentium IV PC sampled at 44.1 kHz. Words were embedded in multitalker babble noise (taken from the noise channel of SPIN (Bilger, Nuetzel, Rabinowitz, & Rzeczkowski, 1984) at -5 dB SNR (henceforth referred to as the *SNR -5 condition*) in the first condition and at 20 dB SNR (henceforth, *SNR 20*) in the second condition. In the third condition, words were presented with no background noise (henceforth, *quiet*). The noise accompanying each word was identical in length to the word. RMS amplitude of the stimuli was calculated and normalized using Level 16 (Tice & Carrell, 1997). SNR was manipulated by varying the amplitude level of the noise while keeping the amplitude of the spoken words constant across the three conditions. The RMS amplitude of the speech material was set at 65 dB SPL, with the noise added at 70 dB SPL and 45 dB SPL for the SNR -5 and SNR 20 conditions, respectively. As the focus of the present study is on how speech perception is affected by noise and the involvement of higher cortical areas, we intentionally manipulated only the noise level to examine how changing the noise level affects the perception of speech presented at a constant level, as is the case in related studies (e.g., Binder, Liebenthal, Possing, Medler, & Ward, 2004; Zekveld et al., 2006). The combined output of the two headphones was measured for all stimuli across the three conditions to ascertain the variation in overall signal level. Specifically, peak pressure for each stimulus token was measured using a flat-plate coupler and a Larson-Davis sound level meter with a 500-gm weight on one headphone at a time. The recorded peak pressure output of the two headphones was combined

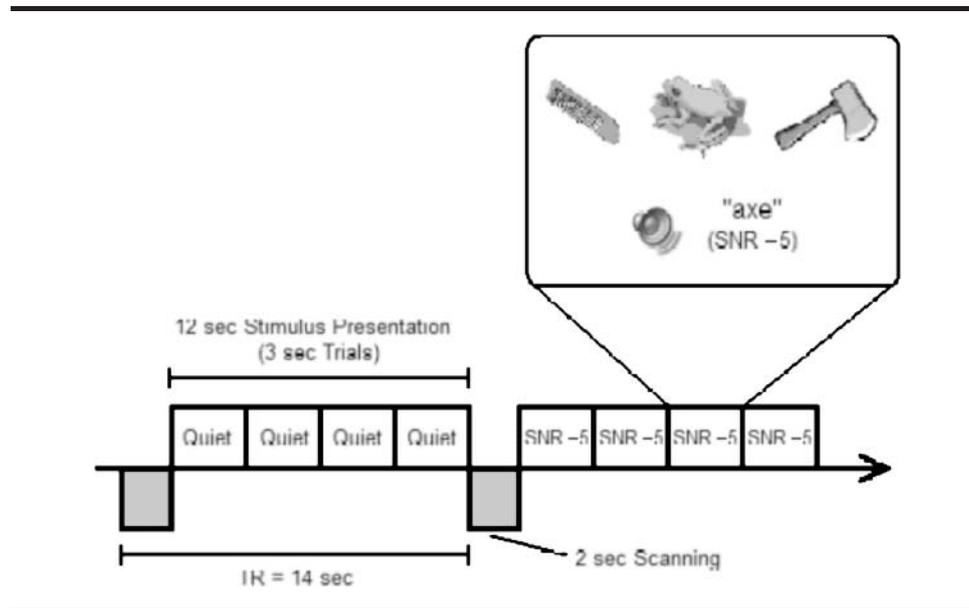
and converted back to dB SPL for comparison across conditions. For the quiet condition, the mean peak level of the stimuli was 81.6 dB SPL ($SD = 1.96$, range = 77.3–85.3); for the SNR 20 condition, the mean was 81.7 dB SPL ($SD = 2.01$, range = 75.6–87.2); and for the SNR -5 condition, the mean was 85.2 dB SPL ($SD = 1.57$, range = 78.7–88.7). In other words, the sound level differences across the three conditions were minimal (not greater than 3.7 dB in the most extreme case) and certainly were not sufficient to induce neurophysiologic differences detectable by fMRI (see Discussion section for further information).

Figure 1 details the stimulus presentation and fMRI sequence. As shown in Figure 1, scanning occurred every 14 s (indicated by the shaded boxes below the line). Scanning itself lasted 2 s when no stimuli were presented. Between scanning, four trials of stimuli from the same condition (quiet, SNR 20, or SNR -5) were presented. Each trial lasted 3 s, including the presentation of an auditory word (the stimulus target word; see Appendix A for the list of words used) that lasted no more than 700 ms and three simultaneous picture choices that remained for the entire 3-s period. One of the three picture choices matched the target stimulus (auditory word). Each of the open boxes above the line in Figure 1 represents a 3-s trial. The zoomed-in box represents one trial with three picture choices and one auditorily presented target word (in this case, *axe*). Participants were asked to identify the stimulus by pressing the button (left, middle, or right) corresponding to the picture (left, middle, or right); in the example depicted by Figure 1, the participant should press the right button. These procedures are similar to Fallon, Trehub, and Schneider (2000) and are highly suitable for the fMRI environment, as talking and movement of the head is minimized.

Each condition (quiet, SNR 20, and SNR -5) contained the same 20 target words. Within each set, the 20 words occurred once at the three picture positions (i.e., 60 trials total for each set); however, the picture distracters, randomly chosen from an independent set of 42, were different for each occurrence of the target. In other words, each of the 60 trials within a set of stimuli had a unique configuration of picture distracters. Moreover, the configurations were different across the three conditions, thereby minimizing practice effects. The 60 trials within each condition were presented four times, totaling 240 trials per set.

Before noise was added to any stimuli, 8 normal-hearing participants were asked to identify these three sets of stimuli in quiet (the configurations of the distracters as well as the order of presentation were identical to the actual fMRI experiment with noise). No reliable differences in accuracy and reaction time among the conditions were found, suggesting that (a) there are

Figure 1. Stimulus presentation and image acquisition. Scanning occurred every 14 s. Each scan lasts 2 s (shaded boxes below the line). Between scanning, participants were presented with four trials (open boxes above the line) of stimuli from the same condition—that is, either listening to speech with no background noise (quiet), speech embedded in noise with a signal-to-noise ratio (SNR) of 20 dB (SNR 20) or an SNR of -5 dB (SNR -5 ; only two blocks are illustrated), with two distractor picture choices and one target picture choice (see zoomed-in box). There were also 30 null trials when no stimuli (speech or noise) were presented.



minimal inherent differences in their identifiability and (b) any differences in behavioral and hemodynamic responses in the actual experiment were likely due to the addition of noise. Furthermore, it suggests that any possible repetition or priming effects are distributed equally across the conditions. Participants listened to the stimuli binaurally via headphones that were custom made for MRI environments (Avotec). The stimuli were played at the intervals specified above using the E-prime software program (Psychology Software Tools, Pittsburgh, PA).

MRI Acquisition

The present experiment used *fMRI*, a neuroimaging technology that has been used in research in communication sciences for more than a decade (e.g., Binder et al., 1996; Fridriksson, Morrow, Moser, & Baylis, 2006; Hugdahl et al., 2004). There are numerous authoritative sources that aimed at providing comprehensive information about this method (e.g., Aguirre & D'Esposito, 2000; Buckner & Braver, 2000; Chen & Ogawa, 2000; Friston, 1997; Huettel, Song, & McCarthy, 2004). For the current study, MR images were acquired at the Center for Advanced MRI in the Department of Radiology at Northwestern University using a Siemens 3T Trio. A high-resolution, three-dimensional anatomical (T1-weighted) volume was acquired sagittally and was used in conjunction with the functional activation maps to

localize the anatomic regions involved. Functional (T2*-weighted) images were acquired axially. Detailed technical information is provided in Appendix B. A sparse sampling method was used, which allowed image acquisition to be halted (i.e., no scanner noise) during stimulus presentation and thus minimized contamination of the speech stimuli. In addition, a TR of 14 s was used, which means that acquisition of images took place every 14 s. Stimuli were presented in this 14-s interval when no scanner noise was present. Because a hemodynamic response function, which is estimated to peak between 6 and 8 s (e.g., Belin et al., 1999) is generated in the brain by auditory stimulation (scanner noise or speech), a long TR provided sufficient time for the scanner-noise-generated hemodynamic response to decay so that its peak was not captured by the images (e.g., Belin et al., 1999; Gaab, Gaser, Zaehle, Jancke, & Schlaug, 2003; Hall et al., 1999; Margulis et al., in press; Wong et al., 2007). Within each TR, scanning occurred during the first 2 s, and participants performed the task for 12 s when no scanning occurred. The 12-s block included the presentation of four stimulus trials, each lasting 3 s. Although the stimulus trials from each condition were randomized across the *fMRI* experiment, within each block, the trials came from only one condition (quiet, SNR 20, or SNR -5). There were 60 blocks of stimuli for each condition, order-randomized across the *fMRI* experiment. In addition, there were 30 null trials of

scanning when no stimuli (speech or noise) were presented. In total, there were 210 (60 × 3 + 30) 14-s TRs lasting 49 min.

fMRI Data Analyses

The functional MR images (also called *time series*) were analyzed by BrainVoyager (Goebel et al., 2006). After standard preprocessing procedures (see Appendix B for technical information), hemodynamic responses were estimated on a voxel-by-voxel basis using a multiple linear regression model. The regressors for the multiple linear regression model consisted of the time points where the three experimental conditions (quiet, SNR 20, and SNR -5) occurred. In multiparticipant analyses, we used normalized beta values acquired from multiple linear regression signifying the fit of the regressors to the functional time series, voxel by voxel, for each condition. The multiparticipant analyses consisted of a random effect analysis and statistical contrasts (i.e., SNR -5 vs. quiet, SNR 20 vs. quiet, and SNR -5 vs. SNR 20) on a voxel-by-voxel basis. Anatomical and functional images from each participant were transformed into a standard stereotaxic space (Talairach & Tournoux, 1988).

Results

Behavioral Results

A repeated measures analysis of variance (ANOVA) showed a main effect of condition, $F(2, 20) = 22.15$, $p < .001$, on response accuracy. Post hoc pairwise comparisons showed that participants were less accurate in the SNR -5 condition than in the quiet ($p < .001$) and SNR 20 ($p < .001$) conditions. No reliable difference

in accuracy between the quiet and SNR 20 condition ($p = .811$) was found. Reaction time data showed similar results. A repeated measures ANOVA performed on reaction time showed a main effect of condition, $F(2, 20) = 22.15$, $p < .001$. Post hoc pairwise comparisons likewise showed that participants were slower in the SNR -5 condition ($p < .001$) than in the quiet and SNR 20 ($p < .001$) conditions, whereas no reliable difference was observed between the quiet and SNR 20 conditions ($p = .674$). In other words, although the stimuli were presented in the presence of background noise in the SNR 20 condition, participants did not show poorer performance relative to the quiet condition, as indicated by both accuracy and reaction time data. However, when the noise level was increased as in the SNR -5 condition, participants' performance was impaired. These results are summarized in Figure 2.

Imaging Results

We report here voxel-wise contrasts from a random effect analysis (Table 1) for clusters exceeding a single-voxel p value of $< .001$ ($t = 4.60$) extending at least 125 mm^3 . These clusters of activation survived cluster correction procedures implemented in BrainVoyager (Goebel et al., 2006), which was calculated on the basis of a Monte Carlo simulation and a family-wise alpha of less than .05. We then report a region-of-interest (ROI) analysis focusing on the auditory cortex and correlational analyses between activation in the ROI and behavioral performance.

Voxel-Wise Contrasts

SNR 20 versus quiet. Listening to speech in a moderately noisy condition (SNR 20), relative to listening to

Figure 2. Participants' behavioral performance. Error bars indicate standard error of the mean. * $p < .05$ (Bonferroni correction applied), based on a pairwise t test between the indicated conditions (results are significant for both accuracy and reaction time measures).

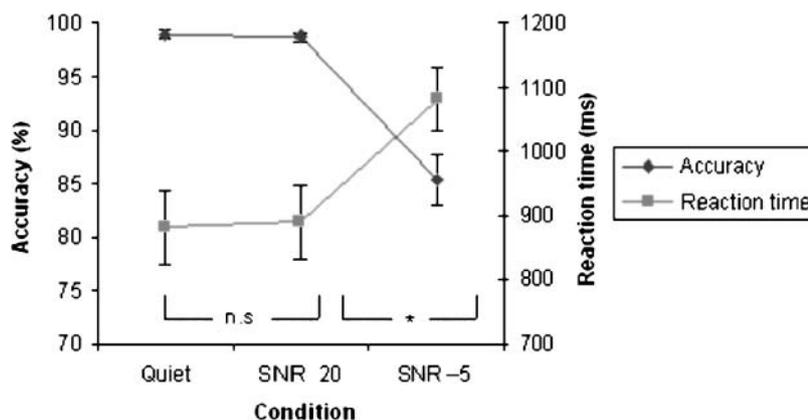


Table 1. Regions of activation.

Area (BA)	x	y	z	t value	Size (mm ³)
SNR 20 dB vs. quiet					
L STG (BA 22)	-57	-16	7	10.88	1930
L STG (BA 22)	-45	-40	13	6.43	333
L IPL/STG (BA 40)	-45	-46	22	8.15	262
L MFG (BA 6)	-30	-4	46	7.79	2752
L fusiform gyrus (BA 37)	-33	-43	-11	6.45	125
L lingual gyrus (BA 18)	-21	-85	-2	5.40	147
L MeFG (BA 9)	-18	35	19	7.30	1401
L posterior lingual gyrus (BA 30)	-12	-64	13	7.07	149
L lingual gyrus (BA 18)	-6	-85	-2	6.94	198
L anterior cingulate (BA 24)	-3	17	22	8.07	344
L thalamus	-18	-13	1	12.49	3318
R mid-STG (BA 22)	51	-19	4	8.46	4650
R pSTG (BA 22)	60	-40	10	6.51	667
R posterior cingulate	9	-43	10	7.49	528
R thalamus	12	-19	1	7.66	234
R globus palladius	18	-4	4	7.47	469
R cingulate	15	-10	34	7.16	711
R lingual gyrus (BA 18)	21	-76	-8	5.20	133
R cingulate gyrus	24	5	31	10.39	3208
R fusiform gyrus (BA 20)	45	-22	-23	6.89	145
SNR -5 dB vs. quiet					
L STG (BA 22)	-66	-19	4	10.02	2745
L IPL (BA 7)	-33	-52	40	5.90	1205
L IFG (BA 9)	-48	7	28	6.70	273
L SFG (BA 6)	-3	14	52	9.81	1884
L MFG (BA 46)	-45	29	22	10.07	559
L anterior insula (BA 13)	-33	17	10	7.53	917
L precentral gyrus (BA 6)	-30	-7	34	11.20	2833
L fusiform gyrus (BA 20)	-39	-40	-17	6.49	579
L precuneus (BA 7)	-27	-61	31	6.18	181
L thalamus	-12	-13	4	8.84	2554
R STG (BA 22)	51	-19	4	7.72	1238
R MTG (BA 22)	39	-46	4	6.24	392
R IFG (BA 46)	36	29	16	6.74	155
R IFG (BA 45)	45	11	19	6.86	139
R precentral gyrus (BA 9)	42	-4	37	11.9	662
R MFG (BA 6)	27	-1	43	8.60	482
R MeFG (BA 6)	9	-1	52	7.39	255
R precuneus (BA 7)	18	-64	28	7.32	639
R thalamus	15	-16	10	8.63	928
R fusiform gyrus (BA 19)	24	-82	-11	6.88	1160

(Continued on the following page)

Table 1 Continued. Regions of activation.

Area (BA)	x	y	z	t value	Size (mm ³)
SNR -5 dB vs. SNR 20 dB					
SNR -5 > SNR 20					
L pSTG (BA 22)	-54	-34	7	7.30	167
L anterior insula (BA 13)	-30	20	13	6.38	231
SNR 20 > SNR -5					
L SOG/MTG (BA 19)	-48	-70	25	-6.70	1287
L parahippocampal gyrus	-21	-4	-17	-6.14	195
L fusiform gyrus (BA 20)	-54	-10	-23	-6.42	147
L aITG (BA 21)	-63	-7	-17	-8.03	403
L posterior cingulate	-9	-52	13	-6.22	427

Note. The coordinates represent the location of the peak voxel for a cluster in Talaraich space. BA = approximate Brodmann's area; L = left; STG = superior temporal gyrus; IPL = inferior parietal lobule; MFG = middle frontal gyrus; MeFG = medial frontal gyrus; R = right; mid-STG = middle superior temporal gyrus; pSTG = posterior superior temporal gyrus; IFG = inferior frontal gyrus; SFG = superior frontal gyrus; MTG = middle temporal gyrus; SOG = superior occipital gyrus; aITG = anterior inferior temporal gyrus.

speech in quiet, activated more strongly an extensive brain network including the STG bilaterally, spatially extensive on the right but higher intensity on the left (Figure 3a). This STG activation extends in the anterior–posterior dimension including the mSTG. Activation was also found in the inferior parietal region (Figure 4a) and middle frontal gyrus (Figure 4c). In addition, attentional areas (e.g., cingulate gyrus, medial frontal gyrus) as well as subcortical structures (basal ganglia and thalamus) were activated. Note that these differences in brain activation were found in the absence of behavioral differences.

SNR -5 versus quiet. Similar to the SNR 20 condition, listening to speech in a highly noisy condition (SNR -5) relative to listening to speech in quiet also

activated an extensive auditory–attentional network. In particular, the left STG was activated with greater intensity and spatial extent than the right STG (Figure 3b). Moreover, additional parietal and frontal regions showed greater activation, including the inferior parietal lobule (Figure 4c) and the lateral prefrontal cortex (Figure 4d) as well as the left anterior insular cortex.

SNR -5 versus SNR 20. A direct comparison between the two noise conditions showed that listening to speech in a highly noisy condition showed increased activation in the left pSTG (Figure 3c) and left anterior insula (Figure 5). No activation in the right hemisphere was found. In the reverse comparison (SNR 20 > SNR -5), activation included that in the inferior temporal region

Figure 3. Brain activation revealed by the (a) SNR 20 vs. quiet, (b) SNR -5 vs. quiet, and (c) SNR -5 vs. SNR 20 statistical contrasts pooled across participants. Arrows indicate activation in the superior temporal region (auditory cortex). Note left lateralized superior temporal gyrus (STG) activation as noise level increased. Activation was projected onto a T1-weighted volume averaged across all participants.

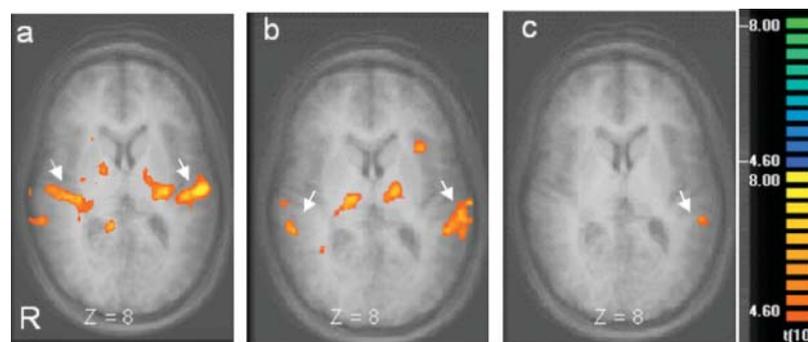
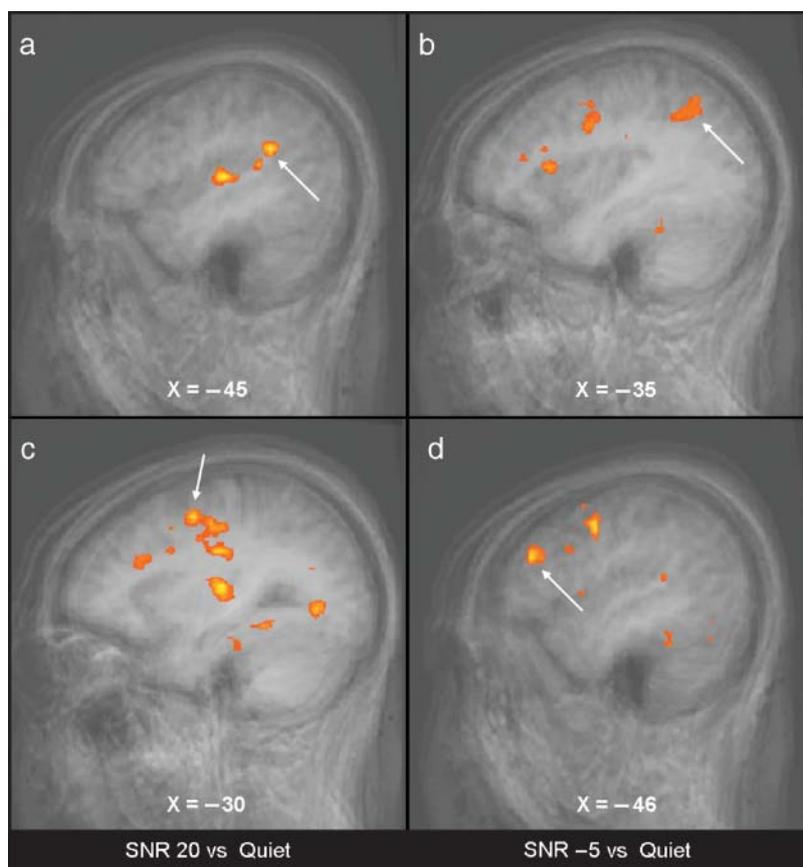


Figure 4. Brain activation revealed by several statistical contrasts pooled across participants. Arrows indicate activation of interests: (a) IPL/STG (BA 40) activation revealed by the SNR 20 vs. quiet contrast, (b) IPL (BA 7) activation revealed by the SNR -5 vs. quiet contrast, (c) MFG (BA 6) activation revealed by the SNR 20 vs. quiet contrast, and (d) MFG (BA 46) activation revealed by the SNR -5 vs. quiet contrast. Color scale from Figure 3 applies here. IPL = inferior parietal lobule; STG = superior temporal gyrus; BA = approximate Brodmann's area; MFG = middle frontal gyrus.



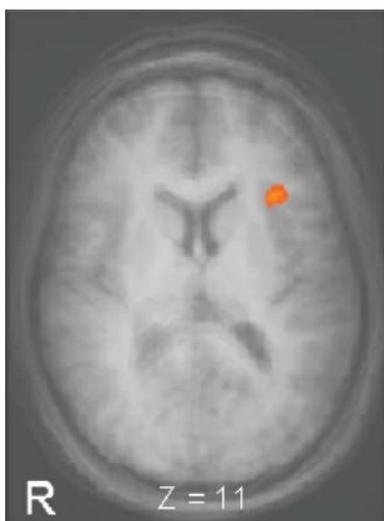
as well as a large cluster that began in the posterior edge of the middle temporal gyrus and extended posteriorly to reside almost entirely in the occipital lobe.

ROI Analysis

The voxel-wise contrasts reported in the previous section suggest the importance of the superior temporal region in perceiving speech in noise. In particular, the left STG appeared to be especially involved when the level of background noise was increased. To further investigate the role of the STG, we performed the following ROI analysis. We identified the strongest 5 mm × 5 mm × 5 mm activating cluster in the superior temporal gyri in both hemispheres when all conditions from all participants were combined; the peak activating voxels were [-63, -22, 4] and [49, -19, 7]. For each participant, we then calculated the percent signal change for each experimental condition in this region and

entered them into a 2 × 3 (Hemisphere × Condition) repeated measures ANOVA. Figure 6 shows the results. We found a main effect of condition, $F(2, 20) = 45.93$, $p < .001$; a main effect of hemisphere, $F(1, 20) = 13.20$, $p < .005$; and a significant Condition × Hemisphere interaction, $F(2, 20) = 9.34$, $p < .01$. Paired samples t tests showed that although there was a significant difference in brain activation between the SNR 20 and SNR -5 conditions in the left hemisphere ($p < .003$), no such difference was found in the right hemisphere ($p = .639$). Significant results survived Bonferroni correction for multiple comparisons with a family-wise alpha of less than .05. In other words, although behavioral differences were observed between the SNR 20 and SNR -5 conditions (SNR 20 being better and faster than SNR -5), only the left hemisphere showed differences in brain activation (SNR -5 > SNR 20). Furthermore, although behavioral differences were not observed between the quiet and SNR 20 conditions, there was

Figure 5. Left anterior insular activation revealed by the SNR -5 vs. SNR 20 contrast. Color scale from Figure 3 applies here.



increased brain activation in the STG bilaterally in the SNR 20 condition.

Correlational Analyses

To further investigate the relationship between brain activation and behavioral performance, we performed a series of correlational analyses between brain activation in the STG in the ROI defined previously and participants' accuracy scores for each experimental condition (e.g., correlation between left STG activation in the SNR -5 condition and accuracy score in that condition). In addition, we calculated correlation between accuracy

score and STG lateralization index ($(\text{left} - \text{right}) / [(\text{left} + \text{right}) / 2]$) (the higher the number, the higher the activation in the left hemisphere relative to the right). The only significant correlations that we found were associated with accuracy in the SNR -5 condition. Although we did not find a significant correlation between accuracy in the SNR -5 condition and left STG activation (Pearson's $r = .178, p = .3$; see Figure 7a), we found a positive correlation between accuracy and brain activation in the right STG (Pearson's $r = .679, p < .011$; see Figure 7b) as well as a negative correlation between accuracy and lateralization index (Pearson's $r = -.575, p < .032$). There was a marginal negative correlation between lateralization index obtained from the SNR 20 condition and behavioral performance in that condition (Pearson's $r = -.508, p < .055$). These results suggest that increased behavioral performance is associated with increased activation in the right STG (with or without considering activation on the left), especially when the background noise is loud.

Discussion

The present study shows that relative to listening to speech in quiet, listening to speech in noise resulted in increased brain activation in a network of brain areas, including the auditory cortex. Such strong involvement of the auditory cortex suggests that speech perception in noise in the human is associated with sensory processes involving the central nervous system. As such, accounts for speech perception in noise deficits would need to consider the contribution of the central auditory nervous system, in addition to higher level cognitive and attentional processes.

Figure 6. STG activation (ROI defined by the strongest activating 5mm^3 cluster pooled across all conditions and participants) in each condition. Error bars indicate standard error of the mean. ROI = region of interest. * $p < .05$ (Bonferroni correction applied), based on a pairwise t test between the indicated conditions.

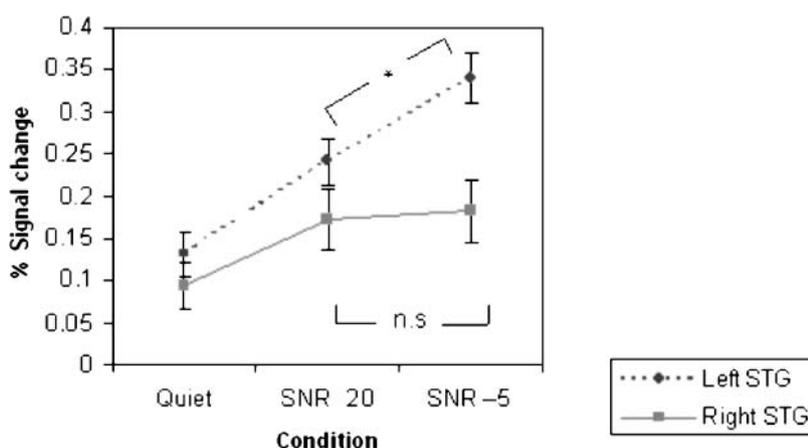
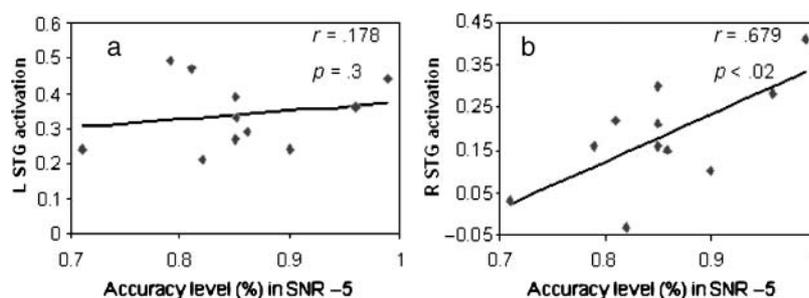


Figure 7. Correlations between accuracy performance in the SNR -5 condition and (a) left STG and (b) right STG.



Our findings are likely a reflection of participants' effort of using temporal and spectral dips to overcome background noise in identifying spoken words, which is likely associated with activation in mSTG as well as increased auditory-motor integration and phonological memory associated with the pSTG. Auditory-motor integration can be interpreted as the activation and integration of both auditory and motor representations of the acoustic-phonetic input for arriving at an accurate phonological representation (see Hickok and Poeppel, 2004, for a detailed discussion). Phonological memory can be viewed as a process in which phonological representations are being maintained temporarily, with the aid of subvocal rehearsal, for the completion of the spoken word-picture matching task (Baddeley, 1992, 2000). Note that despite equal behavioral performance in the quiet and SNR 20 conditions, more brain activation was seen in the SNR 20 condition, implicating a noise-dependent and behavioral-independent effect. This contrast especially shows STG to be activated extensively in the anterior-posterior dimension, including the mSTG, which is likely important for acoustic analysis and phonetic extraction (Hickok & Poeppel, 2004). Only the left pSTG shows increased activation when noise level was increased (when the two noise conditions were compared). In addition to STG, additional activation was observed in the inferior parietal area as well as in the PFC during the noise conditions relative to the quiet condition. The activation follows the posterior auditory pathway proposed by Rauschecker (1998) and is suggestive of possible auditory-motor integration (Hickok & Poeppel, 2004) and a phonological working memory network (Baddeley, 1992), in addition to acoustic analysis, as required by listening to speech in noise. This posterior pathway includes the parabelt region, the parietal area, and ultimately the prefrontal cortex (Rauschecker, 1998). The poor behavioral performance in the SNR -5 condition relative to SNR 20 condition points to its task difficulty and the likely need for additional phonological working memory.

Although we interpreted our results in the context of working memory and auditory-motor integration,

the activation patterns revealed by the current study can be readily mapped onto findings in the auditory attention literature. For example, similar to our study, Lipschutz, Kolinsky, Damhaut, Wikler, & Goldman (2002) and Pugh et al. (1996) found activation in frontal, parietal, and temporal regions during auditory attention tasks, including spatial attention. Moreover, an alternative model to Baddeley's (1992) phonological working framework that is called the embedded-processes model (Cowan, 1995) can potentially account for our results (see Chein et al., 2003, for a review). This latter model, involving the interaction of components such as long-term storage, rehearsal, activated memory, and focus of attention, can be arguably mapped onto the areas of activation observed in our study. Although it is not the scope of our current study to experimentally test the two models (Baddeley's vs. Cowan's), future research can more specifically differentiate them in a hypothesis-driven manner.

Note that one recent study reported SNR-dependent STG changes associated with the perception of simple nonsense syllables /ba/ and /da/ embedded in different levels of white noise (Binder et al., 2004). The /ba/ and /da/ syllables were synthetic stimuli differing only in the initial formant trajectories. Unlike our study, Binder et al. found that behavioral accuracy levels closely followed STG activation levels bilaterally so that the more accurate the participants were in perceiving /ba/ and /da/, the higher the STG activation. These results were interpreted to reflect the association between STG activation and the quality of sensory information (the nonsense syllables were of a higher auditory quality in high SNR conditions). Unlike their study, our experiment required participants to actively identify various spoken words that differed substantially in spectral and temporal characteristics over the entire stimulus, not to mention lexical-semantic differences. Thus, we speculate that our findings are not only a reflection of the quality of sensory information but listeners' active effort in overcoming noisy listening situations for achieving optimal spoken word identification.

We found that activation in some of the brain regions was independent of task difficulty, as indicated by both accuracy and reaction data as well as semantic processing effects (participants found spoken words in both quiet and SNR 20 conditions to be equally intelligible). Activation in these brain regions, including bilateral auditory cortex, the thalamus, the cingulate gyrus, and various frontal areas, suggest the recruitment of additional neural resources for extracting the speech signal in the presence of noise. In the case of the moderately noisy condition (SNR 20), the engagement of these brain areas successfully overcame the presence of background noise to achieve equal performance (relative to quiet) in identifying the speech signal. However, we found that as noise level increased to an SNR of -5 dB, behavioral performance was impaired. This decrease in behavioral performance was accompanied by an increase in the left pSTG (relative to quiet and SNR 20) but a steady level of right STG activation (relative to SNR 20).

The differences between the left and right STG activation in the SNR -5 condition relative to the SNR 20 condition can be interpreted in at least two ways. First, it can be argued that activation in the left STG prevented behavioral performance from being more impaired. After all, depending on the control condition, left STG has been observed to be associated with speech processing in numerous speech perception studies (e.g., Scott & Wise, 2004; Wong et al., 2004a, 2007). Second, it can be argued that the lack of further increase in right STG activation in the SNR -5 condition was driving the relatively poor behavioral performance in that condition. The results from the present study seem to support the latter. Correlational analyses revealed that participants who showed greater activation in the right STG were more accurate in identifying spoken words in that condition, whereas no such correlation was found in the left STG. The involvement of the right hemisphere in accurately performing difficult behavioral tasks has been found in numerous recent neuroimaging studies. For example, Reiterer et al. (2005) found that in pitch and duration discrimination tasks, increased task performance was associated with increased activation in the right auditory cortex. In visual processing, Grondin and Girard (2005) found that task difficulty is associated with stimuli presented to the left, but not the right, visual field, indicating the additional involvement of the right visual areas. Our study complements these previous studies by showing the possible role of the right auditory cortex in contributing to difficult word recognition conditions.

Recent studies have found left anterior STG to be associated with the perception of intelligible speech (e.g., Narain et al., 2003; Scott, Blank, Rosen, & Wise, 2000). Note that the concept of intelligibility is especially tied to the quality of the auditory sensation as

opposed to active engagement in delineating the signal from noise (which likely involves auditory–motor integration and pSTG). Specifically, Scott et al. (2004) found that when all masking (speech or noise maskers) and SNR conditions were considered and averaged, activation in the anterior STG was correlated with how well participants identified the spoken words in a pre-testing behavioral experiment. We identified a $5\text{ mm} \times 5\text{ mm} \times 5\text{ mm}$ voxel cluster in the STG centered at $[-58, 0, 0]$, which Scott et al. (2004) found to be associated with intelligibility. Averaged brain activation in this region across the three experimental conditions were then extracted for each participant, and correlation analyses were performed on averaged accuracy and averaged reaction time data across the experimental conditions. Although we did not find a significant correlation between accuracy and anterior STG activation, we found a significant negative correlation between reaction time and anterior STG activation so that the faster the response, the higher the STG activation. This provides some modest evidence for the association between this region and task difficulty (participants were more likely to respond faster on easier/more intelligible trials).

Previous research has suggested that increased task demand and effort results in increased brain activation (Just, Carpenter, Keller, Eddy, & Thulborn, 1996; Lecas, 1995). More specifically, Wong et al. (2004a) found increased activation in the superior temporal region when participants identified words in mixed- and blocked-talker conditions (the former being a more difficult task). Therefore, it is not surprising to find increased superior temporal activation in the current study. However, it is also likely that the superior temporal region is specifically associated with separating noise from signal (mSTG) and auditory–motor integration (pSTG), including the additional attentional resources required for such processes. Although we do not have direct evidence suggesting that our participants were using temporal cues in their attempt to extract the speech signal from noise, it has been argued that the left STG is particularly involved in temporal processing (e.g., Zatorre, 2001). In addition to increased effort, there are likely changes in processing strategies for listening to speech in noise that could account for increased activation in other brain areas. Activation in the inferior frontal and prefrontal regions is likely due to increased subvocal rehearsal, as found in other studies (Paulesu, Frith, & Frackowiak, 1993), which may have assisted participants with overcoming the noise and accurately perceiving the speech sounds. Activation in these frontal and attentional areas is consistent with previous studies of speech (spoken word) perception in noise (Salvi et al., 2002; Scott et al., 2004).

Although not specifically investigating the neural correlates of listening in noise, Wong et al. (2004b)

embedded white noise in their stimuli in an attempt to increase task difficulty in a cross-linguistic experiment of speech perception. Depending on the linguistic relevance of the task, Wong et al. found increased anterior insular activation: left for native sounds and right for non-native sounds. The current study, involving words that are native to the participants, similarly found left anterior insular activation. This activation in the insular–frontal operculum area may also be related to the decision processing involved in auditory object perception (Binder et al., 2004), as participants were required to select one out of the three picture choices in the present study.

Activation in the thalamus was also observed in both noise conditions relative to the quiet condition. This activation may be a reflection of the engagement of the corticofugal pathway in separating the auditory signal from noise (e.g., Rajan, 2000; Sun et al., 1989), although the current experiment cannot distinguish the contribution of the afferent and efferent pathways.

It is noteworthy that there were slight differences in the mean sound level of the stimuli across the three conditions. We are equally mindful of the fact that the overall (combined) amplitude of the stimuli might be larger in some of the stimuli in the lower SNR conditions and that if all stimuli were normalized to the same level, the speech signal would have been reduced in the lower SNR conditions. Our procedure of keeping the level of the target signal constant while manipulating the level of the noise to create different SNR conditions has been used in related studies (e.g., Binder et al., 2004; Zekveld et al., 2006). In the most crucial comparison of our current study (SNR 20 vs. quiet), the mean peak level difference was only around 1.1 dB and was no more than 3.7 dB in the most extreme case (SNR –5 vs. quiet). Because the amplitude of our speech stimuli (regardless of the presence of noise) is highly variable and because the impact of speech (including lexicosemantics) on the auditory association area is particularly invasive—evident by the fact that even sinewave speech typically with lower amplitude level (if heard as speech) activates the STG to a similar extent as actual speech (Liebenthal et al., 2003)—it is likely that our results are approximately immune to the minimally variable stimulus amplitude characteristics. Evidence from non-speech auditory perception further points to the fact that primary auditory areas, and also subcortical structures, play a larger role in intensity level perception than auditory association areas and beyond (e.g., Hart, Palmer, & Hall, 2002). In the Hart et al. study, which also used a 3-Tesla MRI scanner, sound level differences greater than 6 dB failed to show differentiable activation in the auditory association areas.

In sum, we found evidence linking the involvement of the auditory cortex, including the pSTG, to listening

to spoken words in noisy conditions. These effects include those that are independent of task difficulty and dependent on noise levels. Future research will investigate neural characteristics of individuals with central auditory symptoms (e.g., the elderly).

Acknowledgments

This work was supported by Northwestern University and by the National Institutes of Health Grants HD051827 and DC007468, awarded to the first and third author, and National Institutes of Health Grant DC005692, awarded to the fourth author. We thank Diane Arcuri, Tasha Dees, Geshri Gunasekera, Kiara Lee, Nondas Leloudas, and Tyler Perrachione for their assistance in this research.

References

- Aguirre, G. K., & D'Esposito, M.** (2000). Experimental design for brain fMRI. In C. Moonen & P. A. Bandettini (Eds.), *Functional MRI* (pp. 369–380). New York: Springer-Verlag.
- Alcantara, J. I., Weisblatt, E. J., Moore, B. C., & Bolton, P. F.** (2004). Speech-in-noise perception in high-functioning individuals with autism or Asperger's syndrome. *The Journal of Child Psychology and Psychiatry*, *45*, 1107–1114.
- Baddeley, A.** (1992, January 31). Working memory. *Science*, *255*, 556–559.
- Baddeley, A.** (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*, 417–423.
- Bamiou, D.-E., Musiek, F. E., & Luxon, L. M.** (2001). Aetiology and clinical presentations of auditory processing disorder: A review. *Archives of Disease in Childhood*, *85*, 1–9.
- Belin, P., Zattore, R. J., Hoge, R., Evans, A. C., & Pike, B.** (1999). Event-related fMRI of the auditory cortex. *Neuroimage*, *10*, 417–429.
- Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., & Rzeczkowski, C.** (1984). Standardization of a test of speech perception in noise. *Journal of Speech, Language, and Hearing Research*, *27*, 32–48.
- Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A., & Ward, B. D.** (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nature Neuroscience*, *7*, 295–301.
- Binder, J. R., Swanson, S. J., Hammeke, T. A., Morris, G. L., Mueller, W. M., Fischer, M., et al.** (1996). Determination of language dominance using functional MRI: A comparison with the Wada test. *Neurology*, *46*, 978–984.
- Bradlow, A. R., Kraus, N., & Hayes, E.** (2003). Speaking clearly for children with learning disabilities: Sentence perception in noise. *Journal of Speech, Language, and Hearing Research*, *46*, 80–97.
- Buckner, R. L., & Braver, T. S.** (2000). Event-related functional MRI. In C. Moonen & P. A. Bandettini (Eds.), *Functional MRI* (pp. 441–452). New York: Springer-Verlag.
- Burton, M. W., Small, S. L., & Blumstein, S. E.** (2000). The role of segmentation in phonological processing: An fMRI investigation. *Journal of Cognitive Neuroscience*, *12*, 679–690.

- Chein, J. M., Ravizza, S. M., & Fiez, J. A.** (2003). Using neuroimaging to evaluate models of working memory and their implications for language processing. *Journal of Neurolinguistics, 16*, 315–339.
- Chen, W., & Ogawa, S.** (2000). Principles of BOLD functional MRI. In C. Moonen & P. A. Bandettini (Eds.), *Functional MRI* (pp. 103–113). New York: Springer-Verlag.
- Chermak, G. D., & Musiek, F. E.** (1997). *Central auditory processing disorders: New perspectives*. San Diego, CA: Singular Publishing Group.
- Chermak, G. D., Somers, E. K., & Seikel, J. A.** (1998). Behavioral signs of central auditory processing disorder and attention deficit disorder. *Journal of the American Academy of Audiology, 9*, 78–84.
- Cooper, J. C., & Gates, G. A.** (1991). Hearing in the elderly—the Framingham cohort, 1983–1985. Part II: Prevalence of central auditory processing disorders. *Ear and Hearing, 12*, 304–311.
- Cowan, N.** (1995). *Attention and memory: An integrated framework*. New York: Oxford University Press.
- Davis, M. H., & Johnsruide, I. S.** (2003). Hierarchical processing in spoken language comprehension. *Journal of Neuroscience, 23*, 3423–3431.
- Devlin, J. T., Raley, J., Tunbridge, E., Lanary, K., Floyer-Lea, A., Narain, C., et al.** (2003). Functional asymmetry for auditory processing in human primary auditory cortex. *Journal of Neuroscience, 23*, 11516–11522.
- Fallon, M., Trehub, S. E., & Schneider, B. A.** (2000). Children's perception of speech in multitalker babble. *The Journal of the Acoustical Society of America, 108*, 3023–3029.
- Fridriksson, J., Morrow, K. L., Moser, D., & Baylis, G. C.** (2006). Age-related variability in cortical activity during language processing. *Journal of Speech, Language, and Hearing Research, 49*, 690–697.
- Friston, K. J.** (1997). Imaging cognitive anatomy. *Trends in Cognitive Sciences, 1*, 21–27.
- Gaab, N., & Schlaug, G.** (2003). The effect of musicianship on pitch memory in performance matched groups. *NeuroReport, 14*, 2291–2295.
- Gaab, N., Gaser, C., Zaehle, T., Jancke, L., & Schlaug, G.** (2003). Functional anatomy of pitch memory—an fMRI study with sparse temporal sampling. *Neuroimage, 19*, 1417–1426.
- Giraud, A. L., Garnier, S., Micheyl, C., Lina, G., Chays, A., & Chéry-Croze, S.** (1997). Auditory efferents involved in speech-in-noise intelligibility. *Neuroreport, 8*, 1779–1783.
- Goebel, R., Esposito, F., & Formisano, E.** (2006). Analysis of functional image analysis contest (FIAC) data with Brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Human Brain Mapping, 27*, 392–401.
- Grondin, S., & Girard, C.** (2005). About hemispheric differences in the processing of temporal intervals. *Brain and Cognition, 58*, 125–132.
- Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R., et al.** (1999). “Sparse” temporal sampling in auditory fMRI. *Human Brain Mapping, 7*, 213–223.
- Hall, D. A., Johnsruide, I. S., Haggard, M. P., Palmer, A. R., Akeroyd, M. A., & Summerfield, A. Q.** (2002). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex, 12*, 140–149.
- Hart, C. H., Palmer, A. R., & Hall, D. A.** (2002). Heschl's gyrus is more sensitive to tone level than non-primary auditory cortex. *Hearing Research, 171*, 177–190.
- Hickok, G., & Poeppel, D.** (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition, 92*, 67–99.
- Huettel, S. A., Song, A. W., & McCarthy, G.** (2004). *Functional magnetic resonance imaging*. Sunderland, MA: Sinauer Associates.
- Hugdahl, K., Gundersen, H., Brekke, C., Thomsen, T., Rimol, L. M., Ersland, L., & Niemi, J.** (2004). fMRI brain activation in a Finnish family with specific language impairment compared with a normal control group. *Journal of Speech, Language, and Hearing Research, 47*, 162–172.
- Hugdahl, K., Heiervang, E., Nordby, H., Smievoll, A. I., Steinmetz, H., Stevenson, J., & Lund, A.** (1998). Central auditory processing, MRI morphometry, and brain laterality: Applications to dyslexia. *Scandinavian Audiology, 27*(Suppl.), 26–34.
- Indefrey, P., & Levelt, W. J. M.** (2000). The neural correlates of language production. In M. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 845–865). Cambridge, MA: The MIT Press.
- Jancke, L., Shah, N. J., Posse, S., Grosse-Ryken, M., & Müller-Gärtner, H. W.** (1998). Intensity coding of auditory stimuli: An fMRI study. *Neuropsychologia, 36*, 875–883.
- Just, M. A., Carpenter, P. A., Keller, T. A., Eddy, W. F., & Thulborn, K. R.** (1996, October 4). Brain activation modulated by sentence comprehension. *Science, 274*, 114–116.
- Kawase, T., Delgutte, B., & Liberman, M. C.** (1993). Anti-masking effects of olivocochlear reflex. II: Enhancement of auditory nerve response to masked tones. *Journal of Neurophysiology, 70*, 2533–2549.
- King, C., Warrier, C. M., Hayes, E., & Kraus, N.** (2002). Deficits in auditory brainstem pathway encoding of speech sounds in children with learning problems. *Neuroscience Letters, 319*, 111–115.
- King, W. M., Lombardino, L. J., Crandell, C. C., & Leonard, C. M.** (2003). Comorbid auditory processing disorder in developmental dyslexia. *Ear and Hearing, 24*, 448–456.
- Lecas, J. C.** (1995). Prefrontal neurons sensitive to increased visual attention in the monkey. *Neuroreport, 7*, 305–309.
- Liebenthal, E., Binder, J. R., Piorkowski, R. L., & Remez, R. E.** (2003). Short-term reorganization of auditory analysis induced by phonetic experience. *Journal of Cognitive Neuroscience, 15*, 1–10.
- Liebenthal, E., Binder, J. R., Spitzer, S. M., Possing, E. T., & Medler, D. A.** (2005). Neural substrates of phonemic perception. *Cerebral Cortex, 15*, 1621–1631.
- Lipschutz, B., Kolinsky, R., Damhaut, P., Wikler, D., & Goldman, S.** (2002). Attention-dependent changes of activation and connectivity in dichotic listening. *Neuroimage, 17*, 643–656.
- Margulis, E. H., Mlsna, L. M., Uppunda, A. K., Parrish, T. B., & Wong, P. C. M.** (in press). Selective neurophysiologic

- responses to music in instrumentalists with different listening biographies. *Human Brain Mapping*, DOI: 10.1022/hbm.20503.
- Martin, B. A., Kurtzberg, D., & Stapells, D. R.** (1999). The effects of decreased audibility produced by high-pass noise masking on N1 and the mismatch negativity to speech sounds /ba/ and /da/. *Journal of Speech, Language, and Hearing Research*, 42, 271–286.
- Muchnik, C., Ari-Even Roth, D., Othman-Jebara, R., Putter-Katz, H., Shabtei, E. L., & Hildesheimer, M.** (2004). Reduced medial olivocochlear bundle system function in children with auditory processing disorders. *Audiology and Neurotology*, 9, 107–114.
- Narain, C., Scott, S. K., Wise, R. J. S., Rosen, S., Leff, A., Iversen, S. D., & Matthews, P. W.** (2003). Defining a left-lateralized response specific to intelligible speech using fMRI. *Cerebral Cortex*, 13, 1362–1368.
- Oldfield, R. C.** (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia*, 9, 97–113.
- Paulesu, E., Frith, C. D., & Frackowiak, R. S.** (1993, March 25). The neural correlates of the verbal component of working memory. *Nature*, 362, 342–345.
- Peters, R. W., Moore, B. C., & Baer, T.** (1998). Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. *The Journal of the Acoustical Society of America*, 103, 577–587.
- Pugh, K. R., Offywitz, 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.
- Rajan, R.** (2000). Centrifugal pathways protect hearing sensitivity at the cochlea in noisy environments that exacerbate the damage induced by loud sound. *Journal of Neuroscience*, 20, 6684–6693.
- Rappaport, J. M., Gulliver, J. M., Phillips, D. P., Van Dorpe, R. A., Maxner, C. E., & Bhan, V.** (1994). Auditory temporal resolution in multiple sclerosis. *The Journal of Otolaryngology*, 23, 307–324.
- Rauschecker, J. P.** (1998). Cortical processing of complex sounds. *Current Opinion in Neurobiology*, 8, 516–521.
- Raymer, A. M., Maher, L. M., Greenwald, M. L., Morris, M., Rothi, L. J. G., & Heilman, K. M.** (1990). *Florida Semantics Battery*. Unpublished test.
- Reiterer, S. M., Erb, M., Droll, C. D., Anders, S., Ethofer, T., Grodd, W., & Wildgruber, D.** (2005). Impact of task difficulty on lateralization of pitch and duration discrimination. *NeuroReport*, 16, 239–242.
- Salvi, R. J., Lockwood, A. H., Frisina, R. D., Coad, M. L., Wack, D. S., & Frisina, D. R.** (2002). PET imaging of the normal human auditory system: Responses to speech in quiet and in background noise. *Hearing Research*, 170, 96–106.
- Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J.** (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, 123, 2400–2406.
- Scott, S. K., Rosen, S., Wickham, L., & Wise, R. J.** (2004). A positron emission tomography study of the neural basis of informational and energetic masking efforts in speech perception. *The Journal of the Acoustical Society of America*, 115, 813–821.
- Scott, S. K., & Wise, R. J. S.** (2004). The functional neuro-anatomy of prelexical processing in speech perception. *Cognition*, 92, 13–45.
- Shtyrov, Y., Kujala, T., Ahveninen, J., Tervaniemi, M., Alku, P., Ilmoniemi, R. J., & Näätänen, R.** (1998). Background acoustic noise and the hemispheric lateralization of speech processing in the human brain: Magnetic mismatch negativity study. *Neuroscience Letters*, 251, 141–144.
- Sun, X. D., Jen, P. H., Sun, D. X., & Zhang, S. F.** (1989). Corticofugal influences on the responses of bat inferior collicular neurons to sound stimulation. *Brain Research*, 495, 1–8.
- Talairach, J., & Tournoux, P.** (1988). *Co-planar stereotaxic atlas of the human brain. Three-Dimensional Proportional System: An Approach to Cerebral Imaging*. New York: Thieme Medical Publishers.
- Tice, B., & Carrell, T.** (1997). *Tone*. Lincoln, NE: University of Nebraska.
- Wright, B. A., Lombardino, L. J., King, W. M., Puranik, C. S., Leonard, C. M., & Merzenich, M.** (1997, May 8). Deficits in auditory temporal and spectral resolution in language-impaired children. *Nature*, 387, 176–178.
- Wong, P. C. M., Nusbaum, H. C., & Small, S. L.** (2004a). Neural bases of talker normalization. *Journal of Cognitive Neuroscience*, 16, 1173–1184.
- Wong, P. C. M., Parson, L. M., Martinez, M., & Diehl, R. L.** (2004b). The role of the insular cortex in pitch pattern perception: The effect of linguistic contexts. *Journal of Neuroscience*, 24, 9153–9160.
- Wong, P. C. M., Perrachione, T. K., & Parrish, T. D.** (2007). Neural characteristics of successful and less successful speech and word learning in adults. *Human Brain Mapping*, 28, 995–1006.
- Zatorre, R. T.** (2001). Neural specialization for tone processing. *Annals of the New York Academy of Sciences*, 930, 193–210.
- Zekveld, A. A., Heslenfeld, D. J., Festen, J. M., & Schoonhoven, R.** (2006). Top-down and bottom-up processes in speech comprehension. *NeuroImage*, 32, 1826–1836.
- Zeng, F. G., Martino, K. M., Linthicum, F. H., & Soli, S. D.** (2000). Auditory perception in vestibular-neurectomy participants. *Hearing Research*, 142, 102–112.

Received June 7, 2007

Accepted December 20, 2007

DOI: 10.1044/1092-4388(2008/075)

Contact author: Patrick C. M. Wong, Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, IL 60208.
E-mail: pwong@northwestern.edu.

Appendix A. List of words used for the experiment taken from Raymer et al. (1990).

anchor
axe
boot
broom
cane
carrot
comb
curtain
drum
frog
glove
hammer
lamb
onion
scarf
shack
snail
spoon
witch
worm

Appendix B. Detailed technical MRI information.

Parameters for acquiring T1-weighted anatomical images. MP-RAGE; TR/TE = 2,100 ms/2.4 ms, flip angle = 8°, TI = 1,100 ms, matrix size = 256 × 256, FOV = 22 cm, and slice thickness = 1 mm.

Parameters for acquiring and analyzing T2-weighted functional images.* Acquired using a susceptibility-weighted EPI pulse sequence while participants performed the behavioral task. TE = 30 ms, TR = 14 s, flip angle = 90°, in-plane resolution = 3.4375 mm², 24 slices with a slice thickness = 3 mm, and zero gap were acquired in an interleaved measurement. The preprocessing procedures involved linear detrending, motion and scan time correction, and spatial smoothing (FWHM 6 mm). Note that even though the TR was 14 s long, image acquisition only occurred during the first 2 s of the TR as stated, as opposed to the entire TR. Thus, the images collected reflected either a stimulus event occurring at one of these time points or a null event (no stimulus presented). Imaging at specific time points relative to stimulus presentation removed the need to convolve the task-related extrinsic waveforms with a hemodynamic response function before statistical analyses, as is commonly done (Wong et al., 2004b).

Note. MP-RAGE = magnetization-prepared rapid gradient-echo; EPI = echo planar imaging; TE = echo time; TR = repetition time; TI = T1 (no full term); FWHM = full-width at half-maximum.