Neural Bases of Listening to Speech in Noise

Patrick C.M. Wong[14], Kiara M. Lee[2], and Todd B. Parrish[23]

[1]Department of Communication Sciences and Disorders, [2]Department of Biomedical Engineering, [3]Department of Radiology, and [4]Northwestern University Institute for Neuroscience
Northwestern University, Evanston, Illinois, U.S.A.
pwong@northwestern.edu

Abstract

Ubiquitous speech processing involves listening to speech in ecological environments where noise is often present. The current study investigates the neural mechanisms involved in perceiving speech in noise using a sparse sampling fMRI method. Subjects were asked to match auditorily presented words with picture choices. The auditory stimuli were either presented in quiet or embedded in multi-talker babble noise without the presence of scanner noise in either condition. Behaviorally, it was found that subjects were slower and less accurate in identifying words presented in noise. Comparison of hemodynamic responses associated with listening to the two types of stimuli revealed increased activation in left prefrontal, inferior frontal, anterior insular, and superior temporal regions when subjects listened to speech in noise. These results confirm the importance of the lateral auditory cortex in complex auditory processing and suggest that the prefrontal cortex is likely to be prominently engaged in subvocal rehearsal when noise is affecting the integrity of the speech signal.

1. Introduction

A ubiquitous property of naturally occurring speech is the presence of noise in the environment. As such, listeners must develop cognitive and neural resources for handling noise to achieve successful speech processing and communication. The current study investigates the neural correlates associated with listening to speech in noise.

Much research investigating the neural bases of speech perception focuses on problems such as the perception of speech and non-speech sounds [1], the perception of consonants [2], and the perception of native and non-native speech sounds [3] 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 listening to speech in noise is a primary symptom of (Central) Auditory Processing Disorders [4], which is estimated to affect between 2% to 3% of children [5] and 22.6% of adults over 60 years of age [6]. It is also a characteristic of many individuals with learning disabilities [7] and dyslexia [8]. These disorders are often attributed to neuroanatomic and neurophysiologic anomalies [8], although no evidence has been provided with regards to how such anomalies affect the individuals' ability to listen to speech in noise.

The current study represents a first step towards a better understanding of how speech perception is accomplished despite the presence of noise. Studying speech perception in the magnetic resonance imaging (MRI) environment presents a special challenge because the noise generated by the scanner is often above 110 dB SPL. Although many of the aforementioned studies of speech perception were performed in such an environment, they did not distinguish how speech perception in quiet differed from speech perception in noise, as stimuli in the experimental and controlled conditions were both contaminated by the scanner noise. In recent years, the "sparse sampling method" has been developed, which allows for the presentation of auditory stimuli when image acquisition is temporarily halted [9]. The current study takes advantage of this method. In this study, subjects were asked to match auditorily presented stimuli with picture choices. Some of the stimuli were presented in quiet while others were presented with imposed multi-talker babble noise (not the scanner noise). Comparison of the two experimental conditions can provide insight into the neural mechanisms that are associated with listening to speech in noise. We hypothesize that similar to other speech perception studies involving different task demands [10], listening in noise, which requires exertion of additional cognitive and/or auditory effort, will show increased activation in the lateral auditory cortex.

2. Method

2.1. Subjects

Subjects were eight young adult native speakers of American English who reported to have no audiologic or neurologic deficits. They were right-handed as assessed by the Edinburgh Handedness Inventory [11].

2.2. Stimuli & Experimental Procedures

Two sets of stimuli, consisting of single words similar to the Auditory Figure Ground portion of SCAN-A [12], were used as stimuli. These stimuli were produced by a native male speaker of American English and were amplitude normalized. One set of the stimuli was embedded in multi-talker babble noise at a signal-to-noise ratio (SNR) of -5 dB; another set was presented in quiet. The multi-talker babble noise was taken from the noise channel of the SPIN test [13]. Stimuli were presented to the subjects binaurally at about 75 dB SPL. During each stimulus trial lasting 2 seconds, subjects were presented with three picture drawings simultaneously, one of which depicted the stimulus. Each auditory word lasted less than 700 msec, while the picture choices remained for the entire 2-second period. Subjects were asked to identify the stimulus by pressing the button corresponding to the picture. These procedures were similar to Fallon et al. [14]. Each set of stimuli consisted of 25 words. Within each set, a target picture occurred once at each of three positions (i.e., 75 trials total for each set). However, the distracters, chosen from a set of 42 pictures, were different for each occurrence of the target.

Before noise was added to any stimuli, three normal hearing subjects were asked to identify these two sets of stimuli in quiet. Subjects correctly identified both sets of stimuli with approximately 92% accuracy. The reaction time difference was less than 19 msec (slightly slower in the quiet set). These results suggest that there were minimal inherent differences between the two sets of stimuli, and any differences in behavioral and hemodynamic responses in the actual experiment are likely to be due to the fact that one set was embedded in noise.

2.3. MRI Procedures

Functional and anatomical MR images were acquired at the Center for Advanced Magnetic Resonance Imaging (CAMRI) in the Northwestern University Department of Radiology using a Siemens 3 Tesla Trio whole body machine. For each subject, a high resolution, T1 weighted 3D volume was acquired in the sagittal plane (MP-RAGE with a TR/TE of 2100ms/2.4ms, flip angle of 8 degrees, TI of 1100ms, matrix size of 256x256, FOV of 22cm, 160 slices, slice thickness of 1mm). T2*-weighted functional images were acquired by using a susceptibility weighted EPI pulse sequence while the subjects performed the behavioral task described. A TE of 30 msec, a TR of 14 seconds, a flip angle of 90 degrees, 38 slices without a slice gap, and a slice thickness of 3 mm was prescribed. Because the current study focused on comparing neural activity associated with listening to speech in noise and listening in quiet, controlling the noise level of the experimental environment was crucial. The 14-second TR (sparse sampling method) allowed for the scanner to be quiet during stimulus presentation and thus minimized contamination of the acoustic stimuli [9, 15]. Images were taken at the first 2 seconds of each TR. Figure 1 shows the timing of stimulus presentation and image acquisition. Each trial of stimulus presentation was no longer than 2 seconds long as stated. According to our pilot data and various studies [15-16], auditory associated hemodynamic response peaks at 3 to 5 seconds, and so these time points were chosen to capture only the peak responses. Data collection at these time points also means that the shortest time between the end of image acquisition (2 seconds from the onset of TR) and the onset of stimulus presentation was 7 seconds (9 minus 2). It has been reported that the hemodynamic response generated by scanner noise peaks at 4 to 5 seconds [9]. By separating image acquisition and stimulus presentation by at least 7 seconds, the current scanning protocol avoids the contamination of the stimulus-associated hemodynamic response by the peak of the hemodynamic response associated with the scanner noise. Imaging occurred at the three time points (3, 4, and 5 seconds relative to the onset of imaging) for each stimulus set (i.e. 75 trials for each stimulus set), with 25 trials at each time point for each stimulus set. The stimuli in noise and quiet occurred randomly within one functional MR session, inter-mixed with 25 trials when no stimuli (null events) were presented. Altogether, there were 175 trials (75 noise + 75 quiet + 25 null). With a TR of 14 seconds, the fMRI session took 2450 seconds (40 minutes and 50 seconds).

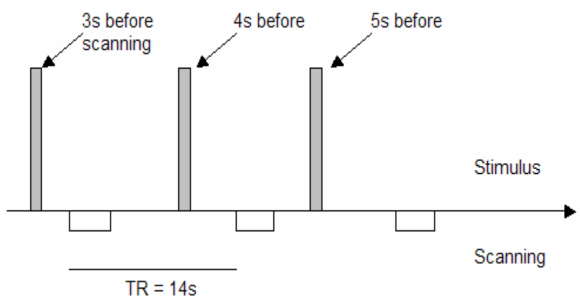

Figure 1: Imaging sequence.

2.4. fMRI Data Analyses

The T2*-weighted functional MR images (time series) were analyzed by BrainVoyager [17]. The data were first linearly detrended, motion and scan time corrected, and spatially smoothed (FWHM 6 mm). After these preprocessing procedures, hemodynamic responses were estimated. Square waves modeling the events of interest were created as extrinsic model waveforms of the task-related hemodynamic response. These events of interest included the 3 time points for each stimulus set (6 events total). Note that even though the TR was 14-second long, image acquisition only occurred during the first two seconds 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 presentation). Imaging at specific time points removed the need to convolve the task-related extrinsic waveforms with a hemodynamic response function before statistical analyses as is commonly done [10]. The waveforms of the modeled events were used as regressors in a multiple linear regression of the voxel-based time series. Beta values signifying the fit of the regressors to the functional scanning series, voxel-by-voxel, for each condition, were obtained for each subject. Anatomical and functional images from each subject were transformed into the Talairach stereotaxic space [18]. Beta values were normalized before entering into a multi-subject general linear model, including contrasting the hemodynamic responses associated with listening in noise with those associated with listening in quiet (3 time points pooled).

3. Results

3.1. Behavioral Results

Relative to listening to speech in quiet, subjects were significantly less accurate [$t(7) = 4.23$, $p < 0.005$] and marginally slower [$t(7) = 2.30$, $p < 0.055$] when listening to speech in noise. Figure 2 shows the response accuracy results (error bars show standard errors).

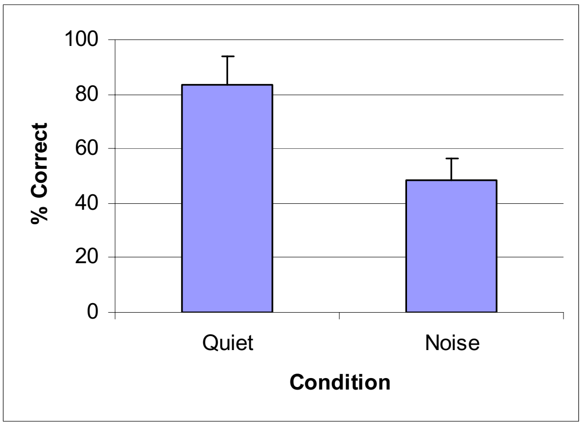

Figure 2: Subjects' response accuracy.

3.2. Imaging Results

Table 1 shows the imaging results contrasting listening to speech in noise with listening to speech in quiet conditions. A false discovery rate (q) of 0.05 for the whole brain was used for correcting multiple comparisons [20]. Relative to listening to speech in quiet, listening to speech in noise had a larger BOLD response in the left prefrontal cortex (Brodmann's Area 9), inferior frontal gyrus (BA 45), anterior insular cortex, superior temporal gyrus extending to the sulcus (STG/STS), and middle occipital gyrus. In addition, there is activation in the right insular region homologous to the activation on the left. Relative to listening in noise, listening in quiet resulted in larger activation in the medial frontal gyrus (MeFG) only. Figure 3 highlights the activation in left superior temporal region (panel A, x = -60), anterior insular cortex (panel B, x = -32), and prefrontal cortex (panel C, x = -41); statistical maps pooled across subjects where projected onto Talairach-normalized anatomical images from one subject for clear visualization.

Table 1: Brain activation comparing listening in quiet with listening in noise.

	Extent (mm^3)	t value
Noise > Quiet		
Left Prefrontal/BA 9	1135	4.26
Left Insula	375	4.17
Left MOG	551	4.00
Left STG/STS	198	3.94
Left IFG/BA 45	35	3.91
Right Insula	35	3.89
Quiet > Noise		
Right MeFG	71	-3.99

4. 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 areas mostly confined to the left hemisphere of the brain. These increased areas of activation include not only the auditory association cortex, which is often associated with complex auditory perception [19], but also areas in the anterior insular cortex, inferior frontal, prefrontal, and occipital areas.

Subjects in this experiment were less accurate and slower when listening to speech in noise. The increased BOLD signal change is likely an indication of increased task demand and subject effort. Previous research has suggested that increased task demand and effort results in increased brain activation [21-22]. More specifically, Wong et al. [10] found increased activation in the superior temporal region when subjects identified words produced by multiple talkers versus single talkers (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. Activation in the occipital region may also be related to increased effort. Even though subjects were primarily performing an auditory task, they may have exerted more effort in attending to the picture choices during the noise trials.

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 [23], which may have assisted subjects with overcoming the noise and accurately perceiving the speech sounds.

Although not specifically investigating the neural correlates of listening in noise, Wong et al. [3] 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 subjects, similarly found left anterior insular activation.

5. Conclusions

We investigated the neural correlates of listening to speech in multi-talker babble noise and found increased activation in a network of brain areas, including superior temporal, inferior frontal, prefrontal, anterior insular, and middle occipital regions. These areas of increased activation may be related to increased effort as well as modification to processing strategies such as subvocal rehearsal. These results provide a foundation for investigating possible neurophysiologic anomalies in individuals who demonstrate difficulty listening in noise such as those who have Auditory Processing Disorders.

6. Acknowledgements

The authors wish to thank Nondas Leloudas, Diane Arcuri, Divya Jain, Tyler Perrachione, Kris Erickson, and Catherine Warrier for their assistance in this research. This work is supported by Northwestern University.

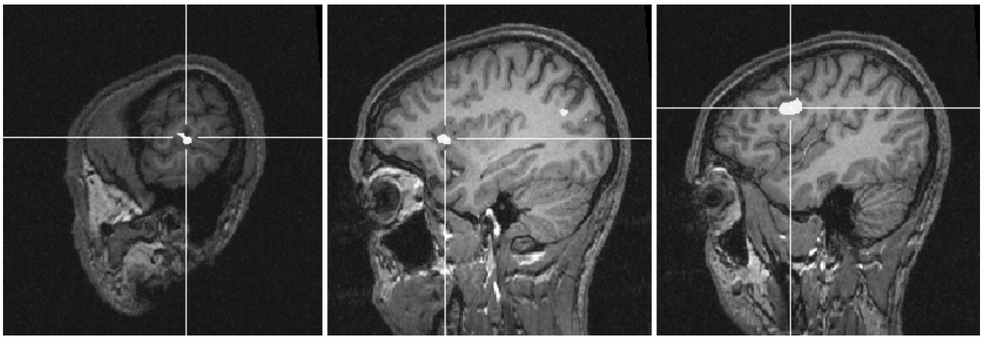

Figure 3: Increased brain activation in left superior temporal (A), anterior insular (B), and prefrontal (C) regions while listening to speech in noise.

7. References

[1] Binder, J. R. et al. "Human temporal lobe activation by speech and nonspeech sounds." *Cerebral Cortex*, 10:512-528, 2000.

[2] Burton, M. W. et al. "The role of segmentation in phonological processing: an fMRI investigation." *J. Cog. Neuro.*, 12:679-690, 2000.

[3] Wong, P. C. M. et al. "The role of the insular cortex in Pitch Pattern Perception: The effect of linguistic contexts." *J. Neuro.*, 24:9153-9160, 2004.

[4] Bamiou, D.-E. "Aetiology and clinical presentations of auditory processing disorder---a review." *Archives of Disease in Childhood*, 85:1-9, 2001.

[5] Chermak, G. D., & Musiek, F.E. *Central auditory processing disorders: New Perspectives*, Singular Publishing Group, San Diego, 1997.

[6] Cooper, J. C., & Gates, G.A. "Hearing in the elderly-the Framingham cohort, 1983-1985: Part II: Prevalence of central auditory processing disorders." *Ear & Hearing*, 12:304-11, 1991.

[7] King, W. M. et al. "Comorbid auditory processing disorder in developmental dyslexia." *Ear & Hearing*, 24:448-456, 2003.

[8] Hugdahl, K. et al. "Central auditory processing, MRI morphometry and brain laterality: Applications to Dyslexia." *Scandinavian Audiology*, 27:26-34, 1998.

[9] Hall, D. A. et al. "Sparse temporal sampling in auditory fMRI." *Human Brain Mapping*, 7:213-223, 1999.

[10] Wong, P. C. M. et al. "Neural Bases of Talker Normalization." *J. Cog. Neuro.*, 16:1173-1184, 2004.

[11] Oldfield, R. C. "The assessment and analysis of handedness: The Edinburgh Inventory." *Neuropsychologia*, 9:97-113, 1971.

[12] Keith, R. W. *SCAN-A: A test for auditory processing disorders in adolescents and adults*, Psychological Corp., Harcourt Brace Jovanovich, Inc., Texas, 1994.

[13] Bilger, R. C. et al. "Standardization of a test of speech perception in noise." *J. Speech & Hearing Res.*, 27:32-48, 1984.

[14] Fallon et al. "Children's perception of speech in multitalker babble." *J. Acoust. Soc. Amer.*, 108:3023-3029, 2000.

[15] Belin, P. et al. "Event-related fMRI of the auditory cortex." *NeuroImage*, 10:417-429, 1999.

[16] Gaab, N., & Schlaug, G. "Functional anatomy of pitch memory---an fMRI study with sparse temporal sampling." *NeuroImage*, 19:1417-1426, 2003.

[17] Goebel. *BrainVoyager QX*, 2004.

[18] Talairach, J., & Tournoux P. *Co-planar Stereotaxic Atlas of the Human Brain. 3-Dimensional Proportional System: An Approach to Cerebral Imaging*, Thieme Medical Publishers, Inc., New York., 1988.

[19] Hall, D. A.et al. "Spectral and temporal processing in human auditory cortex." *Cerebral Cortex*, 12:140-149, 2002.

[20] Genovese, C. R. et al. "Thresholding of statistical maps in functional neuroimaging using the false discovery rate." *NeuroImage*, 15:870-87, 2002.

[21] Lecas, J.-C. "Prefrontal neurons sensitive to increased visual attention in the monkey." *NeuroReport*, 7:305-309, 1995.

[22] Just, M. A. et al. "Brain activation modulated by sentence comprehension." *Science*, 274(5284):114-116, 1996.

[23] Paulesu, E. et al. "The neural correlates of the verbal component of working memory." *Nature*, 362:342-345, 1993.