View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by CiteSeerX

ELSEVIER

Hearing Research 170 (2002) 96–106

HEARING RESEARCH

www.elsevier.com/locate/heares

PET imaging of the normal human auditory system: responses to speech in quiet and in background noise

R.J. Salvi [a,b,*], A.H. Lockwood [a–e], R.D. Frisina [f], M.L. Coad [c–e], D.S. Wack [c–e], D.R. Frisina [g]

[a] *Hearing Research Lab., Center for Hearing and Deafness, 215 Parker Hall, University at Buffalo, Buffalo, NY 14214, USA*
[b] *Department of Neurology, University at Buffalo, Buffalo, NY 14214, USA*
[c] *Center for Positron Emission Tomography, University at Buffalo, Buffalo, NY 14214, USA*
[d] *Department of Nuclear Medicine, University at Buffalo, Buffalo, NY 14214, USA*
[e] *Veterans Administration Western New York Health Care System, Buffalo, NY 14214, USA*
[f] *International Center for Hearing and Speech, Rochester Institute of Technology, Rochester, NY 14623, USA*
[g] *Department of Otolaryngology, University of Rochester, School of Medicine and Dentistry, Rochester, NY 14642-8629, USA*

Received 2 January 2001; accepted 19 March 2002

Abstract

The neural mechanisms involved in listening to sentences, and then detecting and verbalizing a specific word are poorly understood, but most likely involve complex neural networks. We used positron emission tomography to identify the areas of the human brain that are activated when young, normal hearing males and females were asked to listen to a sentence and repeat the last word from the Speech in Noise (SPIN) test. Listening conditions were (1) Quiet, (2) Speech, (3) Noise, and (4) SPIN with stimuli presented monaurally to either the left ear or the right ear. The least difficult listening task, Speech, resulted in bilateral activation of superior and middle temporal gyrus and pre-central gyrus. The Noise and SPIN conditions activated many of the same regions as Speech alone plus additional sites within the cerebellum, thalamus and superior/middle frontal gyri. Comparison of the SPIN condition versus Speech revealed additional activation in the right anterior lobe of the cerebellum and right medial frontal gyrus, near the cingulate. None of the left ear–right ear stimulus comparison revealed any significant differences except for the SPIN condition that showed greater activation in the left superior temporal gyrus for stimuli presented to the right ear. No gender differences were observed. These results demonstrate that repeating the last word in a sentence activates mainly auditory and motor areas of the brain when Speech is presented, whereas more difficult tasks, such as SPIN or multi-talker Noise, activate linguistic, attentional, cognitive, working memory, and motor planning areas. © 2002 Elsevier Science B.V. All rights reserved.

Key words: Speech; Noise; Superior temporal gyrus; Cerebellum; Post-central gyrus; Thalamus; Superior frontal gyrus; Attention; Auditory pathway

---

1. Introduction

Many types of speech tests are used to assess the functional integrity of the auditory system. The most fundamental measure, the speech detection threshold, primarily assesses the sensitivity of the auditory system in the speech frequencies 250–3000 Hz (Carhart and Porter, 1971; Carhart, 1971; Humes et al., 1979; Beattie et al., 1978; Hood and Poole, 1977). A somewhat more realistic and representative listening task involves recognizing speech sounds presented at supra-threshold levels. A common assessment technique is to measure the percentage of correctly identified words, from a phonetically balanced (PB) list, as a function of stimulus intensity, i.e., the performance-intensity (PI) function. In people with normal hearing, word recognition scores increase with level and reach 100% correct (PB

---

[*] Corresponding author. Tel.: +1 (716) 829 2001, x13; Fax: +1 (253) 369 9026.
E-mail address: salvi@buffalo.edu (R.J. Salvi).

Abbreviations: SPM, statistical parametric mapping; SPIN, speech in noise; PB, phonetically balance; PI, performance intensity; rCBF, regional cerebral blood flow; BA, Brodmann area

0378-5955/02/$ – see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0378-5955(02)00386-6

Max) around 30 dB HL. The PI function of patients with sensorineural hearing loss plateaus at less than 100% and may decrease slightly at higher intensities (Persson et al., 2001; Kirk et al., 1997; Shirinian and Arnst, 1982; Dirks et al., 1977; Pascoe, 1975). The PI function of patients with retrocochlear hearing loss initially increases to a PB Max considerably less than 100% and then declines significantly if the level is increased further (Rizzo and Gutnick, 1991; Meyer and Mishler, 1985; Dirks et al., 1977).

A major complaint of hearing-impaired patients diagnosed with sensorineural hearing loss is that they can hear a person speaking to them, but can not recognize or understand what is being said, especially in cases where background noise levels are high. Normal hearing individuals also experience some difficulty understanding speech that is presented in high-level background noise (Persson et al., 2001). Speech recognition testing has been used extensively in audiology for differential diagnosis, assessing the degree of functional hearing impairment and for evaluating the effectiveness of hearing aids, cochlear implants and rehabilitation programs.

Speech signals have also been used in neuroimaging studies aimed at investigating the neural mechanisms underlying speech perception and language (Mirz et al., 1999). However, few studies have evaluated speech perception in background noise, a more difficult listening task that can lead to a significant reduction in hearing performance. Thus, it is unclear how the underlying neural networks responsible for processing and interpreting speech signals change when speech is presented in background noise similar to that encountered in real-world listening conditions. To address this question, we used PET imaging with ^{15}O-labeled water to identify regions of the brain where there were significant changes in regional cerebral blood flow (rCBF), a surrogate measure of neural activity (Posner et al., 1988), during speech in quiet, speech in noise (SPIN), or noise when subject's were asked to determine the last word in a sentence in the SPIN test (Gordon-Salant and Fitzgibbons, 1997; Hutcherson et al., 1979; Schum and Matthews, 1992; Frisina and Frisina, 1997). The primary aim of the study was to determine if the brain regions activated by a difficult (50% correct identification) SPIN listening task were different from those regions activated by speech alone or noise alone.

2. Materials and methods

2.1. Subjects

Ten young adult subjects, five males and five females, ranging in age from 23 to 34 (mean: 25.6) participated in the study. Subjects had normal hearing and were free of tinnitus and neurological disease. To generalize the results of this study to the population at large, we included men and women who were either right- or left-handed. Written informed consent was obtained from all subjects in accordance with the Declaration of Helsinki. All procedures were approved by institutional Human Subjects, Radiation Safety, and Radioactive Drug Research Committees.

2.2. Audiometry

Prior to the PET scans, standard audiometric measures were obtained from all subjects. Measurements were performed in a sound booth using a Grason-Stadler GS 16 audiometer and TDH-49 headphones. Air conduction thresholds at 0.25, 0.5,1, 2, 4, and 8 kHz were all within normal limits (< 20 dB HL). Speech reception thresholds, determined with spondees, were within normal limits for all subjects. Speech discrimination scores obtained in quiet obtained with NU 6 word lists were 100% for all subjects.

2.3. Sentence materials and PET scan conditions

Two versions of 30 context-positive sentences, each adapted from the SPIN test (Kalikow et al., 1977) and later revised (Bilger et al., 1984), served as stimuli during PET data acquisition (Frisina and Frisina, 1997). Different lists, balanced and equivalent from a speech processing point of view, were used for different test conditions. Background noise consisted of multi-talker ($n = 12$) 'babble' particular to each sentence (Bilger et al., 1984). The sentences and noise recorded on audiocassette tape (Cosmos Corp.) were played on a Nakamichi 1000 II cassette deck. The output of the cassette deck was connected to a Grason-Stadler GS 16 audiometer and delivered through calibrated insert earphones (Etymotic ER3A). Insert earphones were placed in both ears to minimize background noise (Kim et al., 2000).

A total of four experimental conditions were used: (1) rest with no stimulus, hereafter referred to as Quiet, and three sound stimulus conditions, (2) SPIN sentences alone, hereafter referred to Speech, (3) SPIN sentences in noise, hereafter referred to as SPIN and (4) noise alone, hereafter referred to as Noise. During the Speech, Noise and SPIN conditions, 30 stimuli were presented at the rate of 1 every 4 s. Stimuli were delivered to the right ear in five subjects and to the left ear in the other five. The noise in the SPIN condition was turned on 500 ms prior to the beginning of each sentence and turned off 500 ms after the end of the sentence. Each condition was presented twice with separate scans for each presentation. The conditions of Speech and of Noise were presented at 80 dB HL. In the

Speech condition, word recognition scores were 100%. For the SPIN condition, speech stimuli were delivered at 80 dB HL with noise stimuli presented at a level where the subject correctly identified approximately 50% of the target words, as determined during prescan testing. For all subjects, the noise level in the SPIN condition was 80 ± 2 dB HL.

2.4. Speech discrimination measures during PET scan

Speech stimuli consisted of sentences containing a target word at the end of the sentence. Identification of the target word in each sentence was aided by sentence context (i.e., context positive). Subjects were asked to repeat aloud the last word in each sentence for the Speech, Noise and SPIN conditions. An example of a sentence, with the target word in italics, is 'Our seats are in the second *row*.' To control for motor speech activity across the Speech, Noise and SPIN conditions, subjects were asked to say "nope" if they were unable to identify the last word at the end of each stimulus. Although sentences were absent from the Noise condition, the multi-talker noise sounds like speech. The subjects, who were unaware of the fact that the Noise alone condition lacked sentences, were nevertheless instructed to try to identify and say a word if they thought they heard it at the end of the Noise stimulus (Bilger et al., 1984).

2.5. Instructions to subjects

After the subject was positioned lying down in the PET scanner, the following instructions were given. "In just a moment I will place the insert earphones in your ears. Before I do, I want to explain what your task will be. Today you are going to experience eight two-minute listening conditions with about 10 to 15 min between each condition. It is important to keep your head still during each of the two-minute segments. Sometimes the sentences will be spoken in quiet. Other times there will be background noise. Sometimes it will be easy to hear the sentences and other times not so easy. Your task is to listen to each sentence and to repeat aloud the last word in each sentence. Remember, listen to each sentence and repeat aloud the last word in each. If you are not certain of the last word, it is OK to guess. Remember, please guess even if you are not certain of the last word. It is important for you to respond after each sentence is presented. If you do not get the last word, respond by saying "nope". Do you have any questions? OK let's practice a few sentences to get the idea of the rate at which each sentence will be spoken and to experience speech alone or the SPIN conditions." Practice sentences were then used to acquaint the person with the task.

Just prior to the start of data acquisition for the PET scans, the following instructions were read to each subject. "When we are ready to begin each segment, you will hear, fifteen seconds, three, two, one, go. At the count of three, I want you to close your eyes and to keep them closed until I tell you to open them."

2.6. Positron emission tomography

Details of PET data acquisition and analysis can be found in our previous auditory studies (Lockwood et al., 1998; Lockwood et al., 1999). Briefly, subjects were positioned in a Siemens ECAT 951/31R tomograph so that the inferior image plane coincided with the canthomeatal line. Head position was fixed by means of a mask. After a 20 min transmission scan, eight emission scans were obtained. Each scan began with the slow i.v. injection (15 second injection followed by a 15 second flush) of 260 MBq or less of ^{15}O–water as a tracer of CBF. Activation procedures began at the beginning of the injection and continued throughout the scan. The initial 60 s of emission data, timed from the arrival of the ^{15}O–water in the brain, were used for image reconstruction (random coincidence correction, measured attenuation, Hann filter, cutoff frequency 0.4 cycles per pixel) and analysis. Thus, the images reflect the neural activity averaged over 60 s of (1) Quiet, (2) Speech alone, (3) Noise alone and (4) SPIN.

For each subject, two scans were obtained in Quiet, two with Speech, two with Noise and two with SPIN. The test series began with a Quiet scan followed by random presentation, without replication, of Speech, Noise and SPIN; this was followed by another random presentation of the three sound stimulus scans and ended with a Quiet scan. To identify potential ear effects, five subjects were tested with stimuli presented to the left ear (two males, three females) while the other five were tested with stimuli delivered to the right ear. In order to generalize the results to the population at large, we included men and women who were either right- or left-handed; half the subjects received sound stimulation to the right ear while the other half were stimulated in the left ear. This non-selective approach seeks to define the 'least common denominator' for ear-specific effects in the data analysis rather than restricting the results to a single highly selected pool of subjects.

Images were converted to the Analyze format and a threshold was set at a level to include all pixels that would be recognized by the additional image processing steps. Images were edited, using visual inspection, on a slice-by-slice basis to remove extracerebral activity (such as scalp, great vessels, muscles, and sinuses) and analyzed by statistical parametric mapping (SPM) using SPM 1999 (Frackowiak et al., 1997). This multi-stage process (1) eliminates between-scan movement, (2) realigns images onto the Talairach stereotaxic framework (Talairach and Tournoux, 1988), (3) smoothes the data with a 15-mm Gaussian kernel to improve the signal-to-noise ratio and reduce the effects of between-subject variations in gyral anatomy, and (4) eliminates between-subject global variations in CBF and normalizes mean blood flow to a common mean value of 50 ml/100 g/min by a by-group analysis of covariance. The SPM $\{Z\}$ images presented here have a resolution of 18 mm (full width at half maximum) and each voxel is $2\times2\times8$ mm.

For the SPM statistical analyses, each scan associated with a particular experimental condition (Quiet, Speech, Noise and SPIN) was treated as individual token. The only averaging that was done is in the summary of the results that represent the sites of activation within the groups for a statistical comparison of particular experimental conditions (e.g., Speech compared to Quiet, Noise compared to Quiet, SPIN compared to Quiet, SPIN compared to Speech, etc.). The SPM software was used to determine if there were statistically significant differences between the experimental conditions (Quiet, Speech, Noise, and SPIN).

The final products, SPM $\{Z\}$ images, are the result of the conversion of pixel-specific t values to Z scores and show significant between-state changes, specified by SPM contrasts. The SPM $\{Z\}$ projections are the result of stacking the individual planes of data generated by the program and projecting the most significant pixel in the three-dimensional set onto sagittal, coronal, and transaxial planes according to the Talairach system (Talairach and Tournoux, 1988). As in our previous papers, individual slices from the SPM $\{Z\}$ image were superimposed on 'glass brain' images of the human brain oriented in transaxial, coronal, or sagittal planes (Lockwood et al., 1998; Lockwood et al., 1999). Z scores in the SPM $\{Z\}$ images were depicted on a gray-scale where threshold scores are shown in gray with progressive increases indicated by changes to black. Statistically significant changes in rCBF were defined in terms of cluster level (number of voxels in a cluster) and voxel level (corrected for multiple comparisons). Thresholds for the display of images was typically set at $P = 0.001$, uncorrected for multiple comparisons.

3. Results

The data analysis focused on three main factors. The first focused on the effects of the four experimental conditions (Quiet, Speech, Noise, and SPIN) irrespective of gender or ear effects. The second looked for gender-related differences across these experimental conditions while the third determined whether there was an ear advantage across experimental conditions.

3.1. Speech versus Quiet

Fig. 1A shows the regions where there were significant increases rCBF, compared to Quiet, when subjects listened to the Speech stimuli and then spoke the word at the end of the sentence. In this analysis, five subjects heard the Speech stimuli in their left ear and five heard it in their right ears. Table 1A identifies the activation sites shown in Fig. 1A, including cluster size and the probability that a cluster of that size is due to chance, coordinates of significant maxima, P values corrected for multiple comparisons, associated Z scores and anatomical information related to the maxima. Each activation site is numbered; this number is used to identify the approximate location of the activation site in frontal, horizontal or sagittal views of the brain in Fig. 1. There were three clusters, containing one or more maxima, where neural activation exceeded the threshold for display. One maximum was observed in the left superior temporal gyrus, Brodmann area (BA) 22; the region of activation containing this site was also significant in terms of its spatial extent (4645 voxels). A second maximum was located in the right superior temporal gyrus, BA22; the activation region containing this maximum was also significant in terms of its spatial extent (5513 voxels). Within the same cluster, a second maximum occurred in the right middle temporal gyrus, BA21. A fourth maximum was observed in the left pre-central gyrus, BA4 within a 418-voxel cluster that was significant in terms of its spatial extent. Finally, a fifth maximum was seen in the right pre-central gyrus, BA4.

3.2. Noise versus Quiet

Following the same format as above (Fig. 1A, Table 1A), Fig. 1B and Table 1B identify sites where significant increases in rCBF occurred, compared to the Quiet condition. One maximum was observed in the right superior temporal gyrus, BA22; the region of activation containing this site was also significant in terms of its spatial extent (5437 voxels). Within this 5437-voxel cluster, a second maximum was centered in the right superior temporal gyrus, BA38. A third maximum was identified in the left superior temporal gyrus, BA22; this maximum occurred within a 3495-voxel cluster that was statistically significant in its spatial extent. A fourth maximum was observed in the right posterior lobe of the cerebellum near the culmen, nodule, and fastigial nucleus; this maximum was located within a 3192 voxel-cluster that was significant in terms of its spatial extent. Within this same cluster, a fifth maximum was seen near the right thalamus and midbrain. A sixth

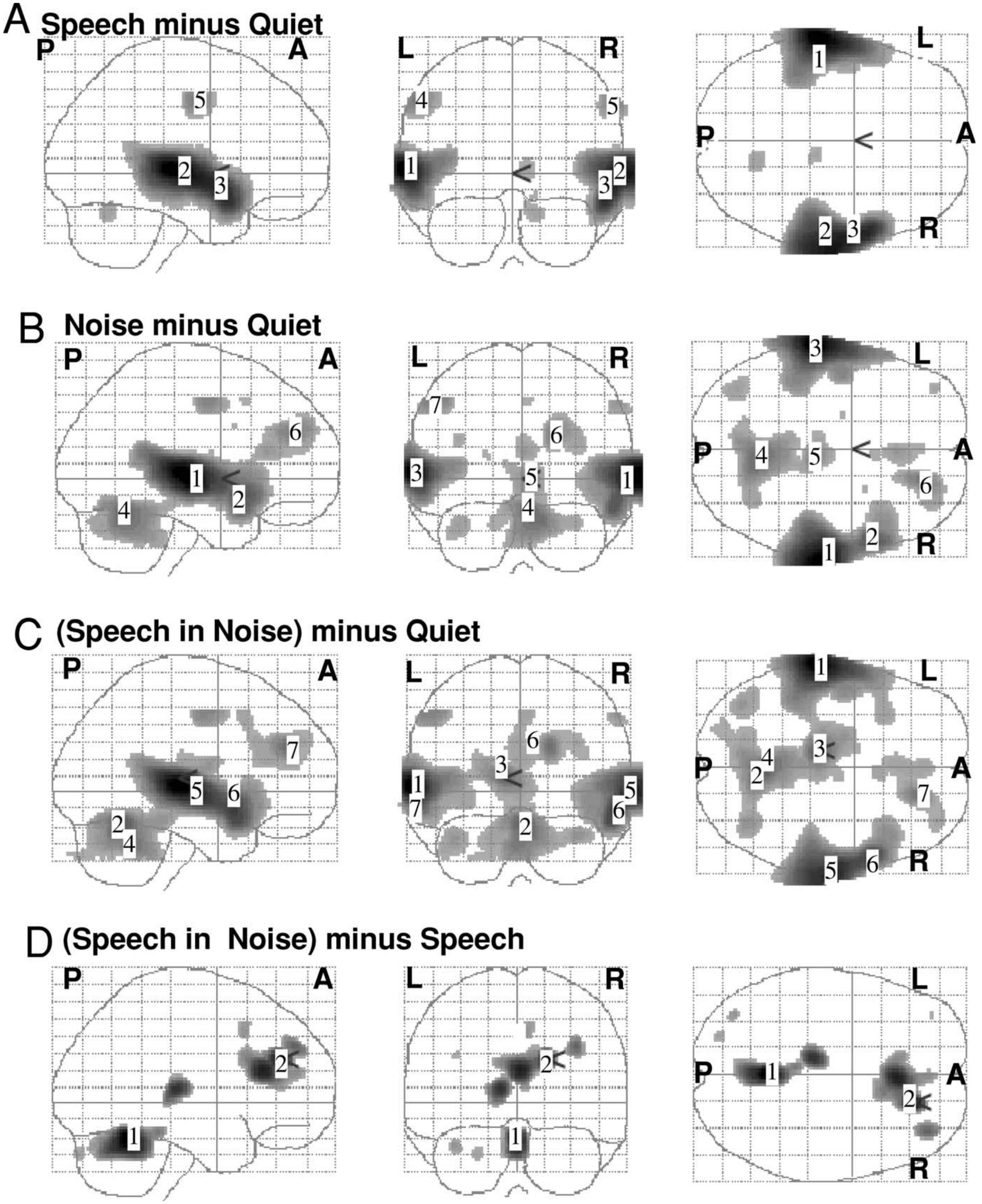
A Speech minus Quiet
P
A
5
2
3
L
R
4
5
1
3
2
L
1
P
A
2
3
R
B Noise minus Quiet
P
A
6
1
2
4
L
R
7
6
3
5
4
1
3
L
P
4
5
A
6
1
2
R
C (Speech in Noise) minus Quiet
P
A
7
5
6
2
4
L
R
6
3
1
7
2
5
6
1
L
P
4
2
3
A
7
5
6
R
D (Speech in Noise) minus Speech
P
A
2
1
L
R
2
1
L
P
1
A
2
R

Table 1

Site #	Cluster size/P	x (−left +right)	y (+anterior −posterior)	z (+rostral −caudal)	Z/P	Location of local maxima
(A) Speech–Quiet						
1	4645 < 0.001	−66	−24	2	> 15/ < 0.001	L superior temporal gyrus, BA22
2	5513 < 0.001	68	−16	0	> 15/ < 0.001	R superior temporal gyrus, BA22
3	5513 < 0.001	60	4	−8	7.71/ < 0.001	R middle temporal gyrus, BA21
4	418 < 0.05	−58	−10	44	4.93/ < 0.005	L pre-central gyrus, BA4
5	None	60	−8	44	4.75/ < 0.01	R pre-central gyrus, BA4
(B) Noise–Quiet						
1	5437 < 0.001	72	−18	0	> 15/ < 0.001	R superior temporal gyrus, BA22
2	5437 < 0.001	56	12	−16	5.68/ < 0.001	R superior temporal gyrus, BA38
3	3495 < 0.001	−66	−24	4	> 15/ < 0.001	L superior temporal gyrus, BA22
4	3192 < 0.001	4	−58	−20	5.17/ < 0.001	R cerebellum posterior lobe, culmen, nodule, near fastigial nucleus
5	3192 < 0.001	6	−22	0	4.36/ < 0.001	R thalamus, midbrain
6	846 < 0.005	20	44	28	4.73/ < 0.001	R superior frontal gyrus, near BA10
7	None	−52	−8	48	4.59/ < 0.001	L pre-central gyrus, BA4
(C) SPIN–Quiet						
1	9871 < 0.001	−66	−24	4	> 15/ < 0.001	L superior temporal gyrus, BA22
2	9871 < 0.001	2	−58	−20	5.71/ < 0.001	R cerebellum, anterior lobe, culmen, near fastigial nucleus
3	9871 < 0.001	−10	−24	8	5.56/ < 0.001	R thalamus near pulvinar
4	9871 < 0.001	−2	−50	−32	4.95/ < 0.005	L cerebellum, posterior lobe, tonsil
5	5411 < 0.001	70	−16	0	7.58/ < 0.001	R superior temporal gyrus, BA22
6	5411 < 0.001	58	10	−12	6.73/ < 0.001	R superior temporal gyrus, BA38
7	1288 < 0.001	18	40	26	5.17/ < 0.001	R medial frontal gyrus near BA9
(D) SPIN–Speech						
1	923 < 0.005	0	−54	−22	4.51/ < 0.001	R cerebellum, anterior lobe, culmen
2	1250 < 0.001	18	38	26	4.33/ < 0.001	R medial frontal gyrus near anterior cingulate

maximum was located in the right superior frontal gyrus, BA10; this maximum was located within an 846-voxel cluster that was significant in terms of its spatial extent. Finally, a seventh maximum was observed in the left pre-central gyrus, BA4.

3.3. *SPIN versus Quiet*

In this condition, the subjects were asked to identify the last word in a sentence when the sentence was presented along with multi-talker noise. Following the same format as above (Fig. 1A, Table 1A), Fig. 1C and Table 1C identify sites where significant increases in rCBF occurred in the SPIN condition compared to the Quiet. One maximum was observed in the left superior temporal gyrus, BA22, within a 9871-voxel cluster that was statistically significant in terms of its spatial extent. Three additional maxima were seen within this large 9871-voxel cluster. The second maximum occurred in the right anterior lobe of the cerebellum near the culmen and fastigial nucleus; the third maximum occurred in the left thalamus (near pulvinar) and the fourth maximum was located in the left posterior lobe of the cerebellum (near tonsil). A fifth maximum occurred in the right superior temporal gyrus, BA22, within a 5411-voxel cluster that was significant in terms of its spatial extent. Within the same cluster, the sixth maximum occurred in the right superior temporal gyrus, BA38. Finally, the seventh maximum was located in the right medial frontal gyrus, BA9, within a 1288-voxel cluster that was significant in terms of its spatial extent.

3.4. *SPIN versus Noise and Noise versus SPIN*

The aim of this contrast was to determine if the SPIN condition produced more activation than Noise alone. None of the activation sites was statistically $(P > 0.01)$

←

Fig. 1. Group data showing regions of significant activation on sagittal (left), frontal (middle) and horizontal (right) views of glass brain. Directions indicated on each view: anterior (A), posterior (P), left (L) and right (R). Row A shows activation sites for the Speech–Quiet contrast. Sites showing a statistically significant increase in activity are numbered; location of each number indicated in Table 1A. Row B shows activation sites for the Noise–Quiet contrast. Numbers show regions with a statistically significant increase in activity (see Table 1B). Row C shows activation sites for the SPIN–Quiet contrast. Numbers show regions with a statistically significant increase in activity (see Table 1C). Row D shows activation sites for the SPIN–Speech contrast. Numbers show regions with a statistically significant increase in activity (see Table 1D).

Table 2
Right ear (SPIN minus Quiet) versus left ear (SPIN minus Quiet)

Site #	Cluster size/P	x (−left +right)	y (+anterior −posterior)	z (+rostral −caudal)	Z/P	Location
1	3475 < 0.001	−46	−22	−2	4.86/ < 0.001	L superior temporal gyrus near BA21

different from the Noise alone condition. Likewise, the Noise–SPIN contrast did not produce any statistically significant ($P > 0.01$) results.

3.5. *SPIN–Speech*

The purpose of this contrast was to identify additional regions of activation that occurred in the SPIN condition versus the Speech alone condition. Following the same format as above (Fig. 1A, Table 1A), Fig. 1D and Table 1D identify sites where significant increases in rCBF occurred relative to quiet. One maximum was identified in the right anterior lobe of the cerebellum near the culmen and vermis; this maximum was located within a 923-voxel cluster that was significant in terms of its spatial extent. A second maximum was observed in the right medial frontal gyrus near the anterior cingulate; this maximum was located within a 1250-voxel cluster that was significant in terms of its spatial extent.

3.6. *Gender difference*

The aim of this analysis was to determine if men and women used different regions of the brain to process Speech, Noise, and SPIN compared to the Quiet condition. Although small differences were observed in the gender analysis, none of these reached statistical significance.

3.7. *Ear effects*

The aim of this analysis was to determine if the activation patterns depended on whether the stimuli were presented to the right or left ear. None of the left–right and right–left contrasts produced statistically significant differences except for right ear (SPIN minus Quiet) versus left ear (SPIN minus Quiet). Following the same format as above (Fig. 1A, Table 1A), Table 2 and Fig. 2 show the results for the right ear (SPIN minus Quiet) versus left ear (SPIN minus Quiet) comparison. The analysis revealed a single cluster with a maximum in the left superior temporal gyrus near BA21 that exceeded the threshold for display. This maximum was located within a 3475-voxel cluster that was significant in terms of its spatial extent.

4. Discussion

The results of the present study, with normal hearing subjects, provide new data on complex neural networks that are activated when subjects listen to Speech, Noise or SPIN relative to neural activity in the Quiet condition. These three sound stimulation conditions impose significantly different neural processing demands on the listener. The Speech in Quiet condition imposed the least demands since subjects were able to clearly hear and identify the last word in the sentence with an accuracy of nearly 100%. At the other extreme, the effort involved in the identification of a word in the Noise condition or the SPIN condition imposed the greatest processing demands since word identification is between 0 and 50% respectively under these conditions. The difficulty of the listening tasks was generally reflected in

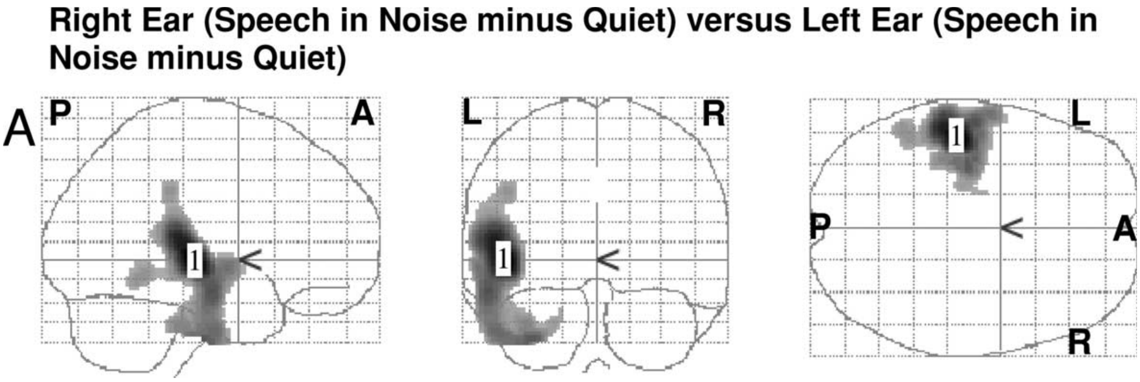
Right Ear (Speech in Noise minus Quiet) versus Left Ear (Speech in Noise minus Quiet)

Fig. 2. Group data showing regions of significant activation on sagittal (left), frontal (middle) and horizontal (right) views of glass brain with anterior (A), posterior (P), left (L) and right (R) directions indicated on each view. Row A shows activation sites for the right ear (SPIN–Quiet) –left ear (SPIN–Quiet) contrast. Sites showing a statistically significant increase in activity are numbered; location of each number indicated in Table 2.

the PET imaging data in terms of the size of the activation sites, the number of activation sites and the regions that are activated (compare Table 1A versus 1B or 1C; see Table 1D). Although word identification in the Noise alone condition was worse than in the SPIN condition, our statistical analyses results failed to show a significant difference between the SPIN condition versus the Noise conditions, suggesting that the processing demands in these two tasks were similar. However, an equally plausible interpretation is that our measurement technique is not sensitive or selective enough to identify the differences in neural processing for these two conditions.

4.1. Speech versus Quiet contrast

For the overall analysis of Speech compared to Quiet, five subjects were presented with sentences in the right ear while the other five heard the stimuli in their left ear. In the Speech versus Quiet contrast, five regions of significant activation were identified. Large activation sites were centered in the left and right superior temporal gyri (Table 1A, Fig. 1A). This region contained local maxima in BA22 and BA21 and also activated nearby regions including auditory association cortex (BA42), Wernicke's area, primary auditory cortex (BA41), and pars opercularis (Broca's area). These regions are similar to those observed in normal hearing listeners and cochlear implant users during monaural presentation of speech stimuli (Okazawa et al., 1996). The sites in the superior temporal gyri activated by our sentences were much larger than those activated by simple stimuli, such as tones and noise, in agreement with earlier studies (Lockwood et al., 1999; Lauter et al., 1985; Frith and Friston, 1996; Mirz et al., 1999; Papathanassiou et al., 2000; Scott et al., 2000; Burton et al., 2001). This increased activation is undoubtedly due to the greater complexity of the present tasks and the associated need for additional neural systems. Unlike stationary tones, spectrally modulated tones, similar to the formant transitions in speech, produce additional bilateral activation in a caudal-lateral belt surrounding the auditory cortex (Thivard et al., 2000). Compared to unintelligible speech sounds, speech stimuli that are intelligible produce additional activation in the anterior portion of the left superior temporal sulcus (Scott et al., 2000). Our speech stimuli also caused significant activation in the right middle temporal gyrus (Fig. 1A, Table 1A). Wong and colleagues reported that speech stimuli presented to the right ear caused bilateral activation in the middle temporal gyrus (Wong et al., 1999) whereas Mirz and colleagues reported that speech presented to the right ear activated the left middle temporal gyrus (Mirz et al., 1999). These differences in the side of activation may be related to several factors. First, the studies by Mirz and by Wong only analyzed data for stimuli presented to the right ear whereas our analysis (Fig. 1A, Table 1A) was carried out on subjects half of whom were stimulated in the right ear and half in the left ear. To test for ear-specific effects, we performed additional analyses for the Speech in Quiet condition, but our tests failed to reveal a significant difference in the middle temporal gyrus (Table 1). Thus, it is unlikely that the difference between our results and Mirz is due to the ear in which the stimuli were presented. A second factor that may account for these differences is the between-subject variability in activation sites reported in speech processing tasks (Burton et al., 2001). Since most PET imaging studies involve small groups of subjects, between-subject differences are likely to have an effect on the average data. The third and perhaps most important factor that could contribute to differences in the side of activation may be the task requirements. Subjects in the Mirz study and Wong study listened passively to the speech stimuli whereas our subjects were required to identify and say the last word in the sentences. The additional requirements of our study would also involve elements of working memory and require activation of networks involved in articulation. The speech in quiet condition also resulted in significant activation in the left and right pre-central gyrus (BA4), in the motor cortex (Fig. 1A); these activation sites are most likely associated with motor activity associated with oral facial movements involved in saying the word at the end of the sentence.

4.2. Noise versus Quiet contrast

The Noise alone condition was one of the more difficult listening tasks and the statistical comparison of Noise versus Quiet resulted in significant increases in activity in many of the same regions as Speech versus Quiet; however, the extent of activation was larger and included some additional activation sites (Fig. 1B, Table 1B). A visual comparison of Fig. 1B versus Fig. 1A suggest additional activity in midline cerebellar and pontine structures and medially located structures in the thalamus (compare Table 1B versus 1A). These sites were not active in the Speech only condition. Like Speech, multi-talker Noise produced broad areas of activation centered in the right and left superior temporal gyrus (compare Fig. 1A and B), but these activation sites were somewhat larger than in the Speech alone condition. Noise also produced significant activation of the cerebellum and thalamus (Fig. 1B). This region may be part of a cerebellar-thalamic network involved in speech perception and production (Petersen and Fiez, 1993; Fox et al., 1996; Muller et al., 1998). Cerebellar-dentate connections are also frequently activated by many cognitive and attentional tasks. Pure word repetition tasks activate midline cerebellar structures whereas more complex tasks, such as those requiring the subject to generate something de novo, activate cerebellar-thalamic-frontal pathways (Fiez, 1996; Kinomura et al., 1996). This is similar to our condition where the listener was asked to say the last word in a sentence after listening to multi-talker noise. Noise also activated the left and right pre-central gyrus (BA4) similar to that seen with the Speech versus Quiet contrast; however, unlike the Speech versus Quiet, Noise versus Quiet only produced a statistically significant increase in the left pre-central gyrus (compare Fig. 1B to 1A). This reduced activation within the right pre-central gyrus may be related to the high degree of uncertainty subjects experienced in trying to identify the last word in the Noise alone listening condition. The new region activated by Noise alone compared to Quiet occurred in the right superior frontal gyrus (BA10); this maximum lies slightly anterior to the medial frontal gyrus and anterior cingulate (Table 1B, Fig. 1B). We have observed activation of the anterior cingulate with low-intensity tone bursts that were difficult to hear (Benedict et al., 1998; Lockwood et al., 1999). Activation of the anterior cingulate and posterior fossa structures in the Noise versus Quiet contrast could reflect greater attention on the part of the subjects in trying to identify a non-existent word in the multi-talker Noise condition. Others have reported activation of the anterior cingulate in tasks requiring selective auditory attention (Benedict et al., 1998) and retrieval from verbal working memory (Jonides et al., 1998; Schumacher et al., 1996; Warburton et al., 1996). Greater activation of the right superior frontal gyrus has been shown to occur with cognitive tasks such as the Wisconsin Card Sort and Spatial Delayed Response Tasks (Van Horn et al., 1996).

One earlier study found that passive listening to continuous white noise delivered to the right ear produced relatively limited activation in the left transverse temporal gyrus (BA41) (Mirz et al., 1999). Similarly, white noise stimulation delivered through a cochlear implant produced activation in the contralateral auditory cortex (Naito et al., 1995). The more extensive activation seen in the present study in the multi-talker Noise condition may be related to two factors. First, based on experimental instructions, the subjects were asked to try to identify and say a non-existent word while listening to the multi-talker noise. The Noise alone condition places greater demands on the listener's attention, motor planning, speech recognition and speech production systems than the Speech alone listening task. Second, while the multi-talker noise was unintelligible, the temporal and spectral features of the Noise alone condition share many of the characteristics of speech and therefore may activate brain regions associated with language processing.

4.3. SPIN versus Quiet contrast

Like the Speech versus Quiet contrast, the SPIN versus Quiet contrast activated the left and right superior temporal gyri; however, the extent of activation in the SPIN listening conditions was larger. The SPIN listening condition produced some activity in the pre-central gyrus; however, unlike the Speech versus Quiet contrast or the Noise versus Quiet contrast, the activation level was not statistically significant. The uncertainty of identifying the last word in the SPIN listening condition could conceivably cause the listener to produce a weaker than normal motor response leading to weaker activation in the pre-motor cortex. The extent of activation in cerebellum and thalamus in the SPIN versus Quiet contrast was greater than in the Speech in Quiet contrast (compare Fig. 1C, Table 1C vs. Fig. 1B, Table 1B). The strong activation seen in the cerebellar-thalamic network may be related to greater cognitive, arousal and attentional demands (Fox et al., 1996; Muller et al., 1998; Fiez, 1996; Kinomura et al., 1996; Petersen and Fiez, 1993). Like the Noise versus Quiet contrast, the SPIN versus Quiet contrast revealed activation in the right frontal lobe, but the extent of activation was somewhat larger and located more medially encompassing the anterior cingulate, a region implicated in attention and processing difficult to detect acoustic signals (Benedict et al., 1998; Lockwood et al., 1999).

4.4. SPIN versus Speech contrast

In theory, identifying specific words in the SPIN listening task should create greater processing demands than listening to Speech alone. This hypothesis was confirmed by the SPIN versus Speech alone contrast that revealed increased activity near the right frontal and cingulate gyri plus increased activation in the right cerebellum (Fig. 1D, Table 1D). Previous studies have reported increased activation in the para-cingulate region during word generation (Crosson et al., 1999). In addition, the anterior cingulate cortex is believed to mediate response selection or allocate attentional resources when faced with competing sources of information, a condition similar to the SPIN listening task (Bush et al., 1998). Moreover, the increased activation in the cerebellum is consistent with earlier reports showing cerebellar activation in several speech production tasks (Roskies et al., 2001; Wildgruber et al., 2001). The increased activation in this region may also be related to greater arousal and attention associated with this task (Benedict et al., 1998; Lockwood et al., 1999).

4.5. Ear-specific effects

To identify potential ear effects, half the subjects were tested with stimuli presented to the left ear while the other five were tested with stimuli delivered to the right ear. We included men and women who were either right- or left-handed so that the data could be generalized to the population at large. Extensive comparisons of left ear versus right ear stimulation failed to reveal significant differences except for the SPIN versus Quiet contrast where right ear stimulation produced greater activation in the left superior temporal gyrus (BA21) than left ear stimulation (Fig. 2, Table 2). The increased activity seen in the left superior temporal gyrus may be related to previous studies showing that neural responses (local field potentials) to voiced and voiceless syllables were lateralized to the left superior temporal gyrus (Liegeois-Chauvel et al., 1999). PET imaging studies have also revealed stronger activation in the left temporal lobe when subjects were required to detect dichotically presented CV-syllables (Hugdahl et al., 1999); this increased activity in the left temporal lobe was associated with greater accuracy in detecting CV syllables that were presented to the right ear.

In summary, our results show that a complex network, involving auditory, cognitive, linguistic, attentional, working memory, motor and motor planning centers, is activated when listeners are asked to verbalize the last word in a sentence that is presented in multitalker noise. The degree of activation and the number of brain regions activated increases with task difficulty, with multi-talker noise and SPIN producing the greater activation than speech in quiet. Finally, in the SPIN condition, stimulation of the right ear caused greater activation in the left superior temporal gyrus than left ear stimulation.

Acknowledgements

Research supported in part by Grants from the National Institutes of Health, DC 3306 and P01 AG 9524, the James H. Cummings Foundation, Buffalo, NY, and The International Center for Hearing and Speech Research, Rochester, NY.

References

Beattie, R.C., Svihovec, D.A., Edgerton, B.J., 1978. Comparison of speech detection and spondee thresholds and half-versus full-list intelligibility scores with MLV and taped presentations of NU-6. J. Am. Aud. Soc. 3, 267–272.

Benedict, R.H., Lockwood, A.H., Shucard, J.L., Shucard, D.W., Wack, D., Murphy, B.W., 1998. Functional neuroimaging of attention in the auditory modality. NeuroReport 9, 121–126.

Bilger, R.C., Nuetzel, J.M., Rabinowitz, W.M., Rzeczkowski, C., 1984. Standardization of a test of speech perception in noise. J. Am. Speech Hear. Assoc. 27, 32–48.

Burton, M.W., Noll, D.C., Small, S.L., 2001. The anatomy of auditory word processing: individual variability. Brain Lang. 77, 119–131.

Bush, G., Whalen, P.J., Rosen, B.R., Jenike, M.A., McInerney, S.C., Rauch, S.L., 1998. The counting Stroop: an interference task specialized for functional neuroimaging-validation study with functional MRI. Hum. Brain Map. 6, 270–282.

Carhart, R., 1971. Observations on relations between thresholds for pure tones and for speech. J. Speech Hear. Disord. 36, 476–483.

Carhart, R., Porter, L.S., 1971. Audiometric configuration and prediction of threshold for spondees. J. Speech Hear. Res. 14, 486–495.

Crosson, B., Sadek, J.R., Bobholz, J.A., Gokcay, D., Mohr, C.M., Leonard, C.M., Maron, L., Auerbach, E.J., Browd, S.R., Freeman, A.J., Briggs, R.W., 1999. Activity in the paracingulate and cingulate sulci during word generation: an fMRI study of functional anatomy. Cereb. Cortex 9, 307–316.

Dirks, D.D., Kamm, C., Bower, D., Betsworth, A., 1977. Use of performance-intensity functions for diagnosis. J. Speech Hear. Disord. 42, 408–415.

Fiez, J.A., 1996. Cerebellar contributions to cognition. Neuron 16, 13–15.

Fox, P.T., Ingham, R.J., Ingham, J.C., Hirsch, T.B., Downs, J.H., Martin, C., Jerabek, P., Glass, T., Lancaster, J.L., 1996. A PET study of the neural systems of stuttering. Nature 382, 158–161.

Frackowiak, R.S.J., Friston, K.J., Frith, C.D., Dolan, R.J., Mazziotta, J.C., 1997. Human Brain Function, New York, pp. 528.

Frisina, D.R., Frisina, R.D., 1997. Speech recognition in noise and presbycusis: relations to possible neural mechanisms. Hear. Res. 106, 95–104.

Frith, C.D., Friston, K.J., 1996. The role of the thalamus in ‘top down’ modulation of attention to sound. Neuroimage 4, 210–215.

Gordon-Salant, S., Fitzgibbons, P.J., 1997. Selected cognitive factors and speech recognition performance among young and elderly listeners. J. Speech. Lang. Hear. Res. 40, 423–431.

Hood, J.D., Poole, J.P., 1977. Improving the reliability of speech audiometry. Br. J. Audiol. 11, 93–101.

Hugdahl, K., Bronnick, K., Kyllingsbaek, S., Law, I., Gade, A., Paulson, O.B., 1999. Brain activation during dichotic presentations of consonant-vowel and musical instrument stimuli: a 15O-PET study. Neuropsychologia 37, 431–440.

Humes, L.E., Schwartz, D.M., Bess, F.H., 1979. Two experiments on subtle mid-frequency hearing loss and its influence on word discrimination in noise-exposed listeners. Audiology 18, 307–319.

Hutcherson, R.W., Dirks, D.D., Morgan, D.E., 1979. Evaluation of the speech perception in noise (SPIN) test. Otolaryngol. Head Neck Surg. 87, 239–245.

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

Kalikow, D.N., Stevens, K.N., Elliott, L.L., 1977. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. J. Acoust. Soc. Am. 61, 1337–1351.

Kim, K., Burkard, R.F., Lockwood, A.H., Salvi, R.J., 2000. Effects of background noise on audiometric thresholds during positron emission tomography: passive and active noise-reduction. Scand. Audiol. 29, 211–216.

Kinomura, S., Larsson, J., Gulyas, B., Roland, P.E., 1996. Activation by attention of the human reticular formation and thalamic intralaminar nuclei. Science 271, 512–515.

Kirk, K.I., Pisoni, D.B., Miyamoto, R.C., 1997. Effects of stimulus variability on speech perception in listeners with hearing impairment. J. Speech. Lang. Hear. Res. 40, 1395–1405.

Lauter, J.L., Herscovitch, P., Formby, C., Raichle, M.E., 1985. Tonotopic organization in human auditory cortex revealed by positron emission tomography. Hear. Res. 20, 199–205.

Liegeois-Chauvel, C., de Graaf, J.B., Laguitton, V., Chauvel, P., 1999. Specialization of left auditory cortex for speech perception in man depends on temporal coding. Cereb. Cortex 9, 484–496.

Lockwood, A.H., Salvi, R.J., Coad, M.L., Arnold, S.A., Wack, D.S., Murphy, B.W., Burkard, R.F., 1999. The functional anatomy of the normal human auditory system: responses to 0.5 and 4.0 kHz tones at varied intensities. Cereb. Cortex 9, 65–76.

Lockwood, A.H., Salvi, R.J., Coad, M.L., Towsley, M.L., Wack, D.S., Murphy, B.W., 1998. The functional neuroanatomy of tinnitus: evidence for limbic system links and neural plasticity. Neurology 50, 114–120.

Meyer, D.H., Mishler, E.T., 1985. Rollover measurements with Auditec NU-6 word lists. J. Speech Hear. Disord. 50, 356–360.

Mirz, F., Ovesen, T., Ishizu, K., Johannsen, P., Madsen, S., Gjedde, A., Pedersen, C.B., 1999. Stimulus-dependent central processing of auditory stimuli: a PET study. Scand. Audiol. 28, 161–169.

Muller, R.A., Chugani, D.C., Behen, M.E., Rothermel, R.D., Muzik, O., Chakraborty, P.K., Chugani, H.T., 1998. Impairment of dentato-thalamo-cortical pathway in autistic men: language activation data from positron emission tomography. Neurosci. Lett. 245, 1–4.

Naito, Y., Okazawa, H., Honjo, I., Hirano, S., Takahashi, H., Shiomi, Y., Hoji, W., Kawano, M., Ishizu, K., Yonekura, Y., 1995. Cortical activation with sound stimulation in cochlear implant users demonstrated by positron emission tomography. Brain Res. Cogn. Brain Res. 2, 207–214.

Okazawa, H., Naito, Y., Yonekura, Y., Sadato, N., Hirano, S., Nishizawa, S., Magata, Y., Ishizu, K., Tamaki, N., Honjo, I., Konishi, J., 1996. Cochlear implant efficiency in pre- and postlingually deaf subjects. A study with H2(15)O and PET. Brain 119, 1297–1306.

Papathanassiou, D., Etard, O., Mellet, E., Zago, L., Mazoyer, B., Tzourio-Mazoyer, N., 2000. A common language network for comprehension and production: a contribution to the definition of language epicenters with PET. Neuroimage 11, 347–357.

Pascoe, D.P., 1975. Frequency responses of hearing aids and their effects on the speech perception of hearing-impaired subjects. Ann. Otol. Rhinol. Laryngol. 84, 1–40.

Persson, P., Harder, H., Arlinger, S., Magnuson, B., 2001. Speech recognition in background noise: monaural versus binaural listening conditions in normal-hearing patients. Otol. Neurotol. 22, 625–630.

Petersen, S.E., Fiez, J.A., 1993. The processing of single words studied with positron emission tomography. Annu. Rev. Neurosci. 16, 509–530.

Posner, M.I., Petersen, S.E., Fox, P.T., Raichle, M.E., 1988. Localization of cognitive operations in the human brain. Sci. 240, 1627–1631.

Rizzo, S.R.Jr., Gutnick, H.N., 1991. Cochlear versus retrocochlear presbyacusis: clinical correlates. Ear Hear. 12, 61–63.

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

Schum, D.J., Matthews, L.J., 1992. SPIN test performance of elderly hearing-impaired listeners. J. Am. Acad. Audiol. 3, 303–307.

Schumacher, E.H., Lauber, E., Awh, E., Jonides, J., Smith, E.E., Koeppe, R.A., 1996. PET evidence for an amodal verbal working memory system. Neuroimage 3, 79–88.

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.

Shirinian, M.J., Arnst, D.J., 1982. Patterns in the performance-intensity functions for phonetically balanced word lists and synthetic sentences in aged listeners. Arch. Otolaryngol. 108, 15–20.

Talairach, J., Tournoux, P., 1988. Co-Planar Stereotaxic Atlas of the Human Brain. Thieme Medical Publishers Inc., New York, 132 pp

Thivard, L., Belin, P., Zilbovicius, M., Poline, J.B., Samson, Y., 2000. A cortical region sensitive to auditory spectral motion. NeuroReport 11, 2969–2972.

Van Horn, J.D., Berman, K.F., Weinberger, D.R., 1996. Functional lateralization of the prefrontal cortex during traditional frontal lobe tasks. Biol. Psychiatry 39, 389–399.

Warburton, E., Wise, R.J., Price, C.J., Weiller, C., Hadar, U., Ramsay, S., Frackowiak, R.S., 1996. Noun and verb retrieval by normal subjects. Studies with PET. Brain 119, 159–179.

Wildgruber, D., Ackermann, H., Grodd, W., 2001. Differential contributions of motor cortex, basal ganglia, and cerebellum to speech motor control: effects of syllable repetition rate evaluated by fMRI. Neuroimage 13, 101–109.

Wong, D., Miyamoto, R.T., Pisoni, D.B., Sehgal, M., Hutchins, G.D., 1999. PET imaging of cochlear-implant and normal-hearing subjects listening to speech and nonspeech. Hear. Res. 132, 34–42.