Introduction

Auditory language comprehension is a multistage process relying on many different cognitive functions as an acoustic signal is perceived, parsed into subphonemic features, and the features synthesized into a semantically coherent utterance. In approximately 96% of right-handers and 70% of left-handers, the left hemisphere is preferentially involved in the later (phonological and semantic) stages of this process, and it is possible that speech processing proceeds asymmetrically from its earliest stages. Indeed, it has been hypothesized that the left hemispheric specialization for speech may be based, in part, on prepotent mechanisms within this hemisphere for the processing of auditory patterns of a fine temporal grain, such as are found in speech. Some phonetic features (such as voicing and place of articulation) are instantiated in spectral changes of an acoustic signal that happen with a time course of tens of milliseconds. For example, the perception of place of articulation contrasts (like /ba/ vs /da/) requires accurate tracking of formation trajectories at constant-vowel transitions which are typically no more than 40 ms long.

Evidence for a left hemispheric specialization for the processing of acoustic transients comes from several sources. First, patients with left hemisphere lesions and aphasias appear to be impaired at discriminating phonemes on the basis of voicing and place of articulation, but such patients do not appear to show impairment of vowel discrimination, which is made on the basis of spectral information present in the auditory signal over a longer period. Second, dichotic listening studies in normal populations demonstrate a greater right ear (left hemisphere) advantage for speech sounds that require perception of short duration spectral changes for their accurate identification than for speech sounds that do not. For example, Shankweiler and Studdert-Kennedy found no differences in accuracy of identification of steady-state vowels between stimuli presented to the left ear and those presented to the right. In contrast, when perception of stop consonant–vowel syllables was examined (/pa, ta, ka, ba, da, and ga/), they found that subjects were more accurate at reporting stimuli presented to the right ear, to a highly significant degree. Finally, two recent positron emission tomography (PET) studies revealed increased activation in left hemisphere areas in response to rapidly changing auditory cues. In a study in which normal subjects listened passively to speech-like syllables incorporating formation transitions, Belin and his colleagues found bilateral asymmetry in activation in auditory areas when the formation transitions were long (200 ms), but a leftward asymmetry in activation in response to short (40 ms) formant transitions. Fiez et al. observed that increases in cerebral blood flow (CBF) in the left frontal operculum were larger when their subjects performed an auditory detection task upon stimuli that incorporated stop consonants than for steady-

Key words: Auditory perception; Cerebellum; Frontal cortex; Fusiform gyrus; Hemispheric specialization; Language; Positron emission tomography; Speech acoustics

Left-hemisphere specialization for the processing of acoustic transients

Ingrid S. Johnsrude, Robert J. Zatorre, Brenda A. Milner and Alan C. Evans

Neuropsychology/Cognitive Neuroscience Unit and McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, 3801 University Street, Montreal, Canada H3A 2B4

CA Corresponding Author
state vowels. Thus, speech signals that are stable or that change only slowly in time (over hundreds of milliseconds) appear to be processed bilaterally. In contrast, speech signals that change rapidly in time (over tens of milliseconds) appear to be preferentially processed by left hemisphere mechanisms.

The present PET activation study was designed to examine the idea that providing fine-grained temporal resolution to the auditory percept may be a general specialization of the left hemisphere and may apply to the processing of any sound that changes rapidly in time, not just sounds with linguistic relevance. If this is the case, then the processing of non-speech pure tones that incorporate frequency glides with the temporal characteristics of stop consonant formant transitions ought to involve left hemisphere regions preferentially, when compared with processing of pure tone stimuli incorporating frequency glides of a longer duration. Speech signals are spectrally complex however, and the left hemispheric specialization observed in the studies described above may not be related to the temporal structure of acoustic transients per se, but may instead depend primarily on their spectral characteristics. The stimuli used in this experiment are qualitatively different from speech stimuli in their spectral properties. It was hoped that this would be a stronger test of the hypothesis that the left hemisphere possesses specialized mechanisms for the processing of fine-grained temporal structure.

Materials and Methods

The experimental protocol was approved by the ethical review committee of the Montreal Neurological Institute and Hospital. Twelve right-handed neurologically normal participants (six men, six women; mean age 23 years) were recruited from the McGill community, and trained on 80 trials of each of the experimental tasks (see below). Any subject who did not score at least 75% correct on both tasks during the last 40 trials was not scanned. Two subjects (one male, one female) were thus excluded.

Two sets of sound stimuli were synthesized on an IBM-compatible computer. The ‘short glide’ set consisted of eight items; 220 ms steady-state pure tone at one of four frequencies (1250, 1600, 1950, or 2300 Hz), preceded by a 30 ms 350 Hz ascending or descending linear frequency glide. The chosen steady-state frequencies, the absolute frequency difference subtended by the glide, and the duration of the glide were all within the range observed in the second formant of English stop consonants. The set of ‘long glide’ stimuli was almost identical, except that the duration of the glide component was extended to 100 ms, and the duration of the steady-state component was decreased to 150 ms, so that the absolute duration of the stimuli was held constant at 250 ms. Similarly, the absolute frequency difference subtended by the glide was held constant at 350 Hz.

Subjects underwent PET scans in two separate conditions, presented in a random order. In both conditions, subjects listened to the glide stimuli, presented in pairs, through insert earphones (EarTone ER3A) at 75 dB SPL (A). The ‘short glide’ stimuli were used in one condition, whilst the ‘long glide’ stimuli were used in the other. In both conditions, subjects were asked to make same–different judge-ments on the pairs, indicating their choice by a key press response with the right hand. Within each pair, the steady-state positions of the stimuli were identical in frequency. In half the trials, however, the glide...
component was different (i.e. ascending to the steady-state tone in one item, and descending to that tone in the other item). Pairs were presented with a 3 s inter-trial interval, and the interval between items of a pair was 1 s (Fig. 1). Subjects were scanned for 60 s, and each task commenced approximately 40 s before scanning, and continued for 45 s after the scan was completed (32 pairs presented).

PET data were acquired using a Scanditronix PC-2048 system, which produces 15 slices at an intrinsic resolution of 5.0 × 5.0 × 6.0 mm. Using the bolus \( \text{H}_2\text{O} \) method without blood sampling, relative distribution of CBF was measured in the two conditions. PET images were reconstructed using an 18 mm full-width half-maximum filter and normalized for global CBF value. A high resolution MR image (160 slices, 1 mm thick) was also obtained for each subject (Philips Gyroscan 1.5T) and each pair of MR and PET data sets was resampled into a standardized stereotaxic coordinate system. The MR and PET images were averaged across subjects for each activation state to obtain an average MRI and a mean CBF change image. Significant focal CBF changes were identified in the averaged PET volume by converting it to a t-statistic volume (by dividing each voxel by the mean standard deviation in normalized CBF for all intracerebral voxels). The functional images were merged with the averaged MR image to allow localization of significant t-statistic peaks, which were identified by application of an intensity threshold to the t-statistic images.

The presence of significant focal changes was tested by a method based on three-dimensional Gaussian random-field theory. Values of \( t > 3.5 \) were deemed statistically significant (\( p < 0.0002 \), one-tailed, uncorrected).
Results

Subjects were significantly more accurate at making same–different judgements on long glide pairs than on short glide pairs (t = 4.72; p < 0.001), although both tasks were performed well (mean 95% correct for the long glide task and 86% correct for the short glide task). There was no difference in mean response latency in the two tasks (t = 1.84; p > 0.05; mean long glide latency 893 ms, mean short glide latency 973 ms).

When the pattern of CBF observed in the long glide condition was subtracted from that in the short glide condition, significant cortical activations were observed in only three regions: the left orbitofrontal cortex, fusiform gyrus, and in the right cerebellar hemisphere, pars lateralis (Fig. 2). A CBF decrease was observed in right mid-dorsolateral frontal cortex (area 9; x = 27, y = 3, z = 36; t = 3.50).

Discussion

All of the neocortical activation foci were in the left hemisphere. The right lateral cerebellar activation is consistent with these left neocortical foci: the right lateral cerebellum is richly interconnected with the left cerebral hemisphere via the red and pontine nuclei and the thalamus, and several investigators have posited a role for the right cerebellar hemisphere in at least some aspects of language function.17,18 Given that the task demands of the two conditions were very similar, differing only in the duration of the glide component of the sounds, our findings support the contention that are specialized mechanisms for the processing of acoustic transients in the left cerebral hemisphere. Since we used pure tone stimuli of no linguistic value, we conclude that this specialization, while relevant to speech perception, must reflect a more general functional property of the human left hemisphere, and may apply to the processing of any short duration spectral change.

The fusiform gyrus activation was particularly interesting, given that studies of phonemic monitoring consistently show activation in this region.19,20 For example, Zatorre and his colleagues20 performed a PET study in which subjects listened to pairs of monosyllabic English words. When the subjects were asked to monitor the pairs for the occurrence of a particular stop consonant, and activation in this region was compared with that obtained when they were simply listening to the words, significant CBF change was observed in area 19/37 (stereotaxic coordinates x = −34, y = −57, z = −9), about 8 mm posterior to the focus observed in the current study. Similarly, Démonet and his colleagues19 asked subjects to perform a stop consonant detection task that had been made perceptually difficult. When activation in this condition was compared with that obtained during performance of a phoneme detection task that was perceptually simpler, the only significant CBF change observed was in the left area 19 (stereotaxic coordinates x = −32, y = −74, z = −12), about 2.4 cm posterior to that observed in the present study. Thus, the posterior fusiform region in the left hemisphere may be involved in processing the acoustic manifestations of some phonetic features, as would be necessary for the correct identification of a target stop consonant.

Given that electrophysiological and neuromagnetic studies in cats, monkeys and humans have demonstrated the existence of neurones in the region of primary auditory cortex that are apparently able to preserve the temporal structure of transient auditory signals, and to encode the direction of a transient change,21–25 we expected to see some differences in activation in the region of auditory cortex. This was not the case. Either the two types of stimuli used in this study are processed similarly at early stages, or, as seems more likely, the sensitivity of the PET subtractive technique in combination with the discriminative task we used is insufficient to distinguish between the relevant auditory systems. We are currently exploring alternative PET methods which may be more sensitive to blood flow changes in this region.

Conclusion

The present study examined the brain regions involved in processing a single property of speech sounds that typically yield a strong right ear advantage in dichotic listening studies; namely, the temporal structure of stop consonant formant transitions. We did not attempt to respect the spectral properties of such formant transitions, and it is important to note that the speech signal is full of other information at changes more slowly over time. Data from this experiment, however, do provide support for the hypothesis that the left hemisphere possesses specialized mechanisms for the processing of acoustic information of a fine temporal grain.

References

General Summary

The identification of a variety of speech sounds depends on the perception of changes in the sound spectrum that happen over tens of milliseconds, such as occurs in the transition of the second formant in stop-consonants. In order to determine whether the left-hemispheric specialization for language may be based in part on a relative specialization of this hemisphere for the processing of acoustic transients, we performed a PET study in which 10 normal subjects were scanned while discriminating pure tone stimuli incorporating frequency glides of either short, or long duration. A comparison of these two conditions yielded significant blood flow changes in left cortical areas and in right cerebellum. Because non-linguistic stimuli were used, we concluded that this asymmetrical activation (while relevant to speech perception) must reflect a more general specialization of the human left hemisphere that applies to the processing of any short duration spectral change.