

Sensitivity to Auditory Object Features in Human Temporal Neocortex

Robert J. Zatorre,¹ Marc Bouffard,¹ and Pascal Belin²

¹Montreal Neurological Institute, McGill University, Montreal, Quebec H3A 2B4, Canada, and ²Department of Psychology, University of Montreal, Montreal, Quebec, Canada H3C 3J7

This positron emission tomography study examined the hemodynamic response of the human brain to auditory object feature processing. A continuum of object feature variation was created by combining different numbers of stimuli drawn from a diverse sample of 45 environmental sounds. In each 60 sec scan condition, subjects heard either a distinct individual sound on each trial or simultaneous combinations of sounds that varied systematically in their similarity or distinctiveness across conditions. As more stimuli are combined they become more similar and less distinct from one another; the limiting case is when all 45 are added together to form a noise that is repeated on each trial. Analysis of covariation of cerebral blood flow elicited by this parametric manipulation revealed a response in the upper bank of the right anterior superior temporal sulcus (STS): when sounds were identical across trials (i.e., a noise made up of 45 sounds), activity was at a minimum; when stimuli were different from one another, activity was maximal. A right inferior frontal area was also revealed. The results are interpreted as reflecting sensitivity of this region of temporal neocortex to auditory object features, as predicted by neurophysiological and anatomical models implicating an anteroventral functional stream in object processing. The findings also fit with evidence that voice processing may involve regions within the anterior STS. The data are discussed in light of these models and are related to the concept that this functional stream is sensitive to invariant sound features that characterize individual auditory objects.

Key words: auditory cortex; superior temporal sulcus; functional neuroimaging; auditory processing; auditory object

Introduction

The functional organization of the auditory cortex has been likened to that of the visual system (Rauschecker and Tian, 2000), with separate functional streams for the identification of sounds (the ventral “what” pathway) and for localizing their spatial position (the dorsal “where” pathway). This hierarchical model has received support from anatomical tracing (Rauschecker et al., 1997; Romanski et al., 1999) and neurophysiological studies in animals (Rauschecker et al., 1995; Recanzone et al., 2000; Tian et al., 2001), as well as human lesion data (Clarke et al., 2000).

Several human functional neuroimaging studies have explored predictions made by this model, with emphasis on its spatial component. The findings have implicated parietal lobe structures in spatial processing (Bushara et al., 1999; Alain et al., 2001; Maeder et al., 2001), which are likely related to sensorimotor integration and spatial transformations required for active localization tasks (Zatorre et al., 2002b). Auditory cortical areas posterior to A1 are recruited by stimuli moving in space (Baumgart et al., 1999; Griffiths and Green, 1999; Warren et al., 2002) or by situations in which multiple stimuli must be disambiguated on the basis of spatial cues (Zatorre et al., 2002b). These findings and others have led to the suggestion that posterior cortical regions may

perform a computation related to the segregation and matching of spectrotemporal patterns (Griffiths and Warren, 2002).

In contrast to the relative interest in spatial processing, the putative ventral stream has not received much study. Alain et al. (2001) reported several superior temporal gyrus (STG) sites, along with frontal and occipital activity, when subjects actively judged a pitch change compared with a spatial change, but this pitch task may not reflect object-related processing per se. Maeder et al. (2001) reported a similar contrast in a task that involved recognizing a specific class of sounds (animal cries) from among a complex background. Results implicated the anterior STG, but parietal, frontal, parahippocampal, insular, and occipital cortices were also involved. Although both these studies converge in reporting that anterior STG areas may be involved in auditory object processing, the areas reported are quite different, perhaps because these studies involved different and complex active tasks, which likely recruited a variety of cognitive mechanisms.

The present study was focused on identifying the existence and location of brain regions sensitive to sound object features (characteristics that distinguish one sound-producing object from another) in a direct, task-independent way using positron emission tomography (PET). We reasoned that if such regions exist, then they should respond more when a variety of such features are present than when features are repeated. This prediction follows from the idea that neural responses to a given class of stimuli tend to habituate when repeated (Grill-Spector et al., 1999). We used difficult-to-identify sounds to focus on early levels of object processing and not recognition processes (Adams and Janata, 2002). We implemented a parametric manipulation, in which the dimension of interest (distinctiveness of auditory

Received Dec. 11, 2003; revised Feb. 11, 2004; accepted Feb. 12, 2004.

This work was supported by operating grants from the Canadian Institutes of Health Research, the Natural Sciences and Engineering Research Council of Canada, and the McDonnell-Pew Cognitive Neuroscience Program. We thank P. Ahad, Dr. A. C. Evans, and the staff of the McConnell Brain Imaging Centre for assistance.

Correspondence should be addressed to Robert J. Zatorre, Montreal Neurological Institute, 3801 University Street, Montreal, Quebec H3A 2B4, Canada. E-mail: robert.zatorre@mcgill.ca.

DOI:10.1523/JNEUROSCI.5458-03.2004

Copyright © 2004 Society for Neuroscience 0270-6474/04/243637-06\$15.00/0

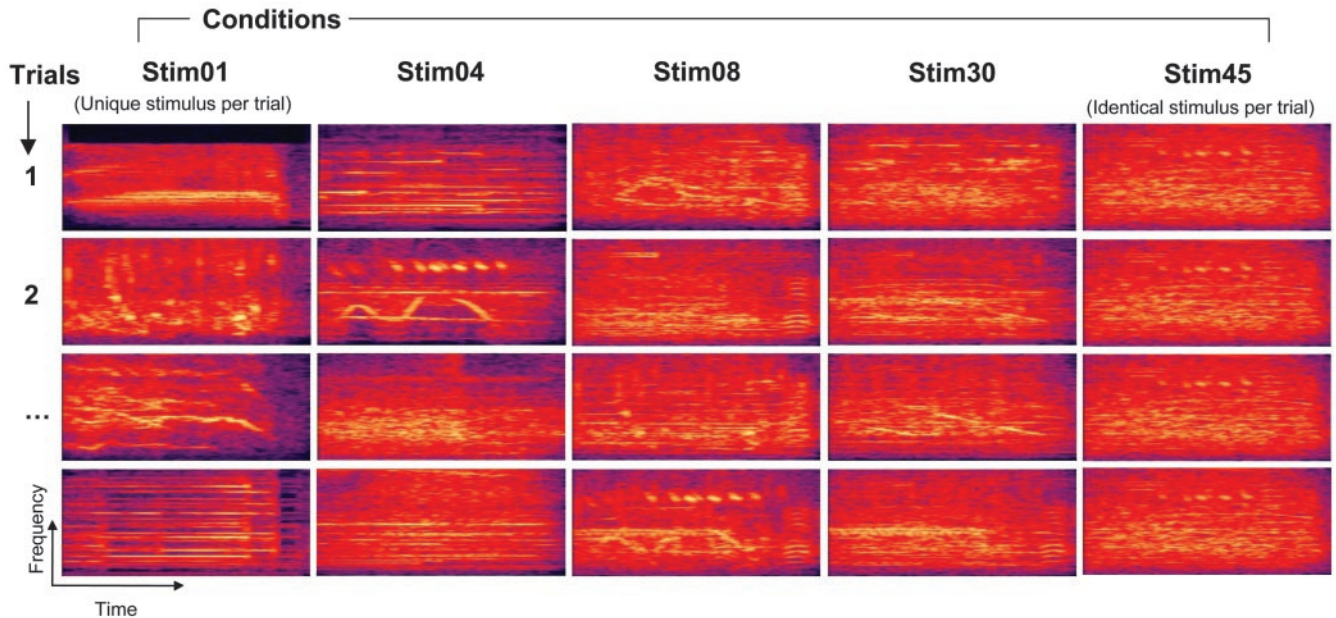


Figure 1. Schematic of the stimulus conditions. Each panel shows a spectrogram (frequency as a function of time; amplitude shown in the color scale) of one stimulus. The five columns correspond to the five conditions used in the neuroimaging study. Each row corresponds to examples of different trials. The leftmost column shows condition Stim01 in which a unique sound is presented on each trial; note that the spectrograms are correspondingly distinct from one another. The second column shows condition Stim04 in which 4 of the 45 stimuli are randomly selected and added together; the spectrograms become less differentiated from one another across trials in this condition. As more stimuli are added together, the stimuli become less distinct from one another; hence, there is more repetition across trials. The limiting case is condition Stim45, in which all stimuli are added together to produce a single noise; in this condition, the stimuli are identical across trials. Samples of these stimuli may be heard at www.zlab.mcgill.ca/JNeurosci2004.

objects) was varied in a graded rather than categorical manner, predicting that anteroventral portions of the temporal neocortex should respond to this dimension.

Materials and Methods

Subjects. Ten healthy right-handed volunteers (half of each sex; mean age, 23 years) were tested after written informed consent was obtained in accord with guidelines approved by the Montreal Neurological Institute Ethics and Research Committee. All had normal hearing as determined via standard audiometry.

Stimuli and behavioral (pilot) data. Forty-five 500 msec excerpts of environmental sounds were chosen; these sounds were drawn from a diversity of categories, including animal cries, environmental noises (such as wind and water), machinery, musical instruments, sirens, and so forth. The sounds were selected to be difficult to identify; in addition, they were time-reversed to render them less likely to elicit a verbal label and hence to contaminate the cognitive processes of interest with verbalization. Each of them was bandpass-filtered (500–8000 Hz; high- and low-pass roll-offs of 6 and 12 dB/octave, respectively) to equate for overall spectral range and equalized for root mean square intensity.

In a pilot phase, combinations of these 45 sounds were created such that different numbers of them were added together to create a sound mixture; the mixtures were each then equalized once again for root mean square intensity. The number of stimuli added was 1, 2, 3, 4, 5, 6, 8, 10, 12, 14, 16, 20, 24, 30, 36, or 45. For each of these categories, a random selection was made on each trial from among the 45 stimuli to create a new stimulus mixture. For example, for the eight-stimulus mixture, eight sounds from among the 45 were added together to produce a mixture for trial 1, then another random selection of eight sounds was made for trial 2, and so forth. This resulted in stimuli that were progressively more indistinct and similar to one another as one goes from 1 (in which a unique stimulus was presented on each trial) to 45 (in which an identical mixture of all stimuli was presented on each trial). Figure 1 shows a schematic of the stimulus design; audio examples of the stimuli are available at www.zlab.mcgill.ca/JNeurosci2004.

To select values along the dimension of interest, 10 normal listeners (who did not participate in the main imaging study) were exposed to

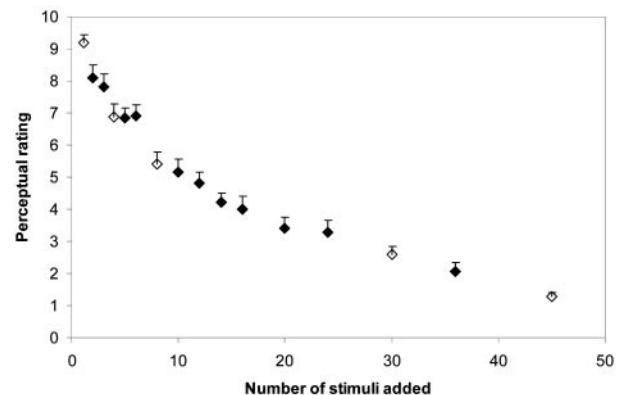


Figure 2. Perceptual similarity ratings obtained in pilot testing. Each data point indicates mean response on a subjective rating scale for pairs of stimuli drawn from a continuum differing in number of sounds added together to form a composite (10, maximally distinctive; 1, maximally similar). Open symbols indicate points chosen for stimuli used in main neuroimaging study. Error bars indicate SEM.

pairs of stimuli drawn from each of the 16 categories enumerated above. Twenty pairs for each of these 16 categories were presented in a random order, and listeners were asked to rate them for overall dissimilarity using a scale of 1 (the two sounds are identical) to 10 (the two sounds are maximally distinct from one another). The results are shown in Figure 2. ANOVA indicated a significant effect of condition [$F_{(15,144)} = 47.63; p < 0.001$]. From these data, we selected five categories of stimulus distinctiveness that were approximately equally spaced perceptually. These were 1, 4, 8, 30, and 45 sounds. *Post hoc* Neuman–Keuls tests were used to verify that ratings from adjacent conditions (e.g., 1 vs 4 or 4 vs 8) did indeed differ significantly ($p < 0.05$). The ratings from the five categories chosen came very close to forming a linear trend ($r = 0.99$), suggesting that they were perceptually evenly distributed, as desired. Thus, in the imaging experiment, subjects experienced stimuli that were either maximally different from one another across trials (condition termed

Stim01), sounds that consisted of mixtures of various stimuli that became gradually more indistinct and hence also more similar to one another (conditions Stim4, Stim8, and Stim30), and, finally, the Stim45 condition, in which the identical noise was repeated from trial to trial. As in the pilot study, a random selection from the set of 45 was made for each trial.

Procedure. The five conditions described above were presented during PET scanning always from a position directly in front of the subject, using the center speaker of a stimulus array, which fit into the scanner (Zatorre et al., 2002b). The interstimulus interval was constant at 833 msec. Stimuli were presented at 66–69 dB sound pressure level (SPL); background noise was 56 dB SPL. A stimulus-free baseline condition was also used. The order of conditions was counterbalanced across subjects. Subjects were asked to close their eyes throughout scanning and to attend to the stimuli, but no explicit task was given. Stimulation started several seconds before scan start and continued until the 60 sec scan period was over.

Neuroimaging parameters. PET scans were obtained with a Siemens AG (Erlangen, Germany) Exact HR+ tomograph operating in the three-dimensional acquisition mode. The distribution of cerebral blood flow (CBF) was measured during each 60 sec scan using the H_2O^{15} water bolus method (Raichle et al., 1983). Magnetic resonance imaging (MRI) scans (160 1-mm-thick slices) were also obtained for each subject with a 1.5 T Philips Medical Systems (Andover, MA) ACS system to provide anatomical detail. CBF images were reconstructed using a 14 mm Hanning filter, normalized for differences in global CBF, and coregistered with the individual MRI data (Evans et al., 1992). Each matched MRI-PET data set was then linearly resampled into a standardized stereotaxic coordinate system based on the Montreal Neurological Institute 305 target, a sample of 305 normal subjects, via an automated feature-matching algorithm (Collins et al., 1994), resulting in a normalized brain space similar to the Talairach and Tournoux (1988) atlas (for additional information, see www.mrc-cbu.cam.ac.uk/Imaging/). Statistical analysis was performed applying the method described by Worsley et al (1992); covariation analysis followed the procedure outlined by Paus et al. (1996).

Results

The principal analysis of interest identified areas of CBF covariation with the parametric change in stimulus composition. This analysis was accomplished by taking the average behavioral ratings of stimulus similarity or distinctiveness (see pilot data above) and regressing them against CBF in the entire brain volume. The behavioral measure was used as the input variable because it best captures the degree to which subjects perceive stimuli in each category as being similar or distinct from one another. Two principal regions emerged in which CBF was greater for highly distinct stimuli and lower for similar stimuli (Fig. 3A): one in the right anterior STG, most likely within the upper bank of the superior temporal sulcus (STS; coordinates, 59, -13, and -9; $t = 4.70$); and the other in the right inferior frontal gyrus (coordinates, 48, 36, and 0; $t = 4.35$). Two weaker foci were also seen: one at the anterior pole of the right temporal lobe, adjacent to the inferior frontal gyrus (coordinates, 46, 24, and -20; $t = 3.46$); and the other in the medial wall of the left parietal lobe (coordinates, -8, -49, and 42; $t = 3.72$). No region emerged in this analysis that was close to primary cortex or to Heschl's gyrus, even allowing for subthreshold activity peaks. Only a single region showed a reverse covariation (i.e., highest CBF for condition Stim45); it was located in the left inferior occipital gyrus (coordinates, -42, -76, and 5; $t = 3.53$).

Extracting the mean CBF values from the right anterior STS location (5 mm spherical region of interest) and plotting them as a function of stimulus composition (Fig. 4) shows a significant correlation between these variables ($r = 0.84$; $p < 0.04$, one-tailed), although a degree of nonlinearity is apparent in that the CBF values seem to reach an asymptote between conditions Stim01 and Stim08. Analysis of individual subject CBF values at this location indicated that 9 of the 10 showed a correlation sim-

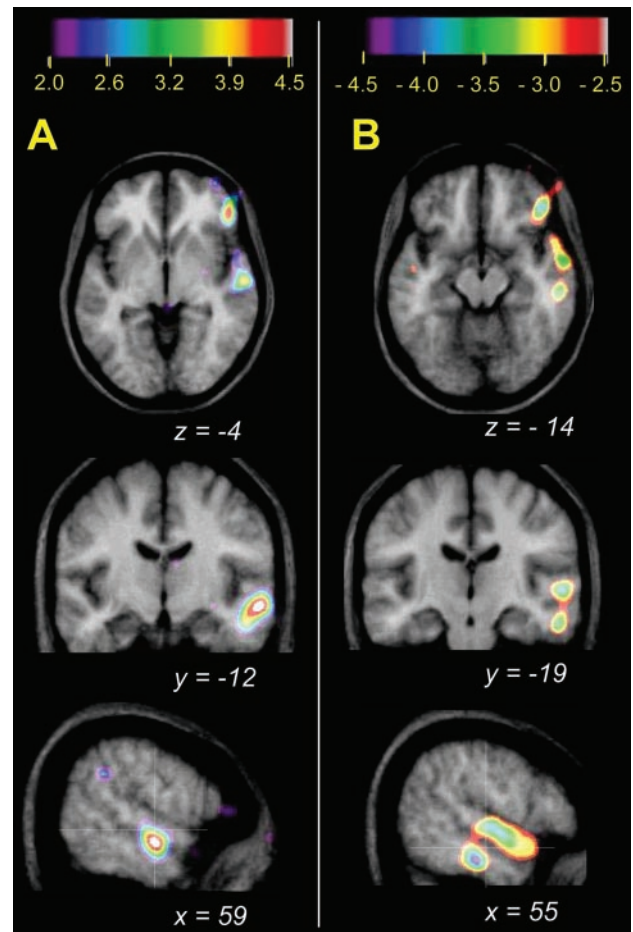


Figure 3. Functional neuroimaging data. *A*, Horizontal (top), coronal (middle), and sagittal (bottom) sections centered on the area of CBF change in the STS identified in the covariation analysis of CBF as a function of stimulus perceptual similarity rating. *B*, Similar sections for the categorical contrast of conditions varying in auditory spatial position (but constant stimulus features) versus conditions varying in stimulus features (but constant spatial position) showing areas with greater activity for stimulus feature variation. Note the similarity of right STS and right inferior frontal sites identified in the two independent analyses. In *B*, an additional infero-temporal region is also identified. Mean CBF data are superimposed on averaged MRI scans for the subjects scanned. A CBF statistical parametric map of t is indicated by the color scale for each analysis. Dotted lines in sagittal sections indicate positions of coronal and horizontal sections. Coordinate values indicate the position in the stereotaxic space at which each section was taken.

ilar to that of the group data. To confirm the finding that this anterior STS region responds to object-related features without making assumptions about linearity, a direct comparison of the two most extreme conditions (Stim01 vs Stim45) was performed; it demonstrated an almost identical set of right frontal and temporal foci as were found in the covariation analysis.

We also performed a subsidiary analysis of the CBF changes associated with the five conditions without using subjective ratings; instead, we regressed CBF against the number of stimuli added together in each condition (45, 30, 8, 4, or 1). Using these values as the input variable resulted in CBF regression maps (data not shown) that were essentially indistinguishable in terms of areas activated and their location to the maps shown in Figure 3A. This regression analysis in fact resulted in a better fit than did the one using behavioral data as the input value (e.g., the STS region yielded a value of $t = 5.33$, and the computed correlation coefficient was $r = 0.96$; $p < 0.001$). This finding provides support for the principal result independently of any assumptions about the validity of the subjective behavioral ratings. In addition to regions

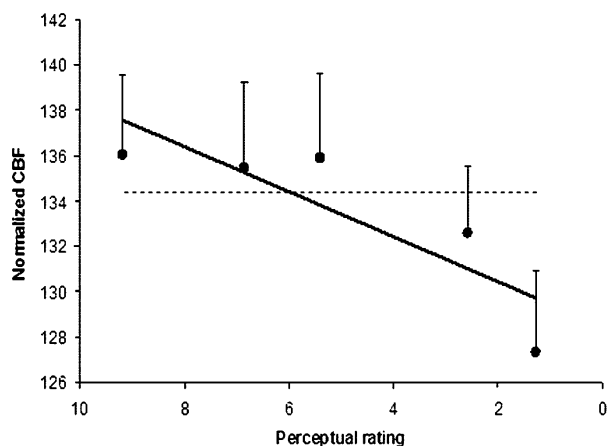


Figure 4. Mean normalized cerebral blood flow (milliliters per 100 gm/min) within the right anterior STS region identified in a whole-brain voxel-wise covariation analysis (Fig. 3A) as a function of perceptual similarity rating. The dashed line indicates average CBF in a no-stimulus baseline condition. Error bars indicate SEM.

similar to those identified in the previous analysis, there was also a subthreshold peak in the right inferotemporal lobe (coordinates, 55, -26, and -21; $t = 3.33$), which is relevant because it also shows up as a significant area in the contrast of object to spatial tasks reported below.

Figure 4 also shows the average CBF value for the baseline condition (dashed line); this suggests that CBF decreases were in part driving the correlation. To explore this aspect of the findings, we compared the two extreme stimulus conditions (Stim 01 and Stim 45) to silence using a categorical contrast. As might be expected, the Stim 01 condition (unique stimulus on each trial) shows significant CBF increases in A1 and surrounding regions bilaterally, but the Stim45 condition (combined noise stimulus repeated on each trial) shows no significant ($t > 3.5$) regions of CBF increase; only relatively weak ($t = 2.9$) activity is detected in areas close to the left A1 in this contrast. On the other hand, significant CBF decreases were seen in comparison of Stim45 with silence in the right anterior STS (coordinates, 58, -11, and -14; $t = 3.97$) and right frontal areas (coordinates, 54, 27, and 5; $t = 3.85$), similar to those revealed by the covariation analysis. Inspection of individual data values at the STS location confirmed that 8 of the 10 subjects showed decreased CBF in the Stim45 condition compared with silence. Thus, the covariation seen in Figure 3A seems to be driven primarily by the deactivations in this region, consistent with the analysis shown in Figure 4.

An additional analysis was performed to compare the CBF pattern associated with variation in stimulus features with the CBF pattern evoked by variation in the spatial position. Because the subjects tested here were the same as those in experiment 1 in the study by Zatorre et al. (2002b), it was possible to contrast the conditions from the present study in which the stimuli were not identical (Stim01–Stim30) to conditions in which the Stim45 noise stimulus was presented from a variety of spatial locations (conditions 2–5 of the earlier study, in which the spatial position was randomly varied from $-15/+15^\circ$ of azimuth to $-60/+60^\circ$). Thus, in this contrast, we pooled conditions Stim1–Stim 30 and compared them with the pooled conditions in which Stim45 was presented at different locations across trials. This analysis therefore allows a direct comparison of stimulus feature variation (holding spatial position constant) to spatial feature variation (holding stimulus features constant). The result is illustrated in Figure 3B. Once again, the right anterior STS region shows significantly increased CBF at a location similar to that seen in the

previous analysis (coordinates, 60, -7, and -3; $t = 4.57$), but in this contrast, an additional right inferior temporal region is also significantly more active for the “object” versus the “spatial” conditions (coordinates, 56, -25, and -20; $t = 4.47$); this region is similar to the one identified earlier, using the number of stimuli as the input parameter. A right inferior frontal region is also revealed (coordinates, 43, 36, and -17; $t = 4.06$) in this object versus spatial contrast, as before. There were no areas of increased CBF associated with the spatial condition compared with the object condition, in accord with the results of experiment 1 in the study by Zatorre et al. (2002b).

Finally, to test whether the right anterior STS focus and the right inferior frontal focus were functionally linked, we performed a region-of-interest regression analysis using all of the CBF volumes from all of the stimulus conditions from the present study. The aim of this procedure is to determine whether activity in a given region predicts activity levels in any other region; the assumption is that if they do, then the areas in question are functionally connected. A target region was identified, centered on the coordinates of the right anterior STS from the first analysis (Fig. 3A), and the analysis sought out any areas in the entire brain volume whose activity covaried with that of the target area. This analysis yielded a significant site of covariation in the right inferior frontal gyrus, close to the location seen in the other analyses (coordinates, 50, 20, and 5; $t = 8.86$), as predicted.

Discussion

The analyses converge on two principal findings: (1) a region of auditory cortex is sensitive to features that distinguish sound sources from one another; and (2) it is located anteriorly within the upper bank of the right STS. This region was the main temporal lobe area to show a systematic change in CBF as a function of perceptual distinctiveness: when sounds were identical across trials, activity was at a minimum; and when stimuli were different from one another activity, was maximal (Figs. 3A, 4). A nearly identical result was obtained when the conditions from the current study were compared with those from a previous experiment (Zatorre et al., 2002b) in which spatial position was varied, indicating that the area in question is specifically sensitive to variations in stimulus-related cues and not to spatial cues (Fig. 3B). In this latter analysis, an additional region of CBF increase in inferotemporal cortex was identified (Fig. 3B). Inferotemporal cortex is usually believed to be exclusively visual (Poremba et al., 2003); this finding may therefore reflect interactions between visual and auditory representations (Gibson and Maunsell, 1997) based on interconnection between STS and inferotemporal regions (Saleem et al., 2000), or it may indicate multimodal object sensitivity in the inferotemporal stream.

Methodological considerations

A recurring problem in neuroimaging studies is defining an appropriate baseline (Friston et al., 1996). Here, we took advantage of a parametric approach (Paus et al., 1996; Zatorre and Belin, 2001), in which brain activity changes are sought that covary with a systematically manipulated input variable. Importantly, our findings hold whether the input variable reflects perceptual similarity, as measured behaviorally via subjective ratings, or the variable simply indicates the number of stimuli added to produce the stimulus. Because the results are similar with these two approaches and with categorical comparisons, the effect seems to be robust and not dependent on specific assumptions about the input variable.

This parametric approach is limited to capturing linear trends; inspection of Figure 4 suggests that there may be some

nonlinearity in the response, such that CBF tends not to change beyond the point at which eight stimuli are added together. This asymptote may or may not represent a true inflection point, but it is of interest that the behavioral pilot data showed some discontinuity near eight stimuli (Fig. 2), which corresponded approximately to the point of emergence of individual stimulus features from the mixture. This interpretation raises the issue that the manipulation used could be construed in two ways: first, the continuum differed in terms of stimulus similarity versus repetitiveness (when stimuli are identical from trial to trial there is maximal repetition and vice versa); and second, the degree to which individual stimuli could be heard out within the mixture also varied from minimal in Stim45 to maximal in Stim01. This feature of the design was purposeful because our aim was to provide a continuum that would drive putative auditory object-related neural populations, and we therefore wanted the strongest possible test of the hypothesis. Having demonstrated the desired effect, however, the next step will be to disentangle the extent to which the response is related to the presence of a mixture of sounds as opposed to the repetitive similarity of one stimulus to the next. On the basis of arguments made below concerning response adaptation in visual studies, we predict that it is the similarity–distinctiveness dimension that is of greatest relevance.

One important detail is that the CBF values for condition Stim45, in which the same sound was repeated, were significantly lower than the values obtained during a baseline condition in which no stimuli were presented (Fig. 4). This response is in contrast to that of primary auditory cortical areas, which showed no CBF decreases in any comparison of stimulus condition with silence. Our interpretation of this aspect of the data is that the response in the anterior STS area reflects adaptation to the repeated stimulus presented in condition Stim45 and release from adaptation when stimulus features are distinct from one another across trials. A similar argument has been made for object-sensitive responses in the visual cortical system. On the basis of neurophysiological findings that repeated exposure to the same object attenuates neural responses in inferotemporal cortex (Miller et al., 1991), several authors have argued that similar decreases in the functional MRI signal reflect adaptation of object-sensitive visual cortices (Grill-Spector and Malach, 2001; Kourtzi and Kanwisher, 2001; Vuilleumier et al., 2002). The important point in these studies and in the present study is the specificity of the adaptation response and its location. We argue similarly to the visual studies that the STS region represents a functionally distinct anteroventral pathway because it adapts to object-related features, whereas primary areas do not.

Auditory object processing versus voices

Another claim for specificity in response to a class of sounds relates to the human voice. Belin et al. (2000, 2002) have reported widespread, bilateral STS responses to human vocal sounds (speech or nonspeech). A more recent study (Belin and Zatorre, 2003) contrasted a sequence of many speakers articulating the same syllable to a sequence of many syllables all spoken by the same voice. A region within the right anterior STS was the only one found to be more active in the former condition, suggesting that adaptation to vocal features had occurred when the syllables were all spoken by the same voice. A similar right STS region was also identified in another recent study in which listeners were asked to recognize a target voice within a sentence, as contrasted with recognizing a target word (von Kriegstein et al., 2003). These results have been taken as evidence for sensitivity to voice information in an anteroventral auditory cortical stream because

voices constitute a particular class of auditory object. As such, these conclusions are compatible with the present study; however, comparison of the location across studies indicates that the voice-related foci are located somewhat more anteriorly within the STS than the region observed in the present study. This difference may represent a degree of domain specificity for voices, as has been claimed for certain classes of visual objects such as faces and scenes (Kanwisher, 2000). However, it may just as well represent a difference related to aspects of the stimuli or the manipulation used, and future studies will be needed to address this point. It is also notable that the findings in the present study were clearly lateralized to the right hemisphere, presumably because the stimuli carried no overt verbal content. The asymmetry is congruent with the voice studies just cited and likely reflects the specialization of a right auditory cortical stream for aspects of processing outside the speech domain (for review, see Zatorre et al., 2002a).

Inferior frontal cortex

Apart from the STS region, another important finding is the systematic recruitment of the right inferior frontal cortex. Not only was this region active in the same analyses that yielded STS activity, but we also found evidence of functional connectivity from the region-of-interest covariation analysis. The functional role of this area remains to be determined, but there is strong evidence that anteroventral temporal and inferior frontal cortical areas are anatomically interconnected (Seltzer and Pandya, 1989; Hackett et al., 1999; Romanski et al., 1999). Neurophysiological evidence indicates that an inferior frontal region is responsive to a variety of complex sounds, including vocalizations in the macaque (Romanski and Goldman-Rakic, 2002). A few neuroimaging studies have also observed inferior frontal activity in relation to object processing (Adams and Janata, 2002; Vuilleumier et al., 2002). It thus seems likely that STS and frontal regions form part of a functional network related to auditory object processing.

Functional significance

The findings outlined in this study constitute one step in understanding the nature of the processing that occurs in the human anterior temporal neocortex. The conclusions are generally compatible with the view that an anteroventral stream is involved in auditory object processing. By analogy to the visual domain, we would argue that processing an auditory object (Kubovy and Van Valkenburg, 2001) entails computing the commonalities or invariances among acoustic features that characterize the object emitting the sound (Belin and Zatorre, 2000; Zatorre and Belin, 2004). For example, one can identify a trumpet regardless of the melody it is playing because it has certain constant features that distinguish it from other sounds; the ventral stream would, in our view, be concerned with extracting those features. In contrast, sound patterns such as frequency-modulated tones and melodies typically engage more dorsal regions of the STG (Thivard et al., 2000; Zatorre and Belin, 2001; Patterson et al., 2002; Hart et al., 2003) (but see Warren et al., 2003). Processing of such patterns is independent of object features to the extent that any given pattern may be produced by objects with different characteristics (e.g., the same melody can be played by different instruments); hence, it is reasonable to propose a distinct processing pathway dedicated to this computation.

The findings of the present study apply primarily to nonverbal sounds; processing of words may entail a specialized pathway, and there is evidence that auditory words may recruit left-lateralized STS areas because of their status as a special class of auditory stimuli (Binder et al., 2000). Our stimuli were not only

nonverbal but also were not readily identifiable; therefore, our results most likely reflect an intermediate stage of object processing, before recognition takes place. Recognition and mnemonic processing likely would involve more anterior areas and interactions with mediotemporal regions, as has been proposed both for auditory (Imaizumi et al., 1997) and visual streams (Nakamura and Kubota, 1995). The advantage of proposing that the anteroventral superior temporal cortex is involved in processing invariant features characteristic of auditory objects is not only its direct relevance to the visual object-processing literature but also that it serves as a useful heuristic for future studies, which will have to address the many issues that remain open.

References

- Adams R, Janata P (2002) A comparison of neural circuits underlying auditory and visual object categorization. *Neuroimage* 16:361–377.
- Alain C, Arnott SR, Hevenor S, Graham S, Grady CL (2001) “What” and “where” in the human auditory system. *Proc Natl Acad Sci USA* 98:12301–12306.
- Baumgart F, Gaschler-Markefski B, Woldorff M, Heinze H-J, Scheich H (1999) A movement-sensitive area in auditory cortex. *Nature* 400:724–725.
- Belin P, Zatorre RJ (2000) What, where, and how in auditory cortex. *Nat Neurosci* 3:965–966.
- Belin P, Zatorre RJ (2003) Adaptation to speaker’s voice in right anterior temporal lobe. *NeuroReport* 14:2105–2109.
- Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000) Voice-selective areas in human auditory cortex. *Nature* 403:309–312.
- Belin P, Zatorre RJ, Ahad P (2002) Human temporal-lobe response to vocal sounds. *Cognit Brain Res* 13:17–26.
- Binder J, Frost J, Hammeke T, Bellgowan P, Springer J, Kaufman J, Possing J (2000) Human temporal lobe activation by speech and nonspeech sounds. *Cereb Cortex* 10:512–528.
- Bushara K, Weeks R, Ishii K, Catalan M-J, Tian B, Rauschecker J, Hallett M (1999) Modality-specific frontal and parietal areas for auditory and visual spatial localization in humans. *Nat Neurosci* 2:759–766.
- Clarke S, Bellman A, Meuli R, Assal G, Steck A (2000) Auditory agnosia and auditory spatial deficits following left hemispheric lesions: evidence for distinct processing pathways. *Neuropsychologia* 38:797–807.
- Collins D, Neelin P, Peters T, Evans A (1994) Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *J Comput Assist Tomogr* 18:192–205.
- Evans A, Marrett S, Neelin P, Collins L, Worsley K, Dai W, Milot S, Meyer E, Bub D (1992) Anatomical mapping of functional activation in stereotaxic coordinate space. *Neuroimage* 1:43–53.
- Friston K, Price C, Fletcher P, Moore C, Frackowiak R, Dolan R (1996) The trouble with cognitive subtraction. *Neuroimage* 4:97–104.
- Gibson JR, Maunsell JHR (1997) Sensory modality specificity of neural activity related to memory in visual cortex. *J Neurophysiol* 78:1263–1275.
- Griffiths T, Warren J (2002) The planum temporale as a computational hub. *Trends Neurosci* 25:348–353.
- Griffiths TD, Green GGR (1999) Cortical activation during perception of a rotating wide-field acoustic stimulus. *Neuroimage* 10:84–90.
- Grill-Spector K, Malach R (2001) fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychol* 107:293–321.
- Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzhak Y, Malach R (1999) Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron* 24:187–203.
- Hackett TA, Stepniowska I, Kaas JH (1999) Prefrontal connections of the parabelt auditory cortex in macaque monkeys. *Brain Res* 817:45–58.
- Hart HC, Palmer AR, Hall DA (2003) Amplitude and frequency-modulated stimuli activate common regions of human auditory cortex. *Cereb Cortex* 13:773–781.
- Imaizumi S, Mori K, Kiritani S, Kawashima R, Sugiura M, Fukuda H, Itoh K, Kato T, Nakamura A, Hatano K, Kojima S, Nakamura K (1997) Vocal identification of speaker and emotion activates different brain regions. *NeuroReport* 8:2809–2812.
- Kanwisher N (2000) Domain specificity in face perception. *Nat Neurosci* 3:759–763.
- Kourtzi Z, Kanwisher N (2001) Representation of perceived object shape by the human lateral occipital complex. *Science* 293:1506–1509.
- Kubovy M, Van Valkenburg D (2001) Auditory and visual objects. *Cognition* 80:97–126.
- Maeder P, Meuli R, Adriani M, Bellmann A, Fornari E, Thiran J-P, Pittet A, Clarke S (2001) Distinct pathways involved in sound recognition and localization: a human fMRI study. *Neuroimage* 14:802–816.
- Miller E, Li L, Desimone R (1991) A neural mechanism for working and recognition memory in inferior temporal cortex. *Science* 254:1377–1379.
- Nakamura K, Kubota K (1995) Mnemonic firing of neurons in the monkey temporal pole during a visual recognition memory task. *J Neurophysiol* 74:162–178.
- Patterson RD, Uppenkamp S, Johnsrude IS, Griffiths TD (2002) The processing of temporal pitch and melody information in auditory cortex. *Neuron* 36:767–776.
- Paus T, Perry DW, Zatorre RJ, Worsley KJ, Evans AC (1996) Modulation of cerebral blood-flow in the human auditory cortex during speech: role of motor-to-sensory discharges. *Eur J Neurosci* 8:2236–2246.
- Poremba A, Saunders RC, Crane AM, Cook M, Sokoloff L, Mishkin M (2003) Functional mapping of the primate auditory system. *Science* 299:568–572.
- Raichle M, Martin W, Herscovitch P, Mintun M, Markham J (1983) Brain blood flow measured with intravenous H₂(15)O. II. Implementation and validation. *J Nucl Med* 24:790–798.
- Rauschecker JP, Tian B (2000) Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proc Natl Acad Sci USA* 97:11800–11806.
- Rauschecker JP, Tian B, Hauser M (1995) Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268:111–114.
- Rauschecker JP, Tian B, Pons T, Mishkin M (1997) Serial and parallel processing in Rhesus monkey auditory cortex. *J Comp Neurol* 382:89–103.
- Recanzone GH, Guard DC, Phan ML, Su T-IK (2000) Correlation between the activity of single auditory cortical neurons and sound-localization behavior in the macaque monkey. *J Neurophysiol* 83:2723–2739.
- Romanski L, Goldman-Rakic P (2002) An auditory domain in primate prefrontal cortex. *Nat Neurosci* 5:15–16.
- Romanski L, Tian B, Fritz J, Mishkin M, Goldman-Rakic P, Rauschecker J (1999) Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat Neurosci* 2:1131–1136.
- Saleem KS, Suzuki W, Tanaka K, Hashikawa T (2000) Connections between anterior inferotemporal cortex and superior temporal sulcus regions in the macaque monkey. *J Neurosci* 20:5083–5101.
- Seltzer B, Pandya D (1989) Frontal lobe connections of the superior temporal sulcus in the rhesus monkey. *J Comp Neurol* 281:97–113.
- Talairach J, Tournoux P (1988) Co-planar stereotaxic atlas of the human brain. New York: Thieme.
- Thivard L, Belin P, Zilbovicius M, Poline J, Samson Y (2000) A cortical region sensitive to auditory spectral motion. *NeuroReport* 11:2969–2972.
- Tian B, Reser D, Durham A, Kustov A, Rauschecker JP (2001) Functional specialization in Rhesus monkey auditory cortex. *Science* 292:290–293.
- von Kriegstein K, Eger E, Kleinschmidt A, Giraud A-L (2003) Modulation of neural responses to speech by directing attention to voices or verbal content. *Cognit Brain Res* 17:48–55.
- Vuilleumier P, Henson R, Driver J, Dolan R (2002) Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. *Nat Neurosci* 5:491–499.
- Warren J, Zielinski B, Green G, Rauschecker J, Griffiths T (2002) Analysis of sound source motion by the human brain. *Neuron* 34:139–148.
- Warren J, Uppenkamp S, Patterson R, Griffiths T (2003) Separating pitch chroma and pitch height in the human brain. *Proc Natl Acad Sci USA* 100:10038–10042.
- Worsley K, Evans A, Marrett S, Neelin P (1992) A three-dimensional statistical analysis for CBF activation studies in human brain. *J Cereb Blood Flow Metab* 12:900–918.
- Zatorre RJ, Belin P (2001) Spectral and temporal processing in human auditory cortex. *Cereb Cortex* 11:946–953.
- Zatorre RJ, Belin P (2004) Auditory cortex processing streams: where are they and what do they do? In: *Plasticity of the central auditory system and processing of complex acoustic signals* (Syka J, ed). London: Plenum, in press.
- Zatorre RJ, Belin P, Penhune VB (2002a) Structure and function of auditory cortex: music and speech. *Trends Cognit Sci* 6:37–46.
- Zatorre RJ, Bouffard M, Ahad P, Belin P (2002b) Where is “where” in the human auditory cortex? *Nat Neurosci* 5:905–909.