Published in final edited form as:

Cortex. 2009 April ; 45(4): 517–526. doi:10.1016/j.cortex.2007.10.008.

Vowel-specific mismatch responses in the anterior superior temporal gyrus:

An fMRI study

Alexander P. Leff^{a,*}, Paul Iverson^b, Thomas M. Schofield^a, James M. Kilner^a, Jennifer T. Crinion^a, Karl J. Friston^a, and Cathy J. Price^a

^aWellcome Trust Centre for Neuroimaging, London, UK

^bDepartment of Phonetics and Linguistics, University College London, London, UK

Abstract

There have been many functional imaging studies that have investigated the neural correlates of speech perception by contrasting neural responses to speech and “speech-like” but unintelligible control stimuli. A potential drawback of this approach is that intelligibility is necessarily conflated with a change in the acoustic parameters of the stimuli. The approach we have adopted is to take advantage of the mismatch response elicited by an oddball paradigm to probe neural responses in temporal lobe structures to a parametrically varied set of deviants in order to identify brain regions involved in vowel processing. Thirteen normal subjects were scanned using a functional magnetic resonance imaging (fMRI) paradigm while they listened to continuous trains of auditory stimuli. Three classes of stimuli were used: ‘vowel deviants’ and two classes of control stimuli: one acoustically similar (‘single formants’) and the other distant (tones). The acoustic differences between the standard and deviants in both the vowel and single-formant classes were designed to match each other closely. The results revealed an effect of vowel deviance in the left anterior superior temporal gyrus (aSTG). This was most significant when comparing all vowel deviants to standards, irrespective of their psychoacoustic or physical deviance. We also identified a correlation between perceptual discrimination and deviant-related activity in the dominant superior temporal sulcus (STS), although this effect was not stimulus specific. The responses to vowel deviants were in brain regions implicated in the processing of intelligible or meaningful speech, part of the so-called auditory “what” processing stream. Neural components of this pathway would be expected to respond to sudden, perhaps unexpected changes in speech signal that result in a change to narrative meaning.

Keywords

ERB; Functional magnetic resonance; imaging; Mismatch responses; Tones; Vowels

1. Introduction

There have been many functional imaging studies over the last decade or so that have investigated the neural correlates of speech perception by contrasting neural responses to speech and “speech-like” but unintelligible control stimuli (Benson et al., 2006; Binder et al., 2000; Crinion et al., 2003; Mottonen et al., 2006; Scott et al., 2000; Specht and Reul,

2003; Zatorre et al., 1992). However, because no single acoustic feature predicts intelligibility of a given auditory stimulus (Arai and Greenberg, 1998), there are a variety of ways of producing unintelligible stimuli that are acoustically matched. One option is to parametrically vary the stimuli between the extremes of intelligibility, perhaps by varying the number of channels used in producing noise vocoded speech (Scott et al., 2006). A potential drawback of this approach is that intelligibility is necessarily conflated with a change in one of the acoustic parameters of the stimuli, in this case spectral complexity, requiring an extra set of stimuli to control for this effect. Another approach, the one we have adopted here, is to keep categorical classes of stimuli separate, but to vary the stimuli within a class in an oddball or mismatch paradigm, relying on the phenomena of automatic change detection to elicit neural responses. The prediction being that oddball responses will serve to identify brain regions involved in the automatic discrimination of auditory changes within a given class. If the acoustic changes that differentiate the deviants within a class are well matched across classes, then the resultant differences in functional magnetic resonance imaging (fMRI) signal are likely to be due to regional differences in speech specific processing.

We tested the hypothesis that mismatch responses would be generated by different regions of the temporal lobe depending on whether speech or non-speech stimuli were used. We wanted to investigate how the mismatch response varied across a range of four deviants. We varied vowels to create our speech deviants and used two classes of control stimuli for the non-speech stimuli. One class was approximately matched for acoustic complexity: single-formant stimuli and the other was more distant: sinusoidal tones. We wished to match the acoustic differences between the standard and deviants across the classes as closely as possible so that mismatch negativity (MMN) responses between the classes could be reasonably compared.

While it has clearly been shown that MMN responses tend to increase with increasing acoustic deviance from the standard in electroencephalography/magnetoencephalography (EEG/MEG) studies (Näätänen, 2001), this type of response is less evident when deviant-related haemodynamic responses are measured; with several studies finding non-linear dependences on parametrically modulated deviance (none examined more than three deviants within a given class: Doeller et al., 2003; Liebenthal et al., 2003; Opitz et al., 2002; Rinne et al., 2005). We wanted to test three different types of deviant response: a physical mismatch response (that reflected the acoustic difference between the deviant and the standard); a psychoacoustic response (based on individual perceptual thresholds) and an equipotent response (all deviants treated equally), within the three stimulus classes.

Mismatch responses have been reported in over 300 EEG/MEG papers since 1999, when the first attempts were made to identify the neural generators of the MMN response using a haemodynamic measure (Celsis et al., 1999; Opitz et al., 1999). However, there are fewer than 20 published reports using either positron emission tomography (PET) or fMRI in this period. Of these studies, the vast majority have employed tone stimuli in their mismatch paradigms, with only one investigating speech sound related responses (Celsis et al., 1999). The relative paucity of haemodynamic-based studies of the MMN paradigm may be due in part to the fact that stimulus presentation paradigms used in classical EEG/MEG experiments cannot be duplicated in fMRI for two reasons. Firstly, the low temporal resolution of PET and fMRI means that deviant responses cannot be differentiated from standard responses unless changes are made to the stimulus train such that there are relatively long (~12-30 sec) periods of time when only standards are presented, a form of block design. EEG/MEG designs also have mini-blocks or runs of standards but these do not have to be so long and can be as little as two consecutive standards before a deviant is presented (Haenschel et al., 2005). Fortunately, one of the strengths of fMRI is that an

improved signal-to-noise ratio can be established with some paradigms such that less deviants are needed to detect MMN responses, sometimes as little as 24 (Schall et al., 2003), compared with the usual number of a 100 or so in EEG/MEG studies. The second difference between haemodynamic and electrophysiological measures of mismatch relates to fMRI scanner noise and potential interference with the standard/deviant-related BOLD auditory responses (Novitski et al., 2001, 2006). One option is to adopt a non-continuous scanning paradigm (sparse or similar “clustered” acquisition), where stimuli are presented in blocks of relative silence (Liebenthal et al., 2003; Muller et al., 2003; Sabri et al., 2006). This option is not ideal though because the change in sound caused by turning on and off the echo-planar imaging (EPI) sequence also produces a MMN response (Kircher et al., 2004). We chose therefore to employ a continuous acquisition paradigm to simulate the classical MMN paradigm used in EEG/MEG experiments and to maximise the efficiency with which we could estimate the stimulus-related haemodynamic response function. We also employed a mixed block/event-related design coupled with an event-related analysis so responses to individual deviants, over and above the response to standards, could be modelled (Schall et al., 2003). Previous work suggests that MMN responses to words are usually left lateralized, while those to tones are right lateralized (Näätänen, 2001); therefore, we employed unilateral anatomically defined volume of interest masks in superior and lateral temporal cortices when interrogating the deviant-related responses.

2. Methods

2.1. Subjects, stimuli and task

Thirteen right-handed subjects with normal hearing, English as their first language and no history of neurological disease took part, seven were female; their mean age was 27.3 years (range = 21–38). All subjects gave informed consent and the study was approved by the local ethical committee.

Three classes of stimuli were created: vowels in consonant-vowel-consonant syllables (vowels), single formants (formants) and tones. Twenty-nine stimuli were produced for each class that deviated systematically from the standard (see Figs. 1 and 2) in a non-linear, monotonic fashion. All the stimuli were used to establish the subjects’ perceptual thresholds behaviourally, but only five (the standard and four deviants, stimuli 4, 12, 20 and 28) were used in the fMRI experiment.

The *vowels* were synthesized stimuli that varied in their first and second (F1 and F2) formant frequencies, and were placed within a /b/-V-/t/ context (e.g., forming the words *Bart* and *beat*). The frequencies of the standard (F1: 628 Hz and F2: 1014 Hz) created a prototypic /a/ vowel (Iverson and Evans, 2007). The frequencies of the four vowel deviants were 620, 596, 507, and 237 Hz for F1 and 1030, 1078, 1287, and 2522 Hz for F2. These frequencies were chosen so that F1 and F2 fell along a vector through the equivalent rectangular bandwidth-transformed vowel space (ERB: a scale based on critical bandwidths in the auditory system, so differences are more linearly related to perception: Glasberg and Moore, 1990), with the formant frequencies of the furthest deviant matching that of a prototypic /i/ (see Iverson and Evans, 2007). The spacing of the deviants was not uniform; the Euclidean distances of the deviants from the standard were .15, .58, 2.32, and 9.30 ERB, respectively (see Fig. 1a). Qualitatively, this meant that the first deviant was below typical behavioural discrimination thresholds (Iverson and Kuhl, 1995), the second deviant was near a discrimination threshold, the third was similar to the distances that would make a categorical difference between vowels in English (Iverson and Evans, 2007), and the fourth represented a stimulus on the other side of the vowel space. The vowels were synthesized using the cascade branch of a Klatt synthesizer (Klatt and Klatt, 1990). The stimuli were based on those from a previous study (Iverson and Evans, 2007), and were designed to

model a male British English speaker (the /b/ burst and the /t/ release were excised from a natural recording of this speaker rather than being synthesized). The duration of each word was 464 msec (260 msec for the vowel, excluding the bursts and /t/ stop-gap). F0 had a falling contour from 152 to 119 Hz. The formant frequencies of F3-F5 were 2500, 3500, and 4500 Hz. The bandwidths of the formants were 100, 180, 250, 300, and 550 for F1-F5.

2.1.1. Single formants—The single-formant stimuli were the same as the F2 formant from the vowel stimuli. They thus reproduced the major spectral difference between the vowel stimuli and had a somewhat speech-like carrier (e.g., the same intonation pattern as in the vowels). However, they fell short of being intelligible, and were somewhat less complex (e.g., lacking stop bursts or other formants). The stimuli were created using the parallel branch of a Klatt synthesizer (Klatt and Klatt, 1990). The standard and four deviants were created by setting the amplitude of F1 and F3-F5 to zero, such that only F2 was produced. The stimuli were 260 msec long (the same as the voiced portion of the vowels, without adding the consonantal bursts for stop-gaps).

2.1.2. Tones—The tones were sinusoids of 234 msec duration that were amplitude modulated by one half cycle of a raised cosine (i.e., leading to a gradually rising then falling amplitude envelope). The stimuli contrasted in frequency, somewhat like that of the vowel and formant stimuli. However, the tones contrasted fundamental frequency (the vowels and formants all had the same pitch, but differed in spectral envelope), and had a timbre that did not sound at all like speech. The standard frequency was the same as for the F2 of the vowel and single-formant standards (1014 Hz). The deviants were 1016, 1021, 1044, and 1138 Hz, which were .03, .13, .50, and 2.00 semitones away from the standard. These step sizes were designed to parallel the progressive deviations of the vowel stimuli (pilot behavioural data indicated that these step sizes could be discriminated about as well as the analogous step sizes used in the other series). The auditory stimuli used in the fMRI experiment can be accessed in the supplementary data.

Perceptual thresholds were measured outside the scanner in a quiet room and before the subjects had been exposed to the experimental paradigm; the order that the three classes of stimuli were tested across subjects was balanced to control for learning effects. Stimuli were delivered via an Axim X50v PocketPC computer using Sennheiser HD 650 headphones. The volume was adjusted to a comfortable level for each subject. Subjects completed an AX same-different discrimination task with a fixed standard; the deviant stimulus was varied adaptively (Levitt, 1971) to find the perceptual threshold (i.e., acoustic difference where subjects correctly discriminated the tones on 71% of the trials). Mean perceptual thresholds in terms of stimulus distance from the standard (in ERB) for vowel and formant stimuli were very similar: .61 (SD .22) and .60 (SD .27). The threshold for tone stimuli was .27 semitones (SD .15). There were no significant within-subject correlations between thresholds for the different classes: V-F, V-T and F-T correlation coefficients all .2.

2.2. In scanner task

Subjects were asked to ignore the auditory stimuli and focus on an incidental visual detection task. Pictures of static outdoor scenes were presented for 60 sec followed by a brief (1.5/sec) picture of either a circle or a square (red, on a grey background). The subjects were asked to press a button (right index finger) for the circles and not to press for the infrequent squares (10%). This is a form of go/no-go task and served to provide evidence that the subjects were awake and attending to the visual stimuli. All subjects performed well on this task with an overall correct response rate of 98% (range: 95-100).

2.3. fMRI scanning and stimulus presentation

A Siemens 1.5T scanner was used to acquire T2*-weighted echo-planar images (EPI) with BOLD contrast. Each EPI comprised 30 axial slices of 2.0 mm thickness with 1 mm interslice interval and 3×3 mm in-plane resolution. Volumes were acquired with an effective repetition time (TR) of 2.7 sec per volume and the first six volumes of each session were discarded to allow for T1 equilibration effects. A total of 178 volume images were acquired in six consecutive sessions, each lasting 8 min (one subject completed only four sessions). After the functional runs, a T1-weighted anatomical volume image was acquired for all subjects.

Stimuli were presented binaurally using E-A-RTONE 3A audiometric insert earphones (Etymotic Research, Inc.: Illinois, USA) that were housed in a lead shielded box at the base of the fMRI scanner and attached to the subject using flexible tubing. The stimuli were presented initially at 85 dB/SPL, and then subjects were allowed to adjust this to a comfortable level, while listening to the stimuli during a test period of EPI scanner noise. Subjects wore noise-attenuating 3 M 1440 ear defenders (Bracknell: UK) over the earphones to provide ~30 dB/SPL attenuation of the scanner noise.

Each session comprised 12 blocks, four of each class of stimuli (vowel, single formant or tone) and one for each type of deviant (four per class). Within each block 30 standards were presented with five deviants of the same type (same class and distance from the standard) at set time points selected to optimize the estimation of deviant-related BOLD changes (Fig. 1b). The deviants were presented towards the end of each block to allow the BOLD response to standards to saturate. Each session (e.g., run) was repeated six times for each subject resulting in a total of 30 events for each of the 12 deviant types (other than the subject who only completed four of the six sessions). The whole scanning period was 48 min. The ordering of the blocks of stimuli both within and across subjects was balanced.

2.4. Statistical parametric mapping

Statistical parametric mapping was performed using the SPM5 software (Wellcome Trust Centre for Neuroimaging: <http://fil.ion.ucl.ac.uk/spm>). All volumes from each subject were realigned and un-warped, using the first as reference and resliced with sinc interpolation. The functional images were then spatially normalized to the standard T2* template within SPM5 normalization software. Functional data were spatially smoothed, with a 6 mm full-width at half-maximum isotropic Gaussian Kernel to compensate for residual variability after spatial normalization and to permit application of Gaussian random field theory for corrected statistical inference. First, the statistical analysis was performed in a subject-specific fashion. To remove low-frequency drifts, the data were high-pass filtered using a set of discrete cosine basis functions with a cut-off period of 128 sec. Each stimulus (standard and deviant) was modelled separately by convolving it with a synthetic haemodynamic response function. The ensuing stimulus-specific parameter estimates were calculated for all brain voxels using the general linear model and contrast images were computed for analysis at the second or between-subject level (Friston et al., 1995).

We analysed two types of contrast image: class effects and deviant effects. Class effects treated standard and deviant stimuli equally comparing (1) vowel stimuli to formants and tones; (2) vowels to formants only; (3) vowels to tones only and (4) formants to tones. We then assessed the three simple effects of deviance under each of three classes: (1) the 'equipotent' contrast treated all deviants equally, comparing them to standards; (2) the 'psychoacoustic' contrast compared deviants above the perceptual threshold (usually deviants three and four) with those below the perceptual threshold (usually deviants one and two) on a subject-specific basis *nota bene* (NB): five subjects had their vowel threshold just

below deviant two, as did six subjects with their formant thresholds; one subject had their tone threshold below deviant two, and one above deviant three) and (3) the ‘physical’ contrast modelled deviancy as the ERB/semitone distance from the standard. These subject-specific contrast images were then smoothed again, to allow for anatomical inter-subject variability, using a 6 mm Gaussian Kernel and entered into second level, one-sample *t*-tests. In a third analysis, deviant specific contrasts from the second analysis were entered into separate one-sample *t*-tests with two regression values entered: the individual subjects’ perceptual threshold and their age.

To focus on the bilateral temporal lobes and correct for multiple statistical comparisons, we limited our analyses of class effects to a volume of interest that included Heschel’s gyrus (HG), planum temporale (PT), the superior temporal gyrus (STG) and superior temporal sulcus (STS). This mask covered 7009 voxels, equal to 40 resolution elements or resels and was created by importing the single-subject canonical T1 brain template into MRICro (Rorden and Brett, 2000) and identifying the regions by hand (APL). From this mask, we also generated a left hemisphere mask (19 resels) used for vowel deviants, and right hemisphere mask (21 resels), used for formant and tone deviants. A lower statistical threshold was set ($p = .001$ uncorrected) for the perceptual correlation analysis.

3. Results

3.1. Imaging analyses

Main effects of class (vowels, V; formants, F and tones, T) averaging over standards and deviants: vowel stimuli were associated with significantly higher activation in the lateral superior temporal lobe bilaterally, compared with both formant and tone stimuli. On the left, two subpeaks were present in the PT, and one at the border between PT and lateral HG. On the right, three peaks were evident with activity spread over a comparable region in the anterior-posterior direction, but lower, placing these peaks in the mid superior temporal sulcus (mSTS) (Table 1). The anatomical labels were assigned with reference to two studies on the boundaries of primary auditory cortex (Penhune et al., 1996) and PT (Westbury et al., 1999). In all regions, left and right, activation declined across the three classes: $V > F > T$, with significant differences between F and T in the left posterior PT and at the border of PT/HG on the same side. There were no voxels within the bilateral mask that survived correction for multiple comparisons for the ‘reverse’ contrasts: $T > F$; $T > V$; $F > V$ and $T + F > V$, and none outside the bilateral mask.

Simple main effects of deviants (St, D1-D4) within class: each deviant class was compared with the appropriate standard using equipotent, psychoacoustic and physical contrasts. Only two of the nine analyses (over all three classes) revealed a significant effect after correction for multiple comparisons. This was for the vowel deviants in the left anterior superior temporal gyrus (aSTG) using the equipotent contrast (Montreal Neurological Institute (MNI) co-ordinates: $-56\ 6\ -10$; $Z = 4.04$; $p = .041$) and for formant deviants in the right posterior superior temporal gyrus (pSTG) using the physical contrast (MNI co-ordinates: $64\ -30\ 18$; $Z = 4.10$; $p = .039$). There was a non-significant trend for tones using the psychoacoustic contrast (MNI co-ordinates: $56\ 2\ -12$; $Z = 3.88$; $p = .093$) in right aSTG (Fig. 3). No other analysis was significant, even if we used the right hemisphere mask for vowels and the left hemisphere mask for tones and formants.

A separate, post-hoc analysis with a search volume of 6 mm radius around the peak voxel of vowel-deviant analysis ($-56\ 6\ -10$) revealed a main effect of vowel deviants over formant and tone deviants in a nearby voxel ($-56\ 10\ -14$; $Z = 2.53$; $p = .006$ uncorrected).

Correlations between deviant-related BOLD activity and perceptual thresholds: guided by the findings in the above analysis which suggested that there may be a vowel-specific mismatch response in left aSTG, we wished to investigate whether there was any correlation between the size of the effect in this region and subjects' auditory discrimination abilities as indexed by their perceptual thresholds. Each subjects' relevant contrast image was regressed against their perceptual cut-off score (in ERBs for vowel and formant stimuli and in semitones for the tone stimuli). Initially we entered the contrast images from the three analyses that favoured each deviant class, but only found an effect for vowels, so then we entered the contrast with all deviants weighted equally (equipotent) for all three classes of stimuli. We used a less stringent statistical threshold, accepting peak voxels that survived $p = .001$ uncorrected within the bilateral mask. The most significant correlation for vowels was in mSTS ($-58 -16 -6$; $Z = 3.83$; $p < .001$ uncorrected), there was also a subpeak in aSTS, within 10 mm of the peak voxel in found in analysis B ($-52 -2 -10$). The correlation between perceptual threshold and deviant-related BOLD activity was not specific to vowels as similar responses could be found in nearby voxels for both the formant ($-58 -8 -2$; $Z = 3.22$; $p = .001$ uncorrected), and tone deviants ($-58 -16 -14$; $Z = 3.83$; $p = .001$ uncorrected) (Fig. 4).

4. Conclusions

Many fMRI studies have shown that speech activates STG/S relative to other acoustically complex stimuli (Belin et al., 2000; Binder et al., 2000; Scott et al., 2000; Specht and Reul, 2003; Vouloumanos et al., 2001; Zatorre et al., 1992). Consistent with these findings, we demonstrate regional brain responses to vowel stimuli in bilateral superior and lateral temporal cortices. There are several possible explanations for the stimulus class effects we observed. For example, both acoustic complexity and the degree to which the stimuli sound 'speech-like' reduced from vowels to formants to tones; moreover, only the vowel stimuli were intelligible, and they were phonologically more complex than the formant stimuli (which lacked consonants). Thus the $V > F > T$ response pattern we observed (see Table 1) could reflect either the acoustic complexity of the stimuli or the degree to which they were speech-like.

To circumvent the difficulties associated with comparisons across different classes of stimuli, we used an oddball paradigm that allowed us to investigate neural responses to changes of a given acoustic parameter within each stimulus class. The mismatch BOLD response for any given deviant was always relative to its own standard; therefore, differences in responses to vowel or formant deviants are more likely to represent differences in the way the brain processes speech sounds rather than variations in the other dimensions that apply to the stimulus class differences.

We found significant effects of deviant stimuli, for the vowel and formant classes. The most novel result was finding that vowel-deviants activated left aSTG. We only used one type of vowel standard and did not use other speech sounds, but there have been no fMRI mismatch studies of vowel deviants, and only one using a consonant deviant ("pa" vs. "ta"). In that study, Celsis et al. (1999) found left-sided deviant responses, but much more posteriorly and superiorly than us, in the dominant supramarginal gyrus. The more posterior location of mismatch activity in the Celsis et al. study may have been due to the much shorter duration of the difference between the standard and deviant, ~30 msec for their consonant stimuli, compared with 260 msec for the vowel stimuli in this study.

The location of the responses to vowel-deviants reported here have similar co-ordinates to those reported in a recent, non-MMN, study of vowel processing (Obleser et al., 2006), and are in a brain region implicated in the processing of intelligible or meaningful speech

(Crinion and Price, 2005; Crinion et al., 2003; Scott et al., 2000, 2006), part of the so-called auditory “what” processing stream (Scott and Wise, 2004). Neural components of this pathway would be expected to respond to sudden, perhaps unexpected, changes in acoustic signal that result in a change to narrative meaning. Our results add to the recent literature suggesting that sub-populations of neurons within the dominant STG and adjacent sulcus show a preferential response to speech sounds.

The response to vowel deviants was significant only when all deviants were weighted equally. We were expecting that the ‘best-fit’ mismatch response for vowel deviants would be psychoacoustic, given previous work on categorical perception of speech sounds (Aaltonen et al., 1993). The BOLD mismatch response to vowels, however, was ‘all-or-nothing’ and was driven equally by any difference between a standard and deviant. This may reflect the fact that vowels are not evenly distributed over vowel space and thus small changes in the first and second formants of vowels can lead to categorical changes; i.e., small acoustic differences can result in similar representational changes as large ones. There is supporting evidence for this from other vowel MMN studies, albeit in different imaging modalities. Using EEG/MEG, Näätänen found that the size of MMN responses to vowel-deviants depended not only on their acoustic distance from the standard; it was also affected by whether the deviants conformed to a prototypical vowel or not (Näätänen et al., 1997). The average perceptual threshold for subjects was just below D2 and it may have been that the D2, D3 and D4 stimuli provoked a similar BOLD response despite being progressively dissimilar acoustically to the standard.

The ‘best-fit’ for formant deviants was when the BOLD responses were predicted with the ERB values; this suggests that the neural generators were responding to the increasing acoustic difference between the deviants and standard (a rise in the F2 formant), which perceptually sounded like a change in timbre. We were expecting a similar response from the tones, but as their semitone values were adjusted to keep them from being perceptually too different from the other two stimulus classes, there was less contrast in these values; as such, the BOLD responses were probably driven by D3 and D4 rather than D1 and D2 (the latter fell below perceptual threshold for all but one of the 13 subjects). Although the formant and tone deviants in this study had different response functions, they were both localized to the right STG/STS, consistent with all fMRI studies of simple/complex tonal deviants which tend to find either bilateral or right-sided deviant-related responses (Molholm et al., 2005; Rinne et al., 2005; Sabri et al., 2004; Doeller et al., 2003; Opitz et al., 2002).

When we correlated each subjects’ perceptual threshold to the BOLD responses to the individual deviants (using ‘equipotent’ contrasts), an effect was seen in the left mid STS and was strongest for vowel deviants. This was not a class specific effect as similar correlations were found in both the formant and tone analyses. Correlations between ‘conscious’ performance on a perceptual discrimination task and the mismatch response in an ‘automatic’ task have been noted many times before in EEG studies (Lang et al., 1990; Novitski et al., 2004), but this is the first time a correlation has been reported using the BOLD mismatch response. Our findings are consistent with work suggesting that neural processes that support active acoustic discrimination may be involved in the neural networks that produce mismatch responses, perhaps by informing or shaping the ‘central sound representations’ thought to be the basis of the mismatch response (Näätänen and Alho, 1997).

Mismatch responses can be measured using fMRI. Although it has been questioned whether these responses are produced by the same neuronal populations responsible for EEG/MEG MMN responses (Nunez and Silberstein, 2000), recent work on the complex relationship

between neural activity and the corresponding changes in regional cerebral blood flow suggests that the BOLD response correlates best with changes in the local field potentials (LFPs) of cortical regions (Logothetis, 2003). LFPs reflect a variety of cellular processes including excitatory and inhibitory post-synaptic potentials that underlie EEG/MEG signals measured at or near the scalp (Hamalainen et al., 1993). While EEG studies of MMN responses in parametrically deviant stimuli tend to show increasing MMN signal (either peak amplitudes or area-under-the-graph) with increasing deviancy from the standard, at least with non-language stimuli (Näätänen et al., 1997), this pattern appears to be less common with fMRI mismatch responses (Doeller et al., 2003; Opitz et al., 2002; Rinne et al., 2005).

In summary, we have taken advantage of the mismatch response elicited by an oddball paradigm to probe the differential responses in temporal lobe structures to a parametrically varied vowel deviant and a close (formant) and distant (tone) set of control stimuli. The results from this study support the notion that there are specific brain regions tuned to detecting acoustic changes in vowel sounds; these regions presumably play a role in the identification of real-world auditory objects that have close acoustic neighbours that alter narrative meaning.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

REFERENCES

- Aaltonen O, Tuomainen J, Laine M, Niemi P. Cortical differences in tonal versus vowel processing as revealed by an ERP component called mismatch negativity (MMN). *Brain and Language*. 1993; 44:139–152. [PubMed: 8428308]
- Arai, T.; Greenberg, S. Speech intelligibility in the presence of cross-channel spectral asynchrony; Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing; 1998; p. 933-936.
- Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B. Voice-selective areas in human auditory cortex. *Nature*. 2000; 403:309–312. [PubMed: 10659849]
- Benson RR, Richardson M, Whalen DH, Lai S. Phonetic processing areas revealed by sinewave speech and acoustically similar non-speech. *NeuroImage*. 2006; 31:342–353. [PubMed: 16530428]
- Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET. Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex*. 2000; 10:512–528. [PubMed: 10847601]
- Celsis P, Boulanouar K, Doyon B, Ranjeva JP, Berry I, Nespoulous JL, Chollet F. Differential fMRI responses in the left posterior superior temporal gyrus and left supramarginal gyrus to habituation and change detection in syllables and tones. *NeuroImage*. 1999; 9:135–144. [PubMed: 9918735]
- Crinion J, Price CJ. Right anterior superior temporal activation predicts auditory sentence comprehension following aphasic stroke. *Brain*. 2005; 128:2858–2871. [PubMed: 16234297]
- Crinion JT, Lambon-Ralph MA, Warburton EA, Howard D, Wise RJ. Temporal lobe regions engaged during normal speech comprehension. *Brain*. 2003; 126:1193–1201. [PubMed: 12690058]
- Doeller CF, Opitz B, Mecklinger A, Krick C, Reith W, Schroger E. Prefrontal cortex involvement in preattentive auditory deviance detection: neuroimaging and electrophysiological evidence. *NeuroImage*. 2003; 20:1270–1282. [PubMed: 14568496]
- Friston KJ, Holmes AP, Worsley KJ, Poline J-B, Frith CD, Frackowiak RSJ. Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping*. 1995; 2:189–210.
- Glasberg BR, Moore BC. Derivation of auditory filter shapes from notched-noise data. *Hearing Research*. 1990; 47:103–138. [PubMed: 2228789]

- Haenschel C, Vernon DJ, Dwivedi P, Gruzelier JH, Baldeweg T. Event-related brain potential correlates of human auditory sensory memory-trace formation. *The Journal of Neuroscience*. 2005; 25:10494–10501. [PubMed: 16280587]
- Hamalainen M, Hari R, Ilmoniemi RJ, Knuutila J, Lounasmaa OV. Magnetoencephalography - theory, instrumentation and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*. 1993; 65:413–497.
- Iverson P, Evans BG. Learning English vowels with different first-language vowel systems I: individual differences in the perception of formant targets, formant movement, and duration. *The Journal of the Acoustical Society of America*. 2007; 122:2842–2854. [PubMed: 18189574]
- Iverson P, Kuhl PK. Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *The Journal of the Acoustical Society of America*. 1995; 97:553–562. [PubMed: 7860832]
- Kircher TT, Rapp A, Grodd W, Buchkremer G, Weiskopf N, Lutzenberger W, Ackermann H, Mathiak K. Mismatch negativity responses in schizophrenia: a combined fMRI and whole-head MEG study. *The American Journal of Psychiatry*. 2004; 161:294–304. [PubMed: 14754779]
- Klatt DH, Klatt LC. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustical Society of America*. 1990; 87:820–857. [PubMed: 2137837]
- Lang, H.; Nyrke, T.; Ek, M.; Aaltonen, O.; Raimo, I.; Näätänen, R. Pitch discrimination performance and auditory event-related potentials. In: Brunia, CHM.; Gaillard, AWK.; Kok, A.; Mulder, G.; Verbaten, MN., editors. *Psychophysiological Brain Research*. Tilburg University Press; Tilburg, The Netherlands: 1990. p. 294–298.
- Levitt H. Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*. 1971; 49(Suppl. 2):467. [PubMed: 5541744]
- Liebenthal E, Ellingson ML, Spanaki MV, Prieto TE, Ropella KM, Binder JR. Simultaneous ERP and fMRI of the auditory cortex in a passive oddball paradigm. *NeuroImage*. 2003; 19:1395–1404. [PubMed: 12948697]
- Logothetis NK. The underpinnings of the BOLD functional magnetic resonance imaging signal. *The Journal of Neuroscience*. 2003; 23:3963–3971. [PubMed: 12764080]
- Molholm S, Martinez A, Ritter W, Javitt DC, Foxe JJ. The neural circuitry of pre-attentive auditory change-detection: an fMRI study of pitch and duration mismatch negativity generators. *Cerebral Cortex*. 2005; 15:545–551. [PubMed: 15342438]
- Mottron R, Calvert GA, Jaaskelainen IP, Matthews PM, Thesen T, Tuomainen J, Sams M. Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. *NeuroImage*. 2006; 30:563–569. [PubMed: 16275021]
- Muller BW, Stude P, Nebel K, Wiese H, Ladd ME, Forsting M, Jueptner M. Sparse imaging of the auditory oddball task with functional MRI. *Neuroreport*. 2003; 14:1597–1601. [PubMed: 14502083]
- Näätänen R. The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*. 2001; 38:1–21. [PubMed: 11321610]
- Näätänen R, Alho K. Mismatch negativity - the measure for central sound representation accuracy. *Audiology & Neurotology*. 1997; 2:341–353. [PubMed: 9390839]
- Näätänen R, Lehtokoski A, Lennes M, Cheour M, Huottilainen M, Iivonen A, Vainio M, Alku P, Ilmoniemi RJ, Luuk A, Allik J, Sinkkonen J, Alho K. Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*. 1997; 385:432–434. [PubMed: 9009189]
- Novitski N, Alho K, Korzyukov O, Carlson S, Martinkauppi S, Escera C, Rinne T, Aronen HJ, Näätänen R. Effects of acoustic gradient noise from functional magnetic resonance imaging on auditory processing as reflected by event-related brain potentials. *NeuroImage*. 2001; 14:244–251. [PubMed: 11525334]
- Novitski N, Maess B, Tervaniemi M. Frequency specific impairment of automatic pitch change detection by fMRI acoustic noise: an MEG study. *Journal of Neuroscience Methods*. 2006; 155:149–159. [PubMed: 16530843]

- Novitski N, Tervaniemi M, Huotilainen M, Näätänen R. Frequency discrimination at different frequency levels as indexed by electrophysiological and behavioral measures. *Brain Research. Cognitive Brain Research*. 2004; 20:26–36. [PubMed: 15130586]
- Nunez PL, Silberstein RB. On the relationship of synaptic activity to macroscopic measurements: does co-registration of EEG with fMRI make sense? *Brain Topography*. 2000; 13:79–96. [PubMed: 11154104]
- Obleser J, Boecker H, Drzezga A, Haslinger B, Hennenlotter A, Roettinger M, Eulitz C, Rauschecker JP. Vowel sound extraction in anterior superior temporal cortex. *Human Brain Mapping*. 2006; 27:562–571. [PubMed: 16281283]
- Opitz B, Mecklinger A, Von Cramon DY, Kruggel F. Combining electrophysiological and hemodynamic measures of the auditory oddball. *Psychophysiology*. 1999; 36:142–147. [PubMed: 10098390]
- Opitz B, Rinne T, Mecklinger A, von Cramon DY, Schroger E. Differential contribution of frontal and temporal cortices to auditory change detection: fMRI and ERP results. *NeuroImage*. 2002; 15:167–174. [PubMed: 11771985]
- Penhune VB, Zatorre RJ, MacDonald JD, Evans AC. Interhemispheric anatomical differences in human primary auditory cortex: probabilistic mapping and volume measurement from magnetic resonance scans. *Cerebral Cortex*. 1996; 6:661–672. [PubMed: 8921202]
- Rinne T, Degerman A, Alho K. Superior temporal and inferior frontal cortices are activated by infrequent sound duration decrements: an fMRI study. *NeuroImage*. 2005; 26:66–72. [PubMed: 15862206]
- Rorden C, Brett M. Stereotaxic display of brain lesions. *Behavioural Neurology*. 2000; 12:191–200. [PubMed: 11568431]
- Sabri M, Kareken DA, Dziedzic M, Lowe MJ, Melara RD. Neural correlates of auditory sensory memory and automatic change detection. *NeuroImage*. 2004; 21:69–74. [PubMed: 14741643]
- Sabri M, Liebenthal E, Waldron EJ, Medler DA, Binder JR. Attentional modulation in the detection of irrelevant deviance: a simultaneous ERP/fMRI study. *Journal of Cognitive Neuroscience*. 2006; 18:689–700. [PubMed: 16768370]
- Schall U, Johnston P, Todd J, Ward PB, Michie PT. Functional neuroanatomy of auditory mismatch processing: an event-related fMRI study of duration-deviant oddballs. *NeuroImage*. 2003; 20:729–736. [PubMed: 14568447]
- Scott SK, Blank CC, Rosen S, Wise RJ. Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*. 2000; 123(Pt 12):2400–2406. [PubMed: 11099443]
- Scott SK, Rosen S, Lang H, Wise RJ. Neural correlates of intelligibility in speech investigated with noise vocoded speech - a positron emission tomography study. *The Journal of the Acoustical Society of America*. 2006; 120:1075–1083. [PubMed: 16938993]
- Scott SK, Wise RJ. The functional neuroanatomy of prelexical processing in speech perception. *Cognition*. 2004; 92:13–45. [PubMed: 15037125]
- Specht K, Reul J. Functional segregation of the temporal lobes into highly differentiated subsystems for auditory perception: an auditory rapid event-related fMRI-task. *NeuroImage*. 2003; 20:1944–1954. [PubMed: 14683700]
- Vouloumanos A, Kiehl KA, Werker JF, Liddle PF. Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech. *Journal of Cognitive Neuroscience*. 2001; 13:994–1005. [PubMed: 11595101]
- Westbury CF, Zatorre RJ, Evans AC. Quantifying variability in the planum temporale: a probability map. *Cerebral Cortex*. 1999; 9:392–405. [PubMed: 10426418]
- Zatorre RJ, Evans AC, Meyer E, Gjedde A. Lateralization of phonetic and pitch discrimination in speech processing. *Science*. 1992; 256:846–849. [PubMed: 1589767]

Fig. 1.

(a) The 29 stimuli plotted against their ERB distance (deviance) from the standard (St) for vowel (top) and against their semitone deviance from the standard for tone stimuli (bottom). The formant stimuli were similar to that for vowels, although only F2 was varied rather than F1 and F2. The four deviant stimuli (D1-D4) for each class that were used in the fMRI experiment are shown in filled circles. Arrows mark the mean perceptual thresholds for each stimulus class and (b) Example of one session (above) of continuous stimulus presentation comprising of 12 blocks of alternating stimulus class. *F*= formant; *T*= tone and *V*= vowel blocks. Each block (below) was composed of 35 stimuli (circles). Fifteen standards (open circles) were presented before five deviants, all of the same level of deviancy from the standard (D4 in the exploded block, filled circles).

Fig. 2. Spectrograms for the standards (left column) and most distant deviants (D4s right column) for the three classes of stimuli. Note that the formant stimuli track the rise in the second formant of the vowel stimuli.

Fig. 3.

First row: fMRI results for main effects of class (threshold at $p < .001$, cluster size > 25 contiguous voxels). Activations are rendered onto the canonical single-subject brain template available in SPM5. L = left and R = right hemisphere responses to vowels > formants and tones. Second row: dotted lines show the 'best-fit' predictions of BOLD responses, i.e., the contrasts that were entered into the individual subjects' design matrices to model each of the three deviant weightings: equipotent, psychoacoustic and physical. Third row: deviant effects (thresholded at $p < .01$, cluster size > 25 contiguous voxels). Forth row: measured, average BOLD responses (arbitrary units) with SEM bars relative to the standards

of the same class. V = vowel deviants (left); F = formant deviants (middle); T = tone deviants (right) and D1-D4 = deviants.

Fig. 4.

First row: fMRI results for the regression analysis of BOLD responses to all vowel deviants with perceptual thresholds (in ERB), across the 13 subjects. Activation is seen along the length of the left STS (threshold at $p < .05$, cluster size > 25 contiguous voxels). Second row: BOLD/ERB plots from the peak voxel for the vowel (V) analysis (mSTS) left; similar plots for the formant (F) and Tone (T) analyses, with the plots extracted from peak voxels near the vowel peak (see text).

Table 1
Co-ordinates, Z and p-values for the peak voxels of the between class activations (analysis A: standards and deviants within each class weighted equally)

Contrast		Region		V > F + T	Z	p	V > F	Z	p	V > T	Z	p	F > T	Z	p
Left	HG/PT			-60 -12 6	4.09	.074	-62 -12 6	2.94	ns	-60 -12 6	4.72	.008	-62 -12 6	4.73	.008
	PT			-64 -12 0	5.04	.002	-64 -12 -2	4.51	.023	-64 -12 -2	4.77	.007	-64 -12 0	2.90	ns
	PT			-54 -24 10	4.57	.017	-54 -24 10	2.97	ns	-52 -24 8	5.12	<.001	-54 -22 6	4.70	.009
Right	mSTS			66 -10 -8	5.75	<.001	66 -12 -10	5.06	.001	66 -10 -10	5.67	<.001	62 -10 -10	3.72	ns
	mSTS			56 -14 -8	5.16	.001	56 -14 -8	3.89	ns	56 -14 -10	5.68	<.001	56 -14 -8	2.94	ns
	mSTS			58 -24 -6	4.12	.057	58 -24 -6	3.25	ns	58 -24 -4	4.05	.096	58 -24 -6	1.28	ns

Voxels with $p < .05$ corrected for multiple comparisons using the bilateral temporal mask as a small volume correction, are in bold, those with $p > .1$ are coded as ns = non-significant. HG = Heschel's gyrus; F = formant stimuli; mSTS = mid superior temporal sulcus; PT = planum temporale; T = tone stimuli and V = vowel stimuli.