cortex 45 (2009) 517–526

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cortex

Research report

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]

[a]*Wellcome Trust Centre for Neuroimaging, London, UK*
[b]*Department of Phonetics and Linguistics, University College London, London, UK*

ARTICLE INFO

Article history:
Received 28 August 2007
Reviewed 22 October 2007
Revised 29 October 2007
Accepted 30 October 2007
Guest editor Stefano Cappa
Published online 7 February 2008

Keywords:
ERB
Functional magnetic resonance imaging
Mismatch responses
Tones
Vowels

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.

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

* *Corresponding author*. Wellcome Trust Centre for Neuroimaging, 12 Queen Square, London WC1N 3BG, UK.
E-mail address: a.leff@fil.ion.ucl.ac.uk (A.P. Leff).
0010-9452/$ – see front matter © 2007 Elsevier Srl. All rights reserved.
doi:10.1016/j.cortex.2007.10.008

(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-planer 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