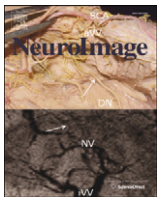
NeuroImage 57 (2011) 1022–1030

Contents lists available at ScienceDirect

NeuroImage

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



Improving the reliability of functional localizers

Keith J. Kawabata Duncan *, Joseph T. Devlin

Cognitive, Perceptual and Brain Sciences & Institute of Cognitive Neuroscience, UCL, London, UK

ARTICLE INFO

Article history:
Received 29 September 2010
Revised 27 April 2011
Accepted 3 May 2011
Available online 10 May 2011

Keywords:
fMRI
Reading
Object recognition
Face recognition
Posterior fusiform gyrus
Occipito-temporal cortex
Region-of-interest

ABSTRACT

A critical assumption underlying the practice of functional localization is that the voxels identified by functional localization are essentially the same as those activated in the main experiment for a particular anatomical area. Violations of this assumption bias the resulting analyses and can dramatically increase the likelihood of both Type I and Type II errors. Here we investigated how the amount of data affects the reliability of a set of common functionally-defined regions-of-interest (fROIs). Four participants were scanned ten times each to functionally localize extrastriate regions sensitive to visually presented words, objects and faces. A within-subject random-effects analysis was used as the "gold standard" for identifying the fROIs and the results were compared to within-subject, fixed-effect analyses typically used for functional localization. By varying the quantity of data included in the analyses, we empirically assessed the amount needed to ensure reliable identification of the fROIs. The results demonstrated that the most consistent fROIs were based on either stringent statistical thresholding ($Z>5.0$) of large quantities of data or on lenient thresholding ($Z>2.3$) of a modest amount of data, with both methods yielding 70–80% overlap between the functional localization results and the "gold standard." Stringent statistical thresholds on typical quantities of localizer data led to the poorest reliability (<20% overlap). These findings suggest that the most reliable and cost-efficient method for functional localization involves collecting a relatively small amount of data (~10 min) and using a lenient statistical threshold to identify all voxels in a given region that are sensitive to the process-of-interest.

Introduction

The practice of identifying a region-of-interest based on its functional response to a class of stimuli is commonly known as functional localization and has become an important methodology in functional magnetic resonance imaging. Typically, multiple scans are acquired in which a participant performs a main task and a different localizer task during a separate, shorter scan, whose sole aim is to identify the functionally defined region-of-interest (fROI). The functional localizer helps to addresses the twin problems facing group studies of inter-subject anatomical variability and the variable correspondence between cognitive function and anatomy (Saxe et al., 2006). In order to be successful, though, it requires that the set of voxels in the fROI is essentially the same as those activated in the main task. If this assumption is violated, then the analysis will be less sensitive to true effects and more susceptible to false positives.

Recent studies have reported surprisingly poor intra-subject consistency between functional localizer scans for a variety of stimulus categories (Kung et al., 2007; Duncan et al., 2009; Berman et al., 2010). For instance, the maximum overlap between fusiform face area (FFA) localizers was only 60% of voxels (Kung et al., 2007; Berman et al., 2010) and the values were similar when localizing other extrastriate regions sensitive to either visual words or objects (Duncan et al., 2009). Although some variability arises from physiological and scanner noise (Aguirre et al., 1998; Kruger and Glover, 2001; Handwerker et al., 2004), the intrinsic reproducibility of the blood-oxygen dependent signal within subjects is high (Neumann et al., 2003; Aron et al., 2006; Bennett and Miller, 2010), suggesting that noise is not a major contributing factor to the variability of fROIs. Other possibilities include factors directly under the experimenter's control such as the choice of tasks during localization and testing, the specific analysis methods used to define the fROIs, and the relatively small quantities of data typically collected during functional localization. Although in some cases, the choice of task appears not to be a significant factor (Berman et al., 2010), the reliability of fROIs is potentially affected by other aspects of the functional localizer design.

The two most common methods for defining an fROI are a sphere centred on the peak voxel (Miller and D'Esposito, 2005; Blankenburg et al., 2006; Pulvermüller et al., 2006; Aleong and Paus, 2010; Tibber et al., 2010; Kühn et al., 2011) and as the cluster of active voxels within a particular anatomical area (Kanwisher et al., 1999; Grill-Spector et al., 2004; Spiridon et al., 2006; Morris et al., 2008; Yovel et al., 2008; Axelrod and Yovel, 2010). Within this broad framework there is considerable variability in the specific methods used to define the fROI. For instance, studies that use a spherical ROI use radii that range from fairly small (e.g. 2 mm Blankenburg et al., 2006) to relatively large (9 mm Taylor et al., 2010). Similarly, a wide range of statistical thresholds are commonly used to define the "active" voxels

* Corresponding author at: Wellcome Trust Centre for Neuroimaging, 12 Queen Square, London, WC1N 3BG, UK. Fax: +44 2074364276.
E-mail address: k.duncan@ucl.ac.uk (K.J. Kawabata Duncan).

1053-8119/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2011.05.009

within a cluster (Morris et al., 2008). Both methods constrain the fROI to fall within an *a priori* anatomical region-of-interest such as within a probabilistically defined region (Eickhoff et al., 2006; Shattuck et al., 2008; Holdstock et al., 2009) or on an anatomical landmark (e.g. fusiform gyrus). Thankfully, only some of these essentially arbitrary decisions affect the reliability of the fROI (e.g. the location of the peak voxel for a spherical ROI and the statistical threshold used to determine "active" voxels in a cluster) while others primarily affect the specificity of the fROI (e.g. the size of the radius and the anatomical priors).

Here we investigated how reproducible peak coordinates and clusters of activation were in relation to the quantity of localizer data collected. This emphasis on the amount of data is motivated by the finding that retinotopic visual fields appear remarkably reproducible within an individual (Sereno et al., 1995; Tootell et al., 1997) and are usually based on much larger data sets than typically collected when localizing category-specific cortical regions. Specifically, we investigated how the quantity of data collected affects the reproducibility of three different extrastriate fROIs sensitive to visually presented words, objects and faces. Four participants completed ten functional localizer scans each over a two-month period. A within-subject random-effects analysis was used as the "gold standard" to define the fROIs and the results were compared to standard functional localizer analyses based on varying quantities of data ranging from (the relatively small) quantities typically used in functional localization to (the fairly large) data sets used in retinotopic mapping. Since the two most common methods for defining an fROI are a sphere with its origin at the peak voxel and active voxels within a particular anatomical ROI, the results are analysed in terms of both the location of the peak voxel within each region and the spatial overlap between the fROIs defined across a range of statistical thresholds. The findings are used to develop heuristic guidelines for optimising functional localization procedures to produce the most reliable and robust effects.

Materials and methods

Participants

4 (2 M, 2 F) healthy, monolingual English speakers volunteered for the study. Their ages ranged from 25 to 39 (mean = 33), and all were right handed with normal or corrected-to-normal vision. None had a personal or family history of any neurological disease, and each gave informed consent after the experimental procedures were explained. This experiment was approved by the UCL Research Ethics Committee.

Experimental procedures

Subjects performed a one-back task with four categories of visual stimuli: written words, pictures of common objects, scrambled pictures of the same objects and faces. Subjects were instructed to press a button if the stimulus was identical to the preceding stimulus and 12.5% of the stimuli were targets. A block design was used to maximize statistical sensitivity. Each block consisted of 16 trials from a single category presented one every second. A trial began with a 650 ms fixation cross, followed by the stimulus for 350 ms. In between blocks, subjects viewed a fixation cross for 16 s. The stimuli were divided equally into two lists and each list comprised 96 stimuli per category per list including targets. Within a run, no stimuli were repeated (except for target trials). Across runs, faces were repeated but words, objects and scrambled objects stimuli were not. In a single scanning session, participants completed two 8-minute runs. Over the course of approximately two months, each participant completed five of these sessions with each session separated by a minimum of one week. The same stimuli were used in each session. In total, each participant completed 10 runs.

Word stimuli (n = 168) were obtained from the MRC Psycholinguistic database (Coltheart, 1981) and consisted of 4 or 5 letter words with regular spellings (e.g. "hope"). All words had familiarity ratings between 300 and 500 (Coltheart, 1981), were either one or two syllables, and had a British English written word frequency value of 40 or less (Baayen et al., 1993). The stimuli in the two runs were fully matched for frequency, familiarity, imageability, number of letters, and number of syllables. Object stimuli consisted of greyscale pictures (200 × 250 pixels) of easily recognizable objects such as a boat, tent, nail, etc. Scrambled objects were generated by dividing the pictures into 10 × 10 pixel squares and permuting their placement within the image. None of the resulting images were recognizable after scrambling. Finally, face stimuli consisted of greyscale images (300 × 300 pixels) of front-view male and female faces of a variety of ethnicities.

Functional imaging

Whole-brain imaging was performed on a Siemens 1.5 Tesla MR scanner at the Birkbeck-UCL Neuroimaging (BUCNI) Centre in London. The functional data were acquired with a gradient-echo EPI sequence (TR = 3000 ms; TE = 50 ms; FOV = 192 × 192, matrix = 64 × 64) giving a notional resolution of 3 × 3 × 3 mm. Each run consisted of 164 volumes (32 volumes per condition) and took 8.2 min. In addition, a high-resolution anatomical scan was acquired (T1-weighted FLASH, TR = 12 ms; TE = 5.6 ms; 1 mm^3 resolution) during the first scanning session for anatomically localising activations in individuals.

Data processing was carried out using FSL 4.0 (www.fmrib.ox.ac.uk/fsl). To allow for T1 equilibrium, the initial two images of each run were discarded. The data were then realigned to remove small head movements (Jenkinson et al., 2002), smoothed with a Gaussian kernel of FWHM 6 mm, and pre-whitened to remove temporal auto-correlation (Woolrich et al., 2001). The resulting images were entered into a general linear model with four conditions of interest corresponding to the four categories of visual stimuli. Blocks were convolved with a double gamma "canonical" hemodynamic response function (Glover, 1999) to generate the main regressors. In addition, the estimated motion parameters were entered as covariates of no interest to reduce structured noise due to minor head motion. First level results were registered to the MNI-152 template using a 12-DOF affine transformation (Jenkinson and Smith, 2001) and all subsequent analyses were conducted in the MNI standard space. Linear contrasts of [words > fixation], [objects > scrambled objects] and [faces > objects] identified reading-, object- and face- sensitive areas, respectively, within the broad anatomical regions-of-interest.

Anatomical region-of-interest masks were drawn for the left and right hemispheres in standard space for each stimulus type. The word mask defined a region of ventral occipito-temporal cortex consistently engaged during visual word recognition (Price et al., 1994, 1996; Herbster et al., 1997; Rumsey et al., 1997; Fiez and Petersen, 1998; Fiez et al., 1999; Shaywitz et al., 2004) using the following standard (i.e. MNI152) space coordinates: X = (±)30 to (±)54, Y = −45 to −70 and Z = −30 to −4. This encompassed the posterior portion of the fusiform gyrus, occipito-temporal sulcus (OTS), and medial parts of the inferior temporal gyrus (ITG). The object mask encompassed lateral posterior fusiform gyrus, posterior OTS and lateral parts of posterior ITG, regions consistently involved in visual object recognition (Malach et al., 1995; Grill-Spector et al., 1999). The standard space coordinates were X = (±)33 to (±)56, Y = −67 to −89 and Z = −20 to +4. Finally, the mask for faces encompassed the "fusiform face area" (Kanwisher et al., 1997), a region similar to the 'words mask' but slightly more anterior. The standard space coordinates were: X = (±)31 to (±)51, Y = −36 to −60 and Z = −31 to Z = −4. These masks were then customised for