NeuroImage 74 (2013) 77–86

Contents lists available at SciVerse ScienceDirect

NeuroImage

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

Separable networks for top-down attention to auditory non-spatial and visuospatial modalities

Rodrigo M. Braga [a,b,*], Liam R. Wilson [a], David J. Sharp [a], Richard J.S. Wise [a], Robert Leech [a,*]

[a] Computational, Cognitive and Clinical Neuroimaging Laboratory C3NL Burlington Danes Building, Hammersmith Hospital Campus, Du Cane Road, Imperial College London, London W12 0NN, UK

[b] MRC Clinical Sciences Centre, Faculty of Medicine, Imperial College London, Hammersmith Hospital Campus, London W12 0NN, UK

ARTICLE INFO

Article history:
Accepted 2 February 2013
Available online 20 February 2013

Keywords:
Top-down attention
Visual
Auditory
Dorsal attention network
Independent component analysis
Endogenous

ABSTRACT

A central question for cognitive neuroscience is whether there is a single neural system controlling the allocation of attention. A dorsal frontoparietal network of brain regions is often proposed as a mediator of top-down attention to all sensory inputs. We used functional magnetic resonance imaging in humans to show that the cortical networks supporting top-down attention are in fact modality-specific, with distinct superior fronto-parietal and fronto-temporal networks for visuospatial and non-spatial auditory attention respectively. In contrast, parts of the right middle and inferior frontal gyri showed a common response to attentional control regardless of modality, providing evidence that the amodal component of attention is restricted to the anterior cortex.

Introduction

The ability to select task-relevant information (top-down or endogenous attention) is central to high-level cognition, perception and behavior (Posner and Petersen, 1990). The assumption that there is a single system mediating top-down attention to all sensory modalities underlies many theoretical accounts of cognitive control (Corbetta et al., 2008; Posner and Petersen, 1990; Spence and Driver, 1997). A frontoparietal network that includes the superior parietal lobe (SPL), frontal eye fields (FEF) and middle frontal gyrus (MFG) is activated during many studies of top-down attention (Kincade et al., 2005; Vossel et al., 2006) and has been labeled the "dorsal attentional network" (DAN; Corbetta et al., 2008). In contrast, a more inferior network that includes the MFG and temporoparietal junction (the "ventral attention network" or VAN) is activated together with the DAN when attention is captured by behaviorally relevant stimuli (bottom-up or exogenous attention), in what has been termed the 'reorienting response' (Corbetta et al., 2008).

The DAN is widely assumed to be amodal, supporting top-down attention to visual, auditory and somatosensory inputs (Driver and Spence, 1998; Langner et al., 2011; Macaluso, 2010; Posner and Petersen, 1990). However, the evidence for this network comes overwhelmingly from visual studies (Corbetta et al., 2008), which agree with reports that the SPL and FEF are strongly involved in visuospatial processing (Behrmann, 2004) and controlling eye movement (Büttner-Ennever and Horn, 1997). For example, the FEF and SPL have been shown to have a strong retinotopic organization both with direct stimulation and functional neuroimaging (Moore et al., 2003; Ruff et al., 2008; Saygin and Sereno, 2008). In vision, both spatial and non-spatial attention tasks have implicated the SPL and FEF (Marois et al., 2000). However, in audition, Shomstein and Yantis (2006) found activation of the SPL but not FEF during spatial attention, and reported SPL deactivation during non-spatial sections of the task. Therefore, although DAN involvement in visual attention is supported by neuropsychological, retinotopic and oculomotor studies, it is less clear whether two core nodes of the DAN, the FEF and SPL, are needed for attending to other sensory modalities such as audition.

Previous functional imaging studies have implicated the full DAN in processing auditory stimuli (Davis et al., 2000; Driver and Spence, 1998; Hallett et al., 1999; Langner et al., 2011; Linden, 1999; Macaluso et al., 2003; Maeder et al., 2001; Mayer et al., 2006; Shomstein and Yantis, 2006; Sridharan et al., 2007; Wu et al., 2007). However, many of these studies focused on crossmodal attention, in which attention to each modality alone cannot be sufficiently separated. For instance, papers that presented visual stimuli to cue auditory attention (Davis et al., 2000; Driver and Spence, 1998; Langner et al., 2011; Macaluso et al., 2003) cannot exclude the effects of visual processing from auditory top-down attention. Along similar

Abbreviations: DAN, Dorsal attention network; MFG, Middle frontal gyrus; IFG, Inferior frontal gyrus; ITG, Inferior temporal gyrus; SPL, Superior parietal lobe; FEF, Frontal eye fields; pMTG, Posterior middle temporal gyrus; VAN, Ventral attention network; fMRI, Functional magnetic resonance imaging; BOLD, Blood oxygenation level dependent; ICA, Independent component analysis; A_p, Attentive phase; P_p, Passive phase.

* Corresponding authors.
E-mail addresses: rodrigo.braga09@imperial.ac.uk (R.M. Braga), r.leech@imperial.ac.uk (R. Leech).

1053-8119
http://dx.doi.org/10.1016/j.neuroimage.2013.02.023

lines, papers that analyzed the period when auditory targets were actually displayed (Linden, 1999; Maeder et al., 2001; Mayer et al., 2006; Shomstein and Yantis, 2006; Sridharan et al., 2007) cannot be said to be looking only at top-down attention, as bottom up and executive networks would be elicited by the presentation of the target. Other papers included an immediate button response to a target (Langner et al., 2011; Mayer et al., 2006; Shomstein and Yantis, 2006) and therefore cannot dissociate the effects of the preparation for and execution of a motor response. These are significant confounds which might evoke DAN activation due to visual or spatial causes. These issues are particularly problematic in studies that use rapid trial times (<5 s; Davis et al., 2000; Hallett et al., 1999; Langner et al., 2011; Macaluso et al., 2003; Maeder et al., 2001; Mayer et al., 2006; Wu et al., 2007; Zatorre et al., 1999), where activations for cues, targets and motor responses are difficult to separate due to the hemodynamic lag. It is therefore hard to say that the previous studies suitably isolated the networks for top-down auditory attention from spatial, crossmodal and executive confounds.

When functional imaging studies have focused on the auditory processing of speech and music, DAN activation is rarely observed. For example, a meta-analysis of 128 language studies showed no activation peaks within SPL and FEF during auditory processing of speech (Vigneau et al., 2011); and see also Cabeza and Nyberg (2000). Similarly, the DAN is not typically observed in studies of music processing (Hickok et al., 2003; Warren, 2008). The neuropsychological evidence also does not support an amodal DAN. Focal parietal lesions which lead to visuospatial neglect (Malhotra et al., 2009) often do not lead to deficits in detecting or identifying sounds, although auditory spatial localization (Pavani et al., 2002) and sustained attention deficits have been reported (Robertson et al., 1997). This suggests that parietal lobe neglect predominantly affects spatial and visual modalities. Hence, although there is compelling evidence for DAN involvement in top-down visuospatial attention, the evidence that the full SPL–FEF–MFG axis is necessary for auditory attention is inconclusive.

Materials and methods

We used functional magnetic resonance imaging (fMRI) to identify networks active during auditory top-down attention in the absence of visual or spatial requirements. A simple non-spatial auditory search task was used (see Fig. 1). Subjects listened to complex natural background sounds and were instructed to listen out for a pitch change that occurred within a pre-trained target sound. The presence of a target divided each trial into three phases: (1) an extended active listening phase (A_p), where subjects listened to the background auditory scenes in order to detect the target sound; (2) a target phase, during which subjects were required to listen to the target and identify whether it contained a pitch change; and (3) a post-target passive listening phase (P_p), where subjects heard the background sounds but had no requirement to listen attentively. Once subjects identified a target they were aware that there was no requirement to listen attentively. We compared the neural activity before and after the target ($A_p > P_p$) to isolate top-down auditory attention. This was anticipated to be high in the attentive listening phase (A_p), when subjects were actively awaiting the target, and lower in the passive listening phase (P_p) after the target. The auditory input during A_p and P_p was equivalent. Importantly, activity associated with motor responses did not affect the critical contrast between A_p and P_p, as the response occurred after each trial. Further, the decision about whether a pitch change had occurred, which could evoke implicit or preparatory motor control, occurred during the elongated target period, and so was isolated from the active or passive listening phases.

Extended trial (40 s) and target (10 s) durations were necessary to allow the activity associated with attentional state and target detection to be clearly separated. Longer conditions such as these can result in reduced signal in standard univariate contrast analyses (e.g. Visscher et al., 2003) due to the attenuation of the repeated neural signal and the transient pattern of activation associated with attentional reorienting occurring within each condition block. Multivariate techniques such as Independent Component Analysis (ICA) are able to decompose the BOLD signal into multiple different components, and therefore isolate different sources of variation in the data that may obscure task-evoked signal over extended durations. ICA is therefore more suited to the current elongated design.

We also recreated the experimental conditions and attentional requirements in an analogous visuospatial search task in a different sample of subjects to confirm the prediction that visual top-down attention would evoke the activation of the DAN, including FEF, SPL and MFG. We hypothesized that the auditory task would activate a top-down attention network that was distinct from the DAN, whereas the full DAN would be activated during the visual task.

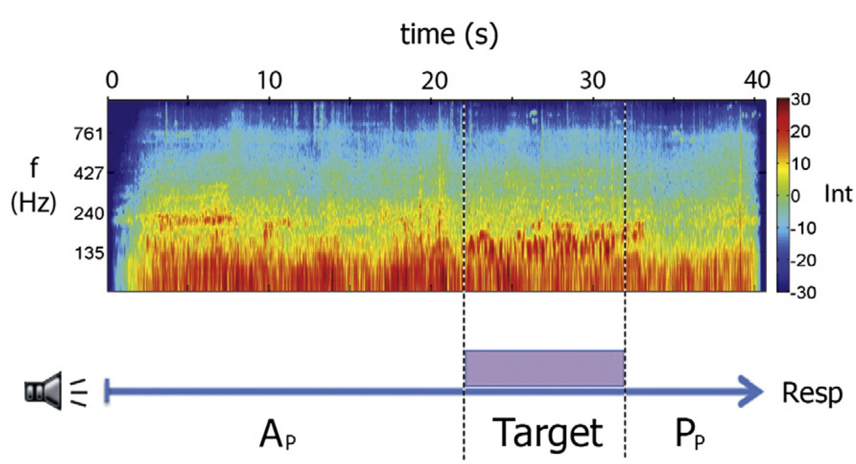

Fig. 1. Auditory search task design. Background sounds (spectrogram and blue arrow) were divided into attentive (A_p) and passive (P_p) listening phases by the presence of a 10 s target foreground sound. The extended trial duration (40 s) allowed the attentional state during A_p and P_p to be clearly separated from target and button response (Resp) evoked activations. The auditory input was equivalent during Ap and P_p. Int: Intensity, f: frequency.