Brain & Language 108 (2009) 89–96

Contents lists available at ScienceDirect

Brain & Language

journal homepage: www.elsevier.com/locate/b&l

Dissociating verbal and nonverbal audiovisual object processing

Julia Hocking [a,b,*], Cathy J. Price [a]

[a] Wellcome Trust Centre for Neuroimaging, UCL, London, UK
[b] The University of Queensland, Centre for Magnetic Resonance, Brisbane, Australia

ARTICLE INFO

Article history:
Accepted 9 October 2008
Available online 19 December 2008

Keywords:
Audiovisual
Integration
Verbal
Nonverbal
Semantic
Conceptual
Phonological
Amodal

ABSTRACT

This fMRI study investigates how audiovisual integration differs for verbal stimuli that can be matched at a phonological level and nonverbal stimuli that can be matched at a semantic level. Subjects were presented simultaneously with one visual and one auditory stimulus and were instructed to decide whether these stimuli referred to the same object or not. Verbal stimuli were simultaneously presented spoken and written object names, and nonverbal stimuli were photographs of objects simultaneously presented with naturally occurring object sounds. Stimulus differences were controlled by including two further conditions that paired photographs of objects with spoken words and object sounds with written words. Verbal matching, relative to all other conditions, increased activation in a region of the left superior temporal sulcus that has previously been associated with phonological processing. Nonverbal matching, relative to all other conditions, increased activation in a right fusiform region that has previously been associated with structural and conceptual object processing. Thus, we demonstrate how brain activation for audiovisual integration depends on the verbal content of the stimuli, even when stimulus and task processing differences are controlled.

1. Introduction

Previous functional imaging studies investigating audiovisual object processing have used either verbal (Bernstein, Auer, Wagner, & Ponton, 2008; Calvert, Campbell, & Brammer, 2000; Kreifelts, Ethofer, Grodd, Erb, & Wildgruber, 2007; Macaluso, George, Dolan, Spence, & Driver, 2004; Raij, Uutela, & Hari, 2000; Sekiyama, Kanno, Miura, & Sugita, 2003; Skipper, Nusbaum, & Small, 2005; Skipper, van Wassenhove, Nusbaum, & Small, 2007; van Atteveldt, Formisano, Blomert, & Goebel, 2007; van Atteveldt, Formisano, Goebel, & Blomert, 2004; Wright, Pelphrey, Allison, McKeown, & McCarthy, 2003) or nonverbal (Beauchamp, Lee, Argall, & Martin, 2004; Sestieri et al., 2006; Taylor, Moss, Stamatakis, & Tyler, 2006) stimuli. By verbal versus nonverbal, we refer to the presence or absence of word stimuli—whether written, spoken or lip-read. The focus of the present paper is on how neuronal activation for audiovisual processing differs for verbal versus nonverbal conceptual stimuli.

Our predictions are based on the following rationale. Verbal and nonverbal conceptual stimuli can access both phonological and semantic processes; however they do so in different ways. For verbal stimuli, phonetic analysis of the input is required before recognition at the semantic level (e.g. Indefrey and Levelt, 2004). By contrast, for nonverbal stimuli, semantic processing is required before phonological retrieval (e.g. Glaser & Glaser, 1989; Seifert, 1997). Consequently, audiovisual matching of two verbal stimuli (i.e. auditory and visual words) can occur at the level of phonology prior to explicit retrieval of semantics whereas audiovisual matching of two nonverbal stimuli (e.g. pictures and sounds of objects) can occur at the level of semantics without explicit retrieval of phonology. This leads us to predict that activation in phonological processing areas may be higher for matching verbal than nonverbal stimuli whereas activation in semantic processing areas may be higher for matching nonverbal than verbal stimuli. In addition, reports of brain damaged patients with selective deficits in either verbal or nonverbal stimuli have suggested that there may be verbal and nonverbal dissociations within the semantic system. Depending on the theoretical perspective taken, this dissociation has been proposed at the level of (i) separate visual and verbal semantic systems (e.g. Ferreira, Giusiano, Ceccaldi, & Poncet, 1997; Warrington, 1975; Warrington & McCarthy, 1994), (ii) a shared distributed semantic system differentiated by the type of knowledge primarily involved during acquisition (e.g. Saffran, Coslett, & Keener, 2003) or (iii) differences at the level that verbal and nonverbal stimuli access a shared semantic system.

Verbal (word) stimuli can either be presented in the form of continuous speech (as in Calvert et al., 2000; Macaluso et al., 2004) or in the form of single words (Ojanen et al., 2005; Olson, Gatenby, & Gore, 2002; Raij et al., 2000; van Atteveldt et al., 2004; van Atteveldt et al., 2007). The verbal stimuli used in this

* Corresponding author. Address: The University of Queensland, Centre for Magnetic Resonance, Brisbane 4072, Australia. Fax: +617 3365 3833.
E-mail address: julia.hocking@cmr.uq.edu.au (J. Hocking).
URL: http://www.fmrilab.net (J. Hocking).

0093-934X/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.bandl.2008.10.005

experiment were written and spoken object names because this permitted a controlled comparison to nonverbal audiovisual stimuli in the form of pictures of objects and the naturally occurring sounds associated with objects. The task was held constant and involved deciding if a visually presented stimulus referred to the same object as a simultaneously presented auditory stimulus. If verbal versus nonverbal audiovisual object matching differ in the relative demands they place on phonological and semantic processing (see above) then we would expect to see fMRI activation differences in areas that have previously been associated with phonological and semantic processing. In contrast if verbal and nonverbal audiovisual object matching depend on different types of semantic processing then we need to consider the results of previous studies that compared verbal and nonverbal semantic/conceptual processing. Below, we briefly review the relevant literature on verbal and nonverbal processing and the influence of these findings on our anatomical expectations.

1.1. Phonological versus semantic processing

Functional imaging studies comparing phonological to semantic processing have associated phonological processing with the left superior temporal sulcus (Binder, 2000; Noppeney, Josephs, Hocking, Price, & Friston, 2008; Scott, Blank, Rosen, & Wise, 2000; Wise et al., 2001), the left supramarginal gyrus and left posterior inferior frontal regions (Booth et al., 2006; Demonet et al., 1992; Devlin, Matthews, & Rushworth, 2003; Gold, Balota, Kirchhoff, & Buckner, 2005; Mummery, Shallice, & Price, 1999; Paulesu, Frith, & Frackowiak, 1993; Price, Mummery, Moore, Frakowiak, & Friston, 1999; Roskies, Fiez, Balota, Raichle, & Petersen, 2001). We therefore predicted that activation in one or more of these regions would be higher for audiovisual matching of verbal relative to nonverbal conceptual stimuli. In contrast, semantic processing has been associated with the left middle temporal gyrus (Binder et al., 1997), left anterior temporal lobe (Scott et al., 2000; Vandenberghe, Price, Wise, Josephs, & Frackowiak, 1996), the angular gyri (Devlin et al., 2003; Mummery et al., 1999; Noppeney et al., 2008) and ventral and anterior frontal regions (Binder et al., 1997; Booth et al., 2006; Devlin et al., 2003; Gold et al., 2005; Poldrack et al., 1999; Roskies et al., 2001). Activation in one or more of these regions was therefore predicted to be higher for audiovisual matching of nonverbal than verbal conceptual stimuli. For a review of phonological and semantic areas, see Vigneau et al. (2006).

1.2. Verbal versus nonverbal semantics

Studies of brain damaged patients have suggested that the left hemisphere may be more engaged in accessing verbal information while the right hemisphere may be more engaged in accessing nonverbal information (Coslett & Saffran, 1992; for reviews see Caramazza, Hillis, Rapp, & Romani, 1990 and Lambon-Ralph et al., 1999). Unfortunately, most of the evidence from patient data comes from a comparison of verbal/nonverbal processing in the visual modality only. Therefore, the conclusions concerning amodal or separable verbal/nonverbal systems are limited.

Recently, functional neuroimaging of normal subjects has provided another source of evidence for a dissociation between verbal and nonverbal processing within either the auditory modality or the visual modality (Adams & Janata, 2002; Bright, Moss, & Tyler, 2004; Chee et al., 2000; Dick et al., 2007; Giraud & Price, 2001; Humphries, Willard, Buchsbaum, & Hickok, 2001; Perani et al., 1999; Thierry, Giraud, & Price, 2003; Thierry & Price, 2006; Vandenberghe et al., 1996; von Kriegstein, Eger, Kleinschmidt, & Giraud, 2003). Critically, however, the areas associated with verbal and nonverbal stimuli differ according to the input modality (visual or auditory). These between-modality differences are difficult to interpret because they are confounded by perceptual differences in the nature of the verbal and nonverbal stimuli. To circumvent perceptual confounds, a study by Thierry and Price (2006) looked for verbal versus nonverbal processing differences that were independent of stimulus modality. Combining data from one experiment using auditory stimuli and another using the corresponding visual stimuli, they reported a left/right double dissociation for verbal/nonverbal material, independent of sensory modality. Specifically, verbal relative to nonverbal material activated anterior and posterior regions of the left superior temporal sulcus and the ventral left inferior frontal gyrus, while nonverbal relative to verbal material activated the right mid fusiform gyrus and right posterior middle temporal gyrus.

The anatomical dissociation reported in Thierry and Price (2006) provides hypotheses for the current experiment. However, it should still be noted that the functional level at which the verbal versus nonverbal differences arise in Thierry and Price (2006) is debatable. For example, the right posterior superior temporal region associated with nonverbal conceptual processing in Thierry and Price (2006) has been associated with spatial localisation in both the auditory (Rauschecker, 1998a; Rauschecker, 1998b; Rauschecker & Tian, 2000) and visual (Milner & Goodale, 1993) domains. Conversely, the auditory and visual verbal stimuli used in Thierry and Price (2006) had a sentence like structure which may have evoked morpho-syntactic associations compared to nonverbal stimuli. Indeed, the left anterior superior temporal cortex that was activated for verbal relative to nonverbal conditions in Thierry and Price (2006) has previously been associated with morpho-syntactic processing (Bornkessel, Zysset, Friederici, von Cramon, & Schlesewsky, 2005; Dronkers, 2000; Dronkers and Ogar, 2004; Friederici & Kotz, 2003; Friederici, Ruschemeyer, Hahne, & Fiebach, 2003; Humphries, Love, Swinney, & Hickok, 2005; Stowe et al., 1999; Vandenberghe, Nobre, & Price, 2002).

In summary, the present study contrasts the effects of matching verbal versus nonverbal simultaneously presented audiovisual pairs by manipulating the type of input material. Purely verbal audiovisual stimuli were simultaneously presented spoken and written object names, purely nonverbal stimuli were photographs of objects simultaneously presented with naturally occurring object sounds. Perceptual differences between verbal and nonverbal stimuli were controlled by including audiovisual conditions that presented one verbal and one nonverbal stimulus (spoken names with photographs or written names with object sounds). The predictions were that, verbal stimuli would increase activation in left hemisphere areas associated with phonological processing whereas nonverbal stimuli would increase activation in semantic processing areas (possibly in the right hemisphere).

2. Materials and methods

2.1. Subjects

Eighteen subjects participated in this Experiment (12 women, 6 men, age range 20–36 years, mean age 26). All were right handed native English speakers with normal or corrected to normal vision and gave informed consent to take part. All had normal neurological and audiological status. The study was approved by the joint ethics committee of the Institute of Neurology and University College London Hospital, London, UK.

2.2. Experimental design

Subjects were presented bimodally with two simultaneously presented conceptual stimuli, one in the visual modality (colour photograph or written object name) and one in the auditory