NeuroImage 52 (2010) 1592–1602

Contents lists available at ScienceDirect

NeuroImage

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

Prior auditory information shapes visual category-selectivity in ventral occipito-temporal cortex

Ruth Adam *, Uta Noppeney

Max Planck Institute for Biological Cybernetics, Spemannstr. 41, 72076 Tuebingen, Germany

A R T I C L E I N F O

Article history:
Received 2 December 2009
Revised 28 April 2010
Accepted 3 May 2010
Available online 7 May 2010

Keywords:
Multisensory integration
Congruency
Categorization
Selective attention
fMRI

A B S T R A C T

Objects in our natural environment generate signals in multiple sensory modalities. This fMRI study investigated the influence of prior task-irrelevant auditory information on visually-evoked category-selective activations in the ventral occipito-temporal cortex. Subjects categorized pictures as landmarks or animal faces, while ignoring the preceding congruent or incongruent sound. Behaviorally, subjects responded slower to incongruent than congruent stimuli. At the neural level, the lateral and medial prefrontal cortices showed increased activations for incongruent relative to congruent stimuli consistent with their role in response selection. In contrast, the parahippocampal gyri combined visual and auditory information additively: activation was greater for visual landmarks than animal faces and landmark-related sounds than animal vocalizations resulting in increased parahippocampal selectivity for congruent audiovisual landmarks. Effective connectivity analyses showed that this amplification of visual landmark-selectivity was mediated by increased negative coupling of the parahippocampal gyrus with the superior temporal sulcus for congruent stimuli. Thus, task-irrelevant auditory information influences visual object categorization at two stages. In the ventral occipito-temporal cortex auditory and visual category information are combined additively to sharpen visual category-selective responses. In the left inferior frontal sulcus, as indexed by a significant incongruency effect, visual and auditory category information are integrated interactively for response selection.

Introduction

Visual object recognition relies on category-selective activations in the ventral occipito-temporal cortices. Most notably, faces and to some extent animal faces (Kanwisher et al., 1997; Kanwisher et al., 1999; Maguire et al., 2001), induce activations in the fusiform gyri (fusiform face area, FFA), while houses, landmarks and scenes are associated with selective responses in the parahippocampal gyri (parahippocampal place area, PPA) (Epstein and Kanwisher, 1998; Maguire et al., 2001). However, in our natural environment, objects commonly emit signals in multiple sensory modalities. For instance, we may see and hear a dog barking. Even landmarks that do not genuinely produce sounds are often associated with characteristic sounds such as the view of a church with the chime of its bell. The human brain is thus challenged to integrate semantic object information from multiple senses to more reliably infer an object's category. This raises the question whether category-selective activations that are commonly observed for visual object stimuli can be influenced by concurrent or prior auditory object input. While numerous studies have investigated category-selective activations in the visual (Epstein and Kanwisher, 1998; Kanwisher et al., 1997; Noppeney et al., 2006; Pitcher et al., 2009; Rhodes et al., 2004; Rotshtein et al., 2005; Tootell et al., 2008) or auditory (Belin et al., 2000; Lewis et al., 2004; Staeren et al., 2009) domains alone, only few studies have focused on influences of category-selective information across sensory modalities (for review on multisensory object processing see Amedi et al., 2005). For instance, familiar voices have been shown to elicit activations not only in temporal voice recognition areas but also in the fusiform face area (von Kriegstein et al., 2005). Furthermore, after being trained with 'face–voice' or 'cellphone–ringing tone' pairs, subjects showed increased activation in the fusiform face area for voices (when presented alone) but not enhanced activation in visual object-selective areas for ringing tones (when presented alone) (von Kriegstein and Giraud, 2006). Yet, in a recent study (Engel et al., 2009), mechanical sounds induced parahippocampal activations without any prior training. Thus, it is currently unclear under which circumstances sounds elicit category-selective activations in the ventral occipito-temporal cortex. In particular, since those studies presented sounds alone and hence, in the focus of attention, the observed activations are susceptible to task-induced strategic effects such as imagery or action simulations. Indeed, voice-induced activations in the fusiform face area were observed primarily during speaker rather than speech recognition (von Kriegstein et al., 2005), suggesting that crossmodal activations

* Corresponding author. Fax: +49 7071 601 616.
E-mail address: ruth.adam@tuebingen.mpg.de (R. Adam).

1053-8119/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2010.05.002

may emerge to facilitate specific task requirements. In contrast, evidence for a more automatic spread of multisensory semantic activations comes from intersensory selective attention paradigms showing that object processing in the task-relevant sensory modality co-activates object representations or associated features in the task-irrelevant sensory modality (Molholm et al., 2007; Murray et al., 2004). Furthermore, audiovisual semantic (in)congruency manipulations demonstrated that task-irrelevant sensory stimuli can modify activations in another sensory modality as a function of semantic congruency (Laurienti et al., 2003; Molholm et al., 2004; Noppeney et al., 2008). For instance, in an intersensory auditory selective attention paradigm, prior incongruent visual information increased activation in the auditory processing stream relative to congruent information (Noppeney et al., 2008). Similarly, in an intersensory visual selective attention paradigm (i.e. presentation of stimuli in visual and auditory modalities with only the visual modality being task-relevant), activations in the ventral occipito-temporal cortex for object pictures (e.g. a cow) were increased in the presence of a semantically conflicting source sounds (e.g. the sound of an alarm clock) (Laurienti et al., 2003).

These (in)congruency effects demonstrate that ventral occipito-temporal 'object' areas are sensitive to auditory inputs and their semantic content even when it is task-irrelevant. They converge with previous studies on response conflict or incongruency in the visual modality alone showing that unattended words induce response amplification in the fusiform gyrus for face recognition (Egner and Hirsch, 2005). However, incongruency effects (as in Egner and Hirsch, 2005; Laurienti et al., 2003; Noppeney et al., 2008) may rely on different neural mechanisms than visual co-activations elicited by auditory stimuli when presented alone (Engel et al., 2009; von Kriegstein and Giraud, 2006). For instance, since incongruent stimulus pairs violate our expectations based on our life-long exposure to natural statistics, they may induce error detection mechanisms. Thus, incongruency effects may actually serve as a prediction error signal (Friston, 2009; Rao and Ballard, 1999). Importantly, in contrast to crossmodal co-activations observed when presented with inputs from one sensory modality alone, congruency effects require that the influence of an auditory object stimulus depends on and interacts with the category of the visual stimulus. Hence, sensory (e.g. auditory) input may influence processing of input from another sensory (e.g. visual) modality using two distinct neurobiological mechanisms. In the additive case, auditory category input induces co-activations in the corresponding visual category-selective areas irrespective of the category of the concurrent visual input resulting in additive effects for visual and auditory category. For instance, a sound associated with a certain landmark will co-activate the parahippocampal place area irrespective of whether it is presented alone, together with a congruent landmark or an incongruent face. In this additive case, the parahippocampal place area would show additive effects of visual category (i.e. landmarks>face pictures) and auditory object category (i.e. landmark sounds>voices) resulting in a stepwise activation profile (see Fig. 1A, additive case). These additive effects do not depend on the relationship (i.e. congruent or incongruent) of visual and auditory object category stimuli. In the interactive case, auditory category input induces activations in the corresponding visual category-selective areas depending on the semantic category of the visual input. For instance, a particular landmark sound may induce an activation increase when paired with a face picture, but an activation decrease when paired with its corresponding landmark picture (or vice versa) resulting in a crossover interaction between visual and auditory factors (n.b. as a consequence, the auditory stimulus may induce no activation increase when averaged across congruent and incongruent trials). This crossover interaction is then formally identical to (in)congruency effects (e.g. an activation increase in the incongruent and a decrease in the congruent case, see Fig. 1A, interactive case).

To summarize, object representations from one sensory modality can influence object processing in another sensory modality via two distinct neurobiological mechanisms, (i) additive co-activations that do not depend on the relationship of the auditory and visual signals and (ii) interactive effects that are determined by the (in)congruency relationship of the two signals.

To our knowledge, none of the previous studies was able to dissociate additive and interactive effects of visual and auditory category, since they used categories that induced responses in mostly overlapping brain areas (e.g. tools and musical instruments). Thus, these studies were only able to characterize interactive (= (in)congruency) effects of visual and auditory category input but could not reveal additive mechanisms of AV interplay, which crucially require a difference in net activation across different categories.

To reveal both additive (i.e. crossmodal co-activations) and interactive (i.e. congruency effects) influences of task-irrelevant semantic sounds on visual processing in the ventral occipito-temporal cortex, the current experiment exploited the fact that faces and landmarks are associated with spatially selective activations in the ventral occipito-temporal cortex (Eger et al., 2005; Epstein and Kanwisher, 1998; Grill-Spector et al., 2004; Kanwisher et al., 1997; Kanwisher et al., 1999; Maguire et al., 2001; Rotshtein et al., 2007). Subjects were presented with pictures of landmarks or animal faces that were preceded by animal vocalizations or landmark sounds. Thus, the experiment factorially manipulated visual (landmark vs. animal face) and auditory (landmark sound vs. animal vocalization) category. In an intersensory visual selective attention paradigm, subjects categorized the pictures as landmarks or animal faces while ignoring the prior congruent or incongruent sounds. This 2×2 factorial design enabled us to dissociate additive and interactive effects of task-irrelevant semantic source sounds on activations in the ventral occipito-temporal cortex and higher order prefrontal cortices. In particular, we investigated whether activations in the fusiform and parahippocampal gyri were influenced by the category of the auditory stimuli in an additive or interactive fashion. In the additive case, the auditory stimulus (e.g. a 'landmark sound') would increase the activation in the corresponding category-selective area within the ventral occipito-temporal cortex (e.g. parahippocampal gyrus) irrespective of the category of the visual stimulus. In the interactive case, the category-selective effect of the auditory stimulus depends on the category of the visual stimulus resulting in activation differences for incongruent relative to congruent stimulus combinations (=(in) congruency effect). Based on previous studies of conflict within and between the senses, we also expected incongruency (=interactive) in the lateral (i.e. inferior frontal sulcus (IFS)) and medial prefrontal cortex (mPFC/anterior cingulate (AC)) (Botvinick et al., 2001; Brown and Braver, 2005; Duncan and Owen, 2000; Hein et al., 2007; Kerns et al., 2004; Noppeney et al., 2008; Noppeney et al., 2010; Paus, 2001). Effective connectivity analyses (i.e. psychophysiologic interaction) were used to further characterize interactions amongst brain regions that mediate the influence of auditory category information on visual object categorization.

Materials and methods

Subjects

28 healthy right-handed students (11 females, mean age 27.2 years, range 21–37 years) participated in this fMRI study. All subjects had normal or corrected to normal vision and reported normal hearing. Subjects gave informed written consent prior to the study which was approved by the joint human research review committee of the Max Planck Society and the University of Tübingen. Data from two female subjects were excluded from the study because of poor performance.