Exp Brain Res (2005) 166: 559–571
DOI 10.1007/s00221-005-2396-5

REVIEW

A. Amedi · K. von Kriegstein · N. M. van Atteveldt
M. S. Beauchamp · M. J. Naumer

Functional imaging of human crossmodal identification and object recognition

Received: 29 July 2004 / Accepted: 12 November 2004 / Published online: 19 July 2005

Abstract The perception of objects is a cognitive function of prime importance. In everyday life, object perception benefits from the coordinated interplay of vision, audition, and touch. The different sensory modalities provide both complementary and redundant information about objects, which may improve recognition speed and accuracy in many circumstances. We review crossmodal studies of object recognition in humans that mainly employed functional magnetic resonance imaging (fMRI). These studies show that visual, tactile, and auditory information about objects can activate cortical association areas that were once believed to be modality-specific. Processing converges either in multisensory zones or via direct crossmodal interaction of modality-specific cortices without relay through multisensory regions. We integrate these findings with existing theories about semantic processing and propose a general mechanism for crossmodal object recognition: The recruitment and location of multisensory convergence zones varies depending on the information content and the dominant modality.

Keywords Object recognition · Crossmodal · Audio-visual · Visuo-tactile · Multisensory · Functional magnetic resonance imaging (fMRI)

A. Amedi, K. von Kriegstein, N. M. van Atteveldt, M. S. Beauchamp and M. J. Naumeri contributed equally to this work

A. Amedi
Laboratory for Magnetic Brain Stimulation,
Department of Neurology, Beth Israel Deaconess Medical Center,
Harvard Medical School, Boston, MA, USA

K. von Kriegstein
Cognitive Neurology Unit, Johann-Wolfgang-Goethe University,
Frankfurt/Main, Germany

K. von Kriegstein · M. J. Naumer
Brain Imaging Center (BIC), Frankfurt/Main, Germany

N. M. van Atteveldt
Department of Cognitive Neuroscience, Faculty of Psychology,
University of Maastricht, Maastricht, The Netherlands

M. S. Beauchamp
Laboratory of Brain and Cognition,
National Institute of Mental Health, Bethesda, MD, USA

M. J. Naumer
Department of Neurophysiology,
Max Planck Institute for Brain Research, Frankfurt/Main,
Germany

M. J. Naumer (✉)
Institute of Medical Psychology, Frankfurt Medical School,
Heinrich-Hoffmann-Strasse 10, 60528 Frankfurt/Main, Germany
E-mail: m.j.naumer@med.uni-frankfurt.de
Tel.: +49-69-63016581
Fax: +49-69-63017606

Introduction

We experience our environment via several sensory modalities at the same time. For example, in an opera or a movie we perceive visual and sound information in parallel. The information provided by these diverse sensory systems is synthesized in our brains to create the coherent and unified percepts we are so familiar with. The multisensory nature of our perceptions has several behavioral advantages—for example, speeded response and improved recognition in noisy environments (e.g. Newell 2004). Until recently, the neural correlates of crossmodal integration and its development have been investigated mainly using (invasive) electrophysiology in a variety of animal species and brain structures (Stein and Meredith 1993; Wallace et al. 2004a). The advent of non-invasive functional neuroimaging techniques led to investigations of crossmodal processes important for human cognition, for example linguistic processing, person identification, and object categorization.

In this paper, we mainly review fMRI investigations of human crossmodal object recognition. We define objects as "something material that may be perceived by the senses" (Merriam-Webster online dictionary). We use this definition in a broader sense by including not only concrete objects such as vehicles, tools, and persons, but also more abstract objects such as letters or

speech with its accompanying lip movements. Because these objects have characteristic attributes in multiple sensory modalities, recognition might strongly benefit from the coordinated interplay of vision, audition, and touch. We will focus on studies of audio-visual (AV) and visuo-tactile (VT) object recognition that have recently produced consistent findings. After a brief introduction to unimodal visual, auditory, and tactile object recognition, we will give an overview of studies including AV and VT processing. In the discussion, we will specifically focus on determinants of crossmodal convergence, on the role of imagery, and, furthermore, describe a model of distributed semantic processing.

Unimodal object recognition

Visual object recognition

On the basis of the vast amount of anatomical data in monkey visual cortex, the picture of a highly diverse and hierarchically structured system has emerged (Felleman and Van Essen 1991). There is a widely accepted notion of two processing streams between early visual areas in the occipital cortex and higher-order processing centers in the temporal and parietal lobes, referred to as the dorsal and ventral streams, respectively (Ungerleider and Haxby 1994). Because the dorsal stream is also involved in visuomotor transformations, Goodale et al. (1994) proposed distinguishing between *vision for action* (dorsal stream) and *vision for perception* (ventral stream). In this review, we will focus on object perception and thus discuss the ventral stream in more detail.

Because the ventral stream contains structures devoted to the fine-detailed analysis of a visual scene such as form and color, it is also known as the "what" pathway. It consists of areas V1–V4 in the occipital lobe and several regions that belong to the lateral and ventral temporal lobe (blue regions in Fig. 1). Electrophysiological studies in non-human primates have identified organizing principles, for example retinotopy, selectivity for simple features like spatial orientation in V1, and selectivity for categories of complex stimuli such as faces or spatial layouts in inferior temporal (IT) cortex. These findings have been confirmed non-invasively in human and non-human primates using fMRI (Tootell et al. 2003; Fig. 1). An active debate is ongoing about the actual organization of the ventral stream (Grill-Spector and Malach 2004; Tootell et al. 2003). One area in the lateral occipital cortex, the lateral occipital complex (LOC; Malach et al. 1995) has also been the subject of intense investigation. LOC responds strongly to pictures of intact objects compared with scrambled objects. In ventral temporal cortex, specialized areas for faces (Kanwisher et al. 1997), scenes (Epstein and Kanwisher 1998), human body parts (Downing et al. 2001; Fig. 1), letters (Gauthier et al. 2000; Polk et al. 2002), and visual word forms (McCandliss et al. 2003) have been described. Developing a theoretical framework for understanding these specialized regions seems problematic, however, and the notion of widely distributed and overlapping cortical object representations remains a possible principle of organization (Haxby et al. 2001). Effects of perceptual expertise for certain object categories (Gauthier et al. 2000) and, more recently, different category-related resolution needs (Hasson et al. 2003) have also been proposed to explain the organization of the human ventral stream.

Auditory object recognition

Auditory objects are complex spectro-temporal sounds that can be either associated with a visual object (e.g. a certain string sound with a cello or a specific voice with a specific face) or can be completely independent (e.g. a melody). By analogy with the visual system, it is thought that auditory "what" information is processed in a ventral as opposed to the dorsal "how/where" stream (Arnott et al. 2004; Belin and Zatorre 2000; Kaas and Hackett 1999; Rauschecker and Tian 2000; Romanski et al. 1999). These processing streams are thought to

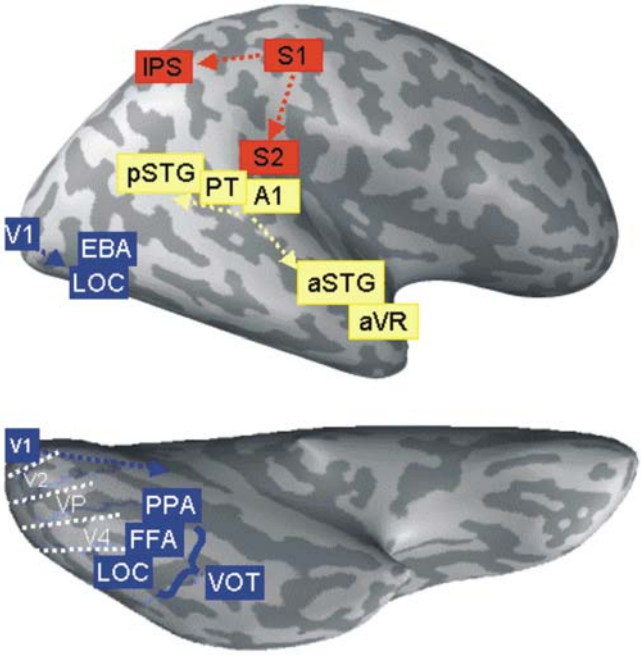

Fig. 1 Human cortical "what" pathways for vision, audition, and touch. Prominent cortical areas/regions that belong to the visual, auditory, and tactile object recognition pathways are shown in *blue*, *yellow*, and *red*, respectively. Their relative locations are indicated on inflated representations of the left (LH) and right (RH) cerebral hemispheres in ventral (LH) and lateral (RH) views, respectively. (Abbreviations: *V1, V2, VP*, and *V4* retinotopic visual areas, *LOC* lateral occipital complex, *FFA* fusiform face area, *PPA* parahippocampal place area, *VOT* ventral occipito-temporal cortex, *EBA* extrastriate body area, *A*1 tonotopic primary auditory cortex, *PT* planum temporale, *aSTG* anterior superior temporal gyrus, *aVR* anterior voice region, *p*STG posterior STG, *SI– SII* somatotopic primary and secondary somatosensory cortices, *IPS* intra-parietal sulcus)

originate at least in part from the planum temporale (PT; Fig. 1), which responds to any kind of complex spectro-temporal sounds (Griffiths and Warren 2002). Auditory processing may be organized in a specialized fashion with regard to different categories of auditory stimulus. Evidence for such organization comes from behavioral and lesion data including cases of auditory agnosia specific to environmental sounds, music, words, and voices (Polster and Rose 1998). Recent functional imaging data show that anterior auditory regions (anterior superior temporal sulcus and gyrus (aSTS/STG); Fig. 1) respond during the discrimination and identification of complex auditory forms (Binder et al. 2004; Zatorre et al. 2004). Furthermore, a region in the aSTS (aVR; Fig. 1) is activated specifically by speaker identity (Belin and Zatorre 2003; von Kriegstein et al. 2003) and thus may be the auditory equivalent to the fusiform face area (FFA; Kanwisher et al. 1997) in visual face recognition. In contrast, stimulation with environmental sounds has been shown to activate posterior regions in the STS and the medial temporal gyrus (MTG; Lewis et al. 2004), which might reflect semantic processing and/or temporal integration (Thierry et al. 2003). Traditionally, the left posterior STG and, more specifically, the PT have been associated with speech perception (Wernicke 1874). Some investigations have suggested that this region is involved in the analysis of rapidly changing acoustic cues that are especially important, but not specific, for speech perception (Jäncke et al. 2002).

Tactile object recognition

Tactile information processing seems to be hierarchically organized demonstrating growing complexity as one moves from area 3b in the primary somatosensory cortex (SI; see also Fig. 1) towards areas located more posterior along the post-central gyrus and sulcus. In non-human primates, there is increasing receptive field complexity as one moves along the anterior-posterior axis of the post-central gyrus (from area 3b, via areas 1 and 2, to areas 5 and 7 that are located around the anterior intraparietal sulcus (IPS)). In humans, there is evidence for a similar functional hierarchy as one moves from areas 3b and 1 (activated by any tactile input) to area 2 (showing selectivity to the attributes of objects such as curvature) and to anterior IPS (showing preference to overall shape compared to shape primitives such as curvature; Bodegard et al. 2001). Activation in IPS was also found in a variety of tasks requiring analysis of object shape, length, and curvature using both simple geometrical shapes (Bodegard et al. 2001; O'Sullivan et al. 1994; Roland et al. 1998) and natural everyday objects (Amedi et al. 2001, 2002; Deibert et al. 1999; Reed et al. 2004). Thus, although it has been less studied than the ventral visual stream, an analogous pathway for tactile object recognition might also exist in primates. It also suggests similarities with the hierarchical organization evident in vision and audition. In general, however, haptic object recognition can be viewed as a more serial process than visual or even auditory object recognition (i.e. in active palpation one needs to integrate the tactile input over time, sampling the different parts of an object). Still, this functional hierarchy is not yet supported by as much anatomical evidence as in the visual system (Felleman and Van Essen 1991).

The possible role of SII in tactile object recognition and the possibility of alternative pathways (a ventro-lateral pathway stretching from SI to inferior parietal regions including SII and the insula; Mishkin 1979; Reed et al. 2004) are still under debate. Early work suggested that there is a division of labor between processing of micro-geometry aspects of the stimulus, for example surface texture properties, in the parietal operculum (SII) and macro-geometry properties, such as shape attributes of objects, in the anterior IPS (O'Sullivan et al. 1994; Roland et al. 1998). This view was challenged recently both for humans (Reed et al. 2004—showing object preference in SII and the insula for tactile object recognition versus palpation of nonsense objects) and in non-human primates (for a review, see Iwamura 1998). In addition, previous studies have shown that lesions to SII and related structures give rise to tactile agnosia (e.g. Reed and Caselli 1994).

Audio-visual object recognition

In everyday object recognition the coordinated interplay of two or more sensory systems is the rule rather than the exception (as illustrated by the influence of lip reading in speech perception or by crossmodal priming in person recognition). However, relatively little is known about multisensory integration in human cerebral cortex. Given the distributed organization of the visual and auditory systems, how do these systems interact? We do not perceive the sight and the sound of a musical instrument as two different events. But what are the underlying neural mechanisms that accomplish this crossmodal sharing and union of information? In the following section, we will describe studies investigating possible neural mechanisms underlying the perception of different categories of AV information such as speech, persons, and common objects.

Linguistic information

Mouth movements of a speaker form a natural visual component of speech perception. Movements and phonetic information are inherently linked by characteristics such as temporal onset, duration, and frequency-amplitude information (Calvert 2001; Munhall and Tohkura 1998). Correspondences between speech sounds and mouth movements (often referred to as "AV speech") are learned implicitly and early in development

by exposure to heard speech together with the sight of the speaker (Kuhl and Meltzoff 1982). In contrast, the visual representation of spoken language by written language is an artificial cultural invention. That is to say, letters and speech sounds exhibit no naturally corresponding characteristics but are related in a completely arbitrary way. Therefore, associations between letters and speech sounds are not learned automatically, but require explicit instruction (Gleitman and Rozin 1977; Liberman 1992).

Viewing the face of a speaker during listening enhances speech perception, especially under noisy conditions (Sumby and Pollack 1954). Lip reading can also change the speech percept, as shown by the McGurk illusion, in which the auditory speech percept is altered by non-corresponding visually perceived speech information (McGurk and MacDonald 1976). A possible neural basis for these perceptual effects is the activation of auditory processing regions by silent lip reading, as is reported by several fMRI studies (Bernstein et al. 2002; Calvert et al. 1997; Olson et al. 2002; Paulesu et al. 2003; Sekiyama et al. 2003). Consistently reported auditory regions include the STS, STG, and auditory association

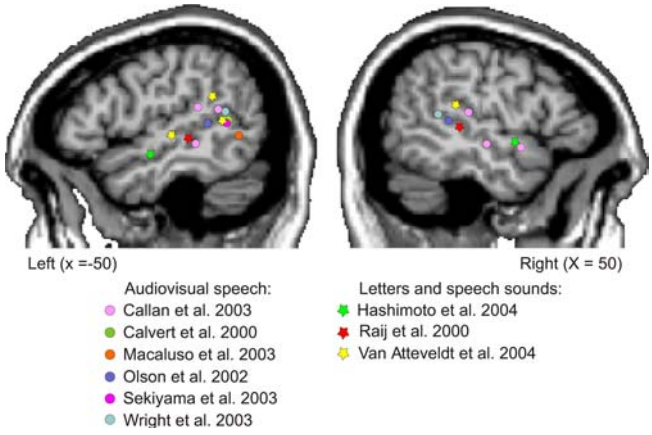

Fig. 2 A summary of the reported superior temporal regions (mainly STG, STS, and MTG) involved in AV integration of linguistic information. The most important activations of the reviewed studies are projected on sagittal slices of the MNI template brain transformed into Talairach space. LH is shown at $x = -50$, RH at $x = 50$. *Circles* reflect results from studies using speech and lip movement stimuli (i.e. naturally related AV stimuli), *stars* reflect results from studies using letters and speech sounds (i.e. artificially related AV stimuli)

Table 1 Putative VT object/form integration sites in humans: references, tasks, and Talairach coordinates

Paper	Task	Regions and Talairach coordinates (x, y, and z)
Hadjikhani and Roland (1998)	DMS of simple ellipsoids varying in curvature (T-V and V-T)	R insula/claustrum (+32, −9, and +13)
Deibert et al. (1999)	TOR: "real" versus "non-real" small objects (recognition versus roughness judgment)	Parietal and occipital areas (Talairach coordinates not available)
Amedi et al. (2001)	Convergence of TOR and VOR (recognition of natural everyday objects versus textures and scrambled objects)	L LOtv: (−45, −62, and −9)
		R LOtv: (Talairach coordinates not reported)
James et al. (2002)	Crossmodal (T-to-V) priming	L&R LOtv (±49, −60, and −1) L&R MO (±49, −74, and −6) L&R MO (±53, −79, and +4)
Amedi et al. (2002)	Convergence of TOR and VOR (man-made objects, models of animals, and vehicles)	L LOtv: (−47, −62, and −10)
		R LOtv: (+45, −54, and −14)
Grefkes et al. (2002)	DMS (T-V and V-T) of arbitrary simple shapes created from wooden spheres	L IPS: (−40, −42, and +36)
Stoesz et al. (2003)	2D macrospatial form discrimination versus microspatial gap detection (T)	L LOtv: (−42, −63, and +3)
		R LOtv: (+60, −57, and +3) L PHG: (−12, −39, and +2) R PHG: (+18, −30, and +9)
Saito et al. (2003)	2D pattern/shape T-V and V-T matching	L post. IPS (−28, −66, and +42) R post. IPS (28, −64, and +44)
Reed et al. (2004)	TOR (everyday objects) versus rest and TOR versus 3D abstract nonsense object palpation	L LOtv: (−50, −59, and −5)
		R LOtv: (+57, −56, and −7)
Prather et al. (2004)	2D macrospatial form versus orientation discrimination (T)	R LOtv: (+51, −59, and −14) R Insula: (+41, +15, and +8)
Pietrini et al. (2004)	Convergence of TOR and VOR (of bottles, shoes, and faces)	L LOtv: (−46, −58, and −10) R LOtv: (+48, −55, and −10)

Abbreviations: TOR tactile object recognition, VOR visual object recognition, T tactile, V visual, DMS delayed match to sample (Please note that in experiments using only tactile tasks, there is the theoretical problem of suggesting their role as possible VT integration sites. Still, we report here also foci activated by TOR only, in cases of activated areas that were classically considered to be unisensory visual regions. These cases should be taken with care as one would ideally want to show VT convergence directly in the same experiment)

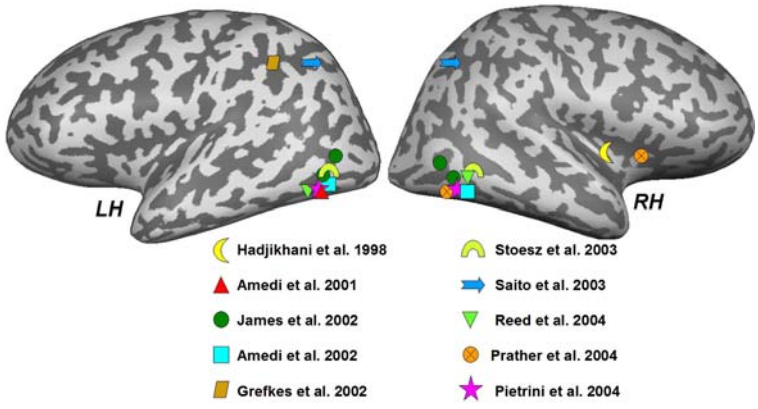

Fig. 3 Putative VT object/form integration sites in humans. Summary of the sites showing multisensory convergence from the different articles reviewed here (see also Table 1). The Talairach coordinates reported in each of the studies were projected on an inflated cerebral representation of a Talairach normalized brain. The approximate locations are denoted by their respective symbol. The relevant putative multisensory areas are found in the bilateral occipito-temporal cortex (centered on LOtv), parietal cortex (centered on the IPS), and the right insula. Please note that in the study of Stoesz et al. (2003) another more ventral cluster in the parahippocampal gyrus was found, but is hidden from view here. The study of Deibert et al. includes several occipital areas but the exact Talairach coordinates are not available and thus are not marked on the figure

regions, for example the PT, that surround the primary auditory cortex (A1). The involvement of A1, as reported by Calvert et al. (1997), is a matter of debate (Bernstein et al. 2002; Paulesu et al. 2003). In addition to studies focusing on the visual component of AV speech, the integration of auditory and visual speech has also been addressed directly. Despite wide variations in methodology, a consistent finding from AV speech studies is the involvement of the left STG, STS, and MTG in the integration of heard and seen speech (Callan et al. 2003; Calvert et al. 2000; Macaluso et al. 2004; Sekiyama et al. 2003; Wright et al. 2003; Fig. 2). Olson et al. (2002) report that the claustrum (but not the STS) integrates AV speech information. Interestingly, crossmodal effects in the STS seem to be strongest when auditory information is degraded by noise (Callan et al. 2001, 2003; Sekiyama et al. 2003), and this is consistent with behavioral findings (Sumby and Pollack 1954). A mechanism for AV speech perception has been proposed in which the auditory and visual speech signals are integrated in STS/STG, followed by modulation of the relevant unisensory cortices (auditory association cortex and MT/V5) via feedback projections (Calvert et al. 1999, 2000).

The neural basis of the associations between speech sounds and their written representations (i.e. letters) is much less extensively investigated. In a recent fMRI study (van Atteveldt et al. 2004) several bilateral STS/STG regions were identified that responded more strongly to bimodal congruent (i.e. matching) letter-sound pairs than to their respective unimodal components (Fig. 2 and the blue regions in Fig. 4). Furthermore, the response to speech sounds in early unisensory auditory regions in Heschl's sulcus (HS) and PT was found to be modulated by the degree of congruency with a simultaneously presented letter. Because these early auditory regions did not respond to unimodal visual letters, the authors interpret this modulation as a feedback effect from integration sites in the STS/STG. An MEG study provides converging time course information for the processing and integration of letters and speech sounds. The estimated source locations also indicate the involvement of the STS in the integration of letters and speech sounds (Raij et al. 2000). More evidence for this function of the STS/STG is provided by a recent learning experiment (Hashimoto and Sakai 2004). Learning of new letter-sound mappings involved a network of IT and parieto-occipital regions, while the STG/MTG was active during processing of already acquired matching letter-sound combinations.

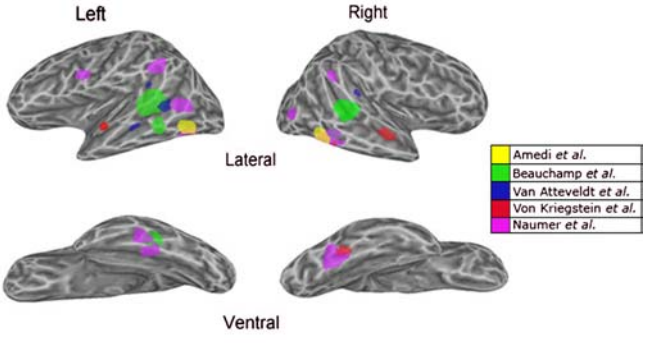

Fig. 4 A summary of regions found in studies of crossmodal object recognition. Active regions are shown in color on lateral and ventral views of an inflated cortical surface model. *Yellow* regions for VT object recognition (from Amedi et al. 2001, 2002). *Green* regions for AV integration of complex object information (from Beauchamp et al. 2004). *Blue* regions for AV letter recognition (from van Atteveldt et al. 2004). *Red* During the task of speaker recognition on familiar speakers' voices the FFA interacts with the middle STS voice area (analysis of functional connectivity from von Kriegstein et al. in press). *Purple* regions for AV recognition of natural or common objects (from Naumer et al. 2004)

Persons: faces and voices

Auditory and visual person-related information can be associated either inherently (e.g. face and voice) or arbitrarily (e.g. person and name). Person identification in general is a special form of subordinate object classification, which consists of binding concurrent information about face, voice, body, and name with a specific person. Studies on unimodal person perception have revealed cortical areas and activation patterns that are more responsive to faces than to other visually or haptically presented object categories (Kanwisher et al. 1997; Haxby et al. 2001; Pietrini et al. 2004) and the same is true for the auditory system where regions in the right STS showed a higher fMRI signal increase in response to voices than to any other category of auditory objects (Belin et al. 2000). These regions are anatomically segregated and the neural mechanisms underlying the combination of visual and auditory person identity information is unclear. An influential model of person recognition (Burton et al. 1990; Ellis et al. 1997) assumes that the combination of unimodal information occurs at a supramodal stage serving as a connecting node for voice, face, and name. However, it has recently been proposed that category-specific visual and auditory association cortices can also interact directly without a relay through supramodal regions (von Kriegstein et al. in press).

In contrast with arbitrary face-name associations, voice and face information is often encoded in parallel over an extended period of time. This close association between facial and vocal information not only enables the ability to read speech from facial movements and emotion from face and voice (Dolan et al. 2001) but also allows inference of information from the identity of a voice to the identity of a face even if voice and face have never been encountered before (Kamachi et al. 2003). When people become familiar with a person, vocal and facial information influences the speed with which voices and faces are recognized (Ellis et al. 1997). In a recent study with familiar and non-familiar speakers' voices and faces, von Kriegstein et al. (in press) showed that if subjects recognize familiar speakers' voices, activity in face-responsive regions is enhanced. According to the person recognition model, crossmodal activation in face-responsive regions should be accomplished via a supramodal node. Possible person identity nodes (e.g. regions regularly responding to familiar faces, voices, and names more than to non-familiar stimuli) are bilateral anterior temporal lobes, temporo-parietal junctions, and precuneus (Gorno-Tempini et al. 1998; Leveroni et al. 2000; Nakamura et al. 2000; Gorno-Tempini and Price 2001; Shah et al. 2001). Interestingly, functional connectivity analyses revealed that these supramodal areas were not functionally connected with the FFA during recognition of familiar persons' voices (von Kriegstein et al. in press). In contrast, connectivity was enhanced to the STS voice regions suggesting a direct interaction of these regions during familiar speaker recognition (red regions in Fig. 4).

Common objects

Although common objects are often perceived audiovisually under natural conditions, only a few crossmodal fMRI studies have employed common object stimuli. Among the cortical regions reported in these studies, three regions seem to be most prominently involved in AV recognition of common objects: lateral temporal cortex (especially the STS), ventral temporal cortex, and frontal cortices.

Lateral temporal cortex

Two recent studies (Beauchamp et al. 2004; Naumer et al. 2004) showed that posterior STS responds most strongly to auditory and visual objects and has an enhanced response when auditory and visual objects are presented together (green and purple regions in Fig. 4). Beauchamp et al. (2004) demonstrated that this is true for different categories of common objects (animals and manipulable tools), stimulus configurations (photograph versus line drawing versus moving video), and behavioral tasks (simple one-back, naming, and semantic judgments). Because multisensory activity in individual subjects always appeared in posterior STS, they named this region the STS multisensory region (STS-MS). This is also consistent with a recent proposal (Wallace et al. 2004a) that crossmodal integration occurs in multisensory zones that arise at the borders between unisensory cortical regions. Although STS-MS shows a preference for semantically consistent, compared with inconsistent, AV stimuli, the exact nature of the computations performed in this region remains unclear.

Ventral temporal cortex

Amedi et al. (2001) recently demonstrated tactile object-related activation in dorsal parts of LOC, an object-related region of the ventral visual pathway (yellow regions in Fig. 4). Because stimulation with sounds related to the same objects did not strongly activate this region (Amedi et al. 2002), they concluded that this lateral occipital tactile-visual (or LOtv) region is devoted to the processing of shape. Other fMRI studies have reported crossmodal AV effects in the ventral visual pathway during subordinate categorization and the detection of semantic (in)consistencies between pictures and sounds (Adams and Janata 2002; Beauchamp et al. 2004; Laurienti et al. 2003; Naumer et al. 2002a, 2004), however. In addition, two recent EEG studies (Molholm et al. 2004; Murray et al. 2004) provided converging time course information for the processing of

common object-related pictures and sounds, and the estimated source locations in these studies also indicated involvement of the ventral visual pathway in object-related AV integration. These findings can be integrated by taking into account the hierarchical organization of the ventral visual stream (Grill-Spector 2003). We assume that regions in ventral occipito-temporal cortex (VOT; Fig. 1) located ventral to LOC are processing more abstract stimulus properties (for example object category) that are also accessible via audition, thus abstracting from those properties that are only accessible via vision and touch (for example 3D structure). Preliminary fMRI data (Naumer et al. 2002b) provided the first empirical support for this assumption by indicating category-related AV convergence in VOT.

Frontal cortices

The AV fMRI studies also revealed contributions of frontal cortices to crossmodal cognitive functions such as object categorization (Adams and Janata 2002) and the detection of semantic inconsistencies (Laurienti et al. 2003; Naumer et al. 2002a, 2004). Adams and Janata (2002) instructed subjects to match either pictures or environmental sounds with simultaneously presented words. During visual and auditory judgments, neural activity in (predominantly left) inferior frontal gyrus (IFG) seemed to reflect the level of categorization (subordinate versus basic), thus indicating a role of the IFG in integrating multisensory object representations with concepts in semantic memory. By varying the amount of AV semantic consistency, Laurienti et al. (2003) found significantly greater activation in anterior cingulate gyrus and medial prefrontal cortex for consistent compared with inconsistent AV objects, whereas the inverse activation profile was found in a fronto-parietal network of regions (bilaterally including the precentral sulci and inferior parietal lobules) in recent studies by Naumer et al. (2002a, 2004).

Visuo-tactile object recognition

The visual and tactile hierarchies

The visual and tactile systems share many similarities. Both systems are arranged in a hierarchical order of increasing complexity. Neurons in V1 and area 3b within SI have relatively small receptive fields of simple features and contralateral specificity (hand or visual field) whereas the receptive field size and complexity (or functional selectivity) increase along the pathways as was detailed above. The two systems also share the principle of topographical organization, that is, adjacent parts of the world (whether in the visual field or on the body) are represented in adjacent parts within retinotopic and somatotopic cortical maps. In both systems the strict topographical representation of the contralateral sensory field at early processing stages decreases as one moves up along the hierarchy. For instance, LOC responds to visual stimuli in both hemifields and areas 1 and 2 have bilateral tactile receptive fields (Iwamura 1998). Finally, the visual system includes parallel pathways to analyze different aspects of the sensory world (Ungerleider and Mishkin 1982) and a similar pathway devoted to tactile form processing might also exist in the somatosensory system (Bodegard et al. 2001; Reed et al. 2004).

Evidence for visuo-tactile integration

The crossmodal delayed match to sample (DMS) task is an approach widely used to study crossmodal interactions. In this task, the cue is presented in one sensory modality and the target in the other. Human PET and fMRI studies employing DMS tasks on ellipsoids with varying curvature (Hadjikhani and Roland 1998), 2D patterns (Saito et al. 2003), or simple arbitrary 3D shapes created from wooden spheres (Grefkes et al. 2002) have all found greater activation after crossmodal versus intramodal matching in the insula/claustrum (Banati et al. 2000; Hadjikhani and Roland 1998), anterior IPS (Grefkes et al. 2002), and posterior IPS (Saito et al. 2003). These results suggest that the insula and IPS play a crucial role in binding visual and tactile information. It was proposed that the insula plays a role in crossmodal matching by serving as a mediating area enabling unisensory areas to communicate and exchange information. This was based on crossmodal matching experiments (Hadjikhani and Roland 1998) and the notion that the insula is highly connected to the various sensory areas and thus might be an ideal candidate to accomplish this function. It was, on the other hand, hypothesized that the IPS was a multisensory convergence site for visual, tactile, and possibly also auditory information. To study the potential role of visual areas in human tactile processing, an early study investigated tactile discrimination of grating orientations (Sathian et al. 1997). They reported robust tactile activation in the parieto-occipital cortex, a region previously regarded as part of the dorsal visual stream. In a later study, these same investigators demonstrated interference in task performance by using transcranial magnetic stimulation (TMS) to disrupt activity in this area (Zangaladze et al. 1999).

Integration of visual and tactile object processing in the ventral visual pathway

VT convergence of object-related information occurs in LOtv (Amedi et al. 2001), which is a subregion within the human LOC. The defining feature of this region is that it is robustly activated during both visual and tactile object recognition. It shows a preference for objects compared with textures and scrambled objects in both modalities and is only weakly activated by the motor,

naming and visual imagery components of object recognition. The category-related specialization for faces, objects, and scenes observed in ventral temporal cortex for visually presented objects is also found when objects are recognized by touch (Amedi et al. 2002; Pietrini et al. 2004). As shown by Pietrini et al. (2004), these category-related responses are correlated across touch and vision, suggesting that a common representation of 3D objects is activated by both these modalities. Stoesz et al. (2003) and Prather et al. (2004) showed that both IPS and LOtv are preferentially activated by macrospatial shape recognition (of imprinted symbol identity) but not during a microspatial task (gap detection or orientation judgment). This demonstrates that 2D shape stimuli activate LOtv and that this activation is maintained without active exploration, thus excluding a contribution from the motor system. James et al. (2002) showed fMRI activation in occipital areas during haptic exploration of novel abstract objects. These objects were used to reduce the potential influence of naming and visual imagery. By demonstrating that the magnitude of tactile-to-visual priming was similar to that of visual-to-visual priming, their findings supported the idea that vision and touch share common representations (see also Easton et al. 1997; Reales and Ballesteros 1999). Reed et al. (2004) used abstract and nonsense 3D shapes combined with familiar and meaningful 3D objects. Lateralized activation was found, suggesting that right LOtv is more involved in recovering the shape of objects while the left LOC is more focused on the association of tactile input with semantic information and possibly also with familiarity knowledge. Table 1 summarizes the different regions revealed by studies on VT form integration. Taking into account differences between stimuli, tasks, contrasts, and analysis techniques, LOtv seems to be highly consistent in location (average Talairach coordinates (x, y and z) -47, -61 and -5 for the left and 51, -56 and -7 for the right hemispheres).

Normally, individuals can easily recognize faces by touch. Interestingly, Kilgour and Lederman (2002) recently reported evidence of an individual with haptic prosopagnosia (i.e. a deficit in recognizing familiar faces by touch) and visual prosopagnosia after lesions to occipital, temporal and prefrontal regions (Kilgour et al. 2004). Some evidence suggests that patients with visual agnosia also suffer from tactile agnosia (Feinberg et al. 1986; Morin et al. 1984; for a short review see Amedi et al. 2002). In an experimental setting TMS can be used to create "virtual lesions" in normal individuals (Pascual-Leone et al. 2000). Zangaladze et al. (1999) used this technique to show interference with a tactile orientation task during TMS of area PO, and Merabet et al. (2004) showed that inhibitory repetitive TMS (at 1 Hz) to the occipital cortex reduced performance in tactile distance judgments but not in roughness judgments, suggesting involvement of the occipital cortex in tactile distance processing. Both studies support the view that the occipital cortex is engaged in tactile tasks requiring fine spatial discriminations. To conclude, three cortical regions seem to be important for VT form integration (Fig. 3)—LOtv and adjacent areas, parietal regions (especially IPS), and the claustrum/insular cortex.

Discussion

Multisensory regions of temporal, parietal, frontal, and insular cortices are involved in crossmodal binding of AV and VT object-related information (Calvert and Lewis 2004; Newell 2004). The fMRI studies reviewed here demonstrate that these regions are not equally responsive to all types of unimodal stimulation and their combinations, but rather show a degree of specialization. After a short summary of the most prominent neural systems involved in AV and VT integration, we will further discuss findings that provide more insight in to what determines the location of crossmodal convergence and the role of imagery in crossmodal object processing. Finally, we will conclude by proposing a model for crossmodal object recognition by distributed semantic processing.

The posterior STS and STG play an important role in the integration of naturally and artificially related AV "what" information (i.e. speech with lip movements or letters, and sounds with pictures of common objects). This is in accordance with the suggested general role of the STS in integration of object identity information (Beauchamp et al 2004; Calvert 2001). The functional role of the STS is also supported by the recent finding that activity in this area is affected by the temporal synchrony of AV speech information, but is insensitive to the spatial displacement of the visual and auditory stimuli (Macaluso et al. 2004). However, the commonalities between vision and audition that elicit STS activation are still elusive. A variety of studies have emphasized the involvement of STS in dynamic aspects of visual object processing (e.g. eye gaze (Hoffman and Haxby 2000), visual speech (Calvert and Campbell 2003), and auditory object processing (von Kriegstein and Giraud 2004; Thierry et al. 2003).

On the other hand, in VT object recognition, the LOtv, IPS, and right insular cortex seem to be prominently involved. It was recently demonstrated that tactile responses in LOtv during object identification are bilateral irrespective of the palpating hand in both sighted and blind subjects (Amedi 2004). This further supports the view that LOC is involved in high-order tactile recognition. Such crossmodal convergence might reflect the processing of features that are common to both vision and touch. Besides their unimodal features (e.g. color and consistency), both contribute to the evaluation of surface characteristics of objects (e.g. geometrical shape). In general, it seems that LOtv and IPS are more involved in object shape analysis and recognition whereas the claustrum/insula region is more involved in the transfer and/or binding of unimodal information. The possible division of labor between these regions and their functional interplay are far from

being understood and thus remain to be addressed in future studies. In contrast with vision and touch, audition contributes much less to evaluation of surface characteristics, but instead, presents the dynamics or energetics of a situation (e.g. whether two objects pass each other or collide; Bregman 1990).

What determines a point of crossmodal convergence?

One determinant of where crossmodal convergence of object information occurs is which modality provides the most accurate information about the object (Welch and Warren 1986; Ernst and Bulthoff 2004). As vision dominates touch and audition in human object perception, crossmodal convergence might occur in the visual "what" pathway. It has, however, been demonstrated in several experimental settings that the dominance of vision is not exclusive. If, for example, the tactile input provides richer information, the tactile rather than the visual input will dominate. In this situation, multisensory convergence might occur in tactile rather than visual cortices. Indeed, some of the studies reviewed here showed VT effects in the parietal cortex (especially the IPS; Banati et al. 2000; Grefkes et al. 2002; Saito et al. 2003) and the right insula (Banati et al. 2000; Hadjikhani and Roland 1998; Prather et al. 2004). Multisensory responses in the parietal cortex were also reported in non-human primates (though their homology in humans is debated. For reviews, see Andersen et al. 1997; Colby and Goldberg 1999).

Another factor that could determine the recruitment of a certain brain region for processing of diverse sensory input might be the task demands of certain computations. This was suggested recently by Pascual-Leone and Hamilton (2001) in their "metamodal theory". This would mean that the crucial aspect of processing is not the kind of sensory input the brain receives, but the contribution that a certain area makes and how the available information from the relevant senses is extracted for successful task performance.

It has recently been proposed that multisensory convergence zones are mainly located at the borders of modality-specific cortical areas (Wallace et al. 2004a). The location of the involved modality-specific regions is, therefore, another important determinant for multisensory convergence zones. The STS/STG is located anatomically between the visual and auditory "what" systems, which makes this region a suitable candidate for integrating auditory and visual information about objects.

A possible role of crossmodally triggered visual imagery

The literature reviewed here demonstrates that crossmodal AV effects often depend on previous associations between unimodal stimulus components. This suggests that crossmodal auditory activation of the ventral visual stream might be different from visual activation of these same regions. The responsiveness of visual cortices to auditory events or tactile objects might correspond to crossmodally triggered visual imagery (i.e. perceptual information accessed from memory; Kosslyn et al. 2001). Along these lines, Sathian et al. (1997) suggested that visual imagery is important for successful performance in tactile tasks.

The question arises, however, whether this association of unimodal information is an epiphenomenon mediated by top-down processes or is actually helpful in the recognition process. The latter would only be observed if unisensory cortices could interact directly before object recognition in unisensory cortices occurs. It has recently been shown (Falchier et al. 2002; Rockland and Ojima 2003) that in the macaque areas V1 and V2 receive direct input from core and parabelt auditory areas. Because connectivity is eccentricity-dependent (peripheral but not central areas receive these inputs), the function may be related to a foveation mechanism designed to reduce behavioral orientation response time to peripheral stimuli. In humans, functional connectivity analyses in a study not explicitly requiring imagery suggest that crossmodal AV coupling of highly familiar stimuli (familiar faces and voices) is subserved by a direct association of unimodal information (von Kriegstein et al. in press). This low-level association leaves open the possibility that even before recognition of the object is accomplished in one modality, information is combined. This effect might explain the behavioral interaction of multisensory information in recognition of unisensory stimuli (Ellis et al. 1997). This is further supported by recent studies in the tactile domain that showed a similar level of LOtv activation during tactile object recognition, even in subjects that have had no visual experience before (i.e. congenitally blind subjects; Pietrini et al. 2004; Amedi 2004).

A distributed model of semantic processing

Independent of their computational function, multisensory regions presumably interact with unisensory cortices via feedback connections and thus subserve a cortical network in which semantic knowledge about objects is represented in a distributed fashion (Martin and Chao 2001). According to this model, brain regions that process a specific type of incoming sensory information also serve as a knowledge store about the attributes of particular objects. For instance, a hammer has a number of salient attributes (for example characteristic color, shape, and sound). Ventral temporal cortex processes incoming information about visual form and color. If the incoming form and color match the stored hammer template, ventral temporal cortex identifies the object as a hammer. Lateral temporal cortex processes information about object motion. If the incoming motion information matches the stored template (e.g. up, down, and impact) then the object is identified as a

hammer. Frontal and parietal regions store information about the hand orientation and arm posture that would be required to grasp the hammer. The representations in these different regions are linked, so that simply seeing a photograph of a static hammer also activates the respective visual motion and motor representations (Beauchamp et al. 2002). Thus, networks of neurons in auditory association cortex that respond to an auditory object are also linked to visual, tactile, and motor information about the same object represented in temporal, parietal, and frontal regions.

The operation of multisensory regions during crossmodal object recognition remains to be determined. Mesulam (1998) suggests that associative areas contain "road maps" for binding distributed information in different modalities. In this view, associative areas do not contain multimodal representations but merely link the unimodal representations to make them accessible to each other. Contrary to this view, recent studies of crossmodal speaker recognition revealed enhanced functional connectivity between face-responsive regions in extrastriate visual cortex and voice regions in anterior/middle STS without involvement of supramodal areas (von Kriegstein and Giraud 2004; von Kriegstein et al. in press). Thus direct interaction of these regions during familiar speaker recognition might rely on direct anatomical connections between lateral and ventral temporal regions of human cortex (Catani et al. 2003).

A possible explanation of the diverse findings regarding the involvement of multisensory associative regions may be that it depends on the type of information that must be bound together. This view is supported by findings of different patterns of anatomical connectivity (feedforward versus feedback) between unisensory and multisensory areas in primates (Falchier et al. 2002; Schroeder et al. 2003). Direct sharing of information between modalities, without recruitment of multisensory regions, might be only expected for information that is important to identify an object quickly and automatically. Examples would be alerting, orienting, and localizing responses, motion perception, and processing of socially important information (e.g. face/voice identity). In contrast, more complex or arbitrarily related information (e.g. AV speech, letters, and common objects) might be integrated through associative nodes (e.g. STS-MS) that are able to form more complex and flexible mappings between information from different modalities.

After placing the reviewed neuroimaging findings on unimodal and crossmodal object processing in the broader context of crossmodal and semantic processing theories, we propose the following tentative model for AV and VT object processing. Modality-specific representations in unisensory association cortices are linked either directly or through a heteromodal binding site depending on the information type and the goal or the requirements of the crossmodal combination. For a mediating heteromodal binding site, this probably reflects an associative node rather than a complex heteromodal object representation (Mesulam 1998). The location of crossmodal convergence, either within a unisensory system or in heteromodal association cortex, is determined by the dominance of one of the modalities, the relevant computational capabilities of a region, and the anatomical location of the unisensory and multisensory cortices relative to each other.

To better understand the rapid computations performed during object perception, future studies should combine high-resolution fMRI with methods that enable examination of the temporal sequence of information processing (for example MEG/EEG or TMS). Such an approach should deepen our understanding of crossmodal object recognition and also shed light on unisensory mechanisms of object processing. The use of more natural stimulation conditions (Bartels and Zeki 2004a, b; Hasson et al. 2004) should substantially contribute to human object recognition theories with increased ecological validity (de Gelder and Bertelson 2003).

Acknowledgements This research was funded by a Horowitz Foundation fellowship (A.A.), the Bundesministerium für Bildung und Forschung (BMBF; K.v.K., M.J.N.), the Volkswagenstiftung (K.v.K.), and the Max Planck Society (M.J.N). The authors thank Nikolas Francis, Axel Kohler (for help with the figures), Lotfi Merabet, Wolf Singer, Lars Muckli, and three anonymous reviewers (for their helpful comments on earlier versions of this paper). Reprint requests and remarks should be addressed to Marcus Johannes Naumer (H.J.Naumer@med.uni-frankfurt.de) or to Amir Amedi (aamedi@bidmc.harvard.edu).

References

Adams RB, Janata P (2002) A comparison of neural circuits underlying auditory and visual object categorization. Neuroimage 16:361–377

Amedi A (2004) Multisensory object-related processing in the visual cortex of sighted and its reversed hierarchical organization in blind humans. In: Presented at the 5th international multisensory research forum in Sitges, Spain, Abstract No. 149

Amedi A, Malach R, Hendler T, Peled S, Zohary E (2001) Visuo-haptic object-related activation in the ventral visual pathway. Nat Neurosci 4:324–330

Amedi A, Jacobson G, Hendler T, Malach R, Zohary E (2002) Convergence of visual and tactile shape processing in the human lateral occipital complex. Cereb Cortex 12:1202–1212

Andersen RA, Snyder LH, Bradley DC, Xing J (1997) Multimodal representation of space in the posterior parietal cortex and its use in planning movements. Annu Rev Neurosci 20:303–330

Arnott SR, Binns MA, Grady CL, Alain C (2004) Assessing the auditory dual-pathway model in humans. Neuroimage 22:401–408

van Atteveldt N, Formisano E, Goebel R, Blomert L (2004) Integration of letters and speech sounds in the human brain. Neuron 43:271–282

Banati RB, Goerres GW, Tjoa C, Aggleton JP, Grasby P (2000) The functional anatomy of visual-tactile integration in man: a study using positron emission tomography. Neuropsychologia 38:115–124

Bartels A, Zeki S (2004a) Functional brain mapping during free viewing of natural scenes. Hum Brain Mapp 21:75–85

Bartels A, Zeki S (2004b) The chronoarchitecture of the human brain—natural viewing conditions reveal a time-based anatomy of the brain. Neuroimage 22:419–433

Beauchamp MS, Lee KE, Haxby JV, Martin A (2002) Parallel visual motion processing streams for manipulable objects and human movements. Neuron 34:149–159

Beauchamp MS, Lee KE, Argall BD, Martin A (2004) Integration of auditory and visual information about objects in superior temporal sulcus. Neuron 41:809–823
Belin P, Zatorre RJ (2000) 'What', 'where', and 'how' in auditory cortex. Nat Neurosci 3:965–966
Belin P, Zatorre RJ (2003) Adaptation to speaker's voice in right anterior temporal lobe. Neuroreport 14:2105–2109
Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000) Voice-selective areas in human auditory cortex. Nature 403:309–312
Bernstein LE, Auer ET Jr, Moore JK, Ponton CW, Don M, Singh M (2002) Visual speech perception without primary auditory cortex activation. Neuroreport 13:311–315
Binder JR, Liebenthal E, Possing ET, Medler DA, Ward BD (2004) Neural correlates of sensory and decision processes in auditory object identification. Nat Neurosci 7:295–301
Binkofski F, Buccino G, Posse S, Seitz RJ, Rizzolatti G, Freund H (1999) A fronto-parietal circuit for object manipulation in man: evidence from an fMRI-study. Eur J Neurosci 11:3276–3286
Bodegard A, Geyer S, Grefkes C, Zilles K, Roland PE (2001) Hierarchical processing of tactile shape in the human brain. Neuron 31:317–328
Bregman AS (1990) Auditory scene analysis. MIT Press, Cambridge, MA
Burton AM, Bruce V, Johnston RA (1990) Understanding face recognition with an interactive activation model. Br J Psychol 81(Pt 3):361–380
Callan DE, Callan AM, Kroos C, Vatikiotis-Bateson E (2001) Multimodal contribution to speech perception revealed by independent component analysis: a single-sweep EEG case study. Brain Res Cogn Brain Res 10:349–353
Callan DE, Jones JA, Munhall K, Callan AM, Kroos C, Vatikiotis-Bateson E (2003) Neural processes underlying perceptual enhancement by visual speech gestures. Neuroreport 14:2213–2218
Calvert GA (2001) Crossmodal processing in the human brain: insights from functional neuroimaging studies. Cereb Cortex 11:1110–1123
Calvert GA, Campbell R (2003) Reading speech from still and moving faces: the neural substrates of visible speech. J Cogn Neurosci 15:57–70
Calvert GA, Lewis JW (2004) Hemodynamic studies of audiovisual interactions. In: Calvert G, Spence C, Stein BE (eds) The handbook of multisensory processes. MIT Press, Cambridge, MA, pp 483–502
Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SC, McGuire PK, Woodruff PW, Iversen SD, David AS (1997) Activation of auditory cortex during silent lipreading. Science 276:593–596
Calvert GA, Brammer MJ, Bullmore ET, Campbell R, Iversen SD, David AS (1999) Response amplification in sensory-specific cortices during crossmodal binding. Neuroreport 10:2619–2623
Calvert GA, Campbell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Curr Biol 10:649–657
Catani M, Jones DK, Donato R, Ffytche DH (2003) Occipito-temporal connections in the human brain. Brain 126:2093–2107
Colby CL, Goldberg ME (1999) Space and attention in parietal cortex. Annu Rev Neurosci 22:319–49
De Gelder B, Bertelson P (2003) Multisensory integration, perception and ecological validity. Trends Cogn Sci 7:460–467
Deibert E, Kraut M, Kremen S, Hart J Jr (1999) Neural pathways in tactile object recognition. Neurology 52:1413–1417
Dolan RJ, Morris JS, de Gelder B (2001) Crossmodal binding of fear in voice and face. Proc Natl Acad Sci USA 98:10006–10010
Downing PE, Jiang Y, Shuman M, Kanwisher N (2001) A cortical area selective for visual processing of the human body. Science 293:2470–2473
Easton RD, Srinivas K, Greene AJ (1997) Do vision and haptics share common representations? Implicit and explicit memory within and between modalities. J Exp Psychol Learn Mem Cogn 23:153–163
Ellis HD, Jones DM, Mosdell N (1997) Intra- and inter-modal repetition priming of familiar faces and voices. Br J Psychol 88:143–156
Epstein R, Kanwisher N (1998) A cortical representation of the local visual environment. Nature 392:598–601
Ernst M, Bülthoff H (2004) Merging the senses into a robust percept. Trends Cogn Sci 8:162–169
Falchier A, Clavagnier S, Barone P, Kennedy H (2002) Anatomical evidence of multimodal integration in primate striate cortex. J Neurosci 22:5749–5759
Feinberg TE, Rothi LJ, Heilman KM (1986) Multimodal agnosia after unilateral left hemisphere lesion. Neurology 36:864–867
Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex 1:1–47
Gauthier I, Skudlarski P, Gore JC, Anderson AW (2000) Expertise for cars and birds recruits brain areas involved in face recognition. Nat Neurosci 3:191–197
Gauthier I, Tarr MJ, Moylan J, Skudlarski P, Gore JC, Anderson AW (2000) The fusiform ''face area'' is part of a network that processes faces at the individual level. J Cogn Neurosci 12:495–504
Gleitman LR, Rozin P (1977) The structure and acquisition of reading I: relations between orthographies and the structure of language. In: Reber A, Scarborough D (eds) Towards a psychology of reading: the proceedings of the CUNY conferences. Lawrence Erlbaum Associates, Hillsdale, NJ
Goodale MA, Meenan JP, Bulthoff HH, Nicolle DA, Murphy KJ, Racicot CI (1994) Separate neural pathways for the visual analysis of object shape in perception and prehension. Curr Biol 4:604–610
Gorno-Tempini ML, Price CJ (2001) Identification of famous faces and buildings: a functional neuroimaging study of semantically unique items. Brain 124:2087–2097
Gorno-Tempini ML, Price CJ, Josephs O, Vandenberghe R, Cappa SF, Kapur N, Frackowiak RS, Tempini ML (1998) The neural systems sustaining face and proper-name processing. Brain 121(Pt 11):2103–2118
Grefkes C, Weiss PH, Zilles K, Fink GR (2002) Crossmodal processing of object features in human anterior intraparietal cortex: an fMRI study implies equivalencies between humans and monkeys. Neuron 35:173–184
Griffiths TD, Warren JD (2002) The planum temporale as a computational hub. Trends Neurosci 25:348–353
Grill-Spector K (2003) The neural basis of object perception. Curr Opin Neurobiol 13:159–166
Grill-Spector K, Malach R (2004) The human visual cortex. Annu Rev Neurosci 27:649–677
Hadjikhani N, Roland PE (1998) Cross-modal transfer of information between the tactile and the visual representations in the human brain: a positron emission tomographic study. J Neurosci 18:1072–1084
Hashimoto R, Sakai KL (2004) Learning letters in adulthood: direct visualization of cortical plasticity for forming a new link between orthography and phonology. Neuron 42:311–322
Hasson U, Harel M, Levy I, Malach R (2003) Large-scale mirror-symmetry organization of human occipito-temporal object areas. Neuron 37:1027–1041
Hasson U, Nir Y, Levy I, Fuhrmann G, Malach R (2004) Inter-subject synchronization of cortical activity during natural vision. Science 303:1634–1640
Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293:2425–2430
Hoffman EA, Haxby JV (2000) Distinct representations of eye gaze and identity in the distributed human neural system for face perception. Nat Neurosci 3:80–84
Iwamura Y (1998) Hierarchical somatosensory processing. Curr Opin Neurobiol 8:522–528
James TW, Humphrey GK, Gati JS, Servos P, Menon RS, Goodale MA (2002) Haptic study of three-dimensional objects activates extrastriate visual areas. Neuropsychologia 40:1706–1714

Jäncke L, Wüstenberg T, Scheich H, Heinze HJ (2002) Phonetic perception and the temporal cortex. Neuroimage 15:733–746
Kaas JH, Hackett TA (1999) 'What' and 'where' processing in auditory cortex. Nat Neurosci 2:1045–1047
Kamachi M, Hill H, Lander K, Vatikiotis-Bateson E (2003) "Putting the face to the voice": matching identity across modality. Curr Biol 13:1709–1714
Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci 17:4302–4311
Kilgour AR, Lederman SJ (2002) Face recognition by hand. Percept Psychophys 64:339–352
Kilgour AR, de Gelder B, Lederman SJ (2004) Haptic face recognition and prosopagnosia. Neuropsychologia 42:707–712
Kosslyn SM, Ganis G, Thompson WL (2001) Neural foundations of imagery. Nat Rev Neurosci 2:635–642
von Kriegstein K, Giraud AL (2004) Distinct functional substrates along the right superior temporal sulcus for the processing of voices. Neuroimage 22:948–955
von Kriegstein K, Eger E, Kleinschmidt A, Giraud AL (2003) Modulation of neural responses to speech by directing attention to voices or verbal content. Brain Res Cogn Brain Res 17:48–55
von Kriegstein K, Kleinschmidt A, Sterzer P, Giraud AL (in press) Interaction of face and voice areas during speaker recognition. J Cog Neurosci
Kuhl PK, Meltzoff AN (1982) The bimodal perception of speech in infancy. Science 218:1138–1141
Laurienti PJ, Wallace MT, Maldjian JA, Susi CM, Stein BE, Burdette JH (2003) Cross-modal sensory processing in the anterior cingulate and medial prefrontal cortices. Hum Brain Mapp 19:213–223
Leveroni CL, Seidenberg M, Mayer AR, Mead LA, Binder JR, Rao SM (2000) Neural systems underlying the recognition of familiar and newly learned faces. J Neurosci 20:878–886
Lewis JW, Wightman FL, Brefczynski JA, Phinney RE, Binder JR, DeYoe EA (2004) Human brain regions involved in recognizing environmental sounds. Cereb Cortex AoP
Liberman AM (1992) The relation of speech to reading and writing. In: Frost R, Katz L (eds) Orthography, phonology, morphology and meaning. Elsevier Science Publishers BV, Amsterdam
Macaluso E, George N, Dolan R, Spence C, Driver J (2004) Spatial and temporal factors during processing of audiovisual speech: a PET study. Neuroimage 21:725–732
Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RB (1995) Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. Proc Natl Acad Sci USA 92:8135–8139
Martin A, Chao LL (2001) Semantic memory and the brain: structure and processes. Curr Opin Neurobiol 11:194–201
McCandliss BD, Cohen L, Dehaene S (2003) The visual word form area: expertise for reading in the fusiform gyrus. Trends Cogn Sci 7:293–299
McGurk H, MacDonald J (1976) Hearing lips and seeing voices. Nature 264:746–748
Merabet L, Thut G, Murray B, Andrews J, Hsiao S, Pascual-Leone A (2004) Feeling by sight or seeing by touch?. Neuron 42:173–179
Mesulam MM (1998) From sensation to cognition. Brain 121(Pt 6):1013–1052
Mishkin M (1979) Analogous neural models for tactual and visual learning. Neuropsychologia 17(2):139–151
Molholm S, Ritter W, Javitt DC, Foxe JJ (2004) Multisensory visual-auditory object recognition in humans: a high-density electrical mapping study. Cereb Cortex 14:452–465
Morin P, Rivrain Y, Eustache F, Lambert J, Courtheoux P (1984) Visual and tactile agnosia. Rev Neurol (Paris) 140:271–277
Munhall KG, Tohkura Y (1998) Audiovisual gating and the time course of speech perception. J Acoust Soc Am 104:530–539
Murray MM, Michel CM, Grave de Peralta R, Ortigue S, Brunet D, Gonzalez AS, Schnider A (2004) Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. Neuroimage 21:125–135
Nakamura K, Kawashima R, Sato N, Nakamura A, Sugiura M, Kato T, Hatano K, Ito K, Fukuda H, Schormann T, Zilles K (2000) Functional delineation of the human occipito-temporal areas related to face and scene processing. A PET study. Brain 123:1903–1912
Naumer MJ, Singer W, Muckli L (2002a) Audio-visual perception of natural objects. OHBM Abstract#15600
Naumer MJ, Wibral M, Singer W, Muckli L (2002b) FMRI-studies of category-specific audio-visual processing—visual cortex. IMRF Abstract#25
Naumer MJ, Petkova V, Havenith MN, Kohler A, Singer W, Muckli L (2004) Paying attention to multisensory objects. OHBM Abstract#TH99
Newell FN (2004) Cross-modal object recognition. In: Calvert G, Spence C, Stein BE (eds) The handbook of multisensory processes. MIT Press, Cambridge, MA, pp 123–139
Ohtake H, Fujii T, Yamadori A, Fujimori M, Hayakawa Y, Suzuki K (2001) The influence of misnaming on object recognition: a case of multimodal agnosia. Cortex 37:175–186
Olson IR, Gatenby JC, Gore JC (2002) A comparison of bound and unbound audio-visual information processing in the human cerebral cortex. Brain Res Cogn Brain Res 14:129–138
O'Sullivan BT, Roland PE, Kawashima R (1994) A PET study of somatosensory discrimination in man. Microgeometry versus macrogeometry. Eur J Neurosci 6:137–148
Pascual-Leone A, Hamilton R (2001) The metamodal organization of the brain. Prog Brain Res 134:427–445
Pascual-Leone A, Walsh V, Rothwell J (2000) Transcranial magnetic stimulation in cognitive neuroscience—virtual lesion, chronometry, and functional connectivity. Curr Opin Neurobiol 10:232–237
Paulesu E, Perani D, Blasi V, Silani G, Borghese NA, De Giovanni U, Sensolo S, Fazio F (2003) A functional-anatomical model for lipreading. J Neurophysiol 90:2005–2013
Pietrini P, Furey ML, Ricciardi E, Gobbini MI, Wu WH, Cohen L, Guazzelli M, Haxby JV (2004) Beyond sensory images: object-based representation in the human ventral pathway. Proc Natl Acad Sci USA 101:5658–5663
Polk TA, Stallcup M, Aguirre GK, Alsop DC, D'Esposito M, Detre JA, Farah MJ (2002) Neural specialization for letter recognition. J Cogn Neurosci 14:145–159
Polster MR, Rose SB (1998) Disorders of auditory processing: evidence for modularity in audition. Cortex 34:47–65
Pons TP, Garraghty PE, Friedman DP, Mishkin M (1987) Physiological evidence for serial processing in somatosensory cortex. Science 237:417–420
Prather SC, Votaw JR, Sathian K (2004) Task-specific recruitment of dorsal and ventral visual areas during tactile perception. Neuropsychologia 42:1079–1087
Raij T, Uutela K, Hari R (2000) Audiovisual integration of letters in the human brain. Neuron 28:617–625
Rauschecker JP, Tian B (2000) Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc Natl Acad Sci USA 97:11800–11806
Reales JM, Ballesteros S (1999) Implicit and explicit memory for visual and haptic objects: cross-modal priming depends on structural descriptions. J Exp Psychol Learn Mem Cog 25:644–663
Reed CL, Caselli RJ (1994) The nature of tactile agnosia: a case study. Neuropsychologia 32:527–539
Reed CL, Shoham S, Halgren E (2004) Neural substrates of tactile object recognition: an fMRI study. Hum Brain Mapp 21:236–246
Rockland KS, Ojima H (2003) Multisensory convergence in calcarine visual areas in macaque monkey. Int J Psychophysiol 50(1–2):19–26

Roland PE, O'Sullivan B, Kawashima R (1998) Shape and roughness activate different somatosensory areas in the human brain. Proc Natl Acad Sci USA 95:3295–3300

Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP (1999) Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. Nat Neurosci 2:1131–1136

Saito DN, Okada T, Morita Y, Yonekura Y, Sadato N (2003) Tactile-visual cross-modal shape matching: a functional MRI study. Brain Res Cogn Brain Res 17:14–25

Sathian K, Zangaladze A, Hoffman JM, Grafton ST (1997) Feeling with the mind's eye. Neuroreport 8:3877–3881

Schroeder CE, Smiley J, Fu KG, McGinnis T, O'Connell MN, Hackett TA (2003) Anatomical mechanisms and functional implications of multisensory convergence in early cortical processing. Int J Psychophysiol 50:5–17

Sekiyama K, Kanno I, Miura S, Sugita Y (2003) Auditory-visual speech perception examined by fMRI and PET. Neurosci Res 47:277–287

Shah NJ, Marshall JC, Zafiris O, Schwab A, Zilles K, Markowitsch HJ, Fink GR (2001) The neural correlates of person familiarity A functional magnetic resonance imaging study with clinical implications. Brain 124:804–815

Stein BE, Meredith MA (1993) The merging of the senses. MIT Press, Cambridge, MA

Stoesz MR, Zhang M, Weisser VD, Prather SC, Mao H, Sathian K (2003) Neural networks active during tactile form perception: common and differential activity during macrospatial and microspatial tasks. Int J Psychophysiol 50:41–49

Sumby WH, Pollack I (1954) Visual contribution to speech intelligibility in noise. J Acoust Soc Am 26:212–215

Thierry G, Giraud AL, Price C (2003) Hemispheric dissociation in access to the human semantic system. Neuron 38:499–506

Tootell RB, Tsao D, Vanduffel W (2003) Neuroimaging weighs in: humans meet macaques in "primate" visual cortex. J Neurosci 23:3981–3989

Ungerleider LG, Haxby JV (1994) 'What' and 'where' in the human brain. Curr Opin Neurobiol 4:157–165

Ungerleider LG, Mishkin M (1982) Two cortical visual streams. In: Ingle DJ, Goodale MA, Mansfield RJW (eds) Analysis of visual behavior. MIT Press, Cambridge, MA

Wallace MT, Ramachandran R, Stein BE (2004a) A revised view of sensory cortical parcellation. Proc Natl Acad Sci USA 101:2167–2172

Wallace MT, Roberson GE, Hairston WD, Stein BE, Vaughan JW, Schirillo JA (2004b) Unifying multisensory signals across time and space. Exp Brain Res [epub ahead of print]

Welch RB, Warren DH (1986) Intersensory interactions. In: Boff KR, Kaufman L, Thomas J (eds) Handbook of perception and human performance. Wiley, New York

Wernicke C (1874) Der aphasische Symptomenkomplex, eine psychologische Studie auf anatomischer Basis. Cohn& Weigert, Breslau

Wright TM, Pelphrey KA, Allison T, McKeown MJ, McCarthy G (2003) Polysensory interactions along lateral temporal regions evoked by audiovisual speech. Cereb Cortex 13:1034–1043

Zangaladze A, Epstein CM, Grafton ST, Sathian K (1999) Involvement of visual cortex in tactile discrimination of orientation. Nature 401:587–590

Zatorre RJ, Bouffard M, Belin P (2004) Sensitivity to auditory object features in human temporal neocortex. J Neurosci 24:3637–3642

Zeki SM (1978) Functional specialization in the visual cortex of the rhesus monkey. Nature 274:423–428