CROSSMODAL AND MULTISENSORY INTERACTIONS BETWEEN VISION AND TOUCH

Simon Lacey⁴ and K. Sathian¹,²,³,⁴

¹Departments of Neurology, Emory University, Atlanta, GA, USA
²Rehabilitation Medicine, Emory University, Atlanta, GA, USA
³Psychology, Emory University, Atlanta, GA, USA
⁴Rehabilitation R&D Center of Excellence, Atlanta VAMC, Decatur, GA, USA

Over the past two decades, there has been growing appreciation of the multisensory nature of perception and its neural basis. Consequently, the concept has arisen that the brain is “metamodal”, with a task-based rather than strictly modality-based organization (Pascual-Leone & Hamilton, 2001; Lacey et al., 2009a; James et al., 2011). Here we focus on interactions between vision and touch in humans, including crossmodal interactions where tactile inputs evoke activity in neocortical regions traditionally considered visual, and multisensory integrative interactions. It is now established that cortical areas in both the ventral and dorsal pathways, previously identified as specialized for various aspects of visual processing, are also routinely recruited during the corresponding aspects of touch (for reviews see Amedi et al., 2005; Sathian & Lacey, 2007; Lacey & Sathian, 2011, 2014). When these regions are in classical visual cortex so that they would traditionally be regarded as unisensory, their engagement is referred to as crossmodal, whereas other regions lie in classically multisensory sectors of the association neocortex. Much of the relevant work concerns haptic perception (active sensing using the hand) of shape; this work is therefore considered in detail. We consider how vision and touch might be integrated in various situations and address the role of mental imagery in visual cortical activity during haptic perception. Finally, we present a model of haptic object recognition and its relationship with mental imagery (Lacey et al., 2014).

Activation of visually responsive cortical regions during touch

The first demonstration that a visual cortical area is active during normal tactile perception came from a positron emission tomographic (PET) study in humans (Sathian et al., 1997). In this study, tactile discrimination of the orientation of gratings applied to the immobilized fingertip, relative to a control task requiring tactile discrimination of grating groove width, activated a focus in extrastriate visual cortex, close to the parieto-occipital fissure. This focus, located in the region of the human V6 complex of visual areas (Pitzalis et al. 2006), is also active during visual discrimination of grating orientation (Sergent et al., 1992).
Similarly, other neocortical regions known to selectively process particular aspects of vision are activated by analogous non-visual tasks: The human MT complex (hMT+), a region well-known to be responsive to visual motion, is also active during tactile motion perception (Hagen et al., 2002; Blake et al., 2004; Summers et al., 2009). This region is sensitive to auditory motion as well (Poirier et al., 2005), but not to arbitrary cues for auditory motion (Blake et al., 2004). Together, these findings suggest that hMT+ functions as a modality-independent motion processor. Parts of early visual cortex and a focus in the lingual gyrus are texture-selective in both vision and touch (Stilla & Sathian, 2008; Sathian et al., 2011; Eck et al., 2013), although one group found that haptically and visually texture-selective regions in medial occipitotemporal cortex were adjacent but non-overlapping (Podrebarac et al., 2014). Further, the early visual regions are sensitive to the congruence of texture information across the visual and haptic modalities (Eck et al., 2013), and information about haptic texture flows from somatosensory regions into these early visual cortical areas (Sathian et al., 2011). Both visual and haptic location judgments involve a dorsally directed pathway comprising cortex along the intraparietal sulcus (IPS) and that constituting the frontal eye fields (FEFs) bilaterally: the IPS is classically considered multisensory while the FEF is now recognized to be so. For both texture and location, several of these bisensory areas show correlations of activation magnitude between the visual and haptic tasks, indicating some commonality of cortical processing across modalities (Sathian et al., 2011).

Most research on visuo-haptic processing of object shape has concentrated on the lateral occipital complex (LOC), an object-selective region in the ventral visual pathway (Malach et al., 1995) that is also object- or shape-selective in touch (Amedi et al. 2001, 2002; James et al., 2002; Stilla & Sathian, 2008). The LOC responds to both haptic 3-D (Amedi et al., 2001, 2002; Stilla & Sathian, 2008) and tactile 2-D stimuli (Stoesz et al., 2003; Prather et al., 2004) but does not respond during auditory object recognition cued by object-specific sounds (Amedi et al., 2002). However, this region is activated when participants listen to the impact sounds made by metal or wood objects and categorize these sounds by the shape of the associated object (James et al., 2011). Auditory shape information can be conveyed by a visual-auditory sensory substitution device using a specific algorithm to convert visual information into an auditory stream or ‘soundscape’. Both sighted and blind humans can learn to recognize objects by extracting shape information from such soundscapes, albeit after extensive training; interestingly, the LOC responds to soundscapes after such training, but not when participants simply learn to arbitrarily associate soundscapes with particular objects (Amedi et al., 2007). Thus, the LOC can be regarded as processing geometric shape information independently of the sensory modality used to acquire it, similar to the view of hMT+ as processing modality-independent motion information (see above).

Apart from the LOC, visuo-haptic responses have also been observed in several, classically multisensory, parietal regions, including multiple loci along the IPS (Grefkes et al., 2002; Saito et al. 2003; Stilla & Sathian, 2008). Given that many of these IPS regions are involved in discrimination of both visual and haptic location of object features, their responsiveness during shape perception may be concerned with reconstruction of global shape representations from object parts (Sathian et al., 2011). The postcentral sulcus (PCS), which corresponds to Brodmann’s area 2 of primary somatosensory cortex (S1) (Grefkes et al., 2001), also shows visuo-haptic shape-selectivity (Stilla & Sathian, 2008).
It is critical to determine whether haptic or tactile recruitment of visual cortical areas is functionally relevant, i.e. whether these regions are actually necessary for task performance. Although research along these lines remains sparse, there is some evidence in support of this idea. Firstly, case studies indicate that the LOC is necessary for both haptic and visual shape perception: A lesion to the left occipito-temporal cortex, which likely included the LOC, resulted in both tactile and visual agnosia even though somatosensory cortex and basic somatosensory function were intact (Feinberg et al., 1986). Another patient with bilateral LOC lesions was unable to learn new objects either visually or haptically (James et al., 2006, 2007). Transcranial magnetic stimulation (TMS) is a technique used to temporarily deactivate specific, functionally defined, cortical areas, i.e. to create ‘virtual lesions’ (Sack, 2006). TMS over the parieto-occipital region active during tactile grating orientation discrimination (Sathian et al., 1997) interfered with performance of this task (Zangaladze et al., 1999) indicating that this area is functionally, rather than epiphenomenally, involved in the task. More work is necessary to fully test the dependence of haptic perception on classic visual cortical areas.

**Why are visual cortical regions active during touch?**

Activation of the LOC and other visual cortical areas during touch might arise from direct, “bottom-up” or “feedforward” somatosensory input. Human electrophysiological studies are consistent with this possibility: activity in somatosensory cortex propagates to the LOC as early as 150ms after stimulus onset (Lucan et al., 2010; Adhikari et al., 2014) in a beta-band oscillatory network (Adhikari et al., 2014). This might reflect cortical pathways between primary somatosensory and visual cortices demonstrated in the macaque (Négyessy et al., 2006). A case study is also illuminating: a patient with bilateral ventral occipito-temporal lesions, with sparing of the dorsal part of the LOC that likely included the multisensory sub-region, showed visual agnosia but intact haptic object recognition associated with activation of the intact dorsal part of the LOC, suggesting that somatosensory input could directly activate this region (Allen & Humphreys, 2009). Thus, neocortical regions classically considered to engage in unisensory visual processing may in actuality integrate multisensory inputs.

Studies of congenitally or early blind individuals are consistent with the notion that many classical visual cortical areas are modality-independent but task-specific. Thus, non-visual stimuli in the early blind, and visual stimuli in the sighted, activate the same extrastriate cortical regions on comparable tasks (reviewed by Sathian, 2014). For instance, an area known as the visual word-form area in the left fusiform gyrus, which responds selectively to visually presented words in the sighted, is also sensitive to Braille words (Reich et al., 2011). Congenitally blind people also engage visual cortical regions that are not active during corresponding tasks in the normally sighted population, most interestingly, visual cortical areas located at the site of primary (and adjacent non-primary) visual cortex of the sighted, i.e. medial occipital cortex (reviewed in Pascual-Leone et al., 2005; Sathian, 2005; Sathian & Lacey, 2007; Pavani & Röder, 2012; Sathian, 2014). As pointed out earlier (Sathian, 2014), it cannot be assumed that primary visual cortex occupies exactly the same anatomical extent in those who are born without vision as in normally sighted people. Across a host of studies, a wide range of non-visual tasks has been reported to recruit medial
occipital cortex in the congenitally blind but not the sighted – these tasks include somatosensory, auditory and language tasks (Pavani & Röder, 2012; Sathian, 2014). However, each study has typically focused on just one or a few tasks, so how these different functional domains are organized in visual cortex of the congenitally blind remains essentially unknown. Further, the idea that blind people are superior to their sighted counterparts on non-visual tasks is not a universal truth; rather, the evidence pooled over many studies suggests that their superiority reflects the specifics of their experience in the absence of vision (Sathian & Lacey, 2007; Pavani & Röder, 2012; Sathian, 2014).

An alternative to feedforward activation of visual cortex by tactile inputs is that haptic perception might evoke visual imagery of the felt object resulting in “top-down” activation of the LOC by “feedback” connections from higher-order areas (Sathian & Zangaladze, 2001). In keeping with this hypothesis, many studies demonstrate LOC activity during visual imagery: During auditorily-cued mental imagery of the shape of familiar objects, both blind and sighted participants show left LOC activation, where shape information would arise mainly from haptic experience for the blind and mainly from visual experience for the sighted (De Volder et al., 2001). The left LOC is also active when geometric and material object properties are retrieved from memory (Newman et al., 2005) and haptic shape-selective activation magnitudes in the right LOC were highly correlated with ratings of visual imagery vividness (Zhang et al., 2004). Even early visual cortical areas have been reported to respond to changes in haptic shape (Snow et al., 2014); however, as with other studies of crossmodal recruitment of visual cortex, it is not possible to exclude visual imagery as an explanation. A counter-argument is that visual imagery cannot explain haptically-evoked LOC activity. In support of this, LOC activity was found to be substantially lower during visual imagery compared to haptic shape perception (Amedi et al., 2001); however, this study did not verify that participants actually engaged in imagery throughout the scan. Another argument against the role of visual imagery is based on the observations that early- as well as late-blind individuals show haptic shape-related LOC activation (Pietrini et al., 2004). While the early blind clearly do not experience visual imagery, these findings do not necessarily rule out a visual imagery explanation in the sighted, given the extensive consequences of visual deprivation on neocortical organization (see above).

Recently, multivariate pattern analyses of voxel-wise activity have been used to demonstrate that activity patterns in primary sensory cortices can differentiate stimuli presented in modalities other than the canonical one. Thus, S1 activity could distinguish between objects in video clips that were being haptically explored, although there was only visual and no somatosensory input (Meyer et al., 2011). Along similar lines, stimulus modality (tactile, pain, auditory or visual) could be decoded in primary sensory cortices (S1, primary visual cortex or primary auditory cortex) regardless of their canonical associations (Liang et al., 2013). These studies provide further evidence of widely distributed multisensory processing in the neocortex; however, it remains uncertain whether the observed non-canonical activity arises from feedforward or feedback projections.
Integration of visual and tactile inputs

Vision and touch share many similarities in the way stimuli are processed. Ahissar and Arieli (2001) proposed that visual and tactile systems use analogous spatiotemporal coding schemes. Consistent with this idea, single neurons in macaques are similarly tuned for curvature direction at intermediate levels of the processing hierarchy in both visual and somatosensory cortex (areas V4 and S2) (Yau et al., 2009). Such similarities in coding lend themselves to multisensory integration. Evidence that visual and tactile inputs are indeed integrated comes from studies showing that the orientation of a tactile grating can disambiguate binocular rivalry (Lunghi et al., 2010), and that tactile motion can bias visually perceived bistable alternations of motion direction (James & Blake, 2004). Ernst and Banks (2002) demonstrated that humans integrate visual and haptic inputs in a manner that turns out to be statistically optimal, with the dominant modality being the one associated with lower variance in its estimates. Thus, vision tends to be dominant when assessing object shape but not surface texture (Klatzky et al., 1987). Statistically optimal integration is probably learnt during development, since it is not apparent in the first decade of life (Gori et al., 2008). A dramatic illustration of the importance of multisensory integration in body ownership is offered by the rubber-hand illusion (RHI), in which a viewed rubber hand is aligned with one’s own hand screened from view; when both the rubber hand and the real hand are tapped synchronously, the rubber hand is rapidly incorporated into the body image and perceived as one’s own (Botvinick and Cohen, 1998). The RHI can be induced in sighted people in the absence of vision, if an experimenter taps the subject on the real hand and synchronously moves the subject’s other hand to tap the rubber hand, suggesting that it is multisensory congruence of body-related information (in this case between tactile and proprioceptive inputs) that is critical, rather than visuo-tactile congruence specifically (Ehrsson et al., 2005). However, this version of the RHI is absent in early blind people, pointing to potential differences in multisensory integrative processes as a consequence of early visual deprivation (Petkova et al., 2012). In a variant of the RHI, one can be induced to perceive a third arm (Guterstam et al., 2011), and visuotactile conflicts can disrupt the feeling of ownership of one’s own limb (Gentile et al., 2013). The illusion has even been extended to the entire body using head-mounted virtual reality displays to create an out-of-body experience (Ehrsson, 2007; Lenggenhager et al., 2007).

There has been some study of the neural processes underlying visuo-tactile integration, although a comprehensive account is not yet feasible. Most neurons in the ventral intraparietal areas (VIP) of macaques exhibit modulation of their responses by bisensory stimulation, even when their overt responses to unisensory stimuli are limited to either vision or touch (Avillac et al., 2007). Similarly, bisensory stimuli in rats augment the somatosensory evoked response and reset the phase of induced network oscillations (Sieben et al., 2013). In the putative human homolog of VIP, topographic maps of tactile and proximal visual stimuli are aligned (Sereno & Huang, 2006), although at a single neuron level in macaques, the reference frame for tactile stimulation is head-centered whereas that for visual stimuli varies between head-centered, eye-centered or intermediate (Avillac et al., 2005). In one study, visuo-haptic responses were enhanced in the LOC and IPS by stimulus salience (Kim & James, 2010), although a subsequent study by the same group showed that
when spatiotemporal congruence was maximized across modalities, the inverse effectiveness pattern characteristic of classic neurophysiologic studies of multisensory integration emerged in the LOC and IPS as well as parietal opercular cortex, all on the left (Kim et al., 2012). Visuo-haptic responses in perirhinal cortex are also sensitive to the congruence of stimuli across modalities (Holdstock et al., 2009). The RHI appears to have a different neural substrate, being associated with activity in ventral premotor cortex (Ehrsson et al., 2004), although IPS and cerebellar activity is also reported (Ehrsson et al., 2004, 2005). Repetitive TMS (rTMS) over the left anterior IPS impairs visual-haptic, but not haptic-visual, shape matching using the right hand (Buelte et al. 2008), while rTMS over occipitotemporal cortex affects the Müller-Lyer illusion regardless of whether it is induced visually, haptically or cross-modally (Mancini et al., 2011) – these studies imply that the multisensory convergence reported in the preceding studies is functionally relevant. Further study of visuo-tactile integration and its neural substrates is warranted.

Individual differences

Two kinds of visual imagery have been described: “object imagery” (involving pictorial images that are vivid and detailed, dealing with the literal appearance of objects in terms of shape, color, brightness, etc.) and “spatial imagery” (involving schematic images more concerned with the spatial relations of objects, their component parts, and spatial transformations) (Kozhevnikov et al., 2002, 2005; Blajenkova et al., 2006). An experimentally important difference is that object imagery includes surface property information while spatial imagery does not. To establish whether object and spatial imagery differences occur in touch as well as vision, we required participants to discriminate shape across changes in texture, and texture across changes in shape (Figure 1), in both visual and haptic within-modal conditions. We found that spatial imagers could discriminate shape despite changes in texture but not vice versa, presumably because their images tend not to encode surface properties. By contrast, object imagers could discriminate texture despite changes in shape, but not the reverse (Lacey et al., 2011), indicating that texture, a surface property, is integrated into their shape representations. In addition, greater preference for object imagery was associated with greater impairment by texture changes (Lacey et al., 2011). Thus, the object-spatial imagery continuum characterizes haptics as well as vision, and individual differences in imagery preference along this continuum affect the extent to which surface properties are integrated into object representations (Lacey et al., 2011).

Cross-modal visuo-haptic object recognition, while fairly accurate, comes at a cost compared to within-modal recognition (Bushnell & Baxt 1999; Casey & Newell, 2007; Lacey et al., 2007). Importantly, while within-modal recognition of a set of previously unfamiliar and highly similar objects is view-dependent in both vision and touch, cross-modal recognition of these objects turns out to be view-independent (Lacey et al., 2007). Moreover, training in either the visual or the haptic modality to induce view-independence in the trained modality automatically confers view-independence in the untrained modality, and cross-modal training yields view-independence in each modality, implying that unisensory, view-dependent representations converge onto a bisensory, view-independent representation, possibly in the LOC (Lacey et al., 2009b). Further, spatial imagery preference correlates with the accuracy of cross-modal object recognition (Lacey et al.,
It appears, then, that the multisensory representation has some features that are stable across individuals, like view-independence, and some that vary across individuals, such as integration of surface property information and individual differences in imagery preference.

**A model of visuo-haptic multisensory object representation**

We now describe a model of visuo-haptic multisensory object representation (Lacey et al., 2009a; Lacey et al., 2014) and review the evidence for this model from studies designed to explicitly test the visual imagery hypothesis discussed above (Lacey et al., 2010; Lacey et al., 2014). In this model, object representations in the LOC can be flexibly accessed either bottom-up or top-down, independently of the input modality, and object familiarity plays a modulatory role. There is no stored representation for unfamiliar objects so that during haptic recognition, an unfamiliar object has to be explored in its entirety in order to compute global shape and to relate component parts to one another. We proposed (Lacey et al., 2009a) that this occurs in a bottom-up pathway from somatosensory cortex to the LOC, with involvement of the IPS in computing part relationships and thence global shape, facilitated by spatial imagery processes. For familiar objects, global shape can be inferred more easily, perhaps from distinctive features or one diagnostic part, and we suggested (Lacey et al., 2009a) that haptic exploration rapidly acquires enough information to trigger a stored visual image and generate a hypothesis about its identity, involving primarily object imagery processes via a top-down pathway from prefrontal cortex to LOC, as has been proposed for vision (e.g., Bar, 2007).

We tested this model by directly comparing activations and effective connectivity during haptic shape perception and both visual object imagery and spatial imagery (Lacey et al., 2010; Lacey et al., 2014), reasoning that reliance on similar processes across tasks would lead to correlations of activation magnitude across participants, as well as similar patterns of effective connectivity across tasks. In contrast to previous studies, we ensured that participants engaged in the desired kind of imagery throughout each scan by using appropriate tasks and recording responses. Participants also performed haptic shape discrimination using either familiar or unfamiliar objects. We found that object familiarity modulated inter-task correlations: Of eleven regions common to visual object imagery and haptic perception of familiar shape, six (including bilateral LOC) showed inter-task correlations of activation magnitude. By contrast, object imagery and haptic perception of unfamiliar shape shared only four regions, only one of which (an IPS region) showed an inter-task correlation (Lacey et al., 2010). Relatively few regions showed inter-task correlations between spatial imagery and haptic perception of either familiar or unfamiliar shape, with parietal foci featuring in both sets of correlations (Lacey et al., 2014). This suggests that spatial imagery is relevant to haptic shape perception regardless of object familiarity (contrary to the initial model), whereas object imagery is more strongly associated with haptic perception of familiar, than unfamiliar, shape (in agreement with the initial model). However, it remains possible that the parietal foci showing inter-task correlations between spatial imagery and haptic shape perception reflect spatial processing more generally, rather than spatial imagery per se (Jäncke et al., 2001; Lacey et al., 2014), or generic imagery processes, e.g., image generation, common to both object and spatial imagery (Lacey et al., 2014; Mechelli et al., 2004).
We also conducted effective connectivity analyses, based on the inferred neuronal activity derived from deconvolving the hemodynamic response out of the observed BOLD signals (Lacey et al., 2014). These analyses supported the broad architecture of the model, showing that the spatial imagery network shared much more commonality with the network associated with unfamiliar, compared to familiar, shape perception, while the object imagery network shared much more commonality with familiar, than unfamiliar, shape perception (Lacey et al., 2014). More specifically, the model proposes that the component parts of an unfamiliar object are explored in their entirety and assembled into a representation of global shape via spatial imagery processes (Lacey et al., 2009a). Consistent with this, in the parts of the network common to spatial imagery and unfamiliar haptic shape perception, the LOC is driven by parietal foci, with complex cross-talk between posterior parietal and somatosensory foci. These findings fit with the notion of bottom-up pathways from somatosensory cortex and a role for cortex in and around the IPS in spatial imagery (Lacey et al., 2014). The IPS and somatosensory interactions were absent from the sparse network shared by spatial imagery and haptic perception of familiar shape. By contrast, the relationship between object imagery and familiar shape perception is characterized by top-down pathways from prefrontal areas reflecting the involvement of object imagery (Lacey et al., 2009a). Supporting this, the LOC was driven bilaterally by the left inferior frontal gyrus in the network shared by object imagery and haptic perception of familiar shape, while these pathways were absent from the sparse network common to object imagery and unfamiliar haptic shape perception (Lacey et al., 2014).

Figure 2 shows the current version of our model for haptic shape perception in which the LOC is driven bottom-up from primary somatosensory cortex as well as top-down via object imagery processes from prefrontal cortex, with additional input from the IPS involving spatial imagery processes (Lacey et al., 2014). We propose that the bottom-up route is more important for haptic perception of unfamiliar than familiar objects, whereas the converse is true of the top-down route, which is more important for haptic perception of familiar than unfamiliar objects. It will be interesting to explore the impact of individual preferences for object vs. spatial imagery on these processes and paths.

ACKNOWLEDGMENTS

Support to KS from the National Eye Institute at the NIH, the National Science Foundation, and the Veterans Administration is gratefully acknowledged.

REFERENCES


James, TW.; James, KH.; Humphrey, GK.; Goodale, MA. Do visual and tactile object representations share the same neural substrate?. In: Heller, MA.; Ballesteros, S., editors. Touch and Blindness: Psychology and Neuroscience. Lawrence Erlbaum Associates; Mahwah, NJ: 2006. p. 139-155.


Figure 1.
(A) Schematic example of Shapes 1 and 2 with original texture schemes (left pair) and the texture schemes exchanged (right pair). (B) Example of Textures 1 and 2 with original shapes (left pair) and the shapes exchanged (right pair). (From Lacey et al., 2011).
Figure 2.
Schematic model of haptic object representation in LOC modulated by object familiarity and imagery type. For unfamiliar more than familiar objects, the LOC is driven bottom-up from somatosensory cortex (S1) with support from spatial imagery processes in the IPS. For familiar more than unfamiliar objects, the LOC is driven top-down from prefrontal cortex (PFC) via object imagery processes. The LOC thus houses an object representation that is flexibly accessible, both bottom-up and top-down, and which is modality- and possibly view-independent. (From Lacey et al., 2014).