RESEARCH REPORT

Thomas Thesen · Jonas F. Vibell · Gemma A. Calvert
Robert A. Österbauer

Neuroimaging of multisensory processing in vision, audition, touch, and olfaction

Abstract The development of neuroimaging methods has had a significant impact on the study of the human brain. Functional MRI, with its high spatial resolution, provides investigators with a method to localize the neuronal correlates of many sensory and cognitive processes. Magneto- and electroencephalography, in turn, offer excellent temporal resolution allowing the exact time course of neuronal processes to be investigated. Applying these methods to multisensory processing, many research laboratories have been successful in describing cross-sensory interactions and their spatio-temporal dynamics in the human brain. Here, we review data from selected neuroimaging investigations showing how vision can influence and interact with other senses, namely audition, touch, and olfaction. We highlight some of the similarities and differences in the cross-processing of the different sensory modalities and discuss how different neuroimaging methods can be applied to answer specific questions about multisensory processing.

Keywords Multisensory · Auditory · Visual · Tactile · Smell

T. Thesen (✉) · G. A. Calvert
University Laboratory of Physiology,
University of Oxford, Oxford, OX1 3PT UK
E-mail: thomas.thesen@physiol.ox.ac.uk

J. F. Vibell
Department of Experimental Psychology,
University of Oxford, Oxford, OX1 3UD, UK

R. A. Österbauer · T. Thesen · G. A. Calvert
Centre for Functional Magnetic Resonance Imaging
of the Brain (FMRIB), University of Oxford,
Oxford, OX3 9DU, UK

Introduction

The human brain samples its surroundings by virtue of a number of distinct peripheral sensors, each sensitive to very different forms of energy and chemicals in the environment. These signals are transduced at the peripheral level into neural impulses that are further relayed to the central nervous system, each sense separately. In the brain, signals from the individual sensory streams are processed simultaneously in a modality-specific fashion to create a sensory representation of the external world.

However, the brain also possesses synergistic properties to detect changes in the environment optimally and resolve ambiguity in a most adaptable fashion. In humans, the burgeoning psychological literature on this topic demonstrates such multisensory effects across many sensory modalities. Investigations to date have mainly focused on interactions between our undoubtedly dominant and most studied sense, vision, and other modalities including audition, touch, and olfaction, leading to the discovery of many perceptual alterations and behavioural performance changes associated with multisensory stimulus processing.

The illumination of the underlying physiological mechanisms by which the human brain accomplishes these tasks has largely been restricted to analogies to animal studies, derived mainly from invasive electro-physiological recordings from individual neurons (Stein and Meredith 1993; King and Palmer 1985). However, recent developments in non-invasive neuroimaging, such as functional magnetic resonance imaging (fMRI; Jezzard et al. 2001) and positon emission tomography (PET; Haxby et al. 1991), both indirect measures of neuronal activity with high spatial and low temporal resolution, and electro- and magnetoencephalography (EEG, MEG; Regan 1989), both direct measures of neuronal population activity with high temporal and low spatial resolution, have yielded a considerable amount of data on multisensory processing in the human brain.

In this review, we discuss recent human neuroimaging data on the interactions of the visual system with other sensory modalities, specifically with audition, touch, and olfaction. We briefly discuss some behavioural and perceptual effects of sensory interactions reported in the psychological literature and then focus on the underlying neural systems and mechanisms that govern them. Where applicable, we will discuss underlying neuroanatomy and neuronal connectivity, as well as research methodology.

Auditory–visual interactions

The interaction and integration of auditory–visual (AV) signals have been well documented behaviourally and show dependence on some coherence in stimulus features, predominantly on temporal and spatial coincidence (Welch and Warren 1986; Stein and Meredith 1993). In this section, we will focus on selected neuroimaging results showing interactions between simple AV stimuli and then summarize recent results from studies of audio-visual speech processing.
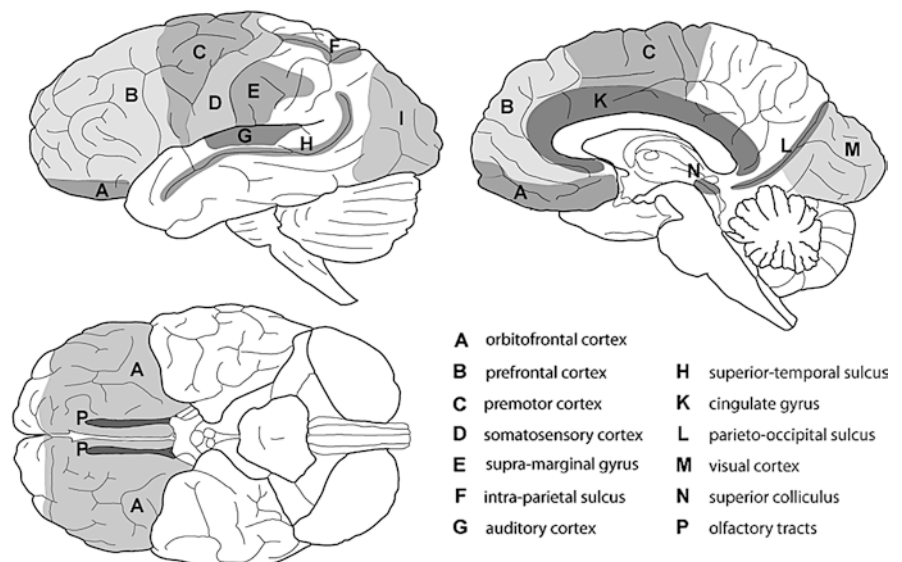
Early versus late modulations

Traditionally, multisensory processing in the cortex has been assumed to occur in specialized cortical modules relatively late in the processing hierarchy and only *after* unimodal sensory processing in the so-called 'sensory-specific' areas. This view developed mainly from the long tradition of studying the sensory systems in isolation but was also supported by evidence from multisensory studies in animals, showing multisensory convergence zones in 'polysensory' areas in parietal (Duhamel et al. 1991), temporal (Bruce et al. 1981), and frontal (Benevento et al. 1977) areas of the brain. However, recent functional imaging studies suggest that the senses can

influence each other even at the earliest levels of cortical processing, that is, at the level of the primary sensory cortices. Many of these experiments employ a redundant target paradigm, which requires subjects to make speeded detection responses to spatially concordant audio–visual targets and to the corresponding unimodal auditory and visual signals (Miller 1982). Measured reaction times (RTs) are shorter to the bimodal event as compared to the unimodal counterparts in isolation, a robust effect that has been consistently replicated (Welch and Warren 1986).

The neural correlates of the redundant target effect during object recognition and detection were investigated by Giard and colleagues (Giard and Peronnet 1999; Fort et al. 2002). For example, subjects discriminated two objects based on their visual, auditory, or auditory–visual features while scalp EEG activity was recorded simultaneously (Fort et al. 2002). As expected, RTs to the bimodal presentation (247 ms) were significantly shorter than to either auditory (276 ms) or visual (310 ms) cues alone. To determine superadditive integration effects, the sum of the event-related potential (ERP) waveforms to both unimodal stimuli was subtracted from the activity to bimodal stimulation, AV−(A + V) [see Calvert and Thesen (2004) for a discussion of different analysis strategies]. The high temporal resolution of EEG allowed the investigators to detect a significant interaction effect in the ERP waveform at around 50 ms post-stimulus over the occipital cortex with scalp distributions typical of activities in the visual cortex (for anatomical regions see Fig. 1; for interaction effects see Fig. 2). The topography and timing of this effect suggest a mechanism for modulation of the visual cortex by auditory input at early stages in visual processing before a complete sensory analysis of the stimulus has been achieved.

Similarly, Molholm et al. (2002) recorded EEG activity during a typical redundant target paradigm. Cortical interactions were detected at 46 ms at posterior
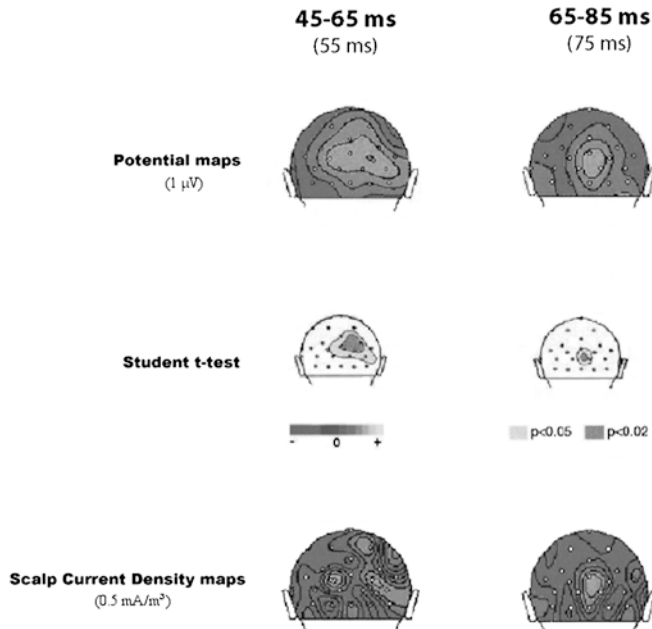
Fig. 1 Schematic drawings of the human brain from three perspectives. *Upper left* Lateral view on the outer surface of the left hemisphere. *Upper right* Sagittal view at the brain midline. *Lower left* Ventral view showing the brain from below. *Shaded areas* correspond to brain structures discussed in the text

A	orbitofrontal cortex		
B	prefrontal cortex	**H**	superior-temporal sulcus
C	premotor cortex	**K**	cingulate gyrus
D	somatosensory cortex	**L**	parieto-occipital sulcus
E	supra-marginal gyrus	**M**	visual cortex
F	intra-parietal sulcus	**N**	superior colliculus
G	auditory cortex	**P**	olfactory tracts

AUDITORY-VISUAL INTERACTIONS
(AV) - (A + V)
Occipital-Parietal Sites

Fig. 2 Event-related brain potential topography and scalp current density (SCD) maps at occipital–parietal sites during time windows of auditory–visual (AV) interaction (maps depicted at 55 ms and 75 ms), including Student *t*-test maps depicting significant AV interactions. Significant positive potential fields between 45 and 85 ms post-stimulus show patterns typical of activities in the visual cortex. Over this period, the SCD maps present two different current patterns (45–65 and 65–85 ms) suggesting the existence of several interaction components. Data from Fort et al. 2002

electrodes over the visual cortex corresponding to the initial visual responses in the visual cortex occurring at around 40–55 ms post-stimulus in the form of the C1 ERP wave (Clark et al. 1995). These effects show that AV multisensory interaction can occur as early in the processing hierarchy as the sensory analysis stage in the primary visual cortex, an area previously thought to process visual information only.

Indeed, anatomical evidence supporting a direct and early influence of audition on visual processing has been reported in a recent tracer study in non-human primates by Falchier et al. (2002) (see also Rockland and Ojima 2001). Retrograde labeling of core and parabelt areas of the auditory cortex revealed previously unknown direct neuronal connections to areas of the primary visual cortex. Interestingly, these projections terminated in an area of the primary visual cortex subserving the peripheral visual field. This finding corresponds well to Shams et al.'s (2001) observations that the illusory flash illusion (described below) is significantly stronger in the visual periphery both psychophysically and electrophysiologically. However, it remains elusive whether the modulation occurs directly in the primary visual cortex

or is a result of even earlier audio–visual interactions in the brain stem.

Under specific stimulus conditions, bisensory stimulation results not only in detection advantages but also in a perceptual modulation of visual experiences by auditory co-stimulation. One such example is the "illusory flash effect", an auditory–visual illusion first reported by Shams et al. (2000) where a phenomenological change in the perception of a visual stimulus is induced by sound. More specifically, when a single flash is presented concurrently with multiple short beeps, the single flash is perceived as multiple flashes. The illusion is so compelling and automatic that even observers informed about the physical nature of the stimulus report a visual experience of seeing multiple flashes. In a follow-up study, Shams et al. (2001) recorded flash-evoked ERPs and introduced a sound to elucidate the temporal and spatial occurrence of the auditory–visual interactions. In this paradigm, sound modulation affected activity at around 170 ms post-stimulus again at the level of visual cortex. However, this effect occurred at a considerably later time window than those observed by Fort et al. (2002) and Molholm et al. (2002) and is therefore suggestive of a feedback modulation from higher-order multisensory areas.

Speech

Similar cross-modal interaction and facilitation has been reported with more complex and socially relevant stimulation. Human speech perception is an inherently multisensory phenomenon, as a speaker's articulatory mouth movement is intrinsically correlated with the time-frequency sound waveform. Having information available from both auditory and visual sources can have profound effects on the resulting speech percept. For example, under noisy acoustic conditions, being able to see the talker's mouth movements can significantly enhance speech comprehension to the equivalent of changing the acoustic signal-to-noise ratio by 11 dB (i.e. increasing the physical intensity by a factor of four (Sumby and Pollack 1954; MacLeod and Summerfield 1987).

Concurrent visual speech information can also influence the resulting auditory speech percept, as is shown in the classical demonstration by McGurk and MacDonald (1976). Here, the incongruous pairing of an auditory /ba/ and a visual /ga/ results in the automatic perception of /da/, producing a very robust effect of auditory–visual integration in speech perception.

In a pioneering MEG study, Sams et al. (1991) showed that the mismatched negativity component of the auditory neuromagnetic response, usually localized within the auditory cortex, was sensitive to mouth movements. Functional MRI studies investigating the neural basis of cross-modal speech processing have shown that the auditory cortex can be activated by silent lipreading (Calvert et al. 1997; MacSweeney et al. 2000),

providing an avenue for visual speech to influence the perception of auditory speech at a pre-lexical stage in the primary sensory cortex.

In an early fMRI study using interaction effects to determine audio–visual integration sites in the human brain, Calvert et al. (2001) presented subjects with both unimodal as well as congruent and incongruent audio–visual speech in a multiplexed block design. In the congruent condition, subjects heard a human voice reading a story while simultaneously seeing the corresponding lip movements through a back-projection screen. In the incongruent condition, the audio and visual streams were shifted in time so that the visual information did not correspond to what the subjects heard. Using the additive model [AV−(A + V)] for determining cross-modal interactions for the first time with fMRI, Calvert et al. (2001) identified a network of brain areas showing super-additive response enhancements and response depressions for synchronous and asynchronous audio–visual speech inputs, including the intra-parietal sulcus, insula, superior colliculus, regions of the medial ventral and dorsal prefrontal cortex, and the superior temporal sulcus (STS). A similar role of the STS in the integration of arbitrary audio–visual information was found in a recent MEG study. Raij et al. (2000) investigated the human brain's audio–visual integration mechanisms for phonetic and graphemic representations of letters, which, in literate people, have been extensively paired through associative learning. Subjects were presented with auditory, visual, and auditory–visual letters in a target identification task while neuromagnetic (MEG) activity was recorded from the skull surface. Reaction time findings showed a clear behavioural advantage for audio–visual stimuli (425 ms) compared to auditory (505 ms) and visual (520 ms) stimuli alone. Time windows of audio–visual interaction were determined by calculating AV−(A + V) from the averaged evoked responses at around 380 ms after stimulus presentation, revealing the left posterior STS as a main area of AV convergence.

In summary, neuroimaging experiments have revealed audio–visual interaction patterns at various cortical locations and latencies extending to primary sensory cortices. Stimulus feature and task requirements play an important role in explaining the variability among findings and underline the dynamic nature of AV bisensory processing. Speech distinguishes itself as a form of multisensory processing not only behaviourally but also neurologically as it recruits specialized cortical areas for AV stimulus integration, but, together with simple AV stimuli, it shows also bisensory interactions in sensory-specific auditory and visual cortices.

Visuo–tactile interactions

This section describes current research focused on the reciprocal influence between touch and vision and its underlying physiological and anatomical connections as evidenced by neuroimaging in humans. Comparatively, visuo–tactile interactions have been studied less than those between vision and sound, but substantially more than visual–olfactory interactions. A sizeable literature is building up and is starting to allow us to gain a clearer picture of the brain regions and the functional connections involved in visuo–tactile interactions.

A simple example that illustrates visual interaction with tactile sensation is the Japanese illusion (Van Riper 1935). When one stretches out the arms straight ahead and turns the palms and the thumbs as far as possible outwards/downwards, then crosses the wrists and intertwines the fingers, one has a reverse view of the fingers. The hands are held as if in prayer except that the left hand is on the right side and vice versa. If the subject is now ordered by pointing to move a specific finger, the subject is very often unable to do so correctly. Van Riper (1935, p. 263) mentioned that he makes 'no attempt to develop a comprehensive theory of explanation' for the above phenomenon, but with today's neuroimaging techniques we are gaining a deeper understanding of where, why, and how such reciprocal influencing between vision and touch occurs.

Behavioural studies

The influence of visual information on other tactile perception has been well documented behaviourally (Tipper et al. 1998; Kennett et al. 2001b). In a temporal order judgement task, Spence et al. (2001) showed that subjects respond by an average of 53 ms faster to tactile than to visual stimulation. However, attention to touch had a greater modulating effect on judgements of simultaneity of visual and tactile stimuli than when vision had to be attended. A potential explanation for this asymmetry was suggested: 'visual dominance'—the inclination to rely more heavily on visual information when simultaneously presented with stimuli from other modalities. Accordingly, directing attention towards visual stimuli has less of an influence on performance than when touch is attended. Whether this bias depends on our greater everyday experience with the visual modality remains to be established.

Generally, vision has a stronger influence on touch than vice versa (Pavani et al. 2000; Maravita et al. 2002), even though touch also has the potential to modulate visual perception (e.g. Macaluso et al. 2000). A behavioural study allowing subjects to look at a non-stimulated part of the arm showed that tactile two-point discrimination thresholds are improved, but not when instead an object was placed at the same location (Kennett et al. 2001b). Interestingly, the performance-enhancing effect was further increased when the area of the arm was magnified.

A study by Tipper et al. (1998) showed that proprioception was not necessary for the enhancement of somatosensory detection by vision. A video camera showed a centralized picture of the hand concurrent with

tactile stimulation. Visual presentation of the stimulated hand improved detection of tactile targets in comparison to when the other hand was presented. Accordingly, it is not just seeing "a hand", but seeing the appropriate hand that facilitates detection. Since tactile information is represented somatotopically in the post-central gyrus while visual input is retinotopically organized in the occipital cortex, the brain must have a way to remap the location of the hand over different visual locations compared to the tactile stimulation. These behavioural studies show interaction of vision and touch and a dynamic relationship between their corresponding spatial maps, but they lack the ability of modern imaging methods to reveal the functional connectivity underlying these bisensory processes.

Neuroimaging

Much of the current visuo–tactile imaging research has focused on the cerebral sources of modalities and their interactions, the temporal dynamics of these interactions, and the reciprocal influence of vision and touch in conjunction with spatial locations. The imaging techniques at hand are differentially suited to tackle such questions and a few recent and notable examples are described below.

A basic fMRI experiment presented by Macaluso and Driver (2001) demonstrated modality-specific and spatially specific areas activated by tactile or visual stimulation. No responses were required while stimuli consisting of tactile stimulation with air-puffs were presented to the left and right thumbs and visual stimuli with LEDs were presented in the left or right hemifields. The activation from right tactile stimulation was subtracted from left tactile stimulation and that from right visual stimulation was subtracted from left visual stimulation. The remaining activations from both modalities were subtracted from each other to reveal spatially specific unimodal areas, by visual stimulation in the contralateral occipital areas and by tactile stimulation mainly in the contralateral post-central sulcus, as would be expected from both these modalities. Bimodal stimulation revealed areas specific to visuo–tactile spatial activation by comparing the sides of stimulation when collapsed across spatial location, while ensuring that both modalities contribute to the bimodal effect. Activation was found in the anterior part of the intraparietal sulcus at the junction with the post-central sulcus. Using the same data, but combining stimulus modality instead of spatial location, revealed activation in the same areas that showed modality-specific spatial effects, but also in a set of areas that did not show any effect of stimulated side. Visual stimulation collapsed across sides showed activation in the occipito–parietal junction and superior parietal gyrus, and tactile stimulation activated the parietal operculum and the insulae. Accordingly, these data show modular representations of both touch and vision in multimodal temporo-parietal areas and

spatially specific representations that are independent of stimulus modality.

Traditionally, the occipital and somatosensory cortices have been considered unimodal, that is, only modality-specific stimulation was thought to activate these areas. However, for the visual and tactile modalities, imaging studies challenge this assumption (Macaluso et al. 2000; Taylor Clarke et al. 2002). Macaluso et al. (2000) used event-related fMRI to look at the modulation of visual cortex by touch. Visual stimuli were presented randomly in either the left or the right hemifield. Half of the visual stimuli were paired with concurrent tactile stimulation to the right hand. The haemodynamic response showed that concurrent non-predictive tactile stimulation amplified the visual activation in what is traditionally considered to be a unimodal visual area, the lingual gyrus in the visual cortex. This amplification could only be observed when the tactile stimuli were presented on the same side as the visual stimulation and was therefore considered spatially specific. Macaluso et al. (2000) hypothesized that modulations of activity in the lingual gyrus originated from back-projections from higher-level multimodal areas. A subsequent test of effective connectivity between activity in the lingual gyrus and that in other brain regions found the strongest correlation with the anterior part of the supramarginal gyrus in the inferior parietal lobe. Accordingly, this area is thought to combine inputs from the unimodal visual areas with those from the somatosensory post-central gyrus. The correlation does not, however, show a direct causal influence, that is, back-projection, especially with the poor temporal resolution of fMRI. EEG or MEG, with their temporal resolution to the order of 1 ms, is better suited to approach this question.

A reciprocal influence of vision on activity in the somatosensory cortex has also been demonstrated in EEG measures. In an ERP study, Taylor Clarke et al. (2002) found that viewing your arm increases ERP activation from simultaneous tactile stimulation during a tactile discrimination task. The first P50 component relating to the initial somatosensory input was not modulated by viewing the arm, but the subsequent N80 component, also initiating in the somatosensory cortex, showed visual modulation. Thus, early interactions between touch and vision do seem to occur even in unimodal areas.

Several different approaches have been adapted to the resolution of the discrepancy in the internal representations of the body between the visual and tactile modalities using a variety of neuroimaging techniques (see Driver and Spence 1998). This issue is of special interest due to the differing representation of the body (somatotopic vs. retinotopic), which still gives a coherent picture of the world. One way to investigate these differences is to compare the neural activations when visual and tactile stimuli interact under varying body postures (Kennett et al. 2001a; Lloyd et al. 2003).

Using ERPs and an orthogonal cuing paradigm, where the response dimension (up/down) differs from

the direction of tactile spatial cuing (left/right), Kennett et al. (2001a) looked at the influence of non-predictive tactile stimuli (taps from solenoids) on subsequent visual events (blinks from LEDs). The arms were either uncrossed or crossed to evaluate tactile stimuli in the same or the opposite visual field. Unsurprisingly, the uncrossed condition showed visual stimuli eliciting larger cuing effects when visual and tactile stimuli occurred on the same side of space. In the crossed arms condition ERP modulation of the visual events was shown to favour not the same side of space as the tactile stimulation as evidenced when the arms were uncrossed, but instead the external location of the tactile stimulation. Concurrent ERP recordings revealed a greater amplitude modulation for the visual event, primarily ipsilateral to the visual stimulation, when it occurred in the same external location as the tactile cue, supporting the behavioural findings. Thus, the external location of the tactile cues as opposed to the hemispheric projection seems to guide the visual modulation, indicating some remapping between the modalities.

Lloyd et al. (2003) presented an interesting study furthering our understanding of the influence of vision upon the internal map of the body during tactile stimulation. Using fMRI, the study showed that different brain regions were activated when the right arm was placed across the body midline during tactile stimulation. When the arm was in the right hemispace and the eyes were closed the right ventral intraparietal sulcus of the parietal cortex was activated. When the participant kept their eyes open activation in the ventral intraparietal sulcus shifted to the contralateral side. Accordingly, visual cues modulated the perceived limb position by activating mirroring brain regions across hemispheres. These findings suggest vision modulates the internal representation of the hand regardless of its external location. In summary, the present neuroimaging studies have helped to reveal a picture of visuo–tactile interactions as a dynamic process involving areas extending into the primary cortices with an adaptive correspondence between their spatial maps mainly dominated by vision.

Visual–olfactory interactions

Behavioural evidence

In this section we discuss the roles that visual cues play in our perception of odours and flavours. The perception of the foods and beverages we consume typically involves the integration of multiple sensory inputs relating to the smell, taste, look, texture, temperature, and trigeminal attributes of a product. Together, these multisensory interactions play a critical role in determining a unified flavour percept and also influence the perceived pleasantness of the foods we consume. However, of the various visual cues that contribute to flavour perception, colour has been found to have an especially potent effect.

Adding colour to an odourous solution, for example, can alter the perceived intensity, pleasantness, and identity of the odour, an effect that has been reported in numerous psychological studies (Moir 1936; Christensen 1985; Morrot et al. 2001). Interestingly, people are even far more likely to report an *odourless* solution as odourous if it is coloured than if it is colourless (Engen 1972), suggesting that people expect a coloured substance to have a smell. Similarly, studies by Zellner and colleagues (Zellner and Kautz 1990; Zellner et al. 1991) found that adding colour (e.g. red) to an odourous liquid (e.g. strawberry) enhances the perceived intensity and pleasantness of its odour. This perceptual effect is so strong that some participants in their study simply refused to believe that coloured and uncoloured solutions of equal odour intensity were actually equally strong.

There are, however, even more dramatic accounts of the effect colouring can have on the pleasantness of food. In one extreme example reported by Wheately (1973), a group of people were given a meal of steak, chips, and peas. The participants initially enjoyed eating the food that looked normal under special lighting that showed the food on the plates in front of them but not its 'real' colour. However, after they had consumed some of the food the lighting was returned to normal, which resulted in people suddenly realizing that the steak they were eating was actually blue, the chips green, and the peas red! According to Wheately (1973), on seeing the food's real colour, almost all of the participants were violently sick. These behavioural examples clearly demonstrate the vital role that colour cues play in olfactory and flavour perception. It is worth noting, however, that it is not just the colour of food that affects its palatability but also other visual features such as shape (Rolls et al. 1982) and the context in which it is presented.

Anatomy and physiology

Firstly, we will briefly describe the anatomical pathways and physiological basis for the effects of visual–olfactory interaction described above, based predominantly on studies in monkeys. Sensory information from the olfactory receptor neurons is relayed to the olfactory bulb from where it reaches the brain in several distinct areas: the piriform cortex located at the junction of the frontal and medial lobes and considered the primary olfactory cortex, the olfactory tubercle from where a projection goes to the thalamus, the amygdala, and the entorhinal cortex. The first three pathways all project to various regions within the orbitofrontal cortex (OFC), which consists of the ventral aspects of the frontal lobes (Barbas 1993). It is thought that the conscious perception of scents and aromas, as well as other higher-order olfactory information processing, occurs within the OFC (Rolls 2001) and it is hence a likely region where visual–olfactory interactions take place. Another main anatomical region for multisensory interactions of

olfaction and vision is the amygdala, which is located in the temporal lobes adjacent to the primary olfactory cortex and receives convergent projections from the retina and from the olfactory bulbs (Cooper et al. 1994).

Both the orbitofrontal cortex and amygdala receive highly processed inputs from all sensory modalities and are also reciprocally connected with each other (Ongur and Price 2000). Anatomical studies in non-human primates have shown that areas 11 and 13 of the OFC receive major visual inputs from the anterior infero-temporal region (Barbas 1988) and that primary olfactory areas such as the piriform cortex have direct projections to posterior orbitofrontal regions (Barbas 1993). Neurophysiological recordings of orbitofrontal neurons in monkeys have also found neurons that respond to both the smell and the sight of food items (Rolls and Baylis 1994). Together, this anatomical and physiological evidence suggests that both the orbitofrontal cortex and amygdala are involved in the multisensory integration of visual and olfactory stimuli and are primarily responsible for the behavioural effects described above.

Neuroimaging

In humans, olfaction has been studied using non-invasive neuroimaging techniques such as fMRI, PET, and EEG. Using unimodal stimulation, it has been found that both amygdala and orbitofrontal cortex are particularly concerned with the reward value of stimuli and activate in response to pleasant and unpleasant visual, olfactory, and gustatory stimulation (Zald et al. 1998; Royet et al. 2000). Amygdala activation has been most frequently reported in studies using aversive stimulation and disgust (Zald 2003), suggesting that this structure may underlie some of the dramatic effects observed in Wheately's (1973) study. However, a recent fMRI study by Anderson et al. (2003) suggests that the amygdala encodes smell intensity, whereas the hedonic valence of an odour is represented in the orbitofrontal cortex. Unfortunately, behavioural ratings of intensity and pleasantness were only collected after scanning and only two different odours (citral and valeric acid) were used at two different concentrations, so that the generality of these findings remains unclear. Nevertheless, the authors conclude that two dissociated circuits process intensity and valence of an olfactory stimulus, and it may well be that the behavioural effect of increased pleasantness and intensity of appropriately coloured odourous solutions and food items have their neural origins within the OFC and the amygdala, respectively.

Investigations of vision and olfaction in a multisensory context using neuroimaging techniques have so far been relatively rare. Grigor et al. (1995, 1999) used event-related potentials (ERPs) to investigate the effect of matching versus mismatching smell–picture combinations with olfaction as a prime for both food-related and non-food-related stimuli. In both studies, the olfactory prime was delivered for 4 s prior to the presentation of either a matching or a mismatching visual stimulus (e.g. the smell of a rose and the picture of a rose vs. the picture of a cricket bat). The results in both cases demonstrated a difference in the negative ERP waveform 400 ms after the onset of the visual stimulus (N400) when the odour and the picture did not match compared to the matching condition. Even though this effect was strongest over frontal electrodes, the authors did not specify whether the source of activation could originate from the orbitofrontal cortex. Additionally, since the two stimuli were presented consecutively and not simultaneously it seems likely that the observed effect could have been caused by stimulus incongruency rather than cross-modal stimulus integration.

A more direct investigation of colour–odour effects has been carried out in our laboratory using fMRI (Osterbauer et al. 2002). In this study, subjects were presented with either a full screen colour image (red, yellow, turquoise, or brown) or one of four smells (strawberry, lemon, spearmint, or caramel) or simultaneously with a combination of both. The participants in this study were asked to rate the 'goodness of match' for the various colour–smell combinations presented to them. A modulation of the fMRI signal according to the perceived congruency of colours and smells was found mainly in the orbitofrontal cortex but also in the insular cortex (Fig. 3). In addition to this, activation for matching colour–smell combinations exceeded the sum of the unimodal stimuli (superadditivity) in medial orbitofrontal areas, whereas incongruent combinations 'suppressed' activity below that of just smell alone. Whether these changes reflect the increase in pleasantness or in intensity or both, as has been reported in the psychological literature for appropriately coloured solutions, unfortunately cannot be determined unambiguously, since behavioural data for these parameters were not collected. Nevertheless, the study has provided preliminary evidence for a similar mechanism of integration for vision and smell in the human OFC as has been previously identified in monkey.

A somewhat separate question that arises in the context of these visual–olfactory interactions is how these cross-modal associations are acquired. A recent event-related fMRI experiment investigated the neuronal mechanisms responsible for olfactory learning with a classical conditioning paradigm (Gottfried et al. 2002). In this study three pictures of 'neutral' faces were repeatedly paired with any one of a pleasant, a neutral, or an unpleasant smell under a 50% reinforcement schedule while a fourth 'neutral' face remained unpaired (control condition). This study demonstrated that the face paired with a pleasant odour evoked significantly stronger activation in medial orbitofrontal regions, the ventral striatum, and the amygdala compared to the neutral (unpaired) face. In contrast, the face paired with the unpleasant odour resulted in peaks of activity in the left lateral and right medial orbital gyri. Even though some brain regions appeared to be specifically engaged
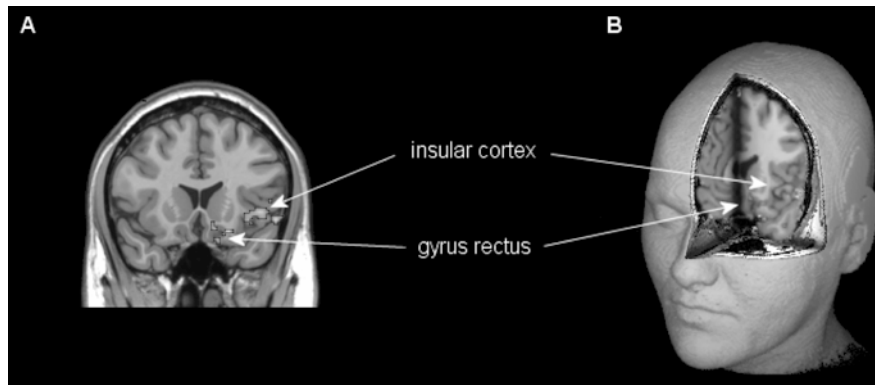
Fig. 3a,b Brain regions showing odour–colour modulation. The activation maps depict brain regions in which the blood oxygen level dependent (BOLD) signal correlates with the perceived odour-colour congruency. **a** Coronal slice at the Montreal Neurological Institute (MNI) standard coordinates of $y = 16$. The activated voxels are located in the left orbitofrontal cortex along the gyrus rectus (x, y, $z = -8$, 16, −24; Z-score = 3.34) and in the left insular cortex (x, y, $z = -32$, 22, −8; Z-score = 2.75). **b** For illustration purposes, the same activation map is rendered on a whole-brain cutout. All active voxels are rendered onto the standard MNI brain and the right side of the picture corresponds to the left cerebral hemisphere

in olfactory learning depending on the valence (pleasant vs. unpleasant) of the odours (such as the ventral striatum and the amygdala), widespread regions of the rostral and caudal OFC activated independently of odour valence. It is thus mainly the OFC that plays a critical role in the formation of associations between odours and visual stimuli. Taken together these neuroimaging studies indicate that the behavioural effects observed in psychological studies such as described above have a neuronal basis mainly within the human OFC and cannot simply be attributed to experimenter effects, demand characteristics, or biases.

Two interesting distinctions between olfactory–visual convergence and interactions and audio–visual or visual–tactile ones have emerged so far: firstly, to date, it appears that visual–odour interactions may be unidirectional in that, unlike the reciprocal influences in visual–auditory and visual–tactile interactions, stronger effects of vision on olfactory perception have been reported in both psychophysical and neuroimaging literature, but so far little is known about whether odours can influence visual perception. Secondly, these interactions appear to take place in secondary sensory areas such as the orbitofrontal cortex and to date we do not know whether there exists a modulation of primary sensory areas via back-projections from higher-order areas or from direct projections from the piriform cortex to visual cortex and vice versa, as has been recently indicated in the case of vision and audition and vision and touch (Falchier et al. 2002; Rockland and Ojima 2002).

A potential reason for this could be the poor spatial resolution of the human olfactory system, which does not enable us to accurately detect the source of an odour solely on the basis of olfactory information. Instead, we heavily rely on visual information to locate the origin of a smell. This of course means that it is not feasible to vary the spatial congruency of olfactory–visual stimuli in a fashion similar to audio–visual or visual–tactile experiments and therefore a direct comparison of the mechanisms of multisensory integration of different sensory combinations is not possible. However, an interesting question for future investigations could be to look at under exactly which stimulus conditions we perceive an odour as emanating from an object and not as an ambient smell.

Discussion

Ample evidence from behavioural studies and single-unit recordings in animals shows unambiguously that the brain is a multisensory processor in which inputs from different senses complement, modulate, and interact with each other. Here, we have presented selected studies on the neuronal correlates of these phenomena in humans. These discoveries are largely due to the recent developments of non-invasive neuroimaging methods such as fMRI and MEG/EEG, which made it possible to investigate sensory processing in alive and awake normal human subjects. Even though each of the sensory modalities we have discussed here (vision, audition, touch, and olfaction) exhibits distinct task-dependent interaction patterns with the other senses, investigators have begun to elucidate some of the general mechanisms by which these modulations manifest themselves in the human brain. Clearly, there are many outstanding questions left to be answered, but the application of neuroimaging methods has contributed to our understanding of polysensory processing and has the potential to contribute much more to the field of multisensory research.

In the functional imaging studies discussed in this article, the STS has been shown to be involved in the simultaneous perception of AV speech signals, but its exact functional architecture and the part it plays in multisensory integration remain elusive. Experiments employing the high spatial resolution of functional MRI can be used to tell us more about the properties and functions of the STS and other related multisensory areas. We have presented studies that showed for the first time that input from the auditory modality influ-

ences stimulus processing in the visual system as early as the primary visual cortex. The high temporal resolution of EEG/MEG can be exploited to determine the exact time course of these events and whether modulation can be due to re-entrant projections from higher multisensory areas or may occur via direct projections from a primary sensory area. Similar interactions can be evaluated further between vision and touch and vision and olfaction. Anatomical locations and temporal relations of functional processes in audiovisual interactions can subsequently be compared to the same processes between visual and tactile and visual and olfactory interactions to gain an understanding of the general processes underlying multisensory integration and facilitation.

A fundamental challenge in neuroscience is how information about the external world, arriving through the different senses and processed in sensory-specific cortical modules, is integrated to form a coherent perception of the external events. This so-called "binding problem", even though applicable equally to both unisensory and multisensory processing of information, has not been addressed adequately outside the visual system. Therefore, it is not yet clear if integration of sensory signals is primarily mediated by the temporal co-stimulation of a particular area by, for example, input from lower-level sensory areas, or if specific cortical mechanisms are necessary to achieve this cross-modal integration. Analysis methods investigating cortical synchrony of large neuronal populations (Singer and Gray 1995), detectable by MEG/EEG, for example, could reveal a likely candidate mechanism by which the brain binds together information across and within the senses.

In summary, further investigations into multisensory processing are likely to lead away from seeing the brain in a modular way where each sense is investigated in isolation, but to emphasize the interactive and integrative properties of the nervous system. Present neuroimaging methods will contribute to that understanding; however, the fast progression in the development of new techniques and their integration will undoubtedly provide us with completely new ways of approaching these questions and help us in our effort to create a better understanding of human sensation, perception, and cognition.

References

Anderson AK, Christoff K, Stappen I, Panitz D, Ghahremani DG, Glover G et al (2003) Dissociated neural representations of intensity and valence in human olfaction. Nat Neurosci 6:196–202

Barbas H (1988) Anatomic organization of basoventral and mediodorsal visual recipient prefrontal regions in the rhesus monkey. J Comput Neurol 276:313–342

Barbas H (1993) Organization of cortical afferent input to orbitofrontal areas in the rhesus monkey. Neuroscience 56:841–864

Benevento LA, Fallon J, Davis BJ, Rezak M (1977) Auditory–visual interaction in single cells in the cortex of the superior temporal sulcus and the orbital frontal cortex of the macaque monkey. Exp Neurol 57:849–872

Bruce C, Desimone R, Gross CG (1981) Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. J Neurophysiol 46:369–384

Calvert GA, Thesen T (2004) Multisensory integration: methodological approaches and emerging principles in the human brain. J Physiol Paris (in press)

Calvert GA, Bullmore ET, Brammer MJ, Campbell R et al (1997) Activation of auditory cortex during silent lipreading. Science 276:593–596

Calvert GA, Hansen PC, Iversen SD, Brammer MJ (2001) Detection of audio–visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. Neuroimage 14:427–438

Christensen CM (1985) Effect of color on judgments of food aroma and flavour intensity in young and elderly adults. Perception 14:755–762

Clark VP, Fan S, Hillyard SA (1995) Identification of early visual evoked potential generators by retinotopic and topographic analyses. Hum Brain Mapp 2:170–187

Cooper HM, Parvopassu F, Herbin M, Magnin M (1994) Neuroanatomical pathways linking vision and olfaction in mammals. Psychoneuroendocrinology 19:623–639

Driver J, Spence C (1998) Attention and the crossmodal construction of space. Trends Cogn Sci 2:254–262

Duhamel JR, Colby CL, Goldberg ME (1991) Brain and space. In: Paillard J (ed). Oxford University Press, New York, pp 223–236

Engen T (1972) The effect of expectation on judgments of odor. Acta Psychol 36:450–458

Falchier A, Clavagnier S, Barone P, Kennedy H (2002) Anatomical evidence of multimodal integration in primate striate cortex. J Neurosci 22:5749–5759

Fort A, Delpuech C, Pernier J, Giard MH (2002) Dynamics of cortico-subcortical cross-modal operations involved in audio–visual object detection in humans. Cereb Cortex 12:1031–1039

Giard MH, Peronnet F (1999) Auditory–visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. J Cogn Neurosci 11:473–490

Gottfried JA, O'Doherty J, Dolan RJ (2002) Appetitive and aversive olfactory learning in humans tudied using event-related functional magnetic resonance imaging. J Neurosci 22:10829–10837

Grigor J (1995) Do the eyes see what the nose knows? An investigation of the effects of olfactory priming on visual event-related potentials. Chem Senses 20:163

Grigor J, Van TS, Behan J, Richardson A (1999) The effect of odour priming on long latency visual evoked potentials of matching and mismatching objects. Chem Senses 24:137–144

Haxby JV, Grady CL, Ungerleider LG, Horwitz B (1991) Mapping the functional neuroanatomy of the intact human brain with brain work imaging. Neuropsychologia 29:539–555

Jezzard P, Matthews P, Smith S (2001) Functional MRI: an introduction to methods. Oxford University Press, Oxford

Kennett S, Eimer M, Spence C, Driver J (2001a) Tactile–visual links in exogenous spatial attention under different postures: convergent evidence from psychophysics and ERPs. J Cogn Neurosci 13:462–478

Kennett S, Taylor Clarke M, Haggard P (2001b) Noninformative vision improves the spatial resolution of touch in humans. Curr Biol 11:1188–1191

King AJ, Palmer AR (1985) Integration of visual and auditory information in bimodal neurones in the guinea-pig superior colliculus. Exp Brain Res 60:492–500

Lloyd DM, Shore DI, Spence C, Calvert GA (2003) Multisensory representation of limb position in human premotor cortex. Nat Neurosci 6:17–18

Macaluso E, Driver J (2001) Spatial attention and crossmodal interactions between vision and touch. Neuropsychologia 39:1304–1316

Macaluso E, Frith CD, Driver J (2000) Modulation of human visual cortex by crossmodal spatial attention. Science 289:1206–1208

MacLeod A, Summerfield Q (1987) Quantifying the contribution of vision to speech perception on noise. Br J Audiol 21:131–141

MacSweeney M, Amaro E, Calvert GA, Campbell R, David AS, McGuire P et al (2000) Silent speechreading in the absence of scanner noise: an event-related fMRI study. Neuroreport 11:1729–1733

Maravita A, Spence C, Sergent C, Driver J (2002) Seeing your own touched hands in a mirror modulates cross-modal interactions. Psychol Sci 13:350–355

McGurk H, MacDonald J (1976) Hearing lips and seeing voices. Nature 264:746–748

Miller J (1982) Divided attention: evidence for coactivation with redundant signals. Cognit Psychol 14:247–279

Moir HC (1936) Some observations on the appreciation of flavour in foodstuffs. J Soc Chem Industry Industry Industry Rev 14:145–148

Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ (2002) Multisensory auditory–visual interactions during early sensory processing in humans: a high-density electrical mapping study. Cogn Brain Res 14:115–128

Morrot G, Brochet F, Dubourdieu D (2001) The color of odors. Brain Lang 79:309–320

Ongur D, Price JL (2000) The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. Cereb Cortex 10:206–219

Osterbauer RA, Matthews PM, Calvert GA (2002) Crossmodal interaction of color and smell. In: Proceedings of the international society for magnetic resonance imaging in medicine.

Pavani F, Spence C, Driver J (2000) Visual capture of touch: out-of-the-body experiences with rubber gloves. Psychol Sci 11:353–359

Raij T, Uutela K, Hari R (2000) Audiovisual integration of letters in the human brain. Neuron 28:617–625

Regan D (1989) Human brain electrophysiology. Elsevier, New York

Rockland K, Ojima H (2001) Calcarine area V1 as a multimodal convergence area. Abstract Soc Neurosci 27:511.20

Rockland KS, Ojima H (2002) Multimodal convergence in calcarine visual areas in macaque monkey. In: International multisensory research forum. Geneva, Switzerland

Rolls BJ, Rowe EA, Rolls ET (1982) How sensory properties of foods affect human feeding behavior. Physiol Behav 29:409–417

Rolls ET (2001) The rules of formation of the olfactory representations found in the orbitofrontal cortex olfactory areas in primates. Chem Senses 26:595–604

Rolls ET, Baylis LL (1994) Gustatory, olfactory, and visual convergence within the primate orbitofrontal cortex. J Neurosci 14:5437–5452

Royet JP, Zald D, Versace R, Costes N, Lavenne F, Koenig O et al (2000) Emotional responses to pleasant and unpleasant olfactory, visual, and auditory stimuli: a positron emission tomography study. J Neurosci 20:7752–7759

Sams M, Aulanko R, Hamalainen M, Hari R, Lounasmaa OV, Lu ST et al (1991) Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. Neurosci Lett 127:141–145

Shams L, Kamitani Y, Shimojo S (2000) Illusions. What you see is what you hear. Nature 408:788

Shams L, Kamitani Y, Thompson S, Shimojo S (2001) Sound alters visual evoked potentials in humans. Neuroreport 12:3849–3852

Singer W, Gray CM (1995) Visual feature integration and the temporal correlation hypothesis. Annu Rev Neurosci 18:555–586

Spence C, Shore DI, Klein RM (2001) Multisensory prior entry. J Exp Psychol 130:799–832

Stein BE, Meredith, MA (1993) Merging of the senses. MIT Press, Cambridge, Mass.

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

Taylor Clarke M, Kennett S, Haggard P (2002) Vision modulates somatosensory cortical processing. Curr Biol 12:233–236

Tipper SP, Lloyd D, Shorland B, Dancer C, Howard LA, McGlone F (1998) Vision influences tactile perception without proprioceptive orienting. Neuroreport 9:1741–1744

Van Riper C (1935) An experimental study of the Japanese illusion. Am J Psychol 47:252–263

Welch RB, Warren DH (1986) Intersensory interactions. In: Kaufman KR, Thomas JP (eds) Handbook of perception and human performance, vol 1. Sensory processes and perception. Wiley, New York, pp 1–36

Wheately J (1973) Putting color into marketing. Marketing 67:24–29

Zald DH (2003) The human amygdala and the emotional evaluation of sensory stimuli. Brain Res Brain Res Rev 41:88–123

Zald DH, Donndelinger MJ, Pardo JV (1998) Elucidating dynamic brain interactions with across-subjects correlational analyses of positron emission tomographic data: the functional connectivity of the amygdala and orbitofrontal cortex during olfactory tasks. J Cereb Blood Flow Metab 18:896–905

Zellner DA, Kautz MA (1990) Color affects perceived odor intensity. J Exp Psychol Hum Percept Perform 16:391–397

Zellner DA, Bartoli AM, Eckard R (1991) Influence of color on odor identification and liking ratings. Am J Psychol 104:547–561