Are the Primary Sensory Cortices Multisensory?

Ciara Devine
Trinity College, Dublin
Correspondence: devineca@tcd.ie

abstract
Historically the brain has been viewed as operating in a modular and hierarchical manner. In particular it was previously assumed that each of the primary sensory cortices processed input from a single sensory modality before feeding this information to higher regions of the brain devoted to multisensory integration. This review however outlines a wealth of findings that have emerged in recent years to suggest that in fact the primary sensory cortices are multisensory and that multisensory integration takes place much earlier in the brain than once believed. These findings have given rise to a shift in how the brain and the sensory systems are viewed, as well as methodological and theoretical debates surrounding these revolutionary findings.

Introduction
Multisensory interactions are a major feature of and requirement for everyday human perception. The brain needs to be able to integrate a wealth of information coming from various different sensory modalities in order to create a coherent representation of the world. Previous models asserted that crossmodal interactions occurred in higher-level multisensory convergence regions of the brain only after extensive processing of unisensory information in lower regions of the brain. The superior temporal sulcus (STS) and intraparietal sulcus (IPS) are two such regions that have been consistently identified as being ‘heteromodal’ (Calvert, 2001; Driver & Noesselt, 2008). However, findings in recent years have led to a radical shift away from this viewpoint with evidence now suggesting that multisensory interactions take place much earlier on, in areas of the brain that were once considered unimodal such as the primary sensory cortices. Thus the focus of this review is on examining evidence relating to crossmodal interactions occurring early on in the primary sensory cortices of the human brain and how consequential these interactions are for perception.
Principles of multisensory interaction

It is important to note that much of this research has been inspired by the pioneering work of Stein and Meredith (1993) who uncovered the main principles governing multisensory processing in single neurons of the superior colliculus (SC) in cats. They demonstrated that many SC neurons respond to input from multiple sensory modalities. These neurons exhibit supraadditive effects whereby they fire significantly more in response to multisensory input than would be predicted by summing together their unimodal responses. Furthermore, such multisensory gains are greatest when the input from each modality is weak, a principle known as the law of inverse effectiveness. These principles have come to guide researchers in investigating human multisensory processing. According to Calvert, Hansen, Iversen and Brammer (2001) response enhancement has become a hallmark feature of multisensory interaction.

Crossmodal interactions in primary sensory cortices

Using neuroimaging techniques such as functional magnetic resonance imaging (fMRI) and positron emission tomography (PET), evidence of crossmodal interactions in the normal human brain, in line with the principles outlined above, has been demonstrated in the primary sensory cortices. A number of studies have demonstrated evidence of audio-visual interactions, showing that the primary visual cortex (V1) is activated in response to simple auditory tones whilst the primary auditory cortex (A1) is activated to simple visual stimuli such as checkerboard patterns and flashes (Martuzzi et al., 2007; McIntosh, Cabeza & Lobaugh, 1998; Noesselt et al., 2007). Audio-visual speech interactions have also been reported in the form of A1 activation during silent lip-reading, indicating that visual speech information can activate A1 in the absence of auditory speech information (Calvert et al., 1997; Pekkola et al., 2005). Sathian, Zangaladze, Hoffman and Grafton (1997) reported evidence also of visuo-tactile interactions in V1. In their PET study, increased blood flow was observed in V1 of sighted participants that were required to judge the orientation of tactile gratings presented to their fingertips. Thus, in contrast to traditionally modular views of the brain, it appears that the primary sensory cortices can be activated to input coming from non-dominant modalities, although it is important to be aware that these activations are weaker than those that occur in response to information from the dominant sense (Martuzzi et al., 2007).

Crossmodal interactions in the form of supraadditive effects have also been demonstrated. Martuzzi et al. (2007) reported greater activation in both A1 and V1 under audio-visual conditions, where checkerboard patterns and auditory tones
were presented simultaneously, compared to unimodal conditions where they were presented independently. This finding has been supported by other audio-visual investigations involving simple stimuli (Meienbrock, Naumer, Doehrmann, Singer & Muckli, 2007; Watkins, Shams, Tanaka, Haynes & Rees, 2006) and speech related stimuli (lip movements and speech sounds) presented congruently (Calvert et al., 1999; Calvert, Campbell & Brammer, 2000; Calvert et al., 2001). Furthermore, Macaluso, Frith and Driver (2000) reported supraadditive effects under visuo-tactile conditions whereby tactile stimulation applied to the same hand towards which vision was being directed resulted in enhanced V1 activation. Thus there is evidence of enhanced neural activation in the primary sensory cortices under congruent multisensory conditions.

Methodological debate

However, a number of constraints relating to neuroimaging techniques have been highlighted. Despite its superior spatial resolution compared to other methods, fMRI measurements are still very limited with each voxel measuring the activity of millions of neurons within the same region (Calvert, 2001). Thus the neuroimaging findings reported here cannot be used to infer that multisensory integration is taking place at the neuronal level, as reported in the superior colliculus (Stein & Meredith, 1993). Rather, these findings could be accounted for by the coexistence of unimodal neurons responding to different modalities in the same cortical area (Calvert, 2001). Furthermore, Laurienti, Perrault, Stanford, Wallace and Stein (2005) highlighted that in some cases large fMRI voxels may be detecting activation in multisensory convergence zones that lie beyond or outside of the primary sensory regions. Klemen and Chambers (2012) stated that our understanding of multisensory processing is constrained by insufficient spatial resolution in many neuroimaging techniques and by the inability of researchers to carry out invasive recording with healthy humans to measure interactions at the neuronal level. Although recent evidence in support of early audio-visual interactions at the neuronal level has emerged from studies on epileptic patients that have intracranial electrodes implanted in A1 (Besle et al., 2008; Reale et al., 2007), such evidence is limited to patient populations.

Are the primary sensory regions truly interacting?

There has been debate in the literature as to whether or not early primary sensory interactions reflect genuine communication between the primary sensory regions or feedback from higher order regions of the brain that underlie higher cognitive abilities.

The role of mental imagery
One possibility is that early primary visual activation is linked to mental imagery (MI) rather than direct interaction between the sensory systems. In Kosslyn and Thompson’s (2003) review it emerged that V1 is activated not only during visual perception but also during MI tasks involving the recall or imagination of shapes. For example, Slotnick, Thompson and Kosslyn (2005) reported V1 activation when participants were asked to imagine a checkerboard stimulus that they had previously been presented with. Furthermore transcranial magnetic stimulation (TMS) studies suggest that V1 may even be crucial for MI (Kosslyn et al., 1999). Such findings have led to the suggestion that MI rather than actual bimodal sensory integration could account for early crossmodal effects (Driver & Noesselt, 2008). From this viewpoint, auditory and tactile sensory information may not be directly processed in V1 but rather non-visual sensory input may elicit mental imagery that in turn gives rise to V1 activation. An example of this comes from Sathian et al. (1997) who suggested that tactile grating stimuli might give rise to visuospatial images that account for V1 activation during tactile perception.

Crucial insight into the MI debate comes from studies of congenitally blind individuals who have never had any visual experience and therefore presumably lack the capacity for visual MI. Such studies have reported activation of V1 during braille reading (Burton, McLaren & Sinclair, 2006; Sadato et al., 1996; 1998) and during auditory speech processing tasks (Burton, Snyder, Diamond & Raichle, 2002; Roder, Stock, Bien, Neville & Rosler, 2002) for congenitally blind participants and participants who became blind early in life prior to gaining much visual experience. Similar V1 activation has also been reported for an early blind individual during echolocation (Arnott, Thaler, Milne, Kish & Goodale, 2013). This implies that MI is not entirely responsible or necessary for crossmodal activation in V1, even though MI does activate V1 in healthy individuals. Furthermore, these findings demonstrate the inherent plasticity of the sensory systems in adapting to different sensory experiences (Renier, DeVolder & Rauschecker, in press). The fact that V1, which is predominantly specialised for visual processing under normal developmental circumstances, can take on a role in processing auditory and tactile information in the absence of visual input provides unique insight into the multisensory nature of this brain region. These findings indicate that the capabilities of V1 are not limited solely to the processing of visual sensory information. Moreover, MI cannot account for V1 activation reported in these studies.

Going forward researchers should seek to distinguish between activation related directly to sensory processing as opposed to MI. In higher-level visual areas
such as the lateral occipital complex, which is devoted to object recognition, researchers have found that imagining tactile stimuli elicits weaker activation than actually perceiving those stimuli (Amedi, Malach, Hendler, Peled & Zohary, 2001). Similar studies could be carried out comparing V1 activation during actual cross-modal perception as compared to during MI tasks in order to provide greater insight into the role of MI in early primary sensory interactions.

**Attentional feedback**

Another possibility is that higher regions of the brain, involved in guiding attention modulate crossmodal interactions by processing input from one modality and sending feedback to primary sensory regions related to another sensory modality. In support of this, measures of directional functional connectivity have produced strong evidence that in the human brain, A1 and V1 are directly influenced by projections from the STS with very little direct connectivity between A1 and V1 (Noesselt et al., 2007). Furthermore, Meienbrock et al. (2007) argued that enhanced V1 activation in response to spatially incongruent audio-visual stimuli may be caused by signals from heteromodal regions, such as the STS and IPS, which detect sensory mismatch and send signals to the V1 in order to redirect attention towards resolving this incongruency. However, most cases of early multisensory interactions reported here have involved congruent rather than incongruent crossmodal stimulation.

Arguing against the role of attentional processes in early primary sensory interactions, Merabet et al. (2006) reported V1 activation during a tactile discrimination task alongside deactivation of the extrastriate cortex, as opposed to enhanced extrastriate activity, which the authors stated is typically associated with attentional processes. Watkins, Shams, Josephs and Rees (2007) also argued against attentional feedback based on their findings of V1 enhancement during the flash illusion, whereby one flash is perceived as two as a result of hearing two auditory tones, and V1 suppression in a contrasting illusion, whereby two flashes are perceived as one as a result of hearing only one auditory tone. If a general attentional process mediated V1 activation then these differences should not have occurred. In both of these studies, activation patterns that would be expected if attentional processes were responsible for early V1 activation were not observed.

**Early electrophysiological interactions**

Some of the strongest evidence that genuine multisensory interactions occur very early on in sensory processing comes from studies using electroencephalography (EEG) and magnetoencephalography (MEG). These recording techniques, although they have poor spatial resolution
(Calvert, 2001), allow researchers to gain precise temporal information about early activity in cortical cell populations that are linked to both unisensory and multisensory processes (Besle, Bertrand & Giard, 2009; Besle, Fort & Giard, 2004).

Giard and Peronnet’s (1999) pioneering study provided the earliest evidence of supraadditive effects occurring at very early latencies over right occipito-parietal regions of the scalp. In their study, participants were required to recognise objects based on audio, visual or audio-visual features. Larger event-related potentials (ERPs), which could not be predicted by summing unimodal ERPs, were recorded as early as 40ms post-stimulus onset in the audio-visual condition along with behavioural facilitation in the form of superior object recognition performance. Similar results have been reported under audio-visual conditions for simple object detection tasks (Fort, Delpuech, Pernier & Giard, 2002; Molholm et al., 2002) and even in the absence of any task (Vidal, Giard, Roux, Barthelemy & Bruneau, 2008), as well as under audio-somatosensory conditions (Foxe et al., 2000). These interactions correspond in terms of latency and topography to ERP components such as the visual P1 and auditory N1, which are associated with early unisensory processing. Studies using illusions have also demonstrated very early crossmodal interactions. Visual ERP signals are modulated by auditory input as early as 35ms post stimulus onset during the flash illusion (Bhattacharya, Shams & Shimojo, 2002; Shams, Iwaki, Chawla & Bhattacharya, 2005; Shams, Kamitani, Thompson & Shimojo, 2001).

The principle of inverse effectiveness has also been investigated in relation to early latency crossmodal interactions. Senkowski, Saint-Amour, Hofle and Foxe (2011) found evidence of audio-visual interactions at 40–60ms when auditory tones and visual gratings were presented at low but not moderate or high intensity levels. In line with this principle, schizophrenia patients, who display natural neurophysiological deficits in unisensory processing, demonstrate greater ERP enhancements than control participants during audio-visual conditions (Stone et al., 2011). This is in line with the notion that multisensory interactions are more beneficial when unisensory signals or processes are insufficient for stimulus detection or recognition.

Researchers argue that these ERP enhancements under crossmodal conditions occur too early to be modulated by top-down signals, such as those relating to attention or MI. In Mottonen, Schurman and Sams’ (2004) audio-visual speech perception study, modulation of the auditory N1 component by visual information relating to lip movements occurred between 150–220ms, much earlier than interactions occurring in the STS between 250–600ms. Additionally, Molholm et al. (2002) repor-
Literature Review

ARE THE PRIMARY SENSORY CORTICES MULTISENSORY

ted that direct connectivity between A1 and V1 is possible given that auditory signals are detected at 15ms, much earlier than visual signals, thereby allowing sufficient time for A1 to project information to V1 in order for integration to take place. Thus early electrophysiological interactions provide indirect evidence in support of the existence of monosynaptic projections between the primary sensory cortices, which to date have only been demonstrated directly in animal tracing studies (Clavagnier et al., 2004; Falchier, Clavagnier, Barone, & Kennedy, 2002; Rockland & Ojima, 2003).

Methodological issues

Although electrophysiological research appears to offer strong support for genuine early multisensory interactions, artefacts may confound these results. Teder-Salejarvi, McDonald, DiRusso and Hillyard (2002) reported evidence of a slow-wave anticipatory potential appearing before stimulus-onset that could account for significant bimodal interactions occurring around 40ms. After filtering the EEG data they reported that true audio-visual interactions might not occur in visual areas until 130ms. According to Driver and Noesselt (2008), very few EEG studies manage to avoid this particular confound, which could indicate that the seemingly strong electrophysiological evidence for early multisensory interactions occurring simultaneously with unisensory processing may merely be the product of methodological confounds. However, some studies have found evidence of genuine early audio-visual interactions when accounting for anticipatory potentials (Klucharev, Mottonen & Sams, 2003; Shams, Iwaki, Chawla & Bhattacharya, 2005). Nevertheless, in using these measures, researchers need to be aware of neural activity and artefacts that may produce extraneous results.

Functional relevance of early interactions

A final question that surrounds early multisensory interactions is whether or not they have an impact on actual perception or merely constitute an epiphenomenon. Numerous studies suggest that crossmodal interactions in the primary sensory cortices are functionally relevant. Researchers investigating the flash illusion have reported greater V1 activation only when participants consciously report experiencing two flashes as opposed to one (Watkins et al., 2006; Shams et al., 2005). This suggests that activity at the neural level is linked to perceptual experiences. In another interesting study, Theoret, Merabet and Pascual-Leone (2004) blindfolded sighted participants who then learned how to read braille letters. Following five days of training the researchers observed increased V1 activation in response to braille that corresponded with increased braille reading abilities. Furthermore, both V1 activation
and behavioural improvements disappeared following blindfold removal thereby indicating that the recruitment of V1 was functionally related to behavioural facilitation effects. A more recent replication of this study found that enhanced braille reading performance was disrupted when TMS was applied to V1 (Merabet et al., 2008). In an earlier study Zangaladze, Epstein, Grafton and Sathian (1999) also applied TMS to V1 in sighted participants and found that they showed impaired performance during a tactile orientation discrimination task. TMS studies provide even stronger support for the functional role of V1 in early crossmodal interactions.

In offering a rationale for why some early interactions may take place in V1, Molholm et al. (2002) suggested that they serve an evolutionary function in facilitating rapid detection of stimuli. In support of this, Van der Burg, Talsma, Olivers, Kickey and Theeuwes (2011) found that participants were faster to locate transforming target visual stimuli among distractors when the transforming target was paired with an auditory tone. Furthermore, the strength of the underlying ERP signals occurring at 50-60ms post stimulus onset correlated with search-time benefits in this study. It is possible that our ancestors relied upon rapid integration of information from different modalities so as to detect threatening stimuli in the environment that may not be easily detected under unimodal conditions, thus giving rise to the evolution of early multisensory capabilities. However, as pointed out by Watkins et al. (2006), despite evidence of early functional interactions, researchers do not yet have a thorough understanding of their consequences for perception.

**Conclusion**

In summary, this review has demonstrated that crossmodal interactions are not limited to higher-level, multisensory convergence zones and in fact take place much earlier on in lower regions of the brain. There is strong evidence suggesting that these interactions are the consequence of direct projections and communication between the primary sensory cortices. Furthermore, early interactions appear to play a functional role in conscious perception although continued research is required in order to fully understand these perceptual consequences. Due to methodological limitations associated with both neuroimaging and electrophysiological recording techniques, human research may continue to rely on animal studies to provide frameworks for multisensory integration at the neuronal level as it is not yet possible to assess this in healthy humans.

**References**


ARE THE PRIMARY SENSORY CORTEXES MULTISENSORY


