BRILL

Seeing and Perceiving 23 (2010) 3–38

brill.nl/sp

Multisensory Processing in Review: from Physiology to Behaviour

David Alais [1,*], **Fiona N. Newell** [2] **and Pascal Mamassian** [3]

[1] School of Psychology, University of Sydney, Australia

[2] School of Psychology and Institute of Neuroscience, Trinity College Dublin, Ireland

[3] Laboratoire Psychologie de la Perception, Université Paris Descartes, France

Received 30 October 2009; accepted 20 January 2010

Abstract

Research in multisensory processes has exploded over the last decade. Tremendous advances have been made in a variety of fields from single-unit neural recordings and functional brain imaging through to behaviour, perception and cognition. These diverse approaches have highlighted how the senses work together to produce a coherent multimodal representation of the external world that enables us to function better by exploiting the redundancies and complementarities provided by multiple sensory modalities. With large numbers of new students and researchers being attracted to multisensory research, and the multi-disciplinary nature of the work, our aim in this review is to provide an overview of multisensory processing that includes all fields in a single review. Our intention is to provide a comprehensive source for those interested in learning about multisensory processes, covering a variety of sensory combinations and methodologies, and tracing the path from single-unit neurophysiology through to perception and cognitive functions such as attention and speech.

© Koninklijke Brill NV, Leiden, 2010

Keywords

Multisensory, perception, behaviour, neuroimaging, neurophysiology

1. Introduction

The recent decade or so has seen an explosion of research activity in multisensory processing. Prior to this, most sensory work, whether cognitive or neurophysiological, focused on single modalities independently of the other senses. This reflected the prevailing view of cortical organisation that each modality initially processed information independently, with sensory integration or 'binding' occurring at a later stage of processing (Treisman and Gelade, 1980), particularly in later association

* To whom correspondence should be addressed. E-mail: davida@psych.usyd.edu.au

© Koninklijke Brill NV, Leiden, 2010

DOI:10.1163/187847510X488603

or 'polysensory' areas of the brain (Penfield and Rasmussen, 1950; Jones and Powell, 1970; Benevento *et al*., 1977; Felleman and Van Essen, 1991). On this view, focusing on unisensory questions was not only a sensible place to begin; it also made research tractable at a time when relatively little was known about cortical processing. It is easy to forget, for example, that the pioneering work on the functional organisation of the primary visual cortex was not done until the 1960s (Hubel and Wiesel, 1962), and similarly for auditory and somatosensory cortices (Mountcastle, 1957; Evans and Whitfield, 1964). A unimodal focus therefore allowed the fundamental principles to be established before the further challenges of binding and integration were tackled.

Recent discoveries, however, have seen this 'unisensory before multisensory' view challenged, with converging evidence from behavioural, neurophysiological and neuroimaging studies suggesting that multisensory processing occurs much earlier than had been supposed. For example, there is growing evidence for multisensory processing occurring in cortical area V1 as a consequence of either auditory stimulation (Burton *et al.*, 2002; Roder *et al*., 2002) or with Braille reading in blind people (Sadato *et al*., 1996; Buchel *et al*., 1998). There is also evidence of multisensory interactions in A1 from somatosensory inputs (Foxe *et al*., 2002). There are even direct connections between early auditory cortex and primary visual cortex that were only recently discovered (Falchier *et al*., 2002; Rockland and Ojima, 2003). Subcortically, too, it is well demonstrated that visual, auditory and somatosensory information is integrated in the superior colliculus (SC) (Stein and Meredith, 1993). Thus, although it is undeniable that various 'later' regions in the frontal and temporal cortices such as the superior temporal sulcus and prefrontal cortex are indeed sites for multisensory convergence, intersensory interactions can occur much earlier than was thought only a decade ago — as early as primary sensory cortices.

The current boom in multisensory research was spurred on by two main factors. First, the time was ripe, as our knowledge of unisensory cortical function had advanced enough to justify detailed investigations of combination and integration. Second, a body of work by Stein and colleagues established some key principles of sensory integration that opened the door for future work (Stein and Meredith, 1993). Their work was multisensory from its inception and was carried out in the deep layers of cat superior colliculus where most cells are multisensory, integrating combinations of visual, auditory and somatosensory inputs. One of these principles, known as superadditivity (discussed below), was adopted as a neural signature of multisensory integration and guided multisensory investigations in cortical areas. Superadditivity also held clear behavioural and perceptual implications and so sparked multisensory work in the cognitive domain.

The aim of this review is to provide a broad sketch of the field. It will begin with an overview of what is known neurophysiologically, both in the midbrain and in various cortices, before moving to a discussion of cognitive aspects such as multisensory perception, attention and speech.

2. Multisensory Neurophysiology

2.1. Superior Colliculus

Multisensory integration has been most thoroughly studied in superior colliculus (SC), a structure common to all mammalian brains that contains spatial maps in retinotopic coordinates. The superior colliculus plays a key role in orienting behaviours, whether overt orienting (e.g., moving the head or eyes) to best capture a stimulus of interest or covert orienting (i.e., allocating spatial attention to a region of interest). The superior colliculus provided an ideal model for examining multisensory integration because it is a low-level structure that receives ascending visual, auditory and somatosensory inputs. Cells in its deep layers (the superficial layers are purely visual) are often bimodal (primarily audio-visual and visual-somatosensory) and may even be trimodal.

A neuron is defined as multisensory if it receives input from more than one sense. In practice this means determining whether a given unit has spatial receptive fields in response to visual and/or auditory and/or somatosensory stimuli. When this process is carried out, the first important principle of organisation in the superior colliculus emerges very clearly: the separately defined receptive fields of collicular multisensory neurons are in 'spatial register'. This means that a neuron's receptive fields overlap so that they respond to stimuli from the same region of space. Across the many cells of a given deep layer, receptive fields of multisensory cells are arranged to provide a functional map of external space (Meredith and Stein, 1990). While the visual receptive fields are dependent on eye position, auditory and somatosensory receptive fields tend to make compensatory location shifts to maintain spatial register (Hartline *et al*., 1995; Groh and Sparks, 1996). Spatial register and topographic organisation reveals that what is important to collicular neurons is *where* a stimulus event occurs, not the sensory modality of that event.

When multisensory collicular neurons are driven by spatially congruent stimuli, they may exhibit interesting non-linear responses. One such response is a multisensory response enhancement that may exceed the sum of the unisensory responses, an effect known as 'superadditivity' (Stein and Meredith, 1993). This kind of response enhancement is most commonly observed when the component inputs are weak and generate only modest responses on their own, which is functionally important in ensuring that weak stimuli are not missed. It is generally observed that superadditivity is increasingly likely to be observed as the salience of the components decreases, a principle known as 'inverse effectiveness' (Stanford *et al*., 2005). There is a sensible functional interpretation of this effect in that if a stimulus event elicits a robust response in each modality of a multisensory neuron then there is no need to enhance it further, and in any case, there is a limit to the response range of any neuron. A final noteworthy point is that simultaneous multisensory stimuli that have spatially disparate locations, one falling within a unit's receptive field and another adjacent to it, will tend to elicit a lower response than either component alone. This has been termed 'response depression' (Stein and Meredith, 1993).

As well as the need for multisensory stimuli to be aligned in space, they must also be approximately aligned in time (Meredith *et al.*, 1987). Interestingly, there is quite a broad temporal window for multisensory interactions to occur in collicular neurons (Meredith *et al.*, 1987; McDonald *et al.*, 2001). The reason for this is probably related to the differences between modalities in terms of transduction times and neural latencies, as well as differences in the speeds of light and sound. These temporal differences mean that the sensory components of multisensory signals will inevitably arrive at different moments in time. In order for the superior colliculus to fulfill its multisensory function, it would need to be able to accommodate considerable temporal variation among incoming signals, even when they are generated by a single multimodal stimulus event.

Finally, studies examining the maturation of collicular function have shown two important findings. First, descending cortical input is necessary for multisensory cells to develop their characteristic non-linear functions, and this descending input is usually unisensory (Jiang *et al.*, 2006). Studies in cat have shown that descending inputs come primarily from the anterior ectosylvian and lateral suprasylvian sulci (Jiang *et al.*, 2001) and that when cortical input to collicular cells is blocked they lose their characteristic multisensory functioning (Stein *et al.*, 2002). That is, while they will continue to respond to stimuli in more than one modality they lose their ability to integrate multisensory inputs in a superadditive manner. A second important finding is that multisensory cells are either absent at birth or are not able to integrate multisensory input (Wallace and Stein, 2001). This slow development is probably due to the need for experience of correlated multisensory input in order to calibrate and register spatial maps.

In summary, multisensory function in collicular neurons is gated by the requirement of spatial and temporal coincidence. The superadditivity of its responses to weak inputs ensures that non-salient stimuli are not likely to be missed, and response depression helps attenuate responses to spuriously simultaneous stimuli. Together, this creates salient peaks on superior colliculus topography, identifying probable locations of external stimuli. Assuming normal development has occurred, orienting behaviours will be faster and more accurate in response to congruent multisensory stimuli, and slower and less accurate in response to spatially disparate stimuli (Calvert *et al.*, 2004; Rowland *et al.*, 2007).

2.2. Multisensory Interactions in Cortex

Recent findings from neurophysiological and neuroimaging studies have found that multisensory interactions are common in the cortex and are far more widespread than was thought even a decade ago. Instances of response superadditivity do occur in cortex, but are more the exception than the rule, although this may have more to do with a special role for collicular superadditivity in distinguishing valid orienting targets from spurious activity. In contrast, cortical multisensory interactions seem to require multisensory congruence, for example sounds that are appropriate to a visual object or action. The most extensive work on cortical multisensory function

has been done in the cat. In the anterior ectosylvian sulcus, which sends descending unisensory inputs to SC, there are also multisensory neurons. These neurons exhibit the same organisation seen in SC of overlapping receptive fields, and may sometimes also show superadditive responses to spatiotemporally correlated multisensory inputs, as well as inverse effectiveness and response depression to disparate inputs (Stein and Wallace, 1996), although these non-linear responses appear to be far less common in cortex than in SC.

In primates, a good deal of work has focused on the posterior parietal cortex (PPC), particularly the lateral intraparietal (LIP) subregion (Graziano, 2001). PPC comprises a number of intraparietal subregions (medial: MIP, ventral: VIP, anterior: AIP) in which multisensory neurons are common. PPC contains variously mapped spatial representations and is involved in attention and goal-directed behaviours such as reaching and gaze direction. To facilitate these functions, multisensory spatial maps in PPC are coded in common coordinate frames, such as auditory-visual or visual-somatosensory maps in eye-centered coordinates (Cohen and Andersen, 2002) which tend to dynamically realign (at least partially) with changes in gaze direction (Avillac *et al*., 2005; Schlack *et al.*, 2005). Very little attention has been paid to examining whether VIP neurons display the non-linear multisensory responses observed in SC, although one study (Avillac *et al*., 2007) has done so using visual-tactile stimuli. They found that most VIP cells were modulated by multisensory stimuli, that they required spatial and temporal coincidence to do so, and that both super- and sub-additive responses were observed. One interesting difference from SC studies was that many cells would show sub-additive responses to spatiotemporally coincident stimuli.

Superior temporal cortical areas are also involved in multisensory processing, and several neuroimaging studies have reported superadditivity in these areas. In one fMRI study, subjects were presented with either a wooden roller to the hand or broadband auditory noise (Foxe *et al.*, 2002). Although the stimuli were presented to separate modalities, the tactile stimulus activated areas of auditory association cortex and left superior temporal gyrus (STG). When the auditory and haptic stimuli were presented simultaneously, superadditivity was observed in the left STG, suggesting it is an area for auditory-tactile integration. A similar study compared cortical activation in response to audiovisual speech presented either simultaneously, asynchronously, or with each modality presented separately (Calvert *et al.*, 2000). Superadditivity was found for simultaneous speech in the left superior temporal sulcus; however, asynchronous presentation reduced activation to the levels seen for the separately presented modalities.

In general, most neuroimaging studies in superior temporal and other areas do not find multisensory interactions strong enough to qualify as superadditivity, although this could simply be due to the stimulus-relevant neurons making up only a fraction of the population driving the BOLD changes. Many studies however report reliable but smaller modulations of multisensory BOLD response (Hein *et al.*, 2007; van Atteveldt *et al.*, 2007). One very recent study that varied signal strength in an

attempt to demonstrate inverse effectiveness with audiovisual stimuli found clearly superadditive responses in STS for weak stimuli (Stevenson and James, 2009). More commonly, however, multisensory interactions are present but weaker than superadditive. For example, multisensory interactions have been found in STS for audio-tactile and audio-visual stimuli, with bimodal stimuli increasing BOLD response on the order of 20% above the maximum unisensory response (Beauchamp *et al.*, 2004, 2008).

One important difference between SC and cortical areas is that the latter are likely to represent perceptually coherent objects or semantic information which means that the congruence of signals between modalities will be important in eliciting multisensory responses. Congruence in this case means that a sound stimulus, for instance, should be an ecologically valid match to a given visual object (e.g., a 'barking' sound would be congruent with an image of a dog; a 'meowing' sound would not). A number of studies have found congruence to be important in eliciting multisensory interactions (Barraclough *et al.*, 2005). In a fMRI study using familiar or novel images and sounds in various pairings, novel audiovisual parings (and incongruent pairings of familiar stimuli) activated inferior frontal cortex, but only familiar stimuli that were congruently paired activated STS (and superior temporal gyrus) (Hein *et al.*, 2007). This study, among a number of others, suggested a role for both semantic congruency and familiarity in object-related audiovisual integration. However, a recent attempt to replicate fMRI congruence findings using careful methods to control factors such as stimuli, task and attention cast doubt over the validity of a number of the reported congruence effects (Hocking and Price, 2008). A related paradigm was used to examine single-unit responses in monkey STS and found that congruence between video clips and sound tracks was essential to produce enhanced multisensory responses to familiar stimuli (Beauchamp *et al.*, 2004).

Much of the research on multisensory function in the cortex has been coloured by what was learnt from studies in the superior colliculus. For example, the quest to find characteristics such as superadditivity was central to many of the early studies. However, it is now clear that semantic and object- or action-related congruence is important in eliciting strong multisensory responses from many cortical areas. Superadditivity is an ideal mechanism for guiding orienting as it highlights regions of spatio-temporal coincidence, yet many of the areas that respond to multisensory information in the cortex are concerned with other functions such as action, language, learning, mirroring and even social perception (Campanella and Belin, 2007). As has been noted, the functional roles of multisensory integration in many cortical areas are still not fully understood, making it difficult to know what sort of response to expect from them (Stein and Stanford, 2008). Unlike the very clear multisensory behaviour in single units of the superior colliculus, some cortical multisensory responses may emerge at a population level and be harder to identify in single units.

2.3. Cross-modal Interactions in Primary Sensory Cortices

A growing number of studies on intersensory interactions have challenged the traditional view that the primary sensory cortices are functionally independent and sensory specific (Schroeder and Foxe, 2002; Fu *et al.*, 2003; Driver and Noesselt, 2008). Anatomical investigations examining patterns of cortical connectivity have found evidence for direct connections between primary sensory cortices, particularly auditory and visual areas (Falchier *et al.*, 2002; Rockland and Ojima, 2003; Clavagnier *et al.*, 2004; Cappe and Barone, 2005). Although the auditory to visual projections terminate primarily in the visual periphery (and in the upper and lower layers, indicative of feedback inputs), these findings nonetheless suggest that the basis for crossmodal interactions to affect perceptual processing is present at very early stages of sensory processing. Other evidence for early audiovisual interactions comes from studies using ERPs to examine the time course of audio-visual interactions in the human brain which report information across these two modalities interacts at very short latencies, and do so in early modality-specific cortical areas, a pattern consistent with feedforward combination rather than feedback.

One of the earliest demonstrations of cross-modal interactions at the level of primary sensory cortices was visual activation of auditory cortex during lip reading (Calvert *et al.*, 1997). Another early study reported primary visual activation during a tactile discrimination task involving oriented gratings (Sathian *et al.*, 1997) and it has been argued that area V1 is crucial for tactile discrimination, since a disruption of V1 activation using TMS impairs performance on this tactile task (Zangaladze *et al.*, 1999). Studies involving practice effects in tactile perception have also found evidence of recruitment of visual areas (Saito *et al.*, 2006). Audiovisual interactions in auditory cortex have been confirmed at the single-unit level (Ghazanfar *et al.*, 2005). Audio-somatosensory interactions are also present in auditory cortex, as shown by evoked potential studies (Foxe *et al.*, 2000), intracranial multicontact depth electrodes examining the time course of activation across cortical laminae (Schroeder *et al.*, 2001), and single-unit studies (Fu *et al.*, 2003).

In support of the idea that sensory cortices are directly connected, some neuroimaging studies have revealed evidence of recruitment of primary sensory areas that have been deprived of normal sensory input either over the long- or short-term. For example, primary visual cortex (V1) in blind individuals is activated during auditory (Kujala *et al.*, 1995), tactile (Sadato *et al.*, 1996; Goyal *et al.*, 2006) and verbal (Burton *et al.*, 2002; Amedi *et al.*, 2003) tasks whereas auditory cortex in deaf individuals is activated during visual tasks (Finney *et al.*, 2001). Moreover, activation in V1 of blind individuals is considered functionally relevant since disruption of processing in V1, either following a local ischemia (Hamilton *et al.*, 2000) or transcranial magnetic stimulation (TMS), interferes with Braille letter recognition (Cohen *et al.*, 1997) and other linguistic tasks (Amedi *et al.*, 2004). Such neural plasticity, involving the recruitment of deafferenated cortical areas, is arguably the causal reason for superior perception in the non-visual senses in blind persons for the purpose of identification (Rice, 1970; Wanet-Defalque *et al.*, 1988)

and spatial localization (Röder *et al.*, 1999; Fortin *et al.*, 2008). Indeed, even temporary loss of sight (e.g., 5 days) is sufficient to induce superior tactile performance in blindfolded, relative to non-blindfolded, sighted individuals irrespective of training (Kauffman *et al.*, 2002) and such performance is thought to be mediated by neural adaptation involving a rapid recruitment of area V1 during tactile perception (Merabet *et al.*, 2008). Furthermore, rapid cortical recruitment (and reversal of this effect) is more likely to be associated with an unmasking of pre-existing inter-cortical connections than a rewiring of the brain.

2.4. Crossmodal Interactions and Sensory Deprivation

A number of studies using neuroimaging technology have provided evidence of cortical reorganisation in humans deprived of vision (Sadato *et al.*, 1996) and audition (Finney *et al.*, 2001; Simon-Dack *et al.*, 2008). For example, more than a decade ago, Sadato *et al.*, used PET to show that the occipital cortex of people blinded at an early age is activated when they read Braille. A second study showed further that TMS over the occipital cortex disrupts the ability to identify Braille letters correctly in persons who are visually impaired (Cohen *et al.*, 1997). This led to the hypothesis that the occipital cortex in the blind is recruited for the purpose of tactile object processing. Although the mechanism which facilitates this cortical recruitment is not known, it is possible that Braille reading is mediated by an expansion of tactile activation from the ventral cortex, specifically the object integration region in the occipito-temporal cortex (Amedi *et al.*, 2001), to earlier retinotopic areas through feedback pathways.

In a fMRI study (Amedi *et al.*, 2003), it was reported that the occipital cortex of the blind (unlike the sighted) is activated during performance relating to verbal tasks. This includes tasks such as verb generation and verbal memory tests, and occurs regardless of the input modality (tactile or auditory), or even without sensory stimulation when retrieving words from memory. Moreover, the magnitude of V1 activation, measured either as percent signal change or volume of activation, was highly correlated with the blind individual's verbal memory capabilities, and in another study with verbal task difficulty (Roder *et al.*, 2002). These findings suggest the additional occipital activation may have a functional role. Generally, the pattern of occipital activation during verbal memory was left-lateralized with clear preference for the ventral pathway. Plausibly, the right occipital hemisphere could also reorganize in the blind for non-verbal tasks such as tactile object recognition, but this has yet to be confirmed.

In a similar way to ventral stream recruitment of occipital cortex for object perception in blind individuals, the dorsal occipito-parietal pathway may be relevant in tasks requiring tactile or auditory spatial memory. Indeed, some studies suggest auditory localisation is superior in early blind than late blind individuals (Collignon *et al.*, 2009b). Other recent studies show evidence of occipital reorganisation during tasks involving spatial localisation of sounds (Voss *et al.*, 2008). Moreover, there is recent evidence that it is specifically the right occipito-parietal stream which reor-

ganises for the purpose of audio-spatial processing (Collignon *et al.*, 2009a). Using TMS over the right dorsal occipital cortex, auditory spatial processing was found to be disrupted in early blind individuals. Together with findings reported by Amedi *et al.*, this suggests that occipital activation to crossmodal tasks is functionally relevant, and that functional distinctions (i.e., 'what' *vs* 'where') are maintained in the reorganisation.

There is evidence that visual experience may be helpful in building spatial representations of the environment. Without visual experience, large-scale spatial knowledge is restricted as tactile spatial inputs are limited to peripersonal space. Consistent with this, recognition of large-scale object layouts is worse in early blind than in late blind individuals (Gaunet and Thinus-Blanc, 1996). Similarly, spatial updating of tactile scenes with observer motion is less efficient in congenitally blind than in either late-blind or sighted individuals (Pasqualotto and Newell, 2007). Others have subsequently argued that early visual experience may be necessary for the development of efficient spatial perception in other modalities (Thinus-Blanc and Gaunet, 1997; Postma *et al.*, 2008), with supporting evidence coming from neurophysiological studies (Carriere *et al.*, 2007; King, 2009).

Overall, the accumulation of evidence indicating intersensory interactions in early cortex has been quite recent but is already substantial. The notion that sensory processing in early cortex is entirely modality specific with interactions occurring in later association areas that was still advocated in a strict form until relatively recently (Jones and Powell, 1970) is no longer tenable. The recent wave of studies showing evidence of early interaction between the senses and reorganization after sensory deprivation has led to the provocative proposal that the cortex may be fundamentally multisensory in nature (Ghazanfar and Schroeder, 2006). The validity of this assertion depends on how much of the documented multisensory interactions in early 'unisensory' cortex is due to feedback, and how much is feedforward. A number of the reports reviewed here, such as those showing feedforward laminar timing and short-latency ERP interactions, are clearly consistent with a feedforward account. Still, evidence is continuing to accumulate for and against and the notion of unisensory cortex is unlikely to be discarded.

3. Multisensory Perception

Once multisensory inputs are encoded at the sensory level, they can be used to understand and interpret the environment. Combining sensory information is a sensible strategy as the senses provide complementary information (Ernst and Bülthoff, 2004; Burr and Alais, 2006). In some cases, especially where the input in one sensory modality is ambiguous, the complementary component may be enough to augment attentional control over the ambiguity (van Ee *et al.*, 2009), and may even alter a percept entirely, as in the stream/bounce illusion (Sekuler *et al.*, 1997). In this case, a pair of disks oscillates back and forth across a video display, beginning from opposite sides so that they move in antiphase. When the disks converge at the centre,

do they collide and bounce apart, or do they stream past each other? The visual input is ambiguous and supports either interpretation; however, simply adding a click sound at the moment of 'impact' is sufficient to bias the interpretation strongly towards the bouncing percept. In other cases, rather than acting to disambiguate perception, information from a second modality can be fundamentally complementary. An example of this is when haptic exploration of a three-dimensional object provides the missing information about the invisible back of the object (Newell *et al.*, 2001). Finding the most reliable and robust interpretation of sensory input is central to our successful interaction with the world (Ernst and Bülthoff, 2004), and combining information is one effective way to achieve this. This section will review multisensory psychophysical research, grouped into two sections focusing on spatial and temporal interactions.

3.1. Spatial Factors

The best-known example of how the perceptual system deals with intersensory spatial conflict is the ventriloquist effect (Howard and Templeton, 1966). In this effect, provided the auditory and visual stimuli are aligned in time (Slutsky and Recanzone, 2001), displacing the visual stimulus over modest distances will usually cause the auditory stimulus to be 'captured' by the visual event (i.e., perceived as co-localized with the visual stimulus). Even over distances too large to produce absolute spatial capture, there is still a clear bias in auditory localization towards the visual stimulus (Welch and Warren, 1980; Bertelson and Aschersleben, 1998; Battaglia *et al.*, 2003). Although ventriloquism is usually cited as an example of vision's dominance over audition for spatial tasks, it is not necessarily so. In cases where the reliability of the visual signal is reduced by blurring, the location of the audiovisual stimulus will be biased towards the location of the auditory component, an example of 'reverse ventriloquism' and a rare example of auditory dominance in spatial localisation (Alais and Burr, 2004b).

The intersensory interactions occurring in spatial localization appear to be automatic. For example, when observers need only to localize the auditory component of a pair of simultaneous but spatially displaced audiovisual signals, their judgments still show a bias towards the visual location (Bertelson and Radeau, 1981). Other studies using a variety of techniques have suggested that ventriloquism occurs automatically (Bertelson and Aschersleben, 1998; Vroomen *et al.*, 2001), and the same conclusion has been drawn for spatial interactions between touch and vision (Caclin *et al.*, 2002; Bresciani *et al.*, 2006). Indeed, spatial biases are not limited to audiovisual combinations and have been reported to occur for visual-tactile (Pavani *et al.*, 2000) and auditory-tactile interactions (Caclin *et al.*, 2002; Guest *et al.*, 2002). It will also occur between vision and proprioception. In a study using lenses that made straight edges appear curved, subjects felt the edge was curved as they ran the fingers along it (Hay *et al.*, 1965).

One influential early study also used lenses to create intersensory conflict to examine perceived size in a visuohaptic context (Rock and Victor, 1964). Observers

felt a square shape that appeared elongated in one dimension when viewed through a cylindrical lens. The perceived size of the bimodal percept was strongly dominated by the visual image, being perceived as rectangular irrespective of whether size was measured by visual matching or haptic matching. This was interpreted in terms of vision being the dominant sense for spatial tasks, often referred to as 'visual capture'. As with the ventriloquist effect, visuo-haptic intergration appears to be automatic (Helbig and Ernst, 2008) in that it occurs even when the lens-distorted hand is visible when exploring the shape (Helbig and Ernst, 2007), and it also depends on spatial proximity of the component stimuli (Gepshtein *et al.*, 2005).

3.2. Temporal Factors

Intersensory interactions also occur in the time domain. One way that temporal interactions have been studied is with sequential pairings of visual flashes and auditory clicks (temporal ventriloquism) and studying the when these stimuli are perceived to occur (Fendrich and Corballis, 2001). To measure the moment when the stimuli occurred, subjects used a pointer which rotated quickly around the 12 hour-points of a 'clock' surrounding the stimuli. When judging the visual event, it was perceived earlier in time when it was preceded by a click, and later in time when it was followed by a click. In other words, the visual event was drawn in time towards the auditory event. In the converse task, when subjects judged the timing of the auditory stimulus, 'capture' effects were also found but were smaller in size. In a similar vein, another study examined the duration of an interval marked by successive visual flashes when those flashes were flanked by brief sounds. If the flanking sound was played just prior to the first flash and just after the second flash, performance on a temporal order task improved, as if the interval duration was greater (Morein-Zamir *et al.*, 2003). This was interpreted as evidence of the visual stimuli being drawn towards the auditory stimuli, in effect temporal ventriloquism. A number of other studies have explored this kind of audiovisual temporal interaction in other contexts and confirmed its generality by showing that sounds can attract the timing of visual events to influence the strength of visual apparent motion (Getzmann, 2007) or of the visual (or cross-modal) flash-lag effect (Alais and Burr, 2003; Vroomen and de Gelder, 2004).

Another clear example of temporal interactions — one which shows a striking influence of audition on vision — is 'auditory driving' (Shipley, 1964). The phenomenon of auditory driving occurs when matching a flickering light and fluttering sound: if the flutter rate is higher than the flicker rate, the subjectively matched flicker rate is biased upwards towards the higher flutter rate (Gebhard and Mowbray, 1959). Using the method of adjustment, the effect can be very powerfully demonstrated: if the flicker and flutter rates are initially matched, the flutter rate can be adjusted upwards by a large degree before the two temporal rates appear to desynchronise (Shipley, 1964). The effect is particularly strong for visual flicker rates above 10 Hz, the range in which visual temporal sensitivity declines from its peak at around 8–10 Hz (De Lange, 1958; Cass and Alais, 2006). Shipley found

that 10 Hz flicker required flutter rates of 14–22 Hz before the desynchrony point was reached. Hysteresis may play a role in this large effect because of the method used, but clearly the effect would remain strikingly large. If the direction is reversed and the 10 Hz flutter rate is decreased, its ability to drive a 10 Hz flicker downward towards lower frequencies is several times weaker, presumably because visual temporal perception is more reliable below 10 Hz.

A more contemporary example similar to auditory driving is the double-flash illusion in which a single light flash is paired with two short sound clicks (Shams *et al.*, 2000). The resulting percept tends to be of two flashes, whereas physically there is only one. Again, this is an example of audition exerting a strong influence over vision, which will usually only occur with rapid visual stimuli, as the visual system is not sensitive to rapid temporal events. A recent careful investigation has studied the effects investigated in the early auditory driving papers in a systematic study and has clearly shown that temporal rate perception is influenced by discrepant auditory rates, even at much lower rates around a 4 Hz standard (Recanzone, 2003). In a similar vein, perceived duration of lights and tones has been studied and when these are discrepant the conflict is resolved towards the duration of the tone (Walker and Scott, 1981). In sensorimotor synchrony tasks, subjects making tapping movements to reproduce the rate of a visual flicker stimulus show a strong bias towards an asynchronous auditory sequence (Aschersleben and Bertelson, 2003) and the variability of their tapping is altered by auditory distractors (Repp and Penel, 2002).

Temporal order judgments have been a common way to examine temporal factors in multisensory research (Hirsch and Sherrick, 1961; Sternberg and Knoll, 1973; Spence *et al.*, 2003; Zampini *et al.*, 2003). Typically this involves presenting two simple stimuli such as light flashes or sound bursts and reducing the temporal asynchrony between them until 75% correct temporal-order performance is reached. Temporal order discrimination thresholds are generally higher for more complex stimuli such as speech than for object actions (Vatakis and Spence, 2006). This study also reported that asynchrony thresholds for music video clips were higher than both these conditions, a finding that is possibly linked to a lack of musical expertise in their subjects, as a subsequent study found that trained musicians are more sensitive to asynchronous drumming sequences than untrained observers (Petrini *et al.*, 2009). Note that any sensitivity differences will tend to decline at higher drumming tempos as asynchrony thresholds decline with tempo, both for visual sequences showing real drummers (Arrighi *et al.*, 2006) and for point-light drummers (Petrini *et al.*, 2009). Action has also been reported to influence auditory perception, with the duration of a sound produced by striking a percussion instrument found to depend on whether the video sequence shows a hard or soft strike (Schutz and Lipscomb, 2007).

There seem to be particularly strong links for temporal processing between touch and audition. In one paper, tactile frequency discrimination for vibrations at particular frequencies was impaired by a masking frequency delivered to the auditory modality (Yau *et al.*, 2009). Interestingly, this masking effect was frequency-tuned

so that tactile discrimination at a given frequency was most impaired by auditory signals modulating at the same frequency. This was demonstrated for standard frequencies of 200 and 400 Hz. This link appears to work bi-directionally as detection of weak stimuli in the tactile domain is enhanced by matching auditory signals (Gescheider *et al.*, 1974). Underscoring the strong temporal links between audition and vision, a recent report comparing temporal resolution for visuo-tactile, audio-visual and audio-tactile stimuli found that audio-tactile resolution was greater than the other two combinations by about a factor of two on a simultaneity judgment, although in an interesting task dependency, this effect was not as strong when measured using a temporal order judgment (Fujisaki and Nishida, 2009).

3.3. Temporal Synchrony

Signals that occur simultaneously in different senses may play an important role in detection and integration of multisensory events. In one study, synchrony between a non-spatialised auditory tone pip and a visual 'target' change was sufficient to guide visual search to the target's location among an array of asynchronously changing visual distractors (van der Burg *et al.*, 2010). Interestingly, this effect was strongest for abrupt (square-wave) synchrony, and did not occur for synchronous gradual (sine-wave) changes. Temporal synchrony is clearly important in multisensory integration at a neural level (Meredith *et al.*, 1987), and these behavioural data show that a synchronous but spatially uninformative auditory event is able to facilitate an efficient visual spatial search, provided the signals are tightly defined in the time domain. The need for sharply defined temporal events may explain why previous studies found sinusoidal temporal modulations supported an upper limit for identifying audiovisual synchrony among a field of asynchronous visual distractors of just 4 Hz or so (Fujisaki *et al.*, 2006).

A number of earlier findings point to an important role for temporal synchrony. In one, the detectability of auditory signals was improved when they were accompanied by a synchronous but task-irrelevant light flash (Lovelace *et al.*, 2003). Importantly, this study used methods from signal detection theory (Green and Swets, 1964) to verify that their results were not simply a case of response bias and instead reflected an early sensory integration rather than a later influence at the decision stage. In analogous studies, visual sensitivity is reported to be enhanced by accompanying sounds (Frassinetti *et al.*, 2002), irrelevant sounds can also improve tactile detection (Gescheider *et al.*, 1974), and irrelevant tactile signals can improve detection of weak auditory signals (Schurmann *et al.*, 2004; Gillmeister and Eimer, 2007). Perceived loudness has also been found to be greater when accompanied by a visual stimulus (Odgaard *et al.*, 2004).

These studies suggest that signals from one modality can increase sensitivity to signals presented to another. Moreover, the perceptual benefit arises from the temporal synchrony of the signals, as the second signal is often not spatially proximal and in any case is not task relevant. In some studies, improved performance in one modality may be due to response biases related to the presence the second signal,

underscoring the importance of using signal detection analysis to verify whether the effects are genuine sensory-level improvements in sensitivity. These improvements in sensitivity may well be due to converging feedforward signals, with the signal from the second, task-irrelevant modality adding to the task relevant signal, effectively boosting the task-relevant signal's strength. One study consistent with this interpretation examined tactile intensity discrimination (Arabzadeh *et al.*, 2008). The study found that adding a visual signal adjacent to the finger-tips receiving the tactile signal improved tactile sensitivity and shifted the tactile intensity discrimination function uniformly to the left, as if the visual and tactile signals simply combined to form a stronger tactile signal.

An important theoretical issue to bear in mind when assessing whether signals from one modality increase sensitivity to signals presented in another modality is that some degree of improvement may be expected simply by a reduction in spatial and temporal uncertainty. That is, to extract a signal from noise, the local spatiotemporal environment has to be sampled, and optimally this would include all of the signal and as little of the noise as possible. Any signal, whether in the same modality or not, that can help delineate the optimal sampling window will increase sensitivity (d' in signal detection terms) because there is reduced uncertainty in the sampling process and hence less noise. It is difficult to exclude this from many of the extant psychophysical studies, especially those using near-threshold stimuli where internal or external noise are strongest. One useful approach is to compare cross-modality with within-modality combinations, as the within-modality condition provides an 'uncertainty reduction' baseline. Any genuine benefit due to cross-modal integration would need to exceed the within-modality baseline, as was suggested originally by Wundt in his discussion of the 'complication experiment'.

3.4. Temporal Limits of Multisensory Integration

There appears to be a wide 'temporal integration window' for multisensory perception. In video sequences of speech for example, the auditory signal can be delayed by as much as 250 ms or more before the desynchrony becomes apparent (Dixon and Spitz, 1980). This rather high estimate may be due to the complex nature of the signals (Vatakis and Spence, 2006) and the temporal correlations between the lip movements and speech sounds helping to maintain the audiovisual relationship. Psychophysical estimates using simpler and briefer stimuli are lower (Hirsch and Sherrick, 1961; Spence *et al.*, 2001b; Zampini *et al.*, 2003). In one study, the point of subjective simultaneity was measured and found to show considerable individual differences from as high as -150 ms (audition leading) to $+20$ ms (audition trailing). Usually, however, the sound must be slightly delayed to produce perceived synchrony, typically on the order of a few tens of milliseconds (Spence *et al.*, 2003; Sugita and Suzuki, 2003; Lewald and Guski, 2004). One of the reasons for vision needing a head-start is that transducing visual signals in the retina is a slower process than auditory transduction by about 30 ms or so (Fain, 2003). For this rea-

son, simple reaction times are also slower in vision than in audition by about this amount (Galton, 1899; Arrighi *et al.*, 2005).

The neural latency difference between audition and vision means that for synchronised audiovisual stimuli in the near-field, the auditory component will activate the brain first. However, as acoustic signals take about 3 ms to travel each metre of distance, its head-start over visual processing declines with distance. At about 10–15 m, auditory and visual signals will activate the brain about simultaneously, and beyond this distance sound will inevitably arrive late. Still, within a distance range of about zero to 25 m, audiovisual signal asynchronies will fall within the range of approximately ± 30 ms. A radius of 25 m is more than enough to cover the most behaviourally relevant spatial scales from peripersonal and near-field to near-distant space, and so it is doubtful that the relatively large window of temporal integration has much at all to do with late-arriving auditory signals due to the slower speed of sound, as is sometimes claimed. A related question is whether the brain can compensate for the slow travel time of acoustic signals in audiovisual synchrony tasks. Results are mixed (Sugita and Suzuki, 2003; Kopinska and Harris, 2004; Lewald and Guski, 2004); however, it does appear that audiovisual synchrony can take into account the distance and travel time of the auditory signal, provided robust cues to auditory distance are available (Alais and Carlile, 2005), and that a period of adaptation to a given asynchrony can reduce distance-related audiovisual delays so that little or no compensation may be required (Heron *et al.*, 2007).

Indeed, the capacity to adapt and recalibrate to temporal asynchronies is important. At the very least, recalibration is needed to deal with changes occurring naturally at a slow time scale, such as the growth of limbs during development, or the increase in head size. For example, the increase in head size significantly alters inter-aural time differences used in sound localization, meaning that effective interactions with the other senses would require recalibration over time. It is easily demonstrated that recalibration between the senses can occur within a short time-frame, as shown by adaptation studies which involve repeated exposure to an intersensory spatial displacement. Adaptation to a synchronous audio-visual signal with an introduced spatial conflict (e.g., produced by a displacement lens) will cause post-adaptation shifts in the localization of unimodal stimuli such that they are biased towards the displaced stimulus (Radeau and Bertelson, 1974; Recanzone, 1998). Repeated exposure to an introduced asynchrony can cause shifts in perceived timing, even causing reversals of temporal order. One psychophysical study examining this found post-adaptation judgments of subjective simultaneity were shifted towards the adapted asynchrony (Fujisaki *et al.*, 2004). In a visuomotor study, a button press that elicited a light flash was manipulated so that the flash occurred with some delay after the action (Stetson *et al.*, 2006). After adapting to this altered 'causal' timing, flashes that were triggered by the button press without delay were perceived as having occurred before the button press. The size of this motor-sensory timing effect was relatively large — larger than shifts reported for adaptation to sensory-sensory asynchrony.

Overall, a large range of multisensory perceptual interactions can be demonstrated in both the spatial and temporal domain, even if eliminating spatial effects from temporal experiments is a challenge. A practical and often quoted rule of thumb in spatial and temporal multisensory perception is Welch and Warren's (1980) 'modality appropriateness hypothesis'. This states that vision dominates audition for spatial tasks, and audition dominates vision for temporal tasks. This is taken to reflect the complementary specialities of each sensory modality and is often described as 'visual capture' and 'auditory capture', respectively. This serves as a useful simplification, but it is an overstatement. In the temporal dimension, a degree of visual attraction of auditory stimuli can occur (Fendrich and Corballis, 2001), and in visuo-haptic tasks, vision may tend to dominate but there remains a small influence of touch (Rock and Victor, 1964). This shows that 'capture' is not absolute, and indeed it is now clear that visual dominance can be reversed in spatial tasks when visual signals are degraded (Alais and Burr, 2004b). As we shall see below, a better model that can flexibly account for all of these findings in an efficient and optimal way is the Maximum Likelihood Estimation model.

3.5. Maximum-Likelihood Estimation

One currently popular model describing how information can be combined from two or more sources is the maximum-likelihood estimation (MLE) model. The model comes from Bayesian probability theory and describes a combination rule that is statistically optimal in that the combined result is the one that is most likely to be true. In essence, MLE is a weighted linear sum that combines two or more signals that are weighted by their reliability. Reliable signals receive a high weight, while unreliable signals receive a low weight. The combination rule is considered statistically optimal in that it always provides the result that is most reliable, where 'most reliable' means most probable or least variable. Separate studies examining visual-tactile and audio-visual integration have shown that human cross-modal perception closely matches predictions from the MLE model (Ernst and Banks, 2002; Alais and Burr, 2004b). More generally, the notion of cross-modal combination being weighted by signal reliability is one that has the potential to explain many of the spatial and temporal cross-modal interactions described above. Detailed reviews of this probabilistic approach can be found elsewhere (Pouget *et al.*, 2002; Kersten *et al.*, 2004).

To illustrate the principles of MLE, consider two modalities (e.g., vision and audition) each providing some information about an attribute (e.g., the location of a common signal source), s_1 and s_2. We assume that these two values (estimated visual and auditory locations) are slightly different. The estimated bimodal attribute resulting from the interaction of the two modalities is then the weighted linear combination:

$$\hat{s}_{\mathrm{B}} = w_1\hat{s}_1 + w_2\hat{s}_2, \tag{1}$$

where w_1 and w_2 are the weights of each modality. The weights represent the relative reliability of each modality to provide relevant information about the attribute

of interest. If r_1 and r_2 represent these reliabilities, the weight of the first modality is defined as:

$$w_1 = r_1/(r_1 + r_2) \tag{2}$$

and similarly for the weight of the second modality (thus the weights sum to one). Under MLE, the reliability of a modality is inversely related to the variability of the estimates it provides. For instance, if σ_1^2 represents the variance of the attribute in the first modality, the corresponding reliability is defined as:

$$r_1 = 1/\sigma_1^2. \tag{3}$$

In other words, the more variable a modality is, the less reliable it is, and the less it will drive the final bimodal percept. The MLE rule is optimal in the sense that it identifies the combined estimate that offers the lowest variance. In particular, the combined estimate will always have a lower variance than either of the unimodal estimates. When the uncertainties of the estimates follow Gaussian distributions, or when the discrepancy between the modalities is very small, the uncertainty of the combined estimate is given by:

$$1/\sigma_B^2 = 1/\sigma_1^2 + 1/\sigma_2^2. \tag{4}$$

It is important to emphasize that the MLE procedure predicts both the mean value of the bimodal estimate ($\hat{s}_B$) and its variance (σ_B^2). These predictions have been verified in a large range of very different situations, indicating that sensory modalities are often integrated in a fashion closely approximating the MLE model (Ernst and Bülthoff, 2004). Examples can be found in a variety of contexts, including audio-visual (Alais and Burr 2004b), visual-tactile (Ernst and Banks, 2002), and even trimodal contexts (Wozny *et al.*, 2008), as well as between independent cues within a single modality (Hillis *et al.*, 2002). MLE integration appears to occur automatically and independently of the level of attention directed to the component stimuli (Helbig and Ernst, 2008). There is also evidence that the perceptual estimates based on each component cue are not lost when MLE takes place across modalities, but that they are lost when MLE integration takes place within a single modality (Hillis *et al.*, 2002).

The MLE model provides a formalization of some older ideas in the cross-modal literature. In particular, the 'modality appropriateness hypothesis', according to which inconsistencies between modalities are resolved in favour of the most relevant modality (Welch and Warren, 1980), appears to be well explained by the MLE model. This hypothesis, for instance, predicts a dominance of vision over audition for all spatial judgments (such as ventriloquism) because spatial sensitivity is higher in the visual domain than in the auditory domain. Conversely, modality appropriateness predicts that audition should dominate vision for temporal tasks (such as auditory driving or Sham's 'double flash' illusion) because the auditory modality is specialised for temporal processing.

The MLE model provides a quantitative and principled alternative to the modality appropriateness hypothesis. More importantly, the MLE model offers two ad-

vantages. First, it is a flexible combination rule, rather than the rigid assumption of visual spatial dominance or auditory temporal dominance of the modality appropriateness hypothesis. As shown by Alais and Burr (2004b), reverse ventriloquism, where audition dominates vision in audiovisual spatial localisation, can occur if visual signals are degraded and made unreliable. This result was predicted by the MLE model, but is not captured by the modality appropriateness hypothesis. Second, it has been clear since early studies (Rock and Victor, 1964) that one sensory modality rarely dominates completely over another one: there is always a residual contribution from the dominated modality. MLE captures this in that the estimate from the less reliable modality is always factored into the combined estimate but is simply down-weighted because of its low reliability. It therefore continues to contribute to the combined estimate, albeit with reduced influence.

3.6. Motion Perception from Audio-Visual Cues

Motion perception offers an acute challenge to multisensory integration because the signals evolve over time (Soto-Faraco and Kingstone, 2004). Proper integration of auditory and visual motion signals not only depends on their spatio-temporal coincidence, but on the ability to track this correlation over time and to accurately weight the contribution of each modality. A number of studies have addressed the question of whether auditory and visual modalities do interact for the perception of motion. Although the balance of evidence suggests there are not specialized audiovisual motion detectors, it is clear that the perception of motion can be influenced by static and motion cues in either modality.

Using clearly visible and audible stimuli, an auditory moving stimulus influences the perceived direction of a visual moving target, and this bias occurs even when the auditory and visual signals come from different locations or move at different speeds (Meyer and Wuerger, 2001). In other words, there was an audiovisual interaction for motion but it was not specific to their particular spatio-temporal characteristics. In contrast, another study found that perception of visual motion was not affected by the presentation of a simultaneous auditory moving stimulus (Soto-Faraco *et al.*, 2004). In addition, these authors found that the perceived direction of the auditory moving stimulus was impaired by the simultaneous presentation of a visual motion and they accounted for this effect with an illusory reversal of perceived direction of sounds. The opposite results between these two studies originate in the different stimulus setups used by these authors, but also on the use of suprathreshold stimuli that potentially offered alternative strategies for performing the task. More recent studies use threshold stimuli to address the issue of the level at which auditory and visual signals interact.

When visual and auditory moving stimuli are presented near threshold, the two signals are found to combine at a decision level, supposedly after the stimuli are processed independently in the visual and auditory pathways (Wuerger *et al.*, 2003). This conclusion is indicated by the fact that the benefit in using both signals was not better than that predicted by a simple probability summation model. In other

words, they failed to find a sensory interaction between auditory and visual motion signals. Another study found a similarly small advantage in detection threshold for an audio-visual motion display relative to unimodal detection that was entirely consistent with a probabilistic combination of the signals (Alais and Burr, 2004a). In contrast to these studies, an investigation examining the automaticity of motion integration argued in favour of sensory-level integration of auditory and visual motion signals (Soto-Faraco *et al.*, 2005). Consistent with this conclusion, motion discrimination performance with auditory and visual signals at threshold agrees with predictions from a neural summation model (Meyer *et al.*, 2005), and in addition, this sensory integration occurred only when the auditory and visual signals were co-localised and moved with the same speed and direction. More recently, a study examining audiovisual speed perception found evidence for sensory integration for co-localised auditory and visual components when they were similar in reliability, but probabilistic (decision-level) combination for components of very different reliabilities (Bentvelzen *et al.*, 2009). In summary, it appears that auditory and visual signals can interact at both sensory and decision levels depending on the experimental conditions (Sanabria *et al.*, 2007).

If auditory and visual signals can interact for motion perception, the next question is how. In a study varying visual signal uncertainty, it was shown that auditory signals had little effect on the localisation of a moving target when visual uncertainty was low, but did exert an influence when visual uncertainty increased (Heron *et al.*, 2004). In their experiment, visual uncertainty was manipulated by varying the size of the target (small targets had a large uncertainty). Results consistent with this reliability-weighted finding were reported in an audiovisual speed discrimination study in which signal reliability was manipulated by adding random positional noise to a series of locations in rapid apparent motion sequences (Bentvelzen *et al.*, 2009). Findings such as these can be interpreted within a Maximum Likelihood Estimation framework in which signals are combined according to their reliabilities (Ernst and Banks, 2002).

Direction is not the only motion attribute that can benefit from an interaction of auditory and visual modalities. A study of the extent to which visual motion can influence the perception of auditory speed found evidence of a strong influence of visual velocity rather than of the more elementary components of visual spatial and temporal frequency (as visual speed is calculated by the ratio of temporal to spatial frequency) (López-Moliner and Soto-Faraco, 2007). In addition, a simple (static) visual cue is sufficient to disambiguate the perceived direction of a rapid circular auditory motion (Lakatos, 1995).

Recent studies have focused on other kinds of motion interaction. Auditory motion can help the detectability of visual biological motion (a complex action revealed purely by visual dynamic cues) presented in noise, but only if the motion direction was congruent with the visual signal (Brooks *et al.*, 2007). Auditory motion can also affect the location of static stimuli in the visual domain. In a cross-modal version of the well-known visual flash-lag effect (Alais and Burr, 2003), it

was shown that a stationary disk briefly flashed at the moment a translating sound passed beside it was perceived as lagging behind its true position. In a study of more complex movements, the planum temporale (PT) has been proposed as being involved in the integration of visual and auditory information of complex motion (Hasegawa *et al.*, 2004). Their results indicate that this cortical area is activated while participants with a good piano training watched the complex sequence of hand movements and tried to match the corresponding sounds to identify the music piece. Musical expertise can also influence the integration of biological motion visual displays of drumming and corresponding sounds (Petrini *et al.*, 2009).

Other types of motion perception benefit from the interaction of auditory and visual modalities, but not necessarily when both modalities are in motion. For instance, an ascending pitch can bias the perceived direction of an ambiguous motion display in the upward direction (Maeda *et al.*, 2004), and the perceived direction of an alternating apparent motion display can be biased simply by presenting non-moving sounds at appropriate times (Getzmann, 2007; Freeman and Driver, 2008). It is as if the timing of the sounds was capturing the visual flashes to induce a more salient motion perception in one direction. Moreover, adaptation to this bimodal display produced a visual motion after-effect in the opposite direction. A visual motion stimulus can also induce an auditory after-effect, with several minutes of adaptation to a square moving in depth causing a steady sound to be perceived as changing in loudness in the opposite direction (Kitagawa and Ichihara, 2002). When a sound was combined with the visual motion during adaptation, the after-effect was stronger when both modalities were consistent and weaker when they were inconsistent. Visual motion can also influence the contingent auditory motion after-effect (Vroomen and de Gelder, 2003).

The search for the neural correlates of the integration of auditory and visual signals for motion perception is currently very active. A cortical area usually involved in visual motion processing (MT+/V5 in humans) was found to be activated by auditory motion in two participants who had been blind since early childhood and whose vision was partially recovered in adulthood (Saenz *et al.*, 2008). In contrast, visually normal controls did not show similar activations by auditory moving stimuli. In normal individuals, audiovisual motion stimuli seem to activate primarily the superior temporal gyrus (Baumann and Greenlee, 2007) and area MT+/V5 (Alink *et al.*, 2008). More work will help determine the specific role of distinct neural structures.

3.7. Attention Across Modalities

Attention is the process that allows us to sort among incoming sensory stimuli to select certain locations or objects of interest from among competing stimuli. Selected stimuli benefit from faster and heightened processing relative to unattended stimuli. Much of the early and pioneering attentional work was done in the auditory modality (Cherry, 1953), although in the recent few decades most research activity on attention has been done in the visual modality (Pashler, 1998). More recently inter-

est has turned to attention in cross-modal contexts. In all cases, the basic principle is the same: to study the process of selection among competing objects, however a cross-modal context offers the scope to ask further interesting questions. For example, is it possible to attend selectivity to one modality instead of another (e.g., to attend to audition rather than vision)? Is it possible to attend selectivity to a visual object among auditory and visual distractors? Does spatial attention to a location cued by one modality also prime responses at that location for targets presented in another modality? Are there separate resources for attention in each modality, or are there common central resources? All of these questions have been explored over the last decade or so and greatly expanded our knowledge of attentional processes (Wright and Ward, 2008).

It is often claimed that vision is the dominant sensory modality, and it may be the case that attention is directed predominantly to vision (Posner *et al.*, 1976). Evidence for this comes from the Colavita effect (Colavita, 1974; Spence, 2009) in which occasional auditory stimuli are missed in a stream of visual targets, even though they are not missed in isolation. However, it is also clear that observers are quite able to direct their attention selectively to a single modality when required, at the expense of other modalities (Spence *et al.*, 2001a; Driver and Spence, 2004). This is also observed for exogenous attention, with transient non-predictive cues in a particular modality briefly priming that modality so that responses to targets are faster for that modality (Turatto *et al.*, 2004).

A related question addressed by a number of studies is whether attentional resources are supramodal or modality specific. Using the attentional blink paradigm, it has been claimed that the post target 'blink' in a rapid stream of stimuli is greater when measured in the same modality compared to another modality (Duncan *et al.*, 1997; Arnell and Jolicoeur, 1999). Results from another temporal attention approach found similar results, with processing deficits being modality specific (Soto-Faraco and Spence, 2002). In a dual-task experiment that involved pairing tasks within or between modalities, little or no cost (relative to single-task performance) was found when doing two tasks cross-modally, but a very large cost was found when doing two tasks within a single modality (Alais *et al.*, 2006). Studies such as these point to there being considerable independence of attentional resources across sensory modalities (Bonnel and Hafter, 1998; Rees *et al.*, 2001; Larsen *et al.*, 2003), as has often been advocated in the human factors literature (Triesman and Davies, 1973; Wickens, 1980; Sarter, 2007).

With evidence for cross-modally linked attentional processes as well as for independent processes, the current challenge is how to accommodate these diverse findings within a single model. There are two opposing and mutually exclusive models. One claims that attention involves a single supramodal system and the other claims that attention comprises separate and independent resources that are specific to each sensory modality (Wickens, 1980; Hancock *et al.*, 2007). Between these extremes there are models which combine elements of both (Posner, 1990; Driver and Spence, 2004). These hybrid models are supported by various findings.

For example, evidence of independent resources across modalities is sometimes strong (Duncan *et al.*, 1997; Alais *et al.*, 2006) but is rarely absolute, suggesting some common attentional links. Also, examining differences between studies suggest that using speeded responses will tend to reveal stronger evidence of common resources than will unspeeded tasks. Another important factor is the type of task and stimulus used, with simple sensory tasks (e.g., pitch discrimination, contrast discrimination) more likely to reveal independent processing, while higher-level tasks (Pashler, 1998) are more likely to reveal common supramodal resources.

Techniques such as neuroimaging and transcranial magnetic stimulation (TMS) will help shed light on these competing models, but there is no consensus at present. For instance, there is neuroimaging evidence for cross-modal (visual-tactile) links in spatial attention (Macaluso *et al.*, 2002; Kida *et al.*, 2007) that supports a proposed supramodal parietal attentional system. However, there is also TMS evidence for modality-specific resources for strategic allocation of spatial attention between touch and vision (Chambers *et al.*, 2004). In that study, disruption of right parietal cortex impaired visual orienting but not somatosensory orienting. Given that TMS can more directly address whether specific areas are critical for a given process, evidence from this technique will make telling contributions to the debate over attentional models. Another TMS study focusing on reflexive allocation of attention has shown that parietal disruption will disturb reflexive attention both within and between modalities (Chambers *et al.*, 2007). Other neuroimaging studies have shown a trade-off in neural activation between sensory modalities. For example, when auditory and visual stimuli are present, attention to one modality raises neural activity associated with that modality while activity related to the unattended modality falls (Kawashima *et al.*, 1995; Johnson and Zatorre, 2005, 2006). Although these findings appear to show that attention can modulate unisensory cortical areas independently, it is also consistent with a central process drawing resources away from one modality to another. Overall, further studies are needed to resolve which model of attention operates in multisensory contexts, with event-related TMS having potential to provide critical input to the debate.

3.8. Audiovisual Speech Perception

The difference between multisensory interactions in the cortex and superior colliculus is well illustrated by the McGurk effect (McGurk and MacDonald, 1976). This well-known audiovisual effect does not require spatially coincident signals in order to be perceived (Spence and Driver, 2004). The McGurk effect is also a very powerful demonstration of audiovisual integration in speech, where visual input completely changes the heard percept, and has long been thought to be automatic. The McGurk effect is present in prelinguistic infants (Burnham and Dodd, 2004), has been shown not to require attention (Bertelson *et al.*, 2000), and is very robust despite considerable spatial and temporal disparity between the auditory and visual signals (Grant *et al.*, 2004). Recent evidence, however, suggests that the McGurk effect does require some attention, and so is not entirely automatic. This was shown

by two studies which had observers do an attentional task on visual images overlaid on the visual sequence of the speech signal (Tiippana *et al.*, 2004; Alsius *et al.*, 2005). Moreover, it appears independent access to the visual and auditory components is possible (Soto-Faraco and Alsius, 2007). Consistent with the McGurk effect not being automatic, a very recent study using a bistable visual stimulus (Rubin's face/vase stimulus) found the effect was reduced when the face percept was not experienced, suggesting conscious awareness of the talker is necessary (Munhall *et al.*, 2009).

The McGurk effect is an example where vision of the talker is available, in addition to the acoustic speech signals. This is known as 'seen speech' and it is one of the most actively studied audiovisual interactions. Seeing speech affords many benefits. Apart from the obvious benefit that would be expected from observing lip movements, a good deal of information is conveyed by other visual information, such as movement of the head and eyebrows (Yehia *et al.*, 2002; Thomas and Jordan, 2004) and even haptic movements have been shown to influence speech perception (Fowler and Dekle, 1991). It has also been shown that the visual kinematics from a talker's face and head correlate with the frequency spectrum of what was spoken (Yehia *et al.*, 2002), and a similar result was found by Munhall *et al.* (Munhall *et al.*, 2004).

Given the audiovisual correlations in seen speech, it is not surprising that seen speech generally leads to better comprehension. In one of the earliest studies to show this, speech comprehension in auditory noise was found to improve when the talker was visible (Sumby and Pollack, 1954). At high noise levels, comprehension improvements were large, equivalent to an improvement in auditory signal-to-noise ratio of about 12 to 15 dB (Sumby and Pollack, 1954; Erber, 1975). More recent studies show that these benefits, although more modest, also occur when the auditory context is not noisy (Remez, 2005). The most recent work suggests that the benefit of seeing speech is most apparent at moderate auditory noise levels (Ross *et al.*, 2007) and varies depending on noise and word complexity in a manner consistent with Bayesian optimal integration (Ma *et al.*, 2009).

In a recent review, it was argued that there are two key aspects to facilitations seen in audiovisual speech perception (Campbell, 2008). One exploits complementary information, and turns on the fact that vision and audition can provide different but complementary sources of information regarding speech. For example, some phonemes are easily distinguished visually, and would provide a basis for comprehension when in acoustic noise or in the case of acoustically confusable phonemes. The other aspect involves correlations between seen and heard speech (Yehia *et al.*, 2002; Munhall *et al.*, 2004; Thomas and Jordan, 2004) which provides redundant information to improve speech perception. Audiovisual redundancies in speech could improve perception through visual signals boosting auditory speech signals (Schroeder *et al.*, 2008) and also enhance cognitive functions more generally as less cognitive resources would need to be allocated for effective speech comprehension.

3.9. Neural Bases of Audiovisual Speech

Detailed reviews of the neural bases of audiovisual speech can be found elsewhere (Callan *et al.*, 2004; Calvert *et al.*, 2004; Capek *et al.*, 2004). Briefly, the superior temporal sulcus, especially the posterior region, appears to be a key area for integration of audiovisual speech (Calvert *et al.*, 2000). Evidence of speech-related neural integration in superior temporal regions occurs relatively early in time (Miller and D'Esposito, 2005), within about 150 ms of the auditory signal's onset (Möttönen *et al.*, 2004; van Wassenhove *et al.*, 2005). STS activation even occurs in response to speaking point-light faces (Santi *et al.*, 2003). Multisensory areas such as the superior temporal sulcus also project back to primary sensory cortices, and probably play a role in the visual modulation of auditory cortex (Kayser *et al.*, 2008) and activity in auditory cortex seen in lip reading of silent speech (Calvert *et al.* 2000, 2004). Feedback probably also underlies effects of seen speech observed in the brain stem (Musacchia *et al.*, 2006).

A recent study employing an fMRI adaptation paradigm studied the McGurk effect and found activity in superior temporal sulcus and intraparietal sulcus relating to the McGurk condition relative to conditions not supporting the McGurk illusion (Benoit *et al.*, 2009). A number of neuroimaging studies have also examined silent speech (lip-reading) and have shown activity in superior temporal regions (Calvert *et al.*, 1997; Bernstein *et al.*, 2002; Hall *et al.*, 2005), in particular the posterior superior temporal sulcus (pSTS). Lip-reading of silent speech has even been shown to produce activity in primary auditory cortex (Pekkola *et al.*, 2005).

4. Concluding Remarks

This review has highlighted the breadth and depth of the multisensory enterprise. It is a vast multidisciplinary endeavour, with much already achieved yet with so much still remaining to be done. For the neurophysiologists there are outstanding questions such as how the senses manage to cooperate despite different coordinate frames, and how multisensory salience is encoded cortically if not by superadditivity. For the latter question, we will need to learn more about how populations combine multisensory information, and how they encode the significance of congruent multisensory stimuli. There is also the matter of how multisensory integration works in natural environments that are typically cluttered with competing unisensory signals. One tantalizing suggestion is that coherent oscillations between corresponding unimodal neural signals may provide a basis for selecting which unimodal signals are related and should be integrated and which should not (Senkowski *et al.*, 2008).

For those investigating perception, there is the ever-present danger of confounding response biases with genuine signatures of multisensory integration. The use of two-interval forced-choice procedures together with analyses based on signal detection theory should ensure that response bias and genuine sensitivity improvements can be separated, so that it can be known whether the benefits of multisensory

stimuli reflect a sensory-level integration *per se*, or a statistical improvement at the decision level due to probability summation or uncertainty reduction. Researchers working in cognition and perception will no doubt move progressively towards more complex and naturalistic stimuli to replace the simple beeps and flashes that are so common in current research. After all, multisensory integration is the key to so much that is fundamental to our lives, including speech, social perception, and music, and far more sophisticated stimuli are needed to approximate these experiences in the laboratory. As the research stimuli continue to evolve, it is likely that models based on MLE will continue to inform basic sensory integration, with the use of priors providing scope to include higher-level aspects of mutlisensory function including learning and expectation.

In addition, research in the time domain is likely to become more critical, both in perceptual and cognitive investigations, as well as in neuroimaging. Knowledge of how multisensory interactions occur over time in the brain will help inform neural models by separating effects due to feedforward convergence from feedback interactions. The technique of TMS may well prove to be the most critical method in guiding our understanding of global neural functioning in multisensory progressing as this 'transient lesioning' technique can illuminate which areas are critical for multisensory integration and which are not. Another key technique will be EEG, as its tight temporal resolution can establish the order and time-course of neural events.

Overall, the last decade has been a prolific one in the field of multisensory processing that has seen many significant advances. Yet, the field is still young and the next decade is poised to deliver even more.

References

Alais, D. and Burr, D. (2003). The 'Flash-Lag' effect occurs in audition and cross-modally, *Curr. Biol.* **13**, 59–63.

Alais, D. and Burr, D. (2004a). No direction-specific bimodal facilitation for audiovisual motion detection, *Brain Res. Cogn. Brain Res.* **19**, 185–194.

Alais, D. and Burr, D. (2004b). The ventriloquist effect results from near-optimal bimodal integration, *Curr. Biol.* **14**, 257–262.

Alais, D. and Carlile, S. (2005). Synchronizing to real events: subjective audiovisual alignment scales with perceived auditory depth and speed of sound, *Proc. Natl Acad. Sci. USA* **102**, 2244–2247.

Alais, D., Morrone, C. and Burr, D. (2006). Separate attentional resources for vision and audition, *Proc. Biol. Sci.* **273**, 1339–1345.

Alink, A., Singer, W. and Muckli, L. (2008). Capture of auditory motion by vision is represented by an activation shift from auditory to visual motion cortex, *J. Neurosci.* **28**, 2690–2697.

Alsius, A., Navarra, J., Campbell, R. and Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands, *Curr. Biol.* **15**, 839–843.

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

Amedi, A., Raz, N., Pianka, P., Malach, R. and Zohary, E. (2003). Early 'visual' cortex activation correlates with superior verbal memory performance in the blind, *Nat. Neurosci.* **6**, 758–766.

Amedi, A., Floel, A., Knecht, S., Zohary, E. and Cohen, L. G. (2004). Transcranial magnetic stimulation of the occipital pole interferes with verbal processing in blind subjects, *Nat. Neurosci.* **7**, 1266–1270.

Arabzadeh, E., Clifford, C. W. and Harris, J. A. (2008). Vision merges with touch in a purely tactile discrimination, *Psychol. Sci.* **19**, 635–641.

Arnell, K. M. and Jolicoeur, P. (1999). Revisiting within-modality and cross-modality attentional blinks: effects of target-distractor similarity, *J. Exper. Psychol., Human Percept. Perform.* **66**, 1147–1161.

Arrighi, R., Alais, D. and Burr, D. (2005). Neural latencies do not explain the auditory and audio-visual flash-lag effect, *Vision Research* **45**, 2917–2925.

Arrighi, R., Alais, D. and Burr, D. (2006). Perceptual synchrony of audiovisual streams for natural and artificial motion sequences, *J. Vision* **6**, 260–268.

Aschersleben, G. and Bertelson, P. (2003). Temporal ventriloquism: crossmodal interaction on the time dimension. 2. Evidence from sensorimotor synchronization, *Intl J. Psychophysiol.* **50**, 157–163.

Avillac, M., Deneve, S., Olivier, E., Pouget, A. and Duhamel, J. R. (2005). Reference frames for representing visual and tactile locations in parietal cortex, *Nature Neurosci.* **8**, 941–949.

Avillac, M., Ben Hamed, S. and Duhamel, J. R. (2007). Multisensory integration in the ventral intraparietal area of the macaque monkey, *J. Neurosci.* **27**, 1922–1932.

Barraclough, N. E., Xiao, D., Baker, C. I., Oram, M. W. and Perrett, D. I. (2005). Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions, *J. Cogn. Neurosci.* **17**, 377–391.

Battaglia, P. W., Jacobs, R. A. and Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization, *J. Opt. Soc. Amer. A, Opt. Image Sci. Vis.* **20**, 1391–1397.

Baumann, O. and Greenlee, M. W. (2007). Neural correlates of coherent audiovisual motion perception, *Cereb. Cortex* **17**, 1433–1443.

Beauchamp, M. S., Lee, K. E., Argall, B. D. and Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus, *Neuron* **41**, 809–823.

Beauchamp, M. S., Yasar, N. E., Frye, R. E. and Ro, T. (2008). Touch, sound and vision in human superior temporal sulcus, *Neuroimage* **41**, 1011–1120.

Benevento, L. A., Fallon, J., Davis, B. J. and 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, *Expere. Neurol.* **57**, 849–872.

Benoit, M., Raij, T., Lin, F., Jääskeläinen, I. and Stufflebeam, S. (2009). Primary and multisensory cortical activity is correlated with audiovisual percepts, *Human Brain Mapping* (in press; published online, DOI: 10.1002/hbm.20884).

Bentvelzen, A., Leung, J. and Alais, D. (2009). Discriminating audiovisual speed: optimal integration of speed defaults to probability summation when component reliabilities diverge, *Perception* **38**, 966–987.

Bernstein, L. E., Auer, E. T., Moore, J. K., Ponton, C. W., Don, M. and Singh, M. (2002). Visual speech perception without primary auditory cortex activation, *Neuroreport* **13**, 311–315.

Bertelson, P. and Aschersleben, G. (1998). Automatic visual bias of perceived auditory location, *Psychonomic Bull. Rev.* **5**, 472–489.

Bertelson, P. and Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance, *Percept. Psychophys.* **29**, 578–584.

Bertelson, P., Vroomen, J., de Gelder, B. and Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention, *Percept. Psychophys.* **62**, 321–332.

Bonnel, A. M. and Hafter, E. R. (1998). Divided attention between simultaneous auditory and visual signals, *Percept. Psychophys.* **60**, 179–190.

Bresciani, J. P., Dammeier, F. and Ernst, M. O. (2006). Vision and touch are automatically integrated for the perception of sequences of events, *J. Vision* **6**, 554–564.

Brooks, A., van der Zwan, R., Billard, A., Petreska, B., Clarke, S. and Blanke, O. (2007). Auditory motion affects visual biological motion processing, *Neuropsychologia* **45**, 523–530.

Buchel, C., Price, C., Frackowiak, R. S. and Friston, K. (1998). Different activation patterns in the visual cortex of late and congenitally blind subjects, *Brain* **121**, 409–419.

Burnham, D. and Dodd, B. (2004). Auditory-visual speech integration by prelinguistic infants: perception of an emergent consonant in the McGurk effect, *Develop. Psychobiol.* **45**, 204–220.

Burr, D. and Alais, D. (2006). Combining visual and auditory information, *Prog. Brain Res.* **155**, 243–258.

Burton, H., Snyder, A. Z., Diamond, J. B. and Raichle, M. E. (2002). Adaptive changes in early and late blind: a FMRI study of verb generation to heard nouns, *J. Neurophysiol.* **88**, 3359–3371.

Caclin, A., Soto-Faraco, S., Kingstone, A. and Spence, C. (2002). Tactile 'capture' of audition, *Percept. Psychophys.* **64**, 616–630.

Callan, D. E., Jones, J. A., Munhall, K., Kroos, C., Callan, A. M. and Vatikiotis-Bateson, E. (2004). Multisensory integration sites identified by perception of spatial wavelet filtered visual speech gesture information, *J. Cogn. Neurosci.* **16**, 805–816.

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., Woodruff, P. W., Iversen, S. D. and David, A. S. (1997). Activation of auditory cortex during silent lipreading, *Science* **276**, 593–596.

Calvert, G. A., Campbell, R. and Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex, *Curr. Biol.* **10**, 649–657.

Calvert, G., Spence, C. and Stein, B. E. (Eds) (2004). *The Handbook of Multisensory Processing*. MIT Press, Cambridge, MA, USA.

Campanella, S. and Belin, P. (2007). Integrating face and voice in person perception, *Trends Cogn. Sci.* **11**, 535–543.

Campbell, R. (2008). The processing of audio-visual speech: empirical and neural bases. *Philosoph. Trans. Royal Soc. London B Biol. Sci.* **363**, 1001–1010.

Capek, C. M., Bavelier, D., Corina, D., Newman, A. J., Jezzard, P. and Neville, H. J. (2004). The cortical organization of audio-visual sentence comprehension: an fMRI study at 4 Tesla, *Brain Res. Cogn. Brain Res.* **20**, 111–119.

Cappe, C. and Barone, P. (2005). Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey, *Eur. J. Neurosci.* **22**, 2886–2902.

Carriere, B. N., Royal, D. W., Perrault, T. J., Morrison, S. P., Vaughan, J. W., Stein, B. E. and Wallace, M. T. (2007). Visual deprivation alters the development of cortical multisensory integration, *J. Neurophysiol.* **98**, 2858–2867.

Cass, J. and Alais, D. (2006). Evidence for two interacting temporal channels in human visual processing, *Vision Research* **46**, 2859–2868.

Chambers, C. D., Stokes, M. G. and Mattingley, J. B. (2004). Modality-specific control of strategic spatial attention in parietal cortex, *Neuron* **44**, 925–930.

Chambers, C. D., Payne, J. M. and Mattingley, J. B. (2007). Parietal disruption impairs reflexive spatial attention within and between sensory modalities, *Neuropsychologia* **45**, 1715–1724.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears, *J. Acoust. Soc. Amer.* **25**, 975–979.

Clavagnier, S., Falchier, A. and Kennedy, H. (2004). Long-distance feedback projections to area V1: implications for multisensory integration, spatial awareness, and visual consciousness, *Cogn. Affect. Behav. Neurosci.* **4**, 117–126.

Cohen, Y. E. and Andersen, R. A. (2002). A common reference frame for movement plans in the posterior parietal cortex, *Nat. Rev. Neurosci.* **3**, 553–562.

Cohen, L. G., Celnik, P., Pascual-Leone, A., Corwell, B., Falz, L., Dambrosia, J., Honda, M., Sadato, N., Gerloff, C., Catala, M. D. and Hallett, M. (1997). Functional relevance of cross-modal plasticity in blind humans, *Nature* **389**, 180–183.

Colavita, F. B. (1974). Human sensory dominance, *Percept. Psychophys.* **16**, 409–412.

Collignon, O., Davare, M., Olivier, E. and De Volder, A. G. (2009a). Reorganisation of the right occipito-parietal stream for auditory spatial processing in early blind humans. A transcranial magnetic stimulation study, *Brain Topogr.* **21**, 232–240.

Collignon, O., Voss, P., Lassonde, M. and Lepore, F. (2009b). Cross-modal plasticity for the spatial processing of sounds in visually deprived subjects, *Exper. Brain Res.* **192**, 343–358.

De Lange, H. (1958). Research into the dynamic nature of the human fovea-cortex systems with intermittent and modulated light. I. Attenuation characteristics with white and colored light, *J. Opt. Soc. Amer.* **48**, 777–784.

Dixon, N. F. and Spitz, L. (1980). The detection of auditory visual desynchrony, *Perception* **9**, 719–721.

Driver, J. and Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments, *Neuron* **57**, 11–23.

Driver, J. and Spence, C. (2004). Crossmodal spatial attention: evidence from human performance, in: *Crossmodal Space and Crossmodal Attention*, Spence, C. and Driver, J. (Eds). Oxford University Press, Oxford, UK.

Duncan, J., Martens, S. and Ward, R. (1997). Restricted attentional capacity within but not between sensory modalities, *Nature* **387**, 808–810.

Erber, N. P. (1975). Auditory-visual perception of speech, *J. Speech Hearing Disorders* **40**, 481–492.

Ernst, M. and Banks, M. (2002). Humans integrate visual and haptic information in a statistically optimal fashion, *Nature* **415**, 429–433.

Ernst, M. O. and Bülthoff, H. H. (2004). Merging the senses into a robust percept, *Trends Cogn. Sci.* **8**, 162–169.

Evans, E. F. and Whitfield, I. C. (1964). Classification of unit responses in the auditory cortex of the unanaesthetized and unrestrained cat, *J. Physiol.* **171**, 476–493.

Fain, G. L. (2003). *Sensory Transduction*. Sinauer Associates, Sunderland, MA, USA.

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

Felleman, D. J. and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex, *Cereb. Cortex* **1**, 1–47.

Fendrich, R. and Corballis, P. M. (2001). The temporal cross-capture of audition and vision, *Percept. Psychophys.* **63**, 719–725.

Finney, E. M., Fine, I. and Dobkins, K. R. (2001). Visual stimuli activate auditory cortex in the deaf, *Nature Neurosci.* **4**, 1171–1173.

Fortin, M., Voss, P., Lord, C., Lassonde, M., Pruessner, J., Saint-Amour, D., Rainville, C. and Lepore, F. (2008). Wayfinding in the blind: larger hippocampal volume and supranormal spatial navigation, *Brain* **131**, 2995–3005.

Fowler, C. A. and Dekle, D. J. (1991). Listening with eye and hand: cross-modal contributions to speech perception, *J. Exper. Psychol., Human Percept. Perform.* **17**, 816–828.

Foxe, J. J., Morocz, I. A., Murray, M. M., Higgins, B. A., Javitt, D. C. and Schroeder, C. E. (2000). Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping, *Brain Res. Cogn. Brain Res.* **10**, 77–83.

Foxe, J. J., Wylie, G. R., Martinez, A., Schroeder, C. E., Javitt, D. C., Guilfoyle, D., Ritter, W. and Murray, M. M. (2002). Auditory-somatosensory multisensory processing in auditory association cortex: an fMRI study, *J. Neurophysiol.* **88**, 540–543.

Frassinetti, F., Bolognini, N. and Ladavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction, *Exper. Brain Res.* **147**, 332–343.

Freeman, E. and Driver, J. (2008). Direction of visual apparent motion driven solely by timing of a static sound, *Curr. Biol.* **18**, 1262–1266.

Fu, K. M., Johnston, T. A., Shah, A. S., Arnold, L., Smiley, J., Hackett, T. A., Garraghty, P. E. and Schroeder, C. E. (2003). Auditory cortical neurons respond to somatosensory stimulation, *J. Neurosci.* **23**, 7510–7515.

Fujisaki, W. and Nishida, S. (2009). Audio-tactile superiority over visuo-tactile and audio-visual combinations in the temporal resolution of synchrony perception, *Exper. Brain Res.* **198**, 245–259.

Fujisaki, W., Shimojo, S., Kashino, M. and Nishida, S. (2004). Recalibration of audiovisual simultaneity, *Nature Neurosci.* **7**, 773–778.

Fujisaki, W., Koene, A., Arnold, D., Johnston, A. and Nishida, S. (2006). Visual search for a target changing in synchrony with an auditory signal, *Proc. Biol. Sci.* **273**, 865–874.

Galton, F. (1899). On instruments for (1) testing perception of differences of tint and for (2) determining reaction time, *J. Anthropolog. Inst.* **19**, 27–29.

Gaunet, F. and Thinus-Blanc, C. (1996). Early-blind subjects' spatial abilities in the locomotor space: exploratory strategies and reaction-to-change performance, *Perception* **25**, 967–981.

Gebhard, J. W. and Mowbray, G. H. (1959). On discriminating the rate of visual flicker and auditory flutter, *Amer. J. Psychol.* **72**, 521–529.

Gepshtein, S., Burge, J., Ernst, M. O. and Banks, M. S. (2005). The combination of vision and touch depends on spatial proximity, *J. Vision* **5**, 1013–1023.

Gescheider, G. A., Kane, M. J., Sager, L. C. and Ruffolo, L. J. (1974). The effect of auditory stimulation on responses to tactile stimuli, *Bull. Psychonomic Soc.* **3**, 204–206.

Getzmann, S. (2007). The effect of brief auditory stimuli on visual apparent motion, *Perception* **36**, 1089–1103.

Ghazanfar, A. A. and Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends Cogn. Sci.* **10**, 278–285.

Ghazanfar, A. A., Maier, J. X., Hoffman, K. L. and Logothetis, N. K. (2005). Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex, *J. Neurosci.* **25**, 5004–5012.

Gillmeister, H. and Eimer, M. (2007). Tactile enhancement of auditory detection and perceived loudness, *Brain Res.* **1160**, 58–68.

Goyal, M. S., Hansen, P. J. and Blakemore, C. B. (2006). Tactile perception recruits functionally related visual areas in the late-blind, *Neuroreport* **17**, 1381–1384.

Grant, K., Greenberg, S., Poeppel, D. and van V. (2004). Effects of spectro-temporal asynchrony in auditory and auditory-visual speech processing, *Seminars in Hearing* **25**, 241–255.

Graziano, M. S. (2001). A system of multimodal areas in the primate brain, *Neuron* **29**, 4–6.

Green, D. M. and Swets, J. A. (1964). *Signal Detection and Recognition by Human Observers*. Wiley, New York, NY, USA.

Groh, J. M. and Sparks, D. L. (1996). Saccades to somatosensory targets. III. Eye-position-dependent somatosensory activity in primate superior colliculus, *J. Neurophysiol.* **75**, 439–453.

Guest, S., Catmur, C., Lloyd, D. and Spence, C. (2002). Audiotactile interactions in roughness perception, *Exper. Brain Res.* **146**, 161–171.

Hall, D. A., Fussell, C. and Summerfield, A. Q. (2005). Reading fluent speech from talking faces: typical brain networks and individual differences, *J. Cogn. Neurosci.* **17**, 939–953.

Hamilton, R., Keenan, J. P., Catala, M. and Pascual-Leone, A. (2000). Alexia for Braille following bilateral occipital stroke in an early blind woman, *Neuroreport* **11**, 237–240.

Hancock, P. A., Oron-Gilad, T. and Szalma, J. L. (2007). Elaborations of the multiple-resource theory of attention, in: *Attention: From Theory to Practice*, Kramer, A. F., Wiegmann, D. A. and Kirlik, A. (Eds), pp. 45–56. Oxford University Press, Oxford, UK.

Hartline, P. H., Vimal, R. L., King, A. J., Kurylo, D. D. and Northmore, D. P. (1995). Effects of eye position on auditory localization and neural representation of space in superior colliculus of cats, *Exper. Brain Res.* **104**, 402–408.

Hasegawa, T., Matsuki, K., Ueno, T., Maeda, Y., Matsue, Y., Konishi, Y. and Sadato, N. (2004). Learned audio-visual cross-modal associations in observed piano playing activate the left planum temporale. An fMRI study, *Brain Res. Cogn. Brain Res.* **20**, 510–518.

Hay, J. C., Pick, H. L. and Ikeda, K. (1965). Visual capture produced by prism spectacles, *Psychonom. Sci.* **2**, 215–216.

Hein, G., Doehrmann, O., Muller, N. G., Kaiser, J., Muckli, L. and Naumer, M. J. (2007). Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas, *J. Neurosci.* **27**, 7881–7887.

Helbig, H. B. and Ernst, M. O. (2007). Knowledge about a common source can promote visual-haptic integration, *Perception* **36**, 1523–1533.

Helbig, H. B. and Ernst, M. O. (2008). Visual-haptic cue weighting is independent of modality-specific attention, *J. Vision* **8**, 1–16.

Heron, J., Whitaker, D. and McGraw, P. V. (2004). Sensory uncertainty governs the extent of audio-visual interaction, *Vision Research* **44**, 2875–2884.

Heron, J., Whitaker, D., McGraw, P. V. and Horoshenkov, K. V. (2007). Adaptation minimizes distance-related audiovisual delays, *J. Vision* **7**, 1–8.

Hillis, J. M., Ernst, M. O., Banks, M. S. and Landy, M. S. (2002). Combining sensory information: mandatory fusion within, but not between, senses, *Science* **298**, 1627–1630.

Hirsch, I. J. and Sherrick, C. E. (1961). Perceived order in different sense modalities, *J. Exper. Psychol.* **62**, 423–432.

Hocking, J. and Price, C. J. (2008). The role of the posterior superior temporal sulcus in audiovisual processing, *Cereb. Cortex* **18**, 2439–2449.

Howard, I. P. and Templeton, W. B. (1966). *Human Spatial Orientation*. Wiley, New York, NY, USA.

Hubel, D. H. and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex, *J. Physiol.* **160**, 106–154.

Jiang, W., Wallace, M. T., Jiang, H., Vaughan, J. W. and Stein, B. E. (2001). Two cortical areas mediate multisensory integration in superior colliculus neurons, *J. Neurophysiol.* **85**, 506–522.

Jiang, W., Jiang, H. and Stein, B. E. (2006). Neonatal cortical ablation disrupts multisensory development in superior colliculus, *J. Neurophysiol.* **95**, 1380–1396.

Johnson, J. A. and Zatorre, R. J. (2005). Attention to simultaneous unrelated auditory and visual events: behavioral and neural correlates, *Cereb. Cortex* **15**, 1609–1620.

Johnson, J. A. and Zatorre, R. J. (2006). Neural substrates for dividing and focusing attention between simultaneous auditory and visual events, *Neuroimage* **31**, 1673–1681.

Jones, E. G. and Powell, T. P. (1970). An anatomical study of converging sensory pathways within the cerebral cortex of the monkey, *Brain* **93**, 793–820.

Kauffman, T., Theoret, H. and Pascual-Leone, A. (2002). Braille character discrimination in blindfolded human subjects, *Neuroreport* **13**, 571–574.

Kawashima, R., O'Sullivan, B. T. and Roland, P. E. (1995). Positron-emission tomography studies of cross-modality inhibition in selective attentional tasks: closing the 'mind's eye', *Proc. Nat. Acad. Sci. USA* **92**, 5969–5972.

Kayser, C., Petkov, C. I. and Logothetis, N. K. (2008). Visual modulation of neurons in auditory cortex, *Cereb. Cortex* **18**, 1560–1574.

Kersten, D., Mamassian, P. and Yuille, A. (2004). Object perception as Bayesian inference, *Ann. Rev. Psychology* **55**, 271–304.

Kida, T., Inui, K., Wasaka, T., Akatsuka, K., Tanaka, E. and Kakigi, R. (2007). Time-varying cortical activations related to visual-tactile cross-modal links in spatial selective attention, *J. Neurophysiol.* **97**, 3585–3596.

King, A. J. (2009). Visual influences on auditory spatial learning, *Philos. Trans. Royal Soc. London B Biol. Sci.* **364**, 331–339.

Kitagawa, N. and Ichihara, S. (2002). Hearing visual motion in depth, *Nature* **416**, 172–174.

Kopinska, A. and Harris, L. R. (2004). Simultaneity constancy, *Perception* **33**, 1049–1060.

Kujala, T., Huotilainen, M., Sinkkonen, J., Ahonen, A. I., Alho, K., Hämäläinen, M. S., Ilmoniemi, R. J., Kajola, M., Knuutila, J. E. and Lavikainen, J. (1995). Visual cortex activation in blind humans during sound discrimination, *Neurosci. Letts* **183**, 143–146.

Lakatos, S. (1995). The influence of visual cues on the localisation of circular auditory motion, *Perception* **24**, 457–465.

Larsen, A., McIlhagga, W., Baert, J. and Bundesen, C. (2003). Seeing or hearing? Perceptual independence, modality confusions, and crossmodal congruity effects with focused and divided attention, *Percept. Psychophys.* **65**, 568–574.

Lewald, J. and Guski, R. (2004). Auditory-visual temporal integration as a function of distance: no compensation for sound-transmission time in human perception, *Neurosci. Letts* **357**, 119–122.

López-Moliner, J. and Soto-Faraco, S. (2007). Vision affects how fast we hear sounds move, *J. Vision* **7**, 1–7.

Lovelace, C. T., Stein, B. E. and Wallace, M. T. (2003). An irrelevant light enhances auditory detection in humans: a psychophysical analysis of multisensory integration in stimulus detection, *Brain Res. Cogn. Brain Res.* **17**, 447–453.

Ma, W. J., Zhou, X., Ross, L. A., Foxe, J. J. and Parra, L. C. (2009). Lip-reading aids word recognition most in moderate noise: a Bayesian explanation using high-dimensional feature space, *PLoS One* **4**, e4638.

Macaluso, E., Frith, C. D. and Driver, J. (2002). Directing attention to locations and to sensory modalities: multiple levels of selective processing revealed with PET, *Cereb. Cortex* **12**, 357–368.

Maeda, F., Kanai, R. and Shimojo, S. (2004). Changing pitch induced visual motion illusion, *Curr. Biol.* **14**, R990–991.

McDonald, J. J., Teder-Salejarvi, W. A., Heraldez, D. and Hillyard, S. A. (2001). Electrophysiological evidence for the 'missing link' in crossmodal attention, *Canad. J. Exper. Psychol.* **55**, 141–149.

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

Merabet, L. B., Hamilton, R., Schlaug, G., Swisher, J. D., Kiriakopoulos, E. T., Pitskel, N. B., Kauffman, T. and Pascual-Leone, A. (2008). Rapid and reversible recruitment of early visual cortex for touch, *PLoS One* **3**, e3046.

Meredith, M. A. and Stein, B. E. (1990). The visuotopic component of the multisensory map in the deep laminae of the cat superior colliculus, *J. Neurosci.* **10**, 3727–3742.

Meredith, M. A., Nemitz, J. W. and Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors, *J. Neurosci.* **7**, 3215–3229.

Meyer, G. F. and Wuerger, S. M. (2001). Cross-modal integration of auditory and visual motion signals, *Neuroreport* **12**, 2557–2560.

Meyer, G. F., Wuerger, S. M., Röhrbein, F. and Zetzsche, C. (2005). Low-level integration of auditory and visual motion signals requires spatial co-localisation, *Exper. Brain Res., Experimentelle Hirnforschung, Expérimentation Cérébrale* **166**, 538–547.

Miller, L. M. and D'Esposito, M. (2005). Perceptual fusion and stimulus coincidence in the cross-modal integration of speech, *J. Neurosci.* **25**, 5884–5893.

Morein-Zamir, S., Soto-Faraco, S. and Kingstone, A. (2003). Auditory capture of vision: examining temporal ventriloquism, *Brain Res. Cogn. Brain Res.* **17**, 154–163.

Möttönen, R., Schürmann, M. and Sams, M. (2004). Time course of multisensory interactions during audiovisual speech perception in humans: a magnetoencephalographic study, *Neurosci. Letts* **363**, 112–115.

Mountcastle, V. B. (1957). Modality and topographic properties of single neurons of cat's somatic sensory cortex, *J. Neurophysiol.* **20**, 408–434.

Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T. and Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility: head movement improves auditory speech perception, *Psychol. Sci.* **15**, 133–137.

Munhall, K. G., Hove, M. W., Brammer, M. and Paré, M. (2009). Audiovisual integration of speech in a bistable illusion, *Curr. Biol.* **19**, 735–739.

Musacchia, G., Sams, M., Nicol, T. and Kraus, N. (2006). Seeing speech affects acoustic information processing in the human brainstem, *Exper. Brain Res., Experimentelle Hirnforschung, Expérimentation Cérébrale* **168**, 1–10.

Newell, F. N., Ernst, M. O., Tjan, B. S. and Bulthoff, H. H. (2001). Viewpoint dependence in visual and haptic object recognition, *Psychol. Sci.* **12**, 37–42.

Odgaard, E. C., Arieh, Y. and Marks, L. E. (2004). Brighter noise: sensory enhancement of perceived loudness by concurrent visual stimulation, *Cogn. Affect. Behav. Neurosci.* **4**, 127–132.

Pashler, H. (1998). *The Psychology of Attention*. MIT Press, Cambridge, MA, USA.

Pasqualotto, A. and Newell, F. N. (2007). The role of visual experience on the representation and updating of novel haptic scenes, *Brain Cogn.* **65**, 184–194.

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

Pekkola, J., Ojanen, V., Autti, T., Jääskeläinen, I. P., Möttönen, R., Tarkiainen, A. and Sams, M. (2005). Primary auditory cortex activation by visual speech: an fMRI study at 3 T, *Neuroreport* **16**, 125–128.

Penfield, W. and Rasmussen, T. (1950). *The Cerebral Cortex of Man*. Macmillan, New York, NY, USA.

Petrini, K., Dahl, S., Rocchesso, D., Waadeland, C. H., Avanzini, F., Puce, A. and Pollick, F. E. (2009). Multisensory integration of drumming actions: musical expertise affects perceived audiovisual asynchrony, *Exper. Brain Res., Experimentelle Hirnforschung, Expérimentation Cérébrale* **198**, 339–352.

Posner, M. I. (1990). Hierarchical distributed networks in the neuropsychology of selective attention, in: *Cognitive Neuropsychology and Neurolinguistics: Advances in Models of Cognitive Function and Impairment*, Carramazza, A. (Ed.), pp. 187–210. Lawrence Erlbaum, Hillsdale, NJ, USA.

Posner, M. I., Nissen, M. J. and Klein, R. M. (1976). Visual dominance: an information-processing account of its origins and significance, *Psychol. Rev.* **83**, 157–171.

Postma, A., Zuidhoek, S., Noordzij, M. L. and Kappers, A. M. (2008). Haptic orientation perception benefits from visual experience: evidence from early-blind, late-blind, and sighted people, *Percept. Psychophys.* **70**, 1197–1206.

Pouget, A., Deneve, S. and Duhamel, J.-R. (2002). A computational perspective on the neural basis of multisensory spatial representations, *Nat. Rev. Neurosci.* **3**, 741–747.

Radeau, M. and Bertelson, P. (1974). The after-effects of ventriloquism, *Quart. J. Exper. Psychol.* **26**, 63–71.

Recanzone, G. H. (1998). Rapidly induced auditory plasticity: the ventriloquism aftereffect, *Proc. Nat. Acad. Sci. USA* **95**, 869–875.

Recanzone, G. H. (2003). Auditory influences on visual temporal rate perception, *J. Neurophysiol.* **89**, 1078–1093.

Rees, G., Frith, C. and Lavie, N. (2001). Processing of irrelevant visual motion during performance of an auditory attention task, *Neuropsychologia* **39**, 937–949.

Remez, R. E. (2005). Three puzzles of multimodal speech perception, in: *Audiovisual Speech*, Vatikiotis-Bateson, E., Bailly, G. and Perrier, P. (Eds), pp. 12–19. MIT Press, Cambridge, MA, USA.

Repp, B. H. and Penel, A. (2002). Auditory dominance in temporal processing: new evidence from synchronization with simultaneous visual and auditory sequences, *J. Exper. Psychol., Hum. Percept. Perform.* **28**, 1085–1099.

Rice, C. (1970). Early blindness, early experience and perceptual enhancement, *Amer. Found. Blind Res. Bull.* **22**, 1–22.

Rock, I. and Victor, J. (1964). Vision and touch: an experimentally created conflict between the two senses, *Science* **143**, 594–596.

Rockland, K. S. and Ojima, H. (2003). Multisensory convergence in calcarine visual areas in macaque monkey, *Intl J. Psychophysiol.* **50**, 19–26.

Röder, B., Teder-Salejarvi, W., Sterr, A., Rosler, F., Hillyard, S. A. and Neville, H. J. (1999). Improved auditory spatial tuning in blind humans, *Nature* **400**, 162–166.

Röder, B., Stock, O., Bien, S., Neville, H. and Rosler, F. (2002). Speech processing activates visual cortex in congenitally blind humans, *Eur. J. Neurosci.* **16**, 930–936.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C. and Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments, *Cereb. Cortex* **17**, 1147–1153.

Rowland, B. A., Quessy, S., Stanford, T. R. and Stein, B. E. (2007). Multisensory integration shortens physiological response latencies, *J. Neurosci.* **27**, 5879–5884.

Sadato, N., Pascual-Leone, A., Grafman, J., Ibanez, V., Deiber, M. P., Dold, G. and Hallett, M. (1996). Activation of the primary visual cortex by Braille reading in blind subjects, *Nature* **380**, 526–528.

Saenz, M., Lewis, L. B., Huth, A. G., Fine, I. and Koch, C. (2008). Visual motion area MT+/V5 responds to auditory motion in muman sight-recovery subjects, *J. Neurosci.* **28**, 5141–5148.

Saito, D. N., Okada, T., Honda, M., Yonekura, Y. and Sadato, N. (2006). Practice makes perfect: the neural substrates of tactile discrimination by Mah-Jong experts include the primary visual cortex, *BMC Neurosci.* **7**, 79.

Sanabria, D., Spence, C. and Soto-Faraco, S. (2007). Perceptual and decisional contributions to audio-visual interactions in the perception of apparent motion: a signal detection study, *Cognition* **102**, 299–310.

Santi, A., Servos, P., Vatikiotis-Bateson, E., Kuratate, T. and Munhall, K. (2003). Perceiving biological motion: dissociating visible speech from walking, *J. Cogn. Neurosci.* **15**, 800–809.

Sarter, N. B. (2007). Multiple-resource theory as a basis for multimodal interface design: success stories, qualifications, and research needs, in: *Attention: from Theory to Practice*, Kramer, A. F., Wiegmann, D. A. and Kirlik, A. (Eds), pp. 187–195. Oxford University Press, Oxford, UK.

Sathian, K., Zangaladze, A., Hoffman, J. M. and Grafton, S. T. (1997). Feeling with the mind's eye, *Neuroreport* **8**, 3877–3881.

Schlack, A., Sterbing-D'Angelo, S. J., Hartung, K., Hoffmann, K. P. and Bremmer, F. (2005). Multisensory space representations in the macaque ventral intraparietal area, *J. Neurosci.* **25**, 4616–4625.

Schroeder, C. E. and Foxe, J. J. (2002). The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex, *Brain Res. Cogn. Brain Res.* **14**, 187–198.

Schroeder, C. E., Lindsley, R. W., Specht, C., Marcovici, A., Smiley, J. F. and Javitt, D. C. (2001). Somatosensory input to auditory association cortex in the macaque monkey, *J. Neurophysiol.* **85**, 1322–1327.

Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S. and Puce, A. (2008). Neuronal oscillations and visual amplification of speech, *Trends Cogn. Sci.* **12**, 106–113.

Schurmann, M., Caetano, G., Jousmaki, V. and Hari, R. (2004). Hands help hearing: facilitatory audiotactile interaction at low sound-intensity levels. *J. Acoust. Soc. Amer.* **115**, 830–832.

Schutz, M. and Lipscomb, S. (2007). Hearing gestures, seeing music: vision influences perceived tone duration, *Perception* **36**, 888–897.

Sekuler, R., Sekuler, A. B. and Lau, R. (1997). Sound alters visual motion perception, *Nature* **385**, 308.

Senkowski, D., Schneider, T. R., Foxe, J. J. and Engel, A. K. (2008). Crossmodal binding through neural coherence: implications for multisensory processing, *Trends Neurosci.* **31**, 401–409.

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

Shipley, T. (1964). Auditory flutter-driving of visual flicker, *Science* **145**, 1328–1330.

Simon-Dack, S. L., Rodriguez, P. D. and Teder-Salejarvi, W. A. (2008). Psychophysiology and imaging of visual cortical functions in the blind: a review, *Behav. Neurol.* **20**, 71–81.

Slutsky, D. A. and Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect, *Neuroreport* **12**, 7–10.

Soto-Faraco, S. and Alsius, A. (2007). Conscious access to the unisensory components of a crossmodal illusion, *Neuroreport* **18**, 347–350.

Soto-Faraco, S. and Kingstone, A. (2004). Multisensory integration of dynamic information, in: *The Handbook of Multisensory Processes*, Calvert, G. A., Spence, C. and Stein, B. E. (Eds). MIT Press, Cambridge, MA, USA.

Soto-Faraco, S. and Spence, C. (2002). Modality-specific auditory and visual temporal processing deficits, *Quart. J. Exper. Psychol. A* **55**, 23–40.

Soto-Faraco, S., Spence, C. and Kingstone, A. (2004). Cross-modal dynamic capture: congruency effects in the perception of motion across sensory modalities, *J. Exper. Psychol., Human. Percept. Perform.* **30**, 330–345.

Soto-Faraco, S., Spence, C. and Kingstone, A. (2005). Assessing automaticity in the audiovisual integration of motion, *Acta Psychologica* **118**, 71–92.

Spence, C. (2009). Explaining the Colavita visual dominance effect, *Prog. Brain Res.* **176**, 245–258.

Spence, C. and Driver, J. (Eds) (2004). *Crossmodal Space and Crossmodal Attention*. Oxford University Press, Oxford, UK.

Spence, C., Nicholls, M. E. and Driver, J. (2001a). The cost of expecting events in the wrong sensory modality, *Percept. Psychophys.* **63**, 330–336.

Spence, C., Shore, D. I. and Klein, R. M. (2001b). Multisensory prior entry, *J. Exper. Psychol. Gen.* **130**, 799–832.

Spence, C., Baddeley, R., Zampini, M., James, R. and Shore, D. I. (2003). Multisensory temporal order judgments: when two locations are better than one, *Percept. Psychophys.* **65**, 318–328.

Stanford, T. R., Quessy, S. and Stein, B. E. (2005). Evaluating the operations underlying multisensory integration in the cat superior colliculus, *J. Neurosci.* **25**, 6499–6508.

Stein, B. E. and Meredith, M. A. (1993). *The Merging of the Senses*. MIT Press, Cambridge, MA, USA.

Stein, B. E. and Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron, *Nature. Rev. Neurosci.* **9**, 255–266.

Stein, B. E. and Wallace, M. T. (1996). Comparisons of cross-modality integration in midbrain and cortex, *Prog. Brain Res.* **112**, 289–299.

Stein, B. E., Wallace, M. W., Stanford, T. R. and Jiang, W. (2002). Cortex governs multisensory integration in the midbrain, *Neuroscientist* **8**, 306–314.

Sternberg, S. and Knoll, R. L. (1973). The perception of temporal order: fundamental issues and a general model, in: *Attention and Performance IV*, Kornblum, S. (Ed.), pp. 629–685. Academic Press, New York, USA.

Stetson, C., Cui, X., Montague, P. R. and Eagleman, D. M. (2006). Motor-sensory recalibration leads to an illusory reversal of action and sensation, *Neuron* **51**, 651–659.

Stevenson, R. A. and James, T. W. (2009). Audiovisual integration in human superior temporal sulcus: inverse effectiveness and the neural processing of speech and object recognition, *Neuroimage* **44**, 1210–1223.

Sugita, Y. and Suzuki, Y. (2003). Audiovisual estimation of sound-arrival time, *Nature* **421**, 911.

Sumby, W. and Pollack, I. (1954). Visual contribution to speech intelligibility in noise, *J. Acoust. Soc. Amer.* **26**, 212–215.

Thinus-Blanc, C. and Gaunet, F. (1997). Representation of space in blind persons: vision as a spatial sense? *Psychol. Bull.* **121**, 20–42.

Thomas, S. M. and Jordan, T. R. (2004). Contributions of oral and extraoral facial movement to visual and audiovisual speech perception, *J. Exper. Psychol., Human Percept. Perform.* **30**, 873–888.

Tiippana, K., Andersen, T. and Sams, M. (2004). Visual attention modulates audiovisual speech perception, *Eur. J. Cogn. Psychol.* **16**, 457–472.

Treisman, A. M. and Gelade, G. (1980). A feature-integration theory of attention, *Cogn. Psychol.* **12**, 97–136.

Triesman, A. M. and Davies, A. (1973). Divided attention to ear and eye, in: *Attention and Performance*, Kornblum, S. (Ed.), Vol. IV, pp. 101–117. Academic Press, New York, USA.

Turatto, M., Galfano, G., Bridgeman, B. and Umilta, C. (2004). Space-independent modality-driven attentional capture in auditory, tactile and visual systems, *Exper. Brain Res.* **155**, 301–310.

van Atteveldt, N. M., Formisano, E., Blomert, L. and Goebel, R. (2007). The effect of temporal asynchrony on the multisensory integration of letters and speech sounds, *Cereb. Cortex* **17**, 962–974.

van der Burg, E., Cass, J., Olivers, C. N., Theeuwes, J. and Alais, D. (2010). Efficient visual search from non-spatial auditory cues requires more than temporal synchrony, *PLoS One* (in press).

van Ee, R., van Boxtel, J. J., Parker, A. L. and Alais, D. (2009). Multisensory congruency as a mechanism for attentional control over perceptual selection, *J. Neurosci.* **29**, 11641–11649.

van Wassenhove, V., Grant, K. W. and Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech, *Proc. Nat. Acad. Sci. USA* **102**, 1181–1186.

Vatakis, A. and Spence, C. (2006). Audiovisual synchrony perception for music, speech, and object actions, *Brain Res.* **1111**, 134–142.

Voss, P., Gougoux, F., Zatorre, R. J., Lassonde, M. and Lepore, F. (2008). Differential occipital responses in early- and late-blind individuals during a sound-source discrimination task, *Neuroimage* **40**, 746–758.

Vroomen, J. and de Gelder, B. (2003). Visual motion influences the contingent auditory motion aftereffect, *Psycholog. Sci.: J. Amer. Psycholog. Soc./APS* **14**, 357–361.

Vroomen, J. and de Gelder, B. (2004). Temporal ventriloquism: sound modulates the flash-lag effect, *J. Exper. Psychol., Human Percept. Perform.* **30**, 513–518.

Vroomen, J., Bertelson, P. and de Gelder, B. (2001). The ventriloquist effect does not depend on the direction of automatic visual attention, *Percept. Psychophys.* **63**, 651–659.

Walker, J. T. and Scott, K. J. (1981). Auditory-visual conflicts in the perceived duration of lights, tones and gaps, *J. Exper. Psychol., Human Percept. Perform.* **7**, 1327–1339.

Wallace, M. T. and Stein, B. E. (2001). Sensory and multisensory responses in the newborn monkey superior colliculus, *J. Neurosci.* **21**, 8886–8894.

Wanet-Defalque, M. C., Veraart, C., De Volder, A., Metz, R., Michel, C., Dooms, G. and Goffinet, A. (1988). High metabolic activity in the visual cortex of early blind human subjects, *Brain Res.* **446**, 369–373.

Welch, R. B. and Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy, *Psycholog. Bull.* **88**, 638–667.

Wickens, C. D. (1980). The structure of attentional resources, in: *Attention and Performance*, Nickerson, R. (Ed.), Vol. VIII, pp. 239–257. Lawrence Erlbaum, Hillsdale, NJ, USA.

Wozny, D. R., Beierholm, U. R. and Shams, L. (2008). Human trimodal perception follows optimal statistical inference, *J. Vision* **8**, 1–11.

Wright, R. D. and Ward, L. M. (2008). *Orienting of Attention*. Oxford University Press, Oxford, UK.

Wuerger, S. M., Hofbauer, M. and Meyer, G. F. (2003). The integration of auditory and visual motion signals at threshold, *Percept. Psychophys.* **65**, 1188–1196.

Yau, J. M., Olenczak, J. B., Dammann, J. F. and Bensmaia, S. J. (2009). Temporal frequency channels are linked across audition and touch, *Curr. Biol.* **19**, 561–566.

Yehia, H., Kuratate, T. and Vatikiotis-Bateson, E. (2002). Linking facial animation, head motion and speech acoustics, *J. Phonetics* **30**, 555–568.

Zampini, M., Shore, D. I. and Spence, C. (2003). Multisensory temporal order judgments: the role of hemispheric redundancy, *Intl J. Psychophysiol.* **50**, 165–180.

Zangaladze, A., Epstein, C. M., Grafton, S. T. and Sathian, K. (1999). Involvement of visual cortex in tactile discrimination of orientation, *Nature* **401**, 587–590.