MINI-REVIEW

On the 'visual' in 'audio-visual integration': a hypothesis concerning visual pathways

**Philip Jaekl · Alexis Pérez-Bellido ·
Salvador Soto-Faraco**

Abstract Crossmodal interaction conferring enhancement in sensory processing is nowadays widely accepted. Such benefit is often exemplified by neural response amplification reported in physiological studies conducted with animals, which parallel behavioural demonstrations of sound-driven improvement in visual tasks in humans. Yet, a good deal of controversy still surrounds the nature and interpretation of these human psychophysical studies. Here, we consider the interpretation of crossmodal enhancement findings under the light of the functional as well as anatomical specialization of magno- and parvocellular visual pathways, whose paramount relevance has been well established in visual research but often overlooked in crossmodal research. We contend that a more explicit consideration of this important visual division may resolve some current controversies and help optimize the design of future crossmodal research.

Introduction

Crossmodal influences on basic visual tasks have been extensively documented in recent years, with evidence spanning across a wide range of methods, animal species and experimental paradigms (see Shams and Kim 2010; Vroomen and Keetels 2010 for reviews). Here, we focus on auditory–visual interactions that lead to enhancement in visual performance in humans. Concerning modality combinations, we consider specifically audio-visual interactions because they have been investigated most comprehensively and are commonly related with the widespread assumption that crossmodal integration confers an adaptive advantage to organisms (Lewkowicz and Kraebel 2004; Bahrick et al. 2004). We consider the term enhancement in a broad sense, describing situations where a sound can cause faster and/ or more accurate and/or more precise perception of a visual event, compared to when there is no concurrent sound. Sound-driven enhancements of vision include reports of decreases in response latencies to visual targets (Miller 1982; Corneil et al. 2002), lowering of detection thresholds (Caclin et al. 2011; Frassinetti et al. 2002; Gleiss and Kayser 2013; Jaekl and Harris 2009; Jaekl and Soto-Faraco 2010; Noesselt et al. 2010), decreases in visual search time (Van der Burg et al. 2008), increases in brightness judgments (Stein et al. 1996), increases in perceived duration of brief visual stimuli (Walker and Scott 1981; Vroomen and de Gelder 2000; Van Wassenhove et al. 2008), faster motion detection (Meyer et al. 2005) and increased visual saliency (Noesselt et al. 2008). Enhancement is only one of

P. Jaekl (✉)
Department of Brain and Cognitive Sciences, Center for Visual Science, University of Rochester, Rochester, NY, USA
e-mail: pjaekl@cvs.rochester.edu

A. Pérez-Bellido
Visió i Control de l'acció (VISCA) Group, Departament de Psicologia Bàsica, Universitat de Barcelona, Barcelona, Spain

A. Pérez-Bellido · S. Soto-Faraco
Multisensory Research Group (MRG), Department de Tecnologies de la Informació i les Comunicacions, Cognition and Brain Centre, Universitat Pompeu Fabra, Barcelona, Spain

S. Soto-Faraco
Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain

several possible outcomes of multisensory integration and is distinguished from other multisensory phenomena conferring what may be considered performance detriments or illusions (e.g. Shams et al. 2000; Sinnett et al. 2008; Thurlow and Jack 1973) or changes in information content (e.g. McGurk and MacDonald 1976). The hypothesis laid out here may well apply to these manifestations of multisensory integration arising from inter-sensory conflict, but fall beyond the scope of the present article. We focus, instead, on crossmodally induced enhancements demonstrated in basic visual judgment tasks because they are often used to underscore direct multisensory interactions occurring at relatively short latencies and in hierarchically early stages of processing. Such phenomena have typically been linked to physiological interactions in subcortical or primary sensory areas—defined as 'early', sensory-based interaction (Driver and Noesselt 2008; Stein and Stanford 2008; Shams and Kim 2010).

Perhaps surprisingly, the interpretation of this common example of multisensory interaction, namely sound-driven enhancement of vision in human behavioural paradigms is not often agreed upon. For example, studies supporting such enhancement include a number of psychophysical audio-visual investigations involving subjective brightness ratings (Stein et al. 1996) along with those using visual detection tasks (Frassinetti et al. 2002; Bolognini et al. 2005; Manjarrez et al. 2007; Andersen and Mamassian 2008; Caclin et al. 2011). Such enhancements have often been measured using paradigms effective for determining sensory-based signal combination independent of higher-level influences (e.g. decision, attentive state—see Ngo and Spence 2012). Sensory-level interactions are consistent with known early, low-level physiological processes (Meredith and Stein 1983; Wilkinson et al. 1996; Wallace et al. 1998; Molholm et al. 2002; Lehmann and Murray 2005; Kayser et al. 2005; Lakatos et al. 2007; Driver and Noesselt 2008; Clemo et al. 2012) and have been sometimes related to the discovery of direct (i.e. monosynaptic) cortico-cortical connections between sensory areas in anatomical studies (Falchier et al. 2002; Rockland and Ojima 2003; Cappe and Barone 2005; Smiley and Falchier 2009; Meredith et al. 2009; also see Lewis and Noppeney 2010 for fMRI-based support). Yet, a considerable number of other psychophysical studies have failed to support this early sensory-based interpretation of sound-driven enhancement of vision (e.g. Meyer and Wuerger 2001; Marks et al. 2003; Odgaard et al. 2003; Alais and Burr 2004; Schnupp et al. 2005; Lippert et al. 2007 see also Kayser and Logothetis 2007). These studies convincingly argue instead for various alternative explanations of enhancement effects, based on other known processes such as attentional orienting, reduction in temporal uncertainty or biases at the level of decision/response (for a relevant discussion see: De Gelder and Bertelson 2003). This category includes simple alerting (see de Boer-Schellekens et al. 2013) based on unspecific subcortical—cortical interactions related to fast changes in arousal (Sturm and Willmes 2001; Maravita and Iriki 2004). For example, findings related to very fast and spatially unspecific crossmodal enhancement have been attributed to such phenomena (Murray et al. 2005). Such an account, however, does not explain enhancements found when auditory stimuli follow visual target onsets (Miller 1986; Andersen and Mamassian 2008; Leone and McCourt 2013), or when enhancements are based on crossmodal correspondences in specific attribute values such as spatial frequency (Pérez-Bellido et al. 2013—see below).

Thus, sensory-level effects are not consistently confirmed and are therefore only observed under certain conditions. What conditions are common to psychophysical experiments supporting sensory-level enhancement? We believe the answer may be integral to demonstrating audio-visual enhancement and, in part, may be present in the existing literature.

Audio-visual enhancement and visual pathways

We contend that sensory interactions facilitating perceptual enhancement do occur and that inconsistencies in the conclusions of previous studies—sensory-level audio-visual enhancement versus alternative explanations—can arise, in part from the characteristics of the different neural mechanisms underlying the very visual processes that are putatively enhanced by sound. In particular, we reason that early, sensory-level crossmodal influences in a variety of psychophysical tasks can depend mostly on the differential involvement of specialized processing channels existing at low-level stages of visual processing (for reviews see: Livingstone and Hubel 1988, Merigan and Maunsell 1993). For example, contrast thresholds (Shapley 1990) and reaction times to visual onsets can be determined by the early magnocellular division of the visual system (M-system) (Breitmeyer 1975). These M-system properties contrast with the early parvocellular division (P-system), the latter being more efficient at processing chromatic information, high spatial frequencies and higher contrasts. The P-system is thought to subserve colour and form/pattern vision leading to object recognition and figure–ground segregation (Livingstone and Hubel 1988; Merigan 1989; Roe et al. 2012). In natural circumstances, both parvocellular and magnocellular pathways are stimulated by objects and events in the visual world, and there is extensive interaction between these pathways at various stages of cortical processing (Maunsell 1992; Schroeder et al. 1998; Saalmann et al. 2007; Nassi and Callaway 2009). Despite the importance of this division in visual processing, its broad mapping

onto putative 'dorsal' and 'ventral' pathway functioning and its well known impact in visual psychophysics, it is surprising that such visual properties are rarely considered explicitly in crossmodal investigations. Here, we expand on previous empirical work relating audio-visual benefit to visual pathways (Jaekl and Soto-Faraco 2010; Pérez-Bellido et al. 2013) and postulate that some discrepancies in previous findings regarding audio-visual enhancements may be resolved by considering the relative level of involvement and effectiveness of processing within these two visual pathways, in the different experimental paradigms. Specifically, relevant investigations both confirming and failing to confirm sensory-level crossmodal interaction that we discuss are likely to critically involve the effectiveness of M-pathway processing.

Magnocellular-based audio-visual interactions

Auditory and visual neural responses are combined into crossmodal signals at various processing stages in different cortical and subcortical areas. An example often cited in multisensory literature is the superior colliculus (SC), a subcortical structure supporting crossmodal sensory integration. The SC plays an integral role in controlling and executing orienting responses towards novel or behaviourally relevant stimuli—namely saccadic orienting (Lee et al. 1988; Roucoux et al. 1980). In mammals, the audio-visual interaction in the SC occurs primarily in neurons within its intermediate and deep layers, receiving input from both auditory and visual modalities (May 2006) and input from higher, extrastriate areas (see Boehnke and Munoz 2008). Importantly, the primary visual afferent to the SC consists of input from magnocellular layers of the lateral geniculate nucleus (Berson and McIlwain 1982; Schiller et al. 1979), via primary visual area, V1 and direct connections from retinal ganglion cells (Garey and Powell 1968; Garey et al. 1968). This visual input subserves detection, localization, attentional orienting (Shen et al. 2011) and is mostly sensitive to transient, low spatial frequency and low-contrast stimulation (Kaplan and Shapley 1986; Plainis and Murray 2005; Schneider and Kastner 2005).

Indeed, evidence for auditory interaction with M-pathway signals in the SC is demonstrated in the temporal pattern of incoming signals. Auditory transduction typically occurs at earlier latencies than those for visual stimuli (Fain 2003). Similarly, auditory SC response latency [typically 10–44 ms—Meredith et al. 1987; Wise and Irvine 1983 (cat studies), 14 ms in primates—Wallace et al. 1996] precedes visual response latencies (typically 40–70 ms in primates—Bell et al. 2006, also see Boehnke and Munoz 2008), and physiological response enhancement in the SC is consistent with overlapping discharge periods of auditory and visual responses (Meredith et al. 1987). Congruent with these physiological findings, behavioural response latencies to audio-visual stimuli have been shown to be significantly speeded up relative to those obtained in a unisensory visual condition, as measured by manual response and saccadic reaction times (Bernstein et al. 1969; Diederich and Colonius 2004; Gielen et al. 1983; Goldring et al. 1996; Harrington and Peck 1998; Hughes et al. 1994; Miller 1982; Perrott et al. 1990; Pérez-Bellido et al. 2013). Audio-visual interaction conferring such reaction time enhancement has been modelled to conform with SC response patterns (Corneil et al. 2002). Such findings would seem to imply a major role of magnocellular input, affecting the sensitivity of these layers as manifested by visual response characteristics.

Audio-visual interactions in the SC are spatially dependent on activity patterns across receptive fields and have accordingly been found to occur most strongly for spatially aligned audio-visual components (Meredith and Stein 1996; Gepshtein et al. 2005; Meyer et al. 2005, but see Spence 2013). However, Stein et al. (1996) and Fiebelkorn et al. (2011) found audio-visual brightness enhancements for spatially discordant stimuli, suggesting such enhancement might instead result from some degree of interaction occurring at a cortical level (Lakatos et al. 2005; Schroeder and Lakatos 2009; see also Romei et al. 2012 for EEG data in humans). In agreement with Stein et al. (1996) and Fiebelkorn et al. (2011), we hypothesize interaction between auditory response and early visual cortical response may contribute to enhancement for spatially disparate stimuli. At early cortical stages, audio-visual correspondences are complicated by the longer response latencies in V1 relative to A1 (V1 latency, 41–55 ms: Clark and Hillyard 1996; Foxe and Simpson 2002; Foxe and Schroeder 2005; A1 latency, 9–15 ms: Celesia 1976; Clark and Hillyard 1996; Molholm et al. 2002). Specifically fast, contrast-sensitive magnocellular responses (Cleland et al. 1971; Cleland and Levick 1973) and their higher temporal resolution (Kulikowski and Tolhurst 1973; Kaplan and Shapley 1982) may be an optimal candidate for an efficient selection of early cortical crossmodal associations concerning contrast enhancement, congruent with psychophysical findings (Jaekl and Soto-Faraco 2010; Pérez-Bellido et al. 2013).

Psychophysical interpretations of sound-induced enhancement of vision

Given the above, investigations set out to determine behavioural enhancements of vision by sound may often be likely to depend on the effective engagement of early, magnocellular processing. It is therefore notable that these studies have frequently utilized visual stimuli not explicitly

designed to optimally engage the M-pathway. For example, commonly used for visual stimulation in such investigations are abrupt stimuli, well above detection threshold. Such stimuli engage the M-system sensitivity to transient onsets, but they may not always confer opportunity for signal enhancement at the level of perceptual influence. Indeed, visual contrast response gain in the lateral geniculate nucleus can be greater than an order of magnitude in magnocellular layers compared to parvocellular responses (Kaplan and Shapley 1986). Specifically, contrast gains computed by Michaelis–Menten saturation functions show that for achromatic stimulation between Michelson contrast values between 0 and 1, magnocellular cells have gained (impulses per second/% change in contrast) values typically between 5 and 8, whereas parvocellular cells are relatively insensitive, with values typically between 0.15 and 0.5 (Kaplan and Shapley 1986, also see Pokorny 2011). Therefore, abrupt visual stimuli of relatively high contrast can easily saturate early magnocellular response levels, leaving contrast discrimination to be primarily determined by activation patterns in the P-system (see Pokorny 2011 for a review). Specifically, graded responses within the early magnocellular system occur only within a narrow contrast range relative to the mean luminance of the display. Thus, although they elicit a large magnocellular response, the properties of higher contrast stimuli easily saturate M-pathway response levels and may preclude the likelihood for multisensory-based improvement in contrast enhancement paradigms for which the level of magnocellular activation plays an integral role. That is, enhancements here are more likely to occur when additional auditory stimulation can boost a relatively weak magnocellular response above the threshold required for detection or discrimination, rather than when responses to stimuli already detectable relative to the adapted background are at most, weakly modulated by sound, if at all.

For example, Marks et al. (2003) and Odgaard et al. (2003) used visual stimuli in a brightness comparison task with dark-adapted participants, for which the lowest luminance level was one just noticeable difference above the 79 % luminance detection threshold and found no crossmodal enhancement. At this level of detection performance, additional crossmodal stimulation provided by concurrent sound may not yield measurable brightness enhancement in a comparison task relative to threshold levels (Wilkinson et al. 1996). Additionally, Caclin et al. (2011), using a criterion-free detection paradigm, showed no audio-visual improvements in detecting foveal, 11.4 cycle-per-degree Gabor patches. According to prior physiological and psychophysical literature, these stimuli were unlikely to optimally engage magnocellular response (Kulikowski and Tolhurst 1973; Legge 1978; Wilson 1980; Tootell et al. 1988; Livingstone and Hubel 1988; Leonova et al. 2003),

although enhancement was, however, observed in a subset of participants with relatively weak performance in a unimodal visual-only condition.

Noesselt et al. (2010) found consistent sensory-level detection advantages attributable to audio-visual integration. The visual stimuli in this study consisted of Gabor patches calibrated to low, 55 and 65 % contrast thresholds, and the effect was only obtained at the lower contrast level. These findings are in agreement with Stein et al. (1996) who used subjective brightness ratings in a comparison task (but see Odgaard et al. 2003). Using a more direct analysis involving a steady/pulsed-pedestal paradigm specifically designed for the purpose of segregating M- and P-based contrast selectivity, Jaekl and Soto-Faraco (2010) have shown that sensory-level audio-visual contrast enhancement of near-threshold stimuli occurs under conditions selectively favouring magnocellular sensitivity to transient, low spatial frequency conditions. Additionally, Pérez-Bellido et al. (2013) found that sound-induced visual enhancement in RTs could be psychophysically dissociated into separate components. One component of the RT enhancement resulted from interactions occurring in post-sensory stages of processing (i.e. uncertainty reduction, speed up of motor reaction by alerting) and affected reaction times across the entire range of visual spatial frequencies tested, while a sensory-based audio-visual RT benefit occurred selectively for low-frequency visual transients configured for optimal magnocellular sensitivity.

Importantly, such sensory-specific interactions conferring enhancement are in line with the principle of inverse effectiveness (Meredith and Stein 1983), a defining principle of sensory integration which implies that relatively weak stimulus intensities lead to stronger crossmodal interaction. This principle is congruent with the findings of Stein et al. (1996) and Noesselt et al. (2010), who reported stronger brightness enhancement at lower stimulus intensities. However, inverse effectiveness alone cannot account for the crossmodal contrast enhancements observed in Jaekl and Soto-Faraco (2010) and Pérez-Bellido et al. (2013) which was shown only for low rather than high spatial frequency stimuli.

Notably, audio-visual improvement to low-contrast stimuli occurs preferentially for transient rather than sustained inputs (Van der Burg et al. 2010; Werner and Noppeney 2011). Transient inputs are defined by both changes from 'off' to 'on' as well as 'on' to 'off' states and are congruently signalled by brief visual responses throughout several stages in the visual system, including responses in subcortical regions (Cleland et al. 1971; Cleland and Levick 1973; Maunsell et al. 1999) as well as primary visual cortex (Horiguchi et al. 2009). In line with these physiological findings, Andersen and Mamassian (2008) demonstrated that for audio-visual stimuli, crossmodal transient

synchrony was sufficient for eliciting sensory enhancements in a luminance change detection paradigm. Additionally, Van der Burg et al. (2010) found that target detection in visual search can be enhanced by sound when auditory and visual stimuli were transiently presented. Conversely, their study also revealed that sustained but temporally correlated signals were ineffective improving this visual search, manifesting that a precise temporal representation of the stimuli is necessary for multisensory integration in these detection paradigms (see also Zannoli et al. 2012). Altogether, these results highlight the importance of optimal magnocellular sensitivity to relatively high temporal frequencies to produce sound-induced enhancement in visual detection tasks.

Influences, above and beyond early sensory-level interactions clearly have also convincingly been demonstrated. Such influences include those putatively arising from reductions in temporal (Lippert et al. 2007) and/or spatial uncertainty (McDonald et al. 2000; Frassinetti et al. 2002; Bolognini et al. 2005) by means of attention orienting, or those promoted by crossmodally induced biases in decision-level processes. In addition, audio-visual integration can also modulate visual perception at processing stages for which visual signals are more integrated between processing streams (e.g. Werner and Noppeney 2010) and in other aspects for which effective parvocellular (rather than magnocellular) involvement may be critical. These modulations can subserve ventral stream processing, functioning to separate figure from ground (Roe et al. 2012) and aid in object perception (Kourtzi and Connor 2011). Such audio-visual interactions have been supported by demonstrations of the influence of sound in brain areas known to receive parvocellular input that contribute to object-related tasks. For example, influences occurring during object naming or categorization (Colombo and Gross 1994; Bookheimer et al. 1998; Tranel et al. 2003). Psychophysical investigations aimed specifically at demonstrating auditory–parvocellular interaction at a sensory level have revealed that non-informative sounds can attenuate the effectiveness of metacontrast masking and influence orientation judgments of high-frequency Gabor patches (Jaekl and Harris 2009). Performance in both these tasks was designed specifically to be dependent upon the effectiveness of parvocellular processing. Importantly, these paradigms differ in objective from those involving visual detection and reaction time tasks exploiting functional aspects of relatively early M-pathway processing.

Conclusion and future directions

We have placed the focus on the discrepancy between studies both confirming and failing to confirm early, sensory-based crossmodal influences in basic visual tasks. Our contention is that such inconsistencies may at least partly be resolved by considering the major anatomical and functional divisions within the early visual system between the magno- and the parvocellular pathways, which broadly map onto putative dorsal and ventral functions. Specifically, we have emphasized those studies which use tasks concerning primarily M-pathway functions—early crossmodal combinatorial processes influencing basic behaviours such as those involved in fast reactions, luminance detection and contrast enhancement—which can be dependent on the effectiveness of early transient magnocellular signals to indicate the presence and location of a near-threshold object or event. If crossmodal influences are to manifest in these tasks, they are mostly like to occur if stimuli are appropriately optimized for magnocellular sensitivity—broadly defined in terms of low-contrast, low spatial frequency transient stimuli. It is interesting that this apparently simple principle has rarely been considered in previous work regarding sensory interaction. We warrant that such consideration is important in future studies concerning audio-visual enhancement, especially those involving saccadic reaction time measurements, stimulus detection and paradigms concerning contrast sensitivity. Carefully designed experiments that measure strictly sensory-level interactions (e.g. unbiased by spatial and/or temporal cueing), conducted with these considerations in mind may most effectively determine the nature of crossmodal enhancement.

References

Alais D, Burr D (2004) No direction-specific bimodal facilitation for audiovisual motion detection. Brain Res Cogn Brain Res 19:185–194. doi:10.1016/j.cogbrainres.2003.11.011

Andersen TS, Mamassian P (2008) Audiovisual integration of stimulus transients. Vis Res 48:2537–2544. doi:10.1016/j.visres.2008.08.018

Bahrick LE, Lickliter R, Flom R (2004) Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy. Curr Dir Psychol Sci 13:99–102. doi:10.1111/j.0963-7214.2004.00283.x

Bell AH, Meredith MA, Van Opstal AJ, Munoz DP (2006) Stimulus intensity modifies saccadic reaction time and visual response latency in the superior colliculus. Exp Brain Res 174:53–59. doi:10.1007/s00221-006-0420-z

Bernstein IH, Clark MH, Edelstein BA (1969) Effects of an auditory signal on visual reaction time. J Exp Psychol 80:567–569. doi:10.1037/h0027444

Berson D, McIlwain J (1982) Retinal Y-cell activation of deep-layer cells in superior colliculus of the cat. J Neurophysiol 47(4):700–714

Boehnke SE, Munoz DP (2008) On the importance of the transient visual response in the superior colliculus. Curr Opin Neurobiol 18:544–551. doi:10.1016/j.conb.2008.11.004

Bolognini N, Frassinetti F, Serino A, Làdavas E (2005) "Acoustical vision" of below threshold stimuli: interaction among spatially converging audiovisual inputs. Exp Brain Res 160:273–282. doi:10.1007/s00221-004-2005-z

Bookheimer SY, Zeffiro TA, Blaxton TA et al (1998) Regional cerebral blood flow during auditory responsive naming: evidence for cross-modality neural activation. NeuroReport 9:2409–2413

Breitmeyer B (1975) Simple reaction time as a measure of the temporal response properties of transient and sustained channels. Vis Res 15:1411–1412

Caclin A, Bouchet P, Djoulah F et al (2011) Auditory enhancement of visual perception at threshold depends on visual abilities. Brain Res 1396:35–44. doi:10.1016/j.brainres.2011.04.016

Cappe C, Barone P (2005) Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. Eur J Neurosci 22:2886–2902. doi:10.1111/j.1460-9568.2005.04462.x

Celesia GG (1976) Organization of auditory cortical areas in man. Brain 99:403–414. doi:10.1093/brain/99.3.403

Clark V, Hillyard S (1996) Spatial selective attention affects early extrastriate but not striate components of the visual evoked potential. J Cogn Neurosci 8:387–402

Cleland BG, Levick WR, Sanderson KJ (1973) Properties of sustained and transient ganglion cells in the cat retina. J Physiol 228(3):649–680

Cleland B, Dubin MW, Levick WR (1971) Sustained and transient neurones in the cat's retina and lateral geniculate nucleus. J Physiol 217:473–496

Clemo HR, Keniston LP, Meredith MA (2012) Structural basis of multisensory processing: convergence. In: Murray MM, Wallace MT (eds) The Neural bases of multisensory processes. Frontiers in Neuroscience, chapter 1. CRC Press, Boca Raton

Colombo M, Gross C (1994) Responses of inferior temporal cortex and hippocampal neurons during delayed matching-to-sample in monkeys (*Macaca fascicularis*). Behav Neurosci 108:443–455

Corneil BD, Van Wanrooij M, Munoz DP, Van Opstal AJ (2002) Auditory–visual interactions subserving goal-directed saccades in a complex scene. J Neurophysiol 88:438–454

De Boer-Schellekens L, Keetels M, Eussen M, Vroomen J (2013) No evidence for impaired multisensory integration of low-level audiovisual stimuli in adolescents and young adults with autism spectrum disorders. Neuropsychologia 51:3004–3013. doi:10.1016/j.neuropsychologia.2013.10.005

De Gelder B, Bertelson P (2003) Multisensory integration, perception and ecological validity. Trends Cogn Sci 7:460–467. doi:10.1016/j.tics.2003.08.014

Diederich A, Colonius H (2004) Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. Percept Psychophys 66:1388–1404

Driver J, Noesselt T (2008) Multisensory interplay reveals cross-modal influences on "sensory-specific" brain regions, neural responses, and judgments. Neuron 57:11–23. doi:10.1016/j.neuron.2007.12.013

Fain GL (2003) Sensory transduction. Sinauer Associates, Sunderland

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

Fiebelkorn IC, Foxe JJ, Butler JS, Molholm S (2011) Auditory facilitation of visual-target detection persists regardless of retinal eccentricity and despite wide audiovisual misalignments. Exp Brain Res. doi:10.1007/s00221-011-2670-7

Foxe JJ, Schroeder CE (2005) The case for feedforward multisensory convergence during early cortical processing. NeuroReport 16:419–423. doi:10.1097/00001756-200504040-00001

Foxe JJ, Simpson GV (2002) Flow of activation from V1 to frontal cortex in humans. A framework for defining "early" visual processing. Exp Brain Res 142:139–150. doi:10.1007/s00221-001-0906-7

Frassinetti F, Bolognini N, Làdavas E (2002) Enhancement of visual perception by crossmodal visuo-auditory interaction. Exp Brain Res 147:332–343. doi:10.1007/s00221-002-1262-y

Garey LJ, Powell TP (1968) The projection of the retina in the cat. J Anat 102:189–222

Garey LJ, Jones EG, Powell TP (1968) Interrelationships of striate and extrastriate cortex with the primary relay sites of the visual pathway. J Neurol Neurosurg Psychiatry 31:135–157

Gepshtein S, Burge J, Ernst MO, Banks MS (2005) The combination of vision and touch depends on spatial proximity. J Vis 5:1013–1023. doi:10.1167/5.11.7

Gielen SC, Schmidt RA, Van den Heuvel PJ (1983) On the nature of intersensory facilitation of reaction time. Percept Psychophys 34:161–168

Gleiss S, Kayser C (2013) Eccentricity dependent auditory enhancement of visual stimulus detection but not discrimination. Front Integr Neurosci 7:1–8

Goldring J, Dorris M, Corneil B et al (1996) Combined eye–head gaze shifts to visual and auditory targets in humans. Exp Brain Res. doi:10.1007/BF00229557

Harrington LK, Peck CK (1998) Spatial disparity affects visual–auditory interactions in human sensorimotor processing. Exp Brain Res 122:247–252

Horiguchi H, Nakadomari S, Misaki M, Wandell BA (2009) Two temporal channels in human V1 identifiedusing fMRI. NeuroImage 47:273–280

Hughes HC, Reuter-Lorenz PA, Nozawa G, Fendrich R (1994) Visual–auditory interactions in sensorimotor processing: saccades versus manual responses. J Exp Psychol Hum Percept Perform 20:131–153

Jaekl PM, Harris LR (2009) Sounds can affect visual perception mediated primarily by the parvocellular pathway. Vis Neurosci 26:477–486. doi:10.1017/S0952523809990289

Jaekl PM, Soto-Faraco S (2010) Audiovisual contrast enhancement is articulated primarily via the M-pathway. Brain Res 1366:85–92. doi:10.1016/j.brainres.2010.10.012

Kaplan E, Shapley R (1982) X and Y cells in the lateral geniculate nucleus of macaque monkeys. J Physiol 330:125–143

Kaplan E, Shapley RM (1986) The primate retina contains two types of ganglion cells, with high and low contrast sensitivity. Proc Natl Acad Sci USA 83:2755–2757

Kayser C, Logothetis NK (2007) Do early sensory cortices integrate cross-modal information? Brain Struct Funct 212:121–132. doi:10.1007/s00429-007-0154-0

Kayser C, Petkov CI, Augath M, Logothetis NK (2005) Integration of touch and sound in auditory cortex. Neuron 48:373–384. doi:10.1016/j.neuron.2005.09.018

Kourtzi Z, Connor CE (2011) Neural representations for object perception: structure, category, and adaptive coding. Annu Rev Neurosci 34:45–67. doi:10.1146/annurev-neuro-060909-153218

Kulikowski J, Tolhurst D (1973) Psychophysical evidence for sustained and transient detectors in human vision. J Physiol 232(1):149–162

Lakatos P, Shah AS, Knuth KH et al (2005) An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. J Neurophysiol 94:1904–1911. doi:10.1152/jn.00263.2005

Lakatos P, Chen C-M, O'Connell MN et al (2007) Neuronal oscillations and multisensory interaction in primary auditory cortex. Neuron 53:279–292. doi:10.1016/j.neuron.2006.12.011

Lee C, Rohrer WH, Sparks DL (1988) Population coding of saccadic eye movements by neurons in the superior colliculus. Nature 332:357–360. doi:10.1038/332357a0

Legge GE (1978) Sustained and transient mechanisms in human vision: temporal and spatial properties. Vis Res 18:69–81

Lehmann S, Murray MM (2005) The role of multisensory memories in unisensory object discrimination. Brain Res Cogn Brain Res 24:326–334. doi:10.1016/j.cogbrainres.2005.02.005

Leone LM, McCourt ME (2013) The roles of physical and physiological simultaneity in audiovisual multisensory facilitation. Iperception 4:213–228. doi:10.1068/i0532

Leonova A, Pokorny J, Smith VC (2003) Spatial frequency processing in inferred PC- and MC-pathways. Vision Res 43:2133–2139. doi:10.1016/S0042-6989(03)00333-X

Lewis R, Noppeney U (2010) Audiovisual synchrony improves motion discrimination via enhanced connectivity between early visual and auditory areas. J Neurosci 30:12329–12339. doi:10.1523/JNEUROSCI.5745-09.2010

Lewkowicz D, Kraebel K (2004) The value of multisensory redundancy in the development of intersensory perception. In: Calvert GA, Spence C, Stein B (eds) Handbook of multisensory process. MIT Press, Cambridge, pp 655–678

Lippert M, Logothetis NK, Kayser C (2007) Improvement of visual contrast detection by a simultaneous sound. Brain Res 1173:102–109. doi:10.1016/j.brainres.2007.07.050

Livingstone M, Hubel D (1988) Segregation of form, color, movement, and depth: anatomy, physiology, and perception. Science 240:740–749. doi:10.1126/science.3283936

Manjarrez E, Mendez I, Martinez L et al (2007) Effects of auditory noise on the psychophysical detection of visual signals: cross-modal stochastic resonance. Neurosci Lett 415:231–236. doi:10.1016/j.neulet.2007.01.030

Maravita A, Iriki A (2004) Tools for the body (schema). Trends Cogn Sci 8:79–86. doi:10.1016/j.tics.2003.12.008

Marks LE, Ben-artzi E, Lakatos S (2003) Cross-modal interactions in auditory and visual discrimination. Int J Psychophysiol 50:125–145. doi:10.1016/S0167-8760

Maunsell JH (1992) Functional visual streams. Curr Opin Neurobiol 2:506–510

Maunsell JH, Ghose GM, Assad JA et al (1999) Visual response latencies of magnocellular and parvocellular LGN neurons in macaque monkeys. Vis Neurosci 16:1–14

May PJ (2006) The mammalian superior colliculus: laminar structure and connections. Prog Brain Res 151:321–378. doi:10.1016/S0079-6123(05)51011-2

McDonald JJ, Teder-Sälejärvi WA, Hillyard SA (2000) Involuntary orienting to sound improves visual perception. Nature 407:906–908. doi:10.1038/35038085

McGurk H, MacDonald J (1976) Hearing lips and seeing voices. Nature 264:746–748. doi:10.1038/264746a0

Meredith M, Stein BE (1983) Interactions among converging sensory inputs in the superior colliculus. Science 221:389–391. doi:10.1126/science.6867718

Meredith MA, Stein BE (1996) Spatial determinants of multisensory integration in cat superior colliculus neurons. J Neurophysiol 75:1843–1857

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

Meredith MA, Allman BL, Keniston LP, Clemo HR (2009) Auditory influences on non-auditory cortices. Hear Res 258:64–71. doi:10.1016/j.heares.2009.03.005

Merigan WH (1989) Chromatic and achromatic vision of macaques: role of the P pathway. J Neurosci 9:776–783

Merigan WH, Maunsell JH (1993) How parallel are the primate visual pathways? Annu Rev Neurosci 16:369–402. doi:10.1146/annurev.ne.16.030193.002101

Meyer GF, Wuerger SM (2001) Cross-modal integration of auditory and visual motion signals. NeuroReport 12:2557–2560

Meyer GF, Wuerger SM, Röhrbein F, Zetzsche C (2005) Low-level integration of auditory and visual motion signals requires spatial co-localisation. Exp Brain Res 166:538–547. doi:10.1007/s00221-005-2394-7

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

Miller J (1986) Timecourse of coactivation in bimodal divided attention. Percept Psychophys 40:331–343

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

Murray MM, Molholm S, Michel CM, Heslenfeld DJ, Ritter W, Javitt DC, Schroeder CE, Foxe JJ (2005) Grabbing your ear: rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. Cereb Cortex 15:963-974

Nassi JJ, Callaway EM (2009) Parallel processing strategies of the primate visual system. Nat Rev Neurosci 10:360–372. doi:10.1038/nrn2619

Ngo MK, Spence C (2012) Facilitating masked visual target identification with auditory oddball stimuli. Exp Brain Res 221:129–136. doi:10.1007/s00221-012-3153-1

Noesselt T, Bergmann D, Hake M et al (2008) Sound increases the saliency of visual events. Brain Res 1220:157–163. doi:10.1016/j.brainres.2007.12.060

Noesselt T, Tyll S, Boehler CN et al (2010) Sound-induced enhancement of low-intensity vision: multisensory influences on human sensory-specific cortices and thalamic bodies relate to perceptual enhancement of visual detection sensitivity. J Neurosci 30:13609–13623. doi:10.1523/JNEUROSCI.4524-09.2010

Odgaard EC, Arieh Y, Marks LE (2003) Cross-modal enhancement of perceived brightness: sensory interaction versus response bias. Percept Psychophys 65:123–132

Pérez-Bellido A, Soto-Faraco S, López-Moliner J (2013) Sound-driven enhancement of vision: disentangling detection-level from decision-level contributions. J Neurophysiol 109:1065–1077. doi:10.1152/jn.00226.2012

Perrott DR, Saberi K, Brown K, Strybel TZ (1990) Auditory psychomotor coordination and visual search performance. Percept Psychophys 48:214–226

Plainis S, Murray IJ (2005) Magnocellular channel subserves the human contrast-sensitivity function. Perception 34:933–940. doi:10.1068/p5451

Pokorny J (2011) Review: Steady and pulsed pedestals, the how and why of post-receptoral pathway separation. J Vis 11:1–23. doi:10.1167/11.5.7.Outline

Rockland KS, Ojima H (2003) Multisensory convergence in calcarine visual areas in macaque monkey. Int J Psychophysiol 50:19–26

Roe AW, Chelazzi L, Connor CE et al (2012) Toward a unified theory of visual area V4. Neuron 74:12–29. doi:10.1016/j.neuron.2012.03.011

Romei V, Gross J, Thut G (2012) Sounds reset rhythms of visual cortex and corresponding human visual perception. Curr Biol 22:807–813. doi:10.1016/j.cub.2012.03.025

Roucoux A, Guitton D, Crommelinck M (1980) Stimulation of the superior colliculus in the alert cat. Exp Brain Res. doi:10.1007/BF00237071

Saalmann YB, Pigarev IN, Vidyasagar TR (2007) Neural mechanisms of visual attention: how top-down feedback highlights relevant locations. Science 316:1612–1615. doi:10.1126/science.1139140

Schiller PH, Malpeli JG, Schein SJ (1979) Composition of geniculostriate input to superior colliculus of the rhesus monkey. J Neurophysiol 42:1124–1133

Schneider KA, Kastner S (2005) Visual responses of the human superior colliculus: a high-resolution functional magnetic resonance imaging study. J Neurophysiol 94:2491–2503. doi:10.1152/jn.00288.2005

Schnupp JWH, Dawe KL, Pollack GL (2005) The detection of multisensory stimuli in an orthogonal sensory space. Exp Brain Res 162:181–190. doi:10.1007/s00221-004-2136-2

Schroeder CE, Lakatos P (2009) Low-frequency neuronal oscillations as instruments of sensory selection. Trends Neurosci 32:9–18. doi:10.1016/j.tins.2008.09.012

Schroeder CE, Mehta AD, Givre SJ (1998) A spatiotemporal profile of visual system activation revealed by current source density analysis in the awake macaque. Cereb Cortex 8:575–592

Shams L, Kim R (2010) Crossmodal influences on visual perception. Phys Life Rev 7:269–284. doi:10.1016/j.plrev.2010.04.006

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

Shapley R (1990) Visual sensitivity and parallel retinocortical channels. Annu Rev Psychol 41:635–658. doi:10.1146/annurev.ps.41.020190.003223

Shen K, Valero J, Day GS, Paré M (2011) Investigating the role of the superior colliculus in active vision with the visual search paradigm. Eur J Neurosci 33:2003–2016. doi:10.1111/j.1460-9568.2011.07722.x

Sinnett S, Soto-Faraco S, Spence C (2008) The co-occurrence of multisensory competition and facilitation. Acta Psychol (Amst) 128:153–161. doi:10.1016/j.actpsy.2007.12.002

Smiley JF, Falchier A (2009) Multisensory connections of monkey auditory cerebral cortex. Hear Res 258:37–46

Spence C (2013) Just how important is spatial coincidence to multisensory integration? Evaluating the spatial rule. Ann NY Acad Sci 1296:31–49. doi:10.1111/nyas.12121

Stein BE, Stanford TR (2008) Multisensory integration: current issues from the perspective of the single neuron. Nat Rev Neurosci 9:255–266. doi:10.1038/nrn2331

Stein BE, London N, Wilkinson LK, Price DD (1996) Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis. J Cogn Neurosci 8:497–506. doi:10.1162/jocn.1996.8.6.497

Sturm W, Willmes K (2001) On the functional neuroanatomy of intrinsic and phasic alertness. Neuroimage 14:S76–S84. doi:10.1006/nimg.2001.0839

Thurlow WR, Jack CE (1973) Some determinants of localization–adaptation effects for successive auditory stimuli. J Acoust Soc Am 53:1573–1577

Tootell RB, Silverman MS, Hamilton SL et al (1988) Functional anatomy of macaque striate cortex. V. Spatial frequency. J Neurosci 8:1610–1624

Tranel D, Damasio H, Eichhorn GR et al (2003) Neural correlates of naming animals from their characteristic sounds. Neuropsychologia 41:847–854. doi:10.1016/S0028-3932(02)00223-3

Van der Burg E, Olivers CNL, Bronkhorst AW, Theeuwes J (2008) Pip and pop: nonspatial auditory signals improve spatial visual search. J Exp Psychol Hum Percept Perform 34:1053–1065. doi:10.1037/0096-1523.34.5.1053

Van der Burg E, Cass J, Olivers CNL et al (2010) Efficient visual search from synchronized auditory signals requires transient audiovisual events. PLoS ONE 5:e10664. doi:10.1371/journal.pone.0010664

Van Wassenhove V, Buonomano DV, Shimojo S, Shams L (2008) Distortions of subjective time perception within and across senses. PLoS ONE. doi:10.1371/journal.pone.0001437

Vroomen J, de Gelder B (2000) Sound enhances visual perception: cross-modal effects of auditory organization on vision. J Exp Psychol Hum Percept Perform 26:1583–1590

Vroomen J, Keetels M (2010) Perception of intersensory synchrony: a tutorial review. Atten Percept Psychophys 72:871–884. doi:10.3758/APP.72.4.871

Walker JT, Scott KJ (1981) Auditory–visual conflicts in the perceived duration of lights, tones and gaps. J Exp Psychol Hum Percept Perform 7:1327–1339

Wallace MT, Wilkinson LK, Stein BE (1996) Representation and integration of multiple sensory inputs inprimate superior colliculus. J Neurophysiol 76:1246–1266

Wallace MT, Meredith MA, Stein BE (1998) Multisensory integration in the superior colliculus of the alert cat. J Neurophysiol 80:1006–1010

Werner S, Noppeney U (2010) Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. J Neurosci 30:2662–2675. doi:10.1523/JNEUROSCI.5091-09.2010

Werner S, Noppeney U (2011) The contributions of transient and sustained response codes to audiovisual integration. Cereb Cortex 21:920–931. doi:10.1093/cercor/bhq161

Wilkinson LK, Meredith MA, Stein BE (1996) The role of anterior ectosylvian cortex in cross-modality orientation and approach behavior. Exp Brain Res. doi:10.1007/BF00227172

Wilson HR (1980) Spatiotemporal characterization of a transient mechanism in the human visual system. Visi Res 20:443–452. doi:10.1016/0042-6989(80)90035-8

Wise L, Irvine D (1983) Auditory response properties of neurons in deep layers of cat superior colliculus. J Neurophysiol 49:674–685

Zannoli M, Cass J, Mamassian P, Alais D (2012) Synchronized audiovisual transients drive efficient visual search for motion-indepth. PLoS ONE 7:e37190. doi:10.1371/journal.pone.0037190