Available online at www.sciencedirect.com
ScienceDirect
Physics of Life Reviews ••• (••••) •••–•••
PHYSICS of LIFE reviews
www.elsevier.com/locate/plrev

Review

Crossmodal influences on visual perception

Ladan Shams [a,b,c,*], Robyn Kim [a]

[a] *Department of Psychology, University of California, Los Angeles, CA 90095, United States*
[b] *Biomedical Engineering, University of California, Los Angeles, CA 90095, United States*
[c] *Interdepartmental Neuroscience Program, University of California, Los Angeles, CA 90095, United States*

Received 19 March 2010; received in revised form 25 March 2010; accepted 25 March 2010

Communicated by L. Perlovsky

Abstract

Vision is generally considered the dominant sensory modality; self-contained and independent of other senses. In this article, we will present recent results that contradict this view, and show that visual perception can be strongly altered by sound and touch, and such alterations can occur even at early stages of processing, as early as primary visual cortex. We will first review the behavioral evidence demonstrating modulation of visual perception by other modalities. As extreme examples of such modulations, we will describe two visual illusions induced by sound, and a visual illusion induced by touch. Next, we will discuss studies demonstrating modulation of activity in visual areas by stimulation of other modalities, and discuss possible pathways that could underpin such interactions. This will be followed by a discussion of how crossmodal interactions can affect visual learning and adaptation. We will review several studies showing crossmodal effects on visual learning. We will conclude with a discussion of computational principles governing these crossmodal interactions, and review several recent studies that demonstrate that these interactions are statistically optimal.

Keywords: Visual perception; Multisensory perception; Visual learning; Multisensory integration

Contents

* Corresponding author at: Department of Psychology, University of California, 7445 Franz Hall, Los Angeles, CA 90095-1563, United States. Tel.: +1 310 206 3630; fax: +1 310 267 2141.
E-mail address: ladan@psych.ucla.edu (L. Shams).

1571-0645/$ – see front matter
doi:10.1016/j.plrev.2010.04.006

Please cite this article in press as: Shams L, Kim R. Crossmodal influences on visual perception. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.04.006

1. Introduction

Modern humans consider themselves visual animals. This may be due to the strong reliance on visual information in our daily lives. The advent of electricity has allowed us to see our surroundings even after dark. New technologies have caused us to rely heavily on images and text for communication and entertainment. This trend towards concentrated use of visual channels has become even more pronounced recently. Radio and telephones, once primary means of communication and entertainment, have given way to text- and image-dominated television, email, text-messaging, social networking sites, web news, and internet blogs.

In the scientific community, too, visual perception has been viewed as the dominant modality, self-contained and unaffected by non-visual information. This view has been consistent with the modular paradigm of brain function in general and perceptual processing in particular, that has dominated for many decades. Even within the visual modality, processing has been considered highly modular, with separate brain areas and mechanisms involved in processing motion, color, stereo, form, and location, etc.

The view of vision as the dominant sensory modality has been reinforced by classic studies of crossmodal interactions, in which experimenters artificially imposed a conflict between visual information and information conveyed through another modality, and reported that the overall percept is strongly dominated by vision. For example, in the ventriloquism effect, the perceived location of sound is captured by the location of the visual stimulus [40,96,107]. Visual capture of location also occurs in relation to proprioceptive and tactile modalities [72]. These effects are quite strong and have been taken as evidence of visual dominance in perception. Even for a function that is generally considered to be an auditory function, namely speech perception, vision has been shown to strongly alter the quality of the auditory percept. For example, pairing the sound of syllable /ba with the video of lips articulating syllable /ga, will induce the percept of syllable /da. This effect is known as the McGurk effect [53].

While the effects of visual signals on other modalities have been appreciated, the influences of other sensory modalities on visual perception have not been acknowledged until recently. The last decade has witnessed a surge of interest in crossmodal interactions, and this, in turn, has resulted in a number of studies that have revealed robust and vigorous influences of non-visual sensory input on both visual perception, and learning. Here, we will review the evidence demonstrating that visual processing is not self-contained and independent of other modalities, with an emphasis on the more recent findings. We will conclude by a discussion of what the advantages of these crossmodal interactions are by reviewing computational models of multisensory perception.

2. Crossmodal modulation of visual perception

Studies have increasingly demonstrated that multisensory stimulation can have a substantial impact not only on cognitive processing, but also on basic visual perception. Non-visual input such as auditory and tactile stimuli can improve visual functioning in a myriad of ways. Here, we discuss how sound and touch can increase perceived brightness, aid detection, improve temporal resolution, guide attention, and affect motion perception in visual processing.

2.1. Perceived brightness and visual detection

Surprisingly, multisensory enhancements can occur even when the extra-modal input does not provide information directly meaningful for the task. A primary example was reported by Stein et al. [93]. Subjects rated the intensity of a visual light higher when it was accompanied by a brief, broad-band auditory stimulus than when it was presented alone. The auditory stimulus produced more enhancement for lower visual intensities, and regardless of the relative location of the auditory cue source. Odgaard et al. [65] later examined whether this enhancement reflected an early-

Please cite this article in press as: Shams L, Kim R. Crossmodal influences on visual perception. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.04.006

stage sensory interaction or a later-stage response bias effect. Response bias refers to the criterion used for decision making [32]. For example, if the criterion for making a "yes" response is made more liberal in presence of sound, that could manifest itself as a higher detection rate or perceived brightness, even when there is no change in the perceptual effect. Indeed, it was found that the effect disappeared when the proportion of trials with sound was reduced, consistent with the response bias explanation [65]. Similarly, Lippert et al. [49] found that sound only aided contrast detection when the sound was informative (though redundant sound sped up reaction times), and only when there was a consistent timing relation between sound and target, of which the subjects were aware. These results support the idea that the crossmodal enhancement of contrast detection largely results from cognitive rather than sensory integration effects.

Evidence for both a decisional and a sensory effect of sound on visual perception was reported by Frassinetti et al. [26] and Bolognini et al. [12] with a visual detection task. Spatially and temporally coincident sound improved visual detection of degraded stimuli; signal detection analysis revealed both a change in decision-making criterion and in perceptual sensitivity (d') [32] caused by sound [26]. Whereas in these studies the auditory cue was spatially and temporally informative, Doyle and Snowden [18] found that even an uninformative auditory signal aided visual processing, in the form of reduced reaction time in a visual identification task. In this case, the effect occurred regardless of the location or spatial validity of the auditory signal. Furthermore, a comparison to the effect of redundant visual signals suggested that the facilitation effect was specific to crossmodal stimuli.

One explanation for the effect of sound in the examples discussed so far is that sound provides a general cueing/priming/alerting function that causes more efficient processing of concurrent stimuli in general. Thus, even though the sound is not specifically relevant for the central task, it nevertheless provides information that can indirectly enhance task performance. But sound can also aid visual perception independent of such temporal cueing effects. For example, Vroomen and de Gelder [101] examined subjects' detection of a masked visual target among a stream of distractor stimuli, when each stimulus (targets and distractors) was accompanied by an auditory tone. When the target was shown with a simultaneous tone that was of higher pitch than the other tones, detection was increased. If this facilitation was a result of a cueing effect, it should also hold when the high tone is presented just before the visual target, but in fact, detection worsened in this case. Furthermore, jittering the interstimulus interval did not reduce the effect of sound, arguing against the idea that rhythmic-based anticipation underlies the facilitation effect. This study suggests that grouping or saliency in the auditory modality can affect saliency or grouping in the visual modality. The tone that was different from the rest of the tones in the sequence of sounds, i.e., the oddball, or the sound that did not get grouped with other sounds, becomes more salient, and renders the visual stimulus accompanying it more salient as well.

A brief sound presented simultaneously with a color change of a visual target can also decrease detection time when searching for a visual target (e.g., a vertical or horizontal line, changing colors at random times) in a complex, dynamic scene consisting of an array of visual distracters (e.g., oblique lines at various orientations, changing colors randomly and at random times) [99]. This is a surprising effect, given that the sound contains no information about the location or identity of the visual object. One may suspect that observers perhaps pay more attention to the visual stimuli when they are accompanied by sound. This top–down control of attention is called endogenous attention. To test this, Van der Burg et al. ran another condition in which the sound was *not* synchronized with the visual target on the majority (80%) of trials. Although in this case, top–down attentional strategy would discourage the use of sound, the effect persisted for the minority of trials in which sound was synchronized with visual target. An endogenous attentional explanation could not explain these findings. Furthermore, a visual alerting cue did not provide the same benefit in search time. Thus, although the facilitation might result from an automatic exogenous (i.e., bottom–up stimulus driven) attentional cueing effect, it must be one that is specific to sound. The authors propose that the visual signal (i.e., the target object) becomes more salient when integrated with the temporal information from the auditory signal, resulting in a pop out effect. Similar effects have been reported in attentional blink and repetition blindness paradigms, wherein a visual stimulus that is usually missed when preceded by certain other visual stimuli becomes more detectable when accompanied by a sound [16,66].

2.2. Temporal processing

Sound can especially affect vision in the temporal domain. This makes adaptive sense, since the auditory modality has much better temporal resolution than the visual modality. The perceived duration of a visual stimulus and interval

Please cite this article in press as: Shams L, Kim R. Crossmodal influences on visual perception. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.04.006

between two visual events can be influenced by sound [102]. In cases of ambiguity or conflict about rate, auditory input strongly affects the perceived rate of a visual stimulus, regardless of spatial congruence or intensity of the auditory stimulus [28,69,87,113]. Furthermore, after a prolonged exposure to auditory and visual stimuli presented at slightly different temporal rates, the influence of the auditory rate produces a long-lasting shift in the perception of visual temporal rate [69]. Sound can also enhance the ability to discriminate the temporal order of two visual stimuli [34,62]. When judging which of two lights appeared first, an irrelevant sound presented slightly before the first light, and another after the second light, improved accuracy. Conversely, two sounds presented between the two visual stimuli worsened performance. The experimenters interpreted this effect as a result of a temporal ventriloquism effect from the temporally disparate auditory stimuli, as if the sounds pulled the perception of the lights earlier or later in time. More recently, however, Hairston et al. [34] have argued against this explanation. They found that rather than slowing down reaction times, as would occur if the visual stimulus were perceived as occurring later in time, a delayed auditory cue sped up reaction times during a visual temporal order judgment task. The authors suggest that in this case, sound is actually enhancing visual temporal acuity, rather than causing a shift in perceived time.

Tactile stimuli have also been demonstrated to affect visual temporal perception. When a static line is visually displayed on a screen, a tactile stimulation of a finger which is placed at a location near one end of the line induces a percept of the line unfolding from the location of finger to the other end [86]. The same effect can also be induced by sounds.

2.3. Attention

Crossmodal signals can also affect sensory processing by directing attention. When you hear a sudden sound, for instance, you tend to visually orient to the location of the sound. That is an example of overt orienting of attention; covert attention (i.e., an internal shift of attention without physical orienting) can also be affected by crossmodal stimuli [20]. Driver and Spence studied the effects of crossmodal signals on covert attention using an orthogonal cueing paradigm in which lights and/or sounds could be presented to the upper left, upper right, lower left, or lower right of central fixation. Subjects judged whether the auditory or visual stimulus was presented above or below central fixation, regardless of which side they were presented. Covert attention was endogenously manipulated by informing the subjects which side was more likely to display the target (e.g., which could be either visual or auditory). Thus, the attentional manipulation (left vs. right) was orthogonal to the discrimination response (up/down), avoiding response bias effects. When an auditory target is expected on a particular side, performance (both speed and accuracy) improves on the expected auditory (attended) side for the visual modality as well, even though the visual target is equally likely to appear on either side [90]. A similar effect of endogenous spatial attention to tactile stimuli was also found on visual performance [92]. Therefore, it seems spatial attention in one modality spreads to other modalities, though the effect is attenuated [90,92]. Additionally, dividing attention across different modalities in different spatial locations reduces the facilitatory effects of attention, providing further evidence that attention in different modalities is not independent [92].

Crossmodal stimuli can also act as exogenous attentional cues; i.e., salient crossmodal stimuli such as sounds can increase visual attention in a certain location, even when they have no predictive value for the visual stimuli [91]. A sudden and unpredictive sound can improve visual signal detection even when the spatial location of the visual stimulus is visually cued, and thus known for certain [52]. This suggests that such involuntary orientation of attention to sound can influence perceptual processing of spatially congruent visual stimuli, even when there is no spatial uncertainty about the location of the visual stimulus.

2.4. Motion perception

In addition to the previous examples of crossmodal influences on static visual perception, many studies have also documented crossmodal influences on visual motion perception. When motion is difficult to perceive visually, for example, in dark or occluded environments, sound and touch can convey information that can contribute to a more accurate perception of motion. Thus, although typically vision dominates over auditory or tactile motion perception [51,88], several studies have reported that visual motion perception itself can also be influenced by other modalities [38,59]. However, these results could be explained by a response bias effect rather than a sensory integration effect. Static auditory stimuli can also influence the perceived direction of visual apparent motion [27], through a temporal

Please cite this article in press as: Shams L, Kim R. Crossmodal influences on visual perception. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.04.006

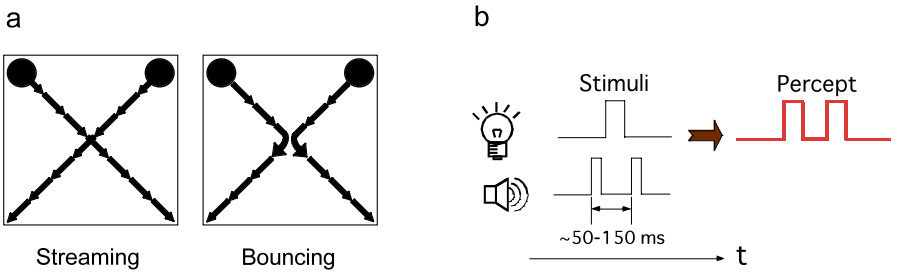

Fig. 1. Visual illusions caused by sound. (a) The stream-bounce illusion reported by Sekuler et al. [79]. Two identical visual objects approach and move away from each other on a screen. In the absence of sound the two objects are often perceived as streaming through each other. However, when a brief sound is presented around the time of visual coincidence of the two objects, the probability of perceiving a bouncing motion is increased. (b) The sound-induced flash illusion reported by Shams et al. [81]. When a brief visual stimulus is accompanied by two brief sounds it is often perceived as two flashes. The same kind of visual illusion is also induced by taps on the finger accompanying flashes.

ventriloquism effect. Sounds can also affect visual motion-in-depth perception (i.e., looming) [98]. More specifically, high-rate acoustic flutter stimuli can reinstate a strong motion after-effect from an otherwise ineffective visual adaptor (e.g, low-rate flicker). The authors propose that sound may fill in sparsely time-sampled visual motion through sound-induced illusory flashes. Evidence for the effect of tactile motion on visual motion perception was recently demonstrated by Konkle et al. [44]. They tested the transfer of motion aftereffects between vision and somatosensation, and found that not only did visual motion cause a tactile motion aftereffect, but tactile motion also induced visual motion after effects. This aftereffect is notable because since the crossmodal stimuli are presented in succession (not concurrently), the crossmodal effects are observed in unisensory contexts, without artifacts of divided attention or competing sensory information.

3. Visual illusions induced by non-visual stimuli

In the previous section, we reviewed how non-visual sensory information can quantitatively modulate visual perception in a variety of domains. Here we will review some findings demonstrating that visual perception can also be qualitatively altered by crossmodal signals.

Sekuler and colleagues [79] showed that the perceived trajectory of visual motion can be altered by sound. When two identical visual objects continuously move towards each other, coincide and move apart from each other in a 2-dimensional display, the objects can be seen either as streaming through each other or bouncing against each other (see Fig. 1a). Nonetheless, the vast majority of observers perceive the two objects as streaming through each other. However, if the visual coincidence of the two objects is accompanied by a brief sound, the visual perception of motion is biased towards the bouncing motion [79]. This is a qualitative change in the visual perception. The underlying mechanism for this change in percept is not clear, and it is possible that it is mediated by cognitive processes (i.e., by higher level knowledge that when objects collide they make a sound), rather than interactions at a perceptual level. The findings of a recent study showing that even a subliminal sound can induce the bias, however, suggest that the illusion does reflect interactions at a perceptual level of processing [21]. Other studies have shown that similar change from streaming to bouncing motion can be induced by other types of transient stimuli, including brief visual events at the time of the coincidence of the two moving objects [108,109]. As a result, it has been suggested that this illusion may be due to a general attentional modulation rather than multisensory integration *per se* [108,109].

Shams and colleagues [81] showed that the perception of brief visual stimuli can be qualitatively altered by concurrent brief sounds. When a single flash of light is accompanied by two or more beeps, its percept often changes from a single flash to two or more flashes [81] (see Fig. 1b). This effect is known as sound-induced flash illusion [81,82]. The reverse illusion can also occur, in which two flashes that are accompanied by a single beep are perceived as a single flash [84,110,115], however this illusion is not always as strong [81,115].

The sound-induced flash illusion has been shown to be associated with a change in perceptual sensitivity (d'), and therefore, it appears to reflect crossmodal interactions at a perceptual level [74,110,111,115]. The illusion is also very robust to changes in many stimulus parameters such as shape, contrast, size, texture, duration of the visual stimulus, frequency, intensity, and duration of sound, and exact relative timing and location of the sounds and flashes ([81,84, 111] and unpublished data). The illusion is also resistant to feedback training. Even providing feedback on each trial

Please cite this article in press as: Shams L, Kim R. Crossmodal influences on visual perception. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.04.006

about the correctness of response does not weaken the illusion [74]. These findings together with the EEG, MEG, and fMRI findings discussed in the next section, indicate that sound-induced flash illusion represents modulation of visual perception by auditory signals at an early perceptual level of processing.

The crossmodal alteration of visual percept is not limited to auditory signals. Tactile stimulation was also found to alter the perceived number of flashes [100,115]. When a single flash is accompanied by two taps, it is often perceived as two flashes. The touch-induced flash illusion is also associated with a change in perceptual sensitivity as measured by d' [100,115].

4. Neural correlates of crossmodal modulation of vision

The pioneering neurophysiological and anatomical work in the field of multisensory integration have revealed that a vast number of neurons in cat superior colliculus integrate information from two or three modalities [56,58,94,95,106]. The response properties, development, the neural circuitry of these neurons and their relationship with orienting behavior have been the topic of extensive research (e.g., [55–57,94,103–106]). In the cortex, however, visual processing areas had been believed to be highly unisensory and unaffected by non-visual input. In the last decade, however, in addition to the abounding behavioral effects discussed above, there has been an accumulating body of electrophysiological and neuroimaging literature confirming the proposition that crossmodal stimulation can affect activity in areas that were previously considered specifically visual. Not only is perception largely multisensory, but even brain regions that were previously considered specifically visual demonstrate multisensory modulation [19,30,41]. In many cases, visual areas functionally specialized for processing certain features can be modulated by crossmodal stimuli conveying analogous features. For example, an fMRI study of haptic object identification (vs. haptic texture identification) found consistent activation of the lateral occipital complex (LOC), a visual object-related region [3]. The LOC area has also been demonstrated to be modulated by auditory experience. An electrical neuroimaging study of visual episodic memory discussed below [63] revealed that differences between audiovisually encoded stimuli and visually encoded stimuli were apparent as early as 60 ms post-stimulus, with changes occurring through generators in the right lateral occipital complex areas, suggesting that multisensory experiences affect unisensory processing at early stages and within "visual" object recognition areas.

Other areas of extrastriate visual cortex have been implicated in crossmodal interactions as well. In a PET study, a tactile grating orientation task (compared to a tactile control task) was associated with greater activation in extrastriate visual cortex [75]. An fMRI study also found that tactile stimulation of a hand spatially congruent with a visual stimulus increased brain response in a visual area (lingual gyrus) compared to visual stimulation alone [50]. Effective connectivity analysis suggested that back-projections from multisensory parietal regions mediated this effect. In a speeded audiovisual reaction time task, Molholm et al. [61] found superadditive ERPs in response to audiovisual stimuli at 45–80 ms post-stimulus at parieto-occipital/occipital locations. The timing and topography of the responses suggest modulation of early visual sensory processing by auditory inputs. Giard and Peronnet [31] found an interaction even earlier, 40 ms post-stimulus, at an occipito-parietal location, and at 90–145 ms in extrastriate cortex.

Area V5/MT+, believed to specialize in processing visual motion [8,14,97], can also be modulated by motion in other modalities. Using fMRI, Lewis et al. [48] examined the effect of auditory motion on visual area MT+, and found that the auditory motion task was associated with a suppression of activity in MT+. More recently, Scheef et al. [77] reported superadditivity (i.e., greater activity for concordant audiovisual stimuli compared to both unimodal conditions) and congruency (i.e., greater effect for concordant audiovisual stimuli than discordant audiovisual stimuli) effects in V5/MT+ using fMRI. Furthermore, Poirier et al. [68] reported that auditory motion activated visual area MT+ in blindfolded subjects performing direction identification. Similarly, Alink et al. [2] studied crossmodal dynamic capture [89] (i.e., auditory motion captured by moving visual stimuli), and found both modulation and activation of MT+ by moving sound. Other fMRI studies have also demonstrated modulation of MT+ by tactile stimuli [11,33,70].

In addition to these effects in extrastriate visual cortex, other studies have demonstrated that crossmodal stimulation can even affect primary visual cortex, previously considered strictly "sensory-specific" [30]. An ERP study of the sound-induced flash illusion [81] discussed above showed a very similar pattern of activity associated with the sound-induced illusory second flash and a physical second flash, suggesting that the modulation of visual activity by sound occurs at early visual cortical areas that are involved in representing a real flash [83]. Notably, the supra-additive auditory–visual interactions were not present for non-illusion audiovisual trials, even though the physical stimulus

Please cite this article in press as: Shams L, Kim R. Crossmodal influences on visual perception. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.04.006

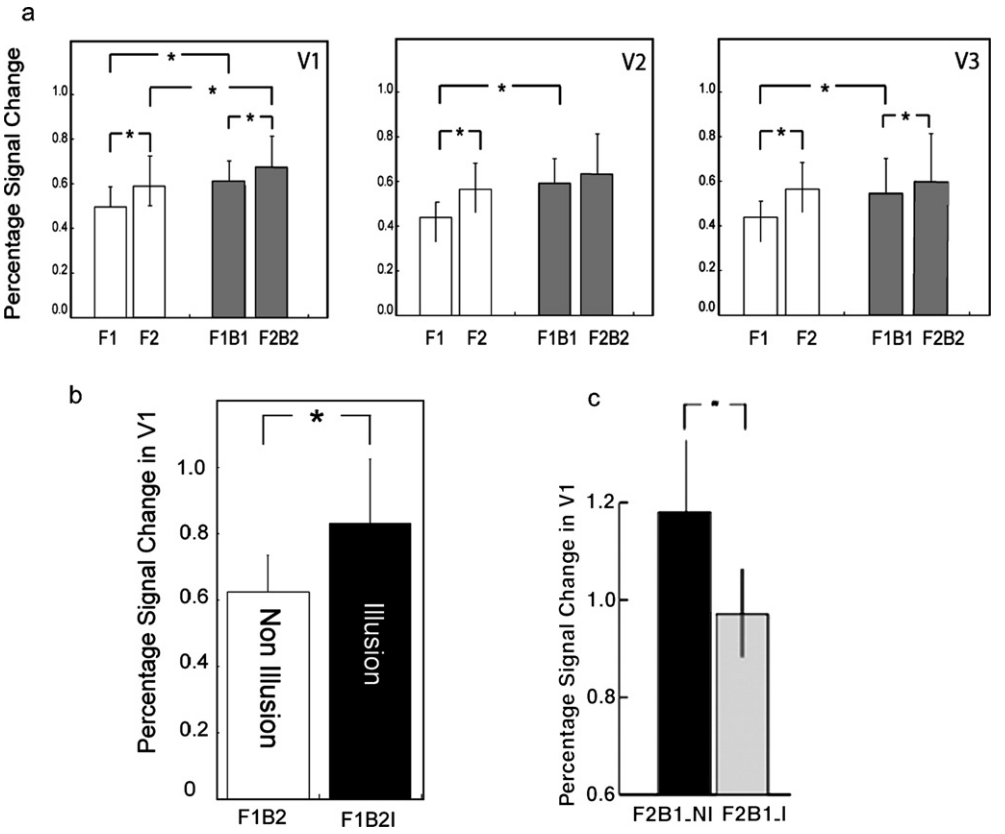

Fig. 2. Modulation of activity in early visual cortical areas in human brain by sound. (a) Data from the fMRI study by Watkins et al. [111] showing activity in retinotopically defined cortical areas V1, V2, and V3 in a variety of stimulus conditions. As can be seen the activity is typically higher in auditory–visual conditions compared to conditions with the exact same visual stimulus but no sound. The number of flashes and beeps presented in each condition are denoted by F*n*B*m* where *n* is the number of flashes and *m* the number of beeps. These results indicate that sound can change the activity in areas V1, V2, V3. (b) Data from Watkins et al. [111] study showing activity in retinotopically defined area V1 in two conditions that are identical in stimuli but differ in the visual perception. In both conditions, one flash was presented with two beeps, however on the illusion trials (F1B2I), two flashes were perceived whereas on non-illusion trials, one flash was perceived. Despite the fact that the stimuli are identical (and hence the attentional effects should be equal), the activity in V1 is higher when the illusion occurs. (c) Data from Watkins et al. [110] showing activity in retinotopically defined V1 in two conditions that are identical in terms of stimuli, and only differ in the reported visual percept. In both conditions, two flashes accompanied with one beep were presented. On illusion trials (F2B1_I), the subjects reported perceiving one flash, whereas on non-illusion trials (F2B1_NI), subjects reported seeing two flashes. The activity in V1 is lower when the illusory single flash is perceived (light bar). The results from (b) and (c) together rule out the role of attention as the underlying factor in observed modulations, and indicate that the modulations of V1 reflect auditory–visual integration.

was the same [10], and the perception of the illusion was associated with an increase in oscillatory and induced gamma band activity [10]. An MEG study of the sound-induced flash illusion reported a superadditive interaction effect in right occipital cortex at 35–65 ms post-stimulus, the short latency suggesting a direct feedforward effect of the auditory input rather than feedback [80]. Consistent with these findings, a more recent ERP study isolated neural activity associated with the illusory second flash and found an early modulation of visual cortex activity at 30–60 ms after the second sound [60]. Also, using the sound-induced flash illusion paradigm, Arden et al. [4] similarly found early modification of visual evoked potentials in occipital channels induced by the beeps [4]. Using functional MRI of the sound-induced flash illusion task (i.e., reporting the number of flashes under a variety of sound conditions), Watkins et al. [110,111] demonstrated that V1 shows significantly greater activity during fission illusion trials (1 flash perceived as two) than physically identical non-illusion trials (see Fig. 2b); during fusion illusion trials (2 flashes perceived as one), activity in V1 decreased [110] (see Fig. 2c). Additionally, irrespective of the illusion, concordant auditory input enhanced activity in V1, V2, and V3 (see Fig. 2a). These findings strongly indicate that human primary visual cortex activity is modulated by sound as a result of auditory–visual integration (as opposed to a general attentional effect).

Please cite this article in press as: Shams L, Kim R. Crossmodal influences on visual perception. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.04.006

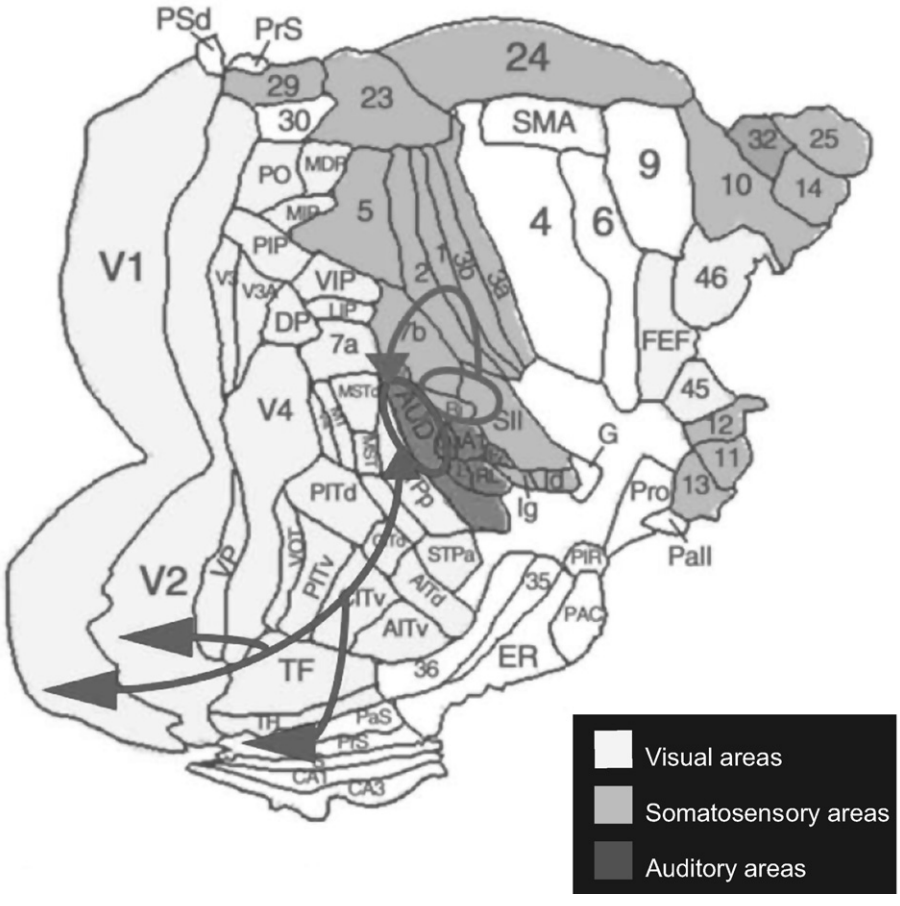

Fig. 3. Anatomical connections among cortical sensory areas in macaque monkey. Flattened representation of macaque monkey cortex, depicting direct connections between different sensory cortical regions (auditory to visual, and somatosensory to auditory) found by Falchier et al. [24]. Reprinted from Cappe C, Rouiller EM, Barone P. Multisensory anatomical pathways. Hearing Research 2009;258:28–36, with permission from Elsevier.

Consistent with these findings, Noesselt et al. [64], using fMRI, recently reported effects of audiovisual temporal correspondence (i.e., for synchronous versus asynchronous audiovisual temporal patterns) in primary visual cortex contralateral to the stimulated visual field [64].

Modulation of V1 activity does not appear to be limited to auditory stimulation. Functional MRI of normally sighted, blindfolded subjects while they were tactually rating raised-dot patterns revealed significant activation in primary visual cortex, simultaneously with deactivation of extrastriate areas V2, V3, V3A, and hV [54].

5. Underlying pathways

These functional effects must be implemented by physical connections between different sensory and associative brain regions. How can the architecture of the brain support such early interactions? Neuroanatomical studies in animals suggest that there are indeed inputs from multisensory areas and other sensory cortices to early sensory areas, including visual cortex. Retrograde tracers injected in peripheral V1 and V2 in monkeys indicated input from both the superior temporal polysensory area and the auditory core and belt and caudal parabelt areas [24] (see Fig. 3). Using anterograde tracers, Rockland and Ojima [73] also found direct connections from auditory cortex as well as parietal association cortex to V1 and V2. Hirokawa et al. [39] recently investigated the functional importance of such lower-order sensory cortices in multisensory integration in rats. Their results suggest area V2L, a secondary visual area in rats (between audio and visual cortices), is responsible for audiovisual reaction time facilitation, since 1) the region was demonstrated to be active for temporally coincident audiovisual stimuli but not temporally inconsistent

Please cite this article in press as: Shams L, Kim R. Crossmodal influences on visual perception. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.04.006

trials, and 2) inactivation of the region with muscimol reduced bimodal reaction time, but not unimodal reaction time [39]. Cappe et al. [15] propose that integration may even occur before primary sensory processing, at the level of the thalamus. Cortico-thalamo-cortical routing could provide a fast feed-forward pathway by which information from remote cortical areas responsive to different sensory modalities could interact. Although they did not test visual areas, they did find evidence for such pathways between auditory, somatosensory, and motor areas in two macaque monkeys, using neuroanatomical tracers.

6. Crossmodal modulation of visual learning and adaptation

Visual perception is highly adaptive even in the mature brain. One type of visual adaptation is in the form of recalibration of visual perception by other modalities. This kind of adaptation has been studied extensively using optical manipulations such as prisms. In these studies, a large-scale radical change in the visual input is caused by wearing prisms that, for example, render the visual input upside down, or laterally invert the image (left-side right). Participants who are initially completely disoriented cannot navigate or carry out simple visual or sensorimotor tasks, quickly adapt and are able to function without help, navigate or even ride a bicycle [35,36,112]. This type of adaptation is based on interaction with the environment [112] and involves using other senses to adapt the visual or visuo-motor system.

Visual recalibration has been reported also using other paradigms. For example, visual stereopsis can be recalibrated by haptic information [6]. The interpretation of depth-from-shading can be modified by touch [1]. And visual temporal processing may be modified by auditory stimuli. After repeated exposure to light leading a sound, the reaction time to lights can be increased [17,37].

For tasks where visual estimation can utilize multiple cues, the weighting of the visual cues can be affected by how consistent each cue is with the non-visual cue. Touch can reweight visual cues for slant [22], and visual cues for shape [5]. In these studies, two visual cues and one haptic cue for object (shape or slant) estimation were available during training. It was shown that whichever visual cue was correlated with the haptic cue during training was given a higher weight later when only visual cues were available. In other words, it seems that the nervous system used the consistency with the tactile cue to measure the reliability of the visual cues, and adjust their weights accordingly.

Crossmodal signals can also enhance visual episodic memory and perceptual learning. The visual recognition of objects can benefit from a multisensory encoding. Murray and colleagues [63] found that when the task is to judge whether an image presented in a sequence is old (presented before) or new (first presentation), the recognition accuracy is superior for images that were initially presented together with their corresponding sounds (e.g., the image of a bell, and the sound “dong”) compared to images that were initially presented without sound, even though all second presentations were in the absence of sound (see Fig. 4a). In other words, the auditory–visual encoding of objects improved the visual retrieval. The facilitation of retrieval only occurred for images that were encoded together with their semantically congruent sounds, and not with any arbitrary sounds [47], suggesting that the facilitation was not due to a general alerting effect of sound, and rather, involved auditory–visual binding.

We recently examined the effects of auditory–visual interactions on visual perceptual learning. We found that multisensory training can enhance visual perceptual learning [78]. A group of participants was trained using a classic perceptual learning paradigm, in which only visual stimuli were used to perform a coherent motion detection and discrimination task. Another group of participants was trained using the exact same visual stimuli, however, with auditory motion accompanying visual coherent motion. A two-interval forced choice paradigm was used, and the task of the observers in both groups was to judge which interval contained coherent motion. Three levels of visual difficulty were combined with three levels of auditory difficulty, and the trials were presented in an interleaved fashion. For the auditory–visual trained group, some trials were intermixed in which there was no auditory motion signal, and therefore the task could only be performed based on visual stimuli. The two groups were compared on trials with no auditory signal. Both groups exhibited improvement in accuracy across the 10 training sessions, however, the group trained with auditory–visual stimuli showed a faster rate of learning and a larger degree of improvement, i.e., their performance asymptoted at a higher level of accuracy [78].

Considering that sound can have an alerting effect, and a higher level of arousal during training may result in better learning, we examined the role of attention in this enhancement effect. If the underlying factor for facilitation of learning is enhanced attention, then a sound that is equally salient, but does not get integrated with the visual stimulus should result in a similar level of facilitation in learning. On the other hand, if the integration between visual and

Please cite this article in press as: Shams L, Kim R. Crossmodal influences on visual perception. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.04.006

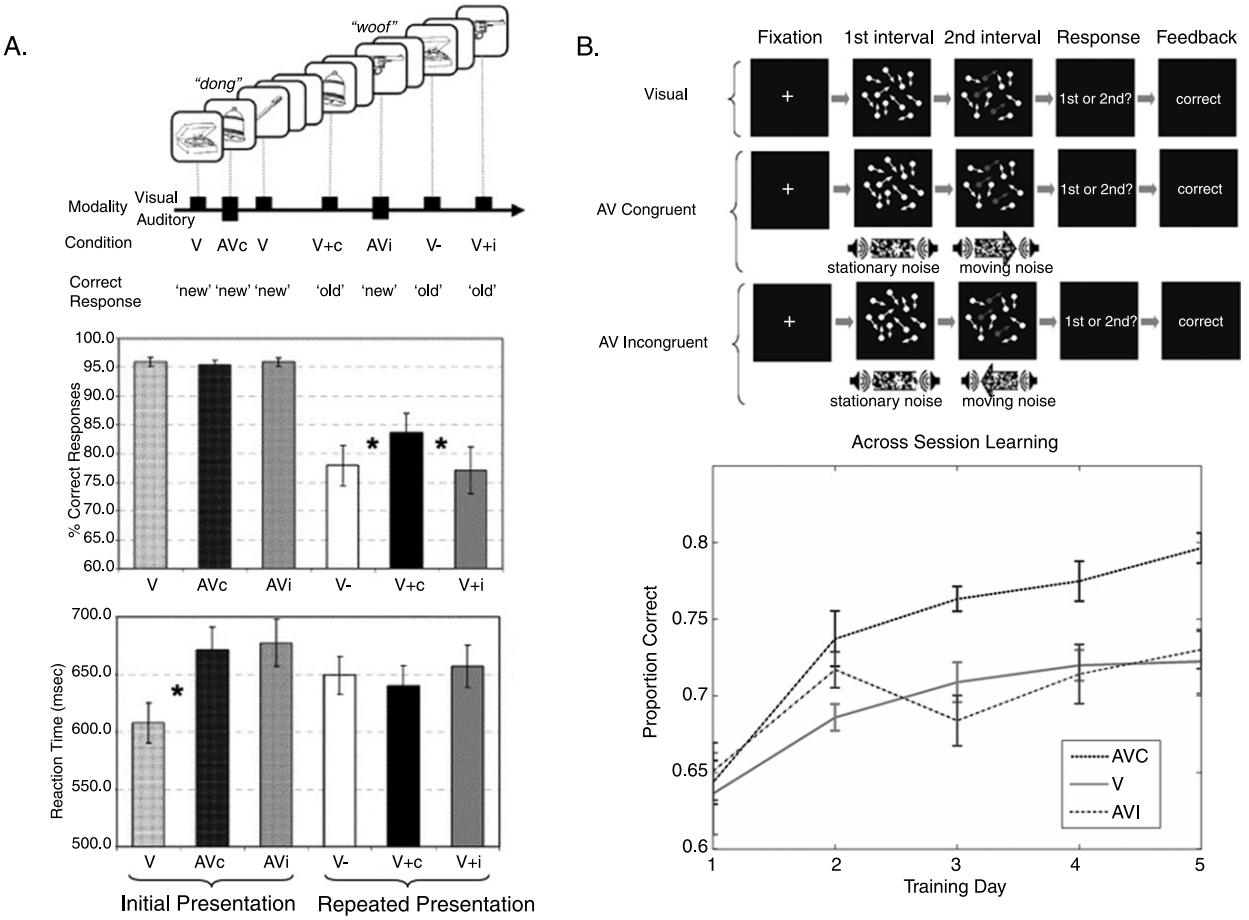

Fig. 4. Auditory effects on visual memory and learning. (a) Experimental design (top) and results (bottom) of a study by Lehmann and Murray [47]. The task was to judge whether each image is "new" or "old" (presented before). The first presentation of some object images was accompanied with the corresponding sound (AVc), some objects with a non-corresponding sound (AVi), and some objects with no sound (V). The second presentation was also in the absence of sound. The condition in which the image is presented for the second time, and the first presentation was without sound is denoted by V−, the condition in which the image is presented for the second time, and the first presentation was with a congruent sound is denoted by V+c, and the condition in which the image is presented for the second time, and the first presentation was with an incongruent sound is denoted by V+i. The data shown below indicates that objects that were presented initially with their congruent sound (V+c) are recognized better than objects that are initially presented without sound or with incongruent sound. (Reprinted from Lehmann S, Murray MM. The role of multisensory memories in unisensory object discrimination. Cognitive Brain Research 2005;24:326–334, with permission from Elsevier.) (b) Experimental design (top) and data (bottom) from Kim et al. [42] study of auditory facilitation of visual perceptual learning. Top panel shows cartoon depiction of one visual (top row) and congruent audiovisual (middle row) trial, and incongruent audiovisual (bottom row) trial. Arrows indicate motion direction of dots, with coherently moving dots represented by darker arrows for illustration purposes (in the second interval). The task was to judge which interval contained coherent motion. Visual-trained group (V) only received visual trials, the congruent auditory–visual trained group (AVC) was trained mostly on congruent AV trials, with some visual trials intermixed. The incongruent AV trained group (AVI) was trained mostly with incongruent AV trials with some visual trials intermixed. Bottom panel shows performance on silent visual trials (no sound) across the training sessions for congruent-audiovisual-trained group (dark dotted line), unisensory-visual-trained group (solid gray line), and incongruent-audiovisual-trained group (light dotted line). Ordinate is proportion correct averaged across three signal levels, abscissa represents training session number. In order to focus on long-term (day-to-day) learning, only the data from the first third of each session is shown, however the results are similar for whole sessions. Error bars reflect within-group standard error.

auditory stimuli is required for the visual learning to be enhanced, then sounds that do not get integrated should not result in facilitation. We compared learning across 5 days among three groups of participants [42] (see Fig. 4b). As in the previous study, one group was trained only with visual stimuli, and one group was trained with congruent auditory and visual stimuli (moving in the same direction). The third group was trained with the same visual stimuli, but paired with incongruent auditory motion, i.e., moving in the opposite direction. In this group, sound was still salient (auditory

Please cite this article in press as: Shams L, Kim R. Crossmodal influences on visual perception. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.04.006

motion had the same coherent levels), and still informative (always in the same interval as the visual coherent motion), but it was always moving in the opposite direction of visual motion. Auditory and visual motion stimuli moving in opposite directions are unlikely to get integrated. All three groups were compared on identical silent visual trials. This study replicated the results of the previous study, showing enhanced learning for the group trained with congruent auditory–visual stimuli compared to the group trained only with visual stimuli. Importantly, the group trained with incongruent auditory–visual stimuli did not show any enhanced learning relative to the unisensory trained group, suggesting that auditory–visual integration is the underlying factor for the observed facilitation of visual learning [42].

These findings altogether suggest that training with multiple correlated sensory inputs is more conducive to learning even for visual tasks. While this may appear counter-intuitive from a certain angle, it is consistent with the sensory experience of humans in nature, which typically involves redundant sensory input across modalities. Therefore, it appears that learning mechanisms are tuned to operate on this type of input and are most effective in a multisensory mode of processing.

7. Computational principles of crossmodal interactions

The findings reviewed so far make it clear that visual processing is not immune to influences from other modalities, and can be affected by non-visual sensory signals in a number of different tasks, in both perception and learning. In some of these interactions, visual performance was improved via crossmodal influences, but in some cases—i.e., illusions—the visual accuracy was found to be reduced as a result of crossmodal influence. Why should the visual system be allowed to be misled by other modalities? This kind of crossmodal interaction appears to be non-adaptive. Do these interactions represent a suboptimality in the human nervous system or are there advantages that justify having such interactions?

Intuitively, it can be seen that if there are two sensory measurements (e.g., auditory and visual) available about an environmental variable (e.g., the timing of an event), then given that sensory measurements are always noisy, it would be beneficial to combine the two measurements to obtain a more informed estimate of the environmental variable. More formally, it can be shown that if there are two noisy observations of the same variable, if both observations are unbiased estimators, then integrating the two measurements can result in a more precise estimate. Therefore, combining a visual observation and an auditory observation can be beneficial for estimating the properties of objects in the environment. On the other hand, if the two sensory signals stem from different objects, for example, the visual signal originates from a cow on one corner of a farm, and the auditory stimulus originates from a rooster in the other corner, then combining the two signals can be misleading, and could result in a large error in estimation of, for example, the location of the object.

Therefore, to make sense of the surrounding environment, the nervous system has to figure out which sensory signals were caused by the same object and should be combined, and which signals were caused by independent objects and should be kept apart. This is quite a non-trivial causal inference problem because the nervous system typically does not have any clue about the causal structure of specific scenes and events at any given time, and thus has to solve this problem purely based on noisy sensory measurements and prior information about the world. In addition to solving this causal inference problem, once multiple signals are inferred to originate from the same object, the nervous system has to figure out *how* to integrate them. This is a problem of multisensory integration. This problem is also non-trivial, because the sensory signals are noisy, and as a result, there is almost always discrepancy between the signals (e.g., the location or time conveyed by visual information and auditory information). The perceptual system has to figure out which signal to trust more, how much to shift which signal towards which, etc. These problems of causal inference and multisensory integration are problems that the perceptual system has to solve at any given moment.

The traditional model of cue combination [46,116] and multisensory integration [23,29,43,114] assumes that the sensory signals are all caused by the same object (see Fig. 5a), and the best estimate of the object is obtained by fusing all the sensory cues. Behavioral studies show, however, that while the sensory signals often get fused when they are largely consistent, the signals that are grossly inconsistent do not interact and are often treated independently of each other by the nervous system [45,84]. Moreover, a moderate degree of conflict between signals sometimes results in a partial integration, i.e., the two percepts get shifted towards each other but do not converge to a single percept

Please cite this article in press as: Shams L, Kim R. Crossmodal influences on visual perception. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.04.006

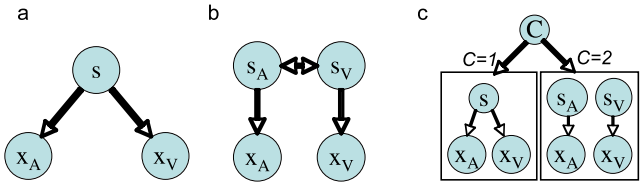

Fig. 5. Different models of cue combination. (a) The traditional model of cue combination. In this model, it is assumed that a single source (s) gives rise to both sensory signals (e.g., auditory signal x_A and visual signal x_V). This model can account for integration, but not segregation of the signals. (b) The model proposed by Shams et al. [80]. This model allows a separate source for each of the sensory signals. The double-arrow between the two sources (s_A and s_V) indicates that the two sources may or may not be independent. This model performs causal inference implicitly, and can account for both integration and segregation as well as partial integration between the signals. (c) The model proposed by Körding et al. [45]. This hierarchical model performs causal inference explicitly, and can make direct predictions about perceived causal structure. It can also account for the entire range of interactions (full integration, partial integration, segregation), as it is a special case of the model shown in (b) (when variable C is integrated out).

[45,84,115]. The traditional model of cue combination does not account for the phenomena of partial integration and segregation.

Shams and colleagues recently introduced a normative model [84] that did not assume a single cause for all sensory signals (see Fig. 5b), and showed that this model can quantitatively account for behavioral data in a wide range of sensory conditions, encompassing the entire spectrum of phenomena ranging from fusion to partial integration to segregation. The model uses Bayesian inference to infer causes from the sensory signals and prior knowledge about the auditory and visual events. Observers were tested in a temporal numerosity judgment task (counting the number of flashes and beeps), wherein the sound-induced flash illusion occurred in some conditions. Importantly, this study showed that the sound-induced flash illusion can be explained by a normative Bayesian causal inference model of multisensory perception.

An extension of this model to three sensory signals was shown to account for interactions (and illusions) between two and three modalities, again accounting for full integration, partial integration, and segregation of the three sensory modalities [115]. Similar models that do not assume forced fusion between modalities have been shown to account for visual–haptic interactions in the numerosity judgment task [13], and auditory–visual rate perception [71]. In all of these models, the interaction between the modalities is captured by the joint probability of the sources/events (s_A and s_V in Fig. 5b), i.e., the (acquired or hard-wired) knowledge about statistical relationship between the events, which affects the perceptual inference in the form of a prior expectation.

The Bayesian model of Shams et al. [80,84] is a non-hierarchical model (see Fig. 5b), and performs causal inference implicitly. The hierarchical Bayesian model shown in Fig. 5c performs causal inference explicitly [45]. In this model, variable C determines the causal structure, and can make predictions about perceived causal inference. This hierarchical model is a special form of the non-hierarchical model of Shams et al. [80,84] (if C is integrated out) [45]. This model was shown to account for auditory–visual interactions in spatial localization, as well as the perceived causal structure of observers [45]. It was also shown that in performing this auditory–visual task, the sensory representations and prior expectations appear to be encoded independently of each other, suggesting that the nervous system indeed follows Bayesian inference in carrying out this perceptual task.

Altogether these findings suggest that in carrying out basic perceptual tasks, the human perceptual system performs causal inference and multisensory integration, and it does so in a fashion highly consistent with a Bayesian observer. This strategy is statistically optimal as it leads to minimizing the average (squared) error of perceptual estimates; however, it results in errors in some conditions, which manifest themselves as illusions.

8. Discussion

For more than a century, brain function in general, and perception in particular, has been viewed to be highly modular [67]. The different sensory modalities have been believed to be organized in separate pathways, without any dialogue and interaction between the pathways, and the unified perception of the world has been believed to be achieved by convergence of the input from these separate pathways at higher levels of processing, after the sensory signals have each been thoroughly processed in their respective unisensory brain areas. This modular view has been

Please cite this article in press as: Shams L, Kim R. Crossmodal influences on visual perception. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.04.006

particularly strong with respect to visual processing, as vision has been considered as the dominant modality, self-contained and independent of input from other modalities.

In contrast with the modular view of perception, and the view of vision as the dominant modality, the accumulating evidence, especially over the last several years has revealed that visual perception can both quantitatively and qualitatively be modified by the input from other modalities. These modulations can take place at a number of different levels of processing, in different perceptual domains, and can be intriguingly strong and robust as evident by some visual illusions. Visual processing can be modulated by non-visual sensory signals even at the earliest stage of cortical processing, primary visual cortex [80,83,110,111], and with a very short latency [25,61,80,83]. The electrophysiological and neuroimaging findings may even underestimate the degree of integration in the brain, given that each method has its own technical limitations. For example, a relatively small proportion of neurons may exhibit super-additivity (which has often been used as a measure of crossmodal interactions in EEG and MEG studies); therefore, physiological recording studies may fail to find such effects due to sampling and signal to noise issues. Additionally, multisensory neurons may be organized in patches amongst unisensory neurons [7], making it difficult to find multisensory effects with the relatively coarse resolution of human brain imaging studies. As research techniques develop, more and more evidence of multisensory integration effects in unexpected regions may become uncovered.

Crossmodal modulations of visual processing are not confined to perception, they seem to play an important role also in visual perceptual learning. Crossmodal sensory signals appear to be used to recalibrate vision [1,6,35,36,112], adjust the relative weight of visual cues [5,22], and to enhance perceptual learning in low-level visual tasks [42,78,85].

Therefore, visual processing does not appear to take place in a module independently of other sensory processes. It appears to interact vigorously with other sensory modalities in a wide variety of domains. These interactions appear to follow a general computational strategy that tries to minimize the error in perceptual estimates on average. In some tasks, observed interactions of visual modality with other modalities have been shown to be consistent with a framework in which the nervous system infers which of the sensory signals are caused by the same objects and integrates those signals [45,76,84,115]. Human multisensory perception appears to perform the tasks of causal inference and sensory integration in a statistically optimal fashion by combining the sensory evidence with prior knowledge [9,13,45,71,76,84,115]. Indeed, visual processing, while an important component of human perception, functions as part of a larger network that takes sensory measurements from a variety of sources and modalities, and tries to come up with an interpretation of the sensory signals that as a whole leads to least amount of error on average.

References

[1] Adams WJ, Graf EW, Ernst MO. Experience can change the 'light-from-above' prior. Nature Neuroscience 2004;7:1057–8.
[2] Alink A, Singer W, Muckli L. Capture of auditory motion by vision is represented by an activation shift from auditory to visual motion cortex. Journal of Neuroscience 2008;28:2690–7.
[3] Amedi A, Malach R, Hendler T, Peled S, Zohary E. Visuo-haptic object-related activation in the ventral visual pathway. Nature Neuroscience 2001;4:324–30.
[4] Arden GB, Wolf JE, Messiter C. Electrical activity in visual cortex associated with combined auditory and visual stimulation in temporal sequences known to be associated with a visual illusion. Vision Research 2003;43:2469–78.
[5] Atkins JE, Fiser J, Jacobs RA. Experience-dependent visual cue integration based on consistencies between visual and haptic percepts. Vision Research 2001;41:449–61.
[6] Atkins JE, Jacobs R, Knill DC. Experience-dependent visual cue recalibration based on discrepancies between visual and haptic percepts. Vision Research 2003;43:2603–13.
[7] Beauchamp MS, Argall BD, Bodurka J, Duyn JH, Martin A. Unraveling multisensory integration: patchy organization within human STS multisensory cortex. Nature Neuroscience 2004;7:1190–2.
[8] Beauchamp MS, Cox RW, DeYoe EA. Graded effects of spatial and featural attention on human area MT and associated motion processing areas. Journal of Neurophysiology 1997;78:516–20.
[9] Beierholm U, Quartz S, Shams L. Bayesian priors are encoded independently of likelihoods in human multisensory perception. Journal of Vision 2009;9:1–9.
[10] Bhattacharya J, Shams L, Shimojo S. Sound-induced illusory flash perception: role of gamma band responses. NeuroReport 2002;13:1727–30.
[11] Blake R, Sobel KV, James TW. Neural synergy between kinetic vision and touch. Psychological Science 2004;15:397–402.
[12] Bolognini N, Frassinetti F, Serino A, Ladavas E. "Acoustical vision" of below threshold stimuli: interaction among spatially converging audiovisual inputs. Experimental Brain Research 2005;160:273–82.
[13] Bresciani JP, Dammeier F, Ernst MO. Vision and touch are automatically integrated for the perception of sequences of events. Journal of Vision 2006;6:554–64.
[14] Britten KH, Shadlen MN, Newsome WT, Movshon JA. The analysis of visual motion: a comparison of neuronal and psychophysical performance. Journal of Neuroscience 1992;12:4745–65.

Please cite this article in press as: Shams L, Kim R. Crossmodal influences on visual perception. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.04.006

[15] Cappe C, Rouiller EM, Barone P. Multisensory anatomical pathways. Hearing Research 2009;258:28–36.
[16] Chen YC, Yeh SL. Catch the moment: multisensory enhancement of rapid visual events by sound. Experimental Brain Research 2009;198:209–19.
[17] Di Luca M, Machulla T-K, Ernst M. Recalibration of multisensory simultaneity: Cross-modal transfer coincides with a change in perceptual latency. Journal of Vision 2009;9:1–16.
[18] Doyle MC, Snowden RJ. Identification of visual stimuli is improved by accompanying auditory stimuli: the role of eye movements and sound location. Perception 2001;30:795–810.
[19] Driver J, Noesselt T. Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. Neuron 2008;57:11–23.
[20] Driver J, Spence C. Crossmodal attention. Current Opinion in Neurobiology 1998;8:245–53.
[21] Dufour A, Touzalin P, Moessinger M, Brochard R, Despres O. Visual motion disambiguation by a subliminal sound. Consciousness and Cognition 2008;17:790–7.
[22] Ernst MO, Banks MS, Bülthoff HH. Touch can change visual slant perception. Nature Neuroscience 2000;3:69–73.
[23] Ernst MO, Bülthoff HH. Merging the senses into a robust percept. Trends in Cognitive Sciences 2004;8:162–9.
[24] Falchier A, Clavagnier S, Barone P, Kennedy H. Anatomical evidence of multimodal integration in primate striate cortex. Journal of Neuroscience 2002;22:5749–59.
[25] Foxe JJ, Schroeder CE. The case for feedforward multisensory convergence during early cortical processing. NeuroReport 2005;16:419–23.
[26] Frassinetti F, Bolognini N, Ladavas E. Enhancement of visual perception by crossmodal visuo–auditory interaction. Experimental Brain Research 2002;147:332–43.
[27] Freeman E, Driver J. Direction of visual apparent motion driven solely by timing of a static sound. Current Biology 2008;18:1262–6.
[28] Gebhard JW, Mowbray GH. On discriminating the rate of visual flicker and auditory flutter. American Journal of Psychology 1959;72:521–8.
[29] Ghahramani Z. Computation and psychophysics of sensorimotor integration. Ph.D. thesis. Cambridge: Massachusetts Institute of Technology; 1995.
[30] Ghazanfar A, Schroeder CE. Is neocortex essentially multisensory? Trends in Cognitive Sciences 2006;10:278–85.
[31] Giard MH, Peronnet F. Auditory–visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. Journal of Cognitive Neuroscience 1999;11:473–90.
[32] Green DM, Swets JA. Signal detection theory and psychophysics. Los Altos, CA: Peninsula Publishing; 1989.
[33] Hagen MC, Franzen O, McGlone F, Essick G, Dancer C, Pardo JV. Tactile motion activates the human middle temporal/V5 (MT/V5) complex. European Journal of Neuroscience 2002;16:957–64.
[34] Hairston WD, Hodges DA, Burdette JH, Wallace MT. Auditory enhancement of visual temporal order judgment. NeuroReport 2006;17:791–5.
[35] Held R, Hein A. Adaptation to disarranged eye-hand coordination contingent upon reafferent stimulation. Perceptual and Motor Skills 1958;8:87–90.
[36] Held R, Hein A. Movement-produced stimulation in the development of visually guided behavior. Journal of Comparative and Physiological Psychology 1963;56:872–6.
[37] Herrar V, Harris L-R. The effect of exposure to asynchronous audio, visual, and tactile stimulus combinations on the perception of simultaneity. Experimental Brain Research 2008;186:517–27.
[38] Hidaka S, Manaka Y, Teramoto W, Sugita Y, Miyauchi R, Gyoba J, et al. Alternation of sound location induces visual motion perception of a static object. PLoS ONE 2009;4:e8188.
[39] Hirokawa J, Bosch M, Sakata S, Sakurai Y, Yamamori T. Functional role of the secondary visual cortex in multisensory facilitation in rats. Neuroscience 2008;153:1402–17.
[40] Howard IPaWBT. Human spatial orientation. London: Wiley; 1966.
[41] Kayser C, Logothetis NK. Do early sensory cortices integrate cross-modal information? Brain Structure & Function 2007;12:121–32.
[42] Kim RS, Seitz AR, Shams L. Benefits of stimulus congruency for multisensory facilitation of visual learning. PLoS ONE 2008;3:e1532.
[43] Knill DC, Pouget A. The Bayesian brain: the role of uncertainty in neural coding and computation. Trends in Neurosciences 2004;27:712–9.
[44] Konkle T, Wang Q, Hayward V, Moore CI. Motion aftereffects transfer between touch and vision. Current Biology 2009;19:745–50.
[45] Körding K, Beierholm U, Ma WJ, Tenenbaum JM, Quartz S, Shams L. Causal inference in multisensory perception. PLoS ONE 2007;2:e943.
[46] Landy MS, Maloney LT, Johnston EB, Young M. Measurement and modeling of depth cue combination: In defense of weak fusion. Vision Research 1995;35:389–412.
[47] Lehmann S, Murray MM. The role of multisensory memories in unisensory object discrimination. Brain Research, Cognitive Brain Research 2005;24:326–34.
[48] Lewis JW, Beauchamp MS, DeYoe EA. A comparison of visual and auditory motion processing in human cerebral cortex. Cerebral Cortex 2000;10:873–88.
[49] Lippert M, Logothetis NK, Kayser C. Improvement of visual contrast detection by a simultaneous sound. Brain Research 2007;1173:102–9.
[50] Macaluso E, Frith CD, Driver J. Modulation of human visual cortex by crossmodal spatial attention. Science 2000;289:1206–8.
[51] Mateeff S, Hohnsbein J, Noack T. Dynamic visual capture: apparent auditory motion induced by a moving visual target. Perception 1985;14:721–7.
[52] McDonald JJ, Teder-Sälejärvi WA, Hillyard SA. Involuntary orienting to sound improves visual perception. Nature 2000;407:906–8.
[53] McGurk H, MacDonald J. Hearing lips and seeing voices. Nature 1976;264:746–8.
[54] Merabet LB, Swisher JD, McMains SA, Halko MA, Amedi A, Pascual-Leone A, et al. Combined activation and deactivation of visual cortex during tactile sensory processing. Journal of Neurophysiology 2007;97:1633–41.
[55] Meredith MA, Nemitz JW, Stein BE. Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. Journal of Neuroscience 1987;10:3215–29.

Please cite this article in press as: Shams L, Kim R. Crossmodal influences on visual perception. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.04.006

[56] Meredith MA, Stein BE. Interactions among converging sensory inputs in the superior colliculus. Science 1983;221:389–91.
[57] Meredith MA, Stein BE. Spatial factors determine the activity of multisensory neurons in cat superior colliculus. Brain Research 1986;365:350–4.
[58] Meredith MA, Stein BE. Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. Journal of Neurophysiology 1986;56:640–62.
[59] Meyer GF, Wuerger SM. Cross-modal integration of auditory and visual motion signals. NeuroReport 2001;12:2557–60.
[60] Mishra J, Martinez A, Sejnowski T, Hillyard SA. Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. Journal of Neuroscience 2007;27:4120–31.
[61] Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ. Multisensory auditory–visual interactions during early sensory processing in humans: a high-density electrical mapping study. Cognitive Brain Research 2002;14:115–28.
[62] Morein-Zamir S, Soto-Faraco S, Kingstone A. Auditory capture of vision: examining temporal ventriloquism. Cognitive Brain Research 2003;17:154–63.
[63] Murray MM, Michel CM, Grave de Peralta R, Ortigue S, Brunet D, Gonzalez Andino S, et al. Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. Neuroimage 2004;21:125–35.
[64] Noesselt T, Rieger JW, Schoenfeld MA, Kanowski M, Hinrichs H, Heinze HJ, et al. Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. Journal of Neuroscience 2007;27:11431–41.
[65] Odgaard EC, Arieh Y, Marks LE. Cross-modal enhancement of perceived brightness: sensory interaction versus response bias. Perception & Psychophysics 2003;65:123–32.
[66] Olivers CN, Van der Burg E. Bleeping you out of the blink: sound saves vision from oblivion. Brain Research 2008;1242:191–9.
[67] Pascual-Leone A, Hamilton R. The metamodal organization of the brain. Progress in Brain Research 2001;134:427–45.
[68] Poirier C, Collignon O, Devolder AG, Renier L, Vanlierde A, Tranduy D, et al. Specific activation of the V5 brain area by auditory motion processing: an fMRI study. Brain Research, Cognitive Brain Research 2005;25:650–8.
[69] Recanzone GH. Auditory influences on visual temporal rate perception. Journal of Neurophysiology 2003;89:1078–93.
[70] Ricciardi E, Vanello N, Sani L, Gentili C, Scilingo EP, Landini L, et al. The effect of visual experience on the development of functional architecture in hMT+. Cerebral Cortex 2007;17:2933–9.
[71] Roach N, Heron J, McGraw P. Resolving multisensory conflict: a strategy for balancing the costs and benefits of audio–visual integration. Proceedings of Biological Sciences 2006;273:2159–68.
[72] Rock J, Victor I. Vision and touch: an experimentally created conflict between the two senses. Science 1964;143:594–6.
[73] Rockland KS, Ojima H. Multisensory convergence in calcarine visual areas in macaque monkey. International Journal of Psychophysiology 2003;50:19–26.
[74] Rosenthal O, Shimojo S, Shams L. Sound-induced flash illusion is resistant to feedback training. Brain Topography 2009;21:185–92.
[75] Sathian K, Zangaladze A, Hoffman JM, Grafton ST. Feeling with the mind's eye. NeuroReport 1997;8:3877–81.
[76] Sato Y, Toyoizumi T, Aihara K. Bayesian inference explains perception of unity and ventriloquism aftereffect: Identification of common sources of audiovisual stimuli. Neural Computation 2007;19:3335–55.
[77] Scheef L, Boecker H, Daamen M, Fehse U, Landsberg MW, Granath DO, et al. Multimodal motion processing in area V5/MT: evidence from an artificial class of audio-visual events. Brain Research 2009;1252:94–104.
[78] Seitz AR, Kim R, Shams L. Sound facilitates visual learning. Current Biology 2006;16:1422–7.
[79] Sekuler R, Sekuler AB, Lau R. Sound alters visual motion perception. Nature 1997;385:308.
[80] Shams L, Iwaki S, Chawla A, Bhattacharya J. Early modulation of visual cortex by sound: An MEG study. Neuroscience Letters 2005;378:76–81.
[81] Shams L, Kamitani Y, Shimojo S. What you see is what you hear. Nature 2000;408:788.
[82] Shams L, Kamitani Y, Shimojo S. Visual illusion induced by sound. Cognitive Brain Research 2002;14:147–52.
[83] Shams L, Kamitani Y, Thompson S, Shimojo S. Sound alters visual evoked potentials in humans. NeuroReport 2001;12:3849–52.
[84] Shams L, Ma WJ, Beierholm U. Sound-induced flash illusion as an optimal percept. NeuroReport 2005;16:1923–7.
[85] Shams L, Seitz A. Benefits of multisensory learning. Trends in Cognitive Sciences 2008;12:411–7.
[86] Shimojo S, Miyauchi S, Hikosaka O. Visual motion sensation yielded by non-visually driven attention. Vision Research 1997;37:1575–80.
[87] Shipley T. Auditory flutter-driving of visual flicker. Science 1964;145:1328–30.
[88] Soto-Faraco S, Kingstone A, Spence C. Multisensory contributions to the perception of motion. Neuropsychologia 2003;41:1847–62.
[89] Soto-Faraco S, Spence C, Kingstone A. Cross-modal dynamic capture: congruency effects in the perception of motion across sensory modalities. Journal of Experimental Psychology: Human Perception and Performance 2004;30:330–45.
[90] Spence C, Driver J. Audiovisual links in endogenous covert spatial attention. Journal of Experimental Psychology: Human Perception and Performance 1996;22:1005–30.
[91] Spence C, Driver J. Audiovisual links in exogenous covert spatial orienting. Perception & Psychophysics 1997;59:1–22.
[92] Spence C, Pavani F, Driver J. Crossmodal links between vision and touch in covert endogenous spatial attention. Journal of Experimental Psychology: Human Perception and Performance 2000;26:1298–319.
[93] Stein BE, London N, Wilkinson LK, Price DD. Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis. Journal of Cognitive Neuroscience 1996;8:497–506.
[94] Stein BE, Meredith MA. The merging of the senses. Cambridge (Mass): MIT Press; 1993.
[95] Stein BE, Meredith MA, Wallace MT. The visually responsive neuron and beyond: multisensory integration in cat and monkey. Progress in Brain Research 1993;95:79–90.
[96] Thurlow WR, Jack CE. Certain determinants of the "ventriloquism effect". Perceptual and Motor Skills 1973;36:1171–84.
[97] Tootell RB, Reppas JB, Kwong KK, Malach R, Born RT, Brady TJ, et al. Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. Journal of Neuroscience 1995;15:3215–30.

Please cite this article in press as: Shams L, Kim R. Crossmodal influences on visual perception. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.04.006

[98] Valjamae A, Soto-Faraco S. Filling-in visual motion with sounds. Acta Psychologica (Amst) 2008;129:249–54.
[99] Van der Burg E, Olivers CN, Bronkhorst AW, Theeuwes J. Pip and pop: nonspatial auditory signals improve spatial visual search. Journal of Experimental Psychology: Human Perception and Performance 2008;34:1053–65.
[100] Violentyev A, Shimojo S, Shams L. Touch-induced visual illusion. NeuroReport 2005;16:1107–10.
[101] Vroomen J, de Gelder B. Sound enhances visual perception: cross-modal effects of auditory organization on vision. Journal of Experimental Psychology: Human Perception and Performance 2000;26:1583–90.
[102] Walker JT, Scott KJ. Auditory–visual conflicts in the perceived duration of lights, tones, and gaps. Journal of Experimental Psychology: Human Perception and Performance 1981;7:1327–39.
[103] Wallace MT, Meredith MA, Stein BE. Converging influences from visual, auditory, and somatosensory cortices onto output neurons of the superior colliculus. Journal of Neurophysiology 1993;69:1797–809.
[104] Wallace MT, Stein BE. Cross-modal synthesis in the mid-brain depends on input from association cortex. Journal of Neurophysiology 1994;71:429–32.
[105] Wallace MT, Stein BE. Development of multisensory neurons and multisensory integration in cat superior colliculus. Journal of Neuroscience 1997;17:2429–44.
[106] Wallace MT, Wilkinson LK, Stein BE. Representation and integration of multiple sensory inputs in primate superior colliculus. Journal of Neurophysiology 1996;76:1246–66.
[107] Warren DH, Welch RB, McCarthy TJ. The role of visual–auditory "compellingness" in the ventriloquism effect: implications for transitivity among the spatial senses. Perception & Psychophysics 1981;30:557–64.
[108] Watanabe K. Crossmodal interaction in humans. Ph.D. thesis. Los Angeles: California Institute of Technology; 2001.
[109] Watanabe K, Shimojo S. Attentional modulation in perception of visual motion events. Perception 1998;27:1041–54.
[110] Watkins S, Shams L, Josephs O, Rees G. Activity in human V1 follows multisensory perception. Neuroimage 2007;37:572–8.
[111] Watkins S, Shams L, Tanaka S, Haynes J-D, Rees G. Sound alters activity in human V1 in association with illusory visual perception. Neuroimage 2006;31:1247–56.
[112] Welch RB. Perceptual modification: Adapting to altered sensory environments. New York: Academic Press; 1978.
[113] Welch RB, DuttonHurt LD, Warren DH. Contributions of audition and vision to temporal rate perception. Perception & Psychophysics 1986;39:294–300.
[114] Witten IB, Knudsen EI. Why seeing is believing: merging auditory and visual worlds. Neuron 2005;48:489–96.
[115] Wozny D, Beierholm U, Shams L. Human trimodal perception follows optimal statistical inference. Journal of Vision 2008;8(3):1–11. doi:10.1167/8.3.24. http://journalofvision.org/8/3/24/.
[116] Yuille AL, Bülthoff HH. Bayesian decision theory and psychophysics. In: Knill DC, Richards W, editors. Perception as Bayesian inference. 1996. p. 123–61.

Please cite this article in press as: Shams L, Kim R. Crossmodal influences on visual perception. Physics of Life Reviews (2010), doi:10.1016/j.plrev.2010.04.006