Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Journal of Physiology - Paris 98 (2004) 265–279

Journal of Physiology Paris

www.elsevier.com/locate/jphysparis

Crossmodal integration for perception and action

Christophe Lalanne [a,b,*], Jean Lorenceau [a]

[a] UNIC, CNRS UPR 2191, 1 avenue de la Terrasse, F91198 Gif-sur-Yvette, France
[b] France Telecom R&D, DIH/UCE, 38-40 rue du Général Leclerc, F92794 Issy-les Moulineaux, France

Abstract

The integration of information from different sensory modalities has many advantages for human observers, including increase of salience, resolution of perceptual ambiguities, and unified perception of objects and surroundings. Several behavioral, electrophysiological and neuroimaging data collected in various tasks, including localization and detection of spatial events, crossmodal perception of object properties and scene analysis are reviewed here. All the results highlight the multiple faces of crossmodal interactions and provide converging evidence that the brain takes advantages of spatial and temporal coincidence between spatial events in the crossmodal binding of spatial features gathered through different modalities. Furthermore, the elaboration of a multimodal percept appears to be based on an adaptive combination of the contribution of each modality, according to the intrinsic reliability of sensory cue, which itself depends on the task at hand and the kind of perceptual cues involved in sensory processing. Computational models based on bayesian sensory estimation provide valuable explanations of the way perceptual system could perform such crossmodal integration. Recent anatomical evidence suggest that crossmodal interactions affect early stages of sensory processing, and could be mediated through a dynamic recurrent network involving backprojections from multimodal areas as well as lateral connections that can modulate the activity of primary sensory cortices, though future behavioral and neurophysiological studies should allow a better understanding of the underlying mechanisms.

Keywords: Crossmodal interaction; Crossmodal binding; Maximum-likelihood estimation; Bayesian perception

1. Introduction

Our everyday life perception is subject to a continuous bombardment of external and internal information, both of which arise as a consequence of our own active and adaptive behavior. Our current understanding of how the brain processes these afferent sensory signals to yield a coherent and unified perception of the world remains, however, far from complete. Indeed, the problem of crossmodal integration appears to be one of the most challenging issues in the study of human perception and sensorimotor control, and several studies have tried to characterize the way human observers combine multiple sensory signals in an efficient way, ensuring both the necessary coalescence of the senses and an adaptive behavior in the constantly evolving environment we experience every day (e.g. reviews in [26,93,102]).

In fact, most living systems are able to integrate information coming from different sensors and to use this information to select and control their active behaviors. These afferent signals code for different aspects of the environment—for example, the eyes detect photons arising on the retina, the ears analyze variations of acoustic pressure in the auditory duct—and are mapped onto different reference frames. Nevertheless, the brain appears to be endowed with a surprising capability of extracting invariant properties among this blend of information, and to selectively combine these space–time distributed signals, in order to get a coherent interpretation of the surroundings.

Such highly-skilled capabilities can be illustrated by the following example. Imagine someone attending to his own occupations, and whose attention is suddenly drawn by the sound made by the boiling water of a saucepan in

[*] Corresponding author. Current address: LENA, CNRS UPR 640, Hôpital de la Salpétrière, 47 boulevard de l'Hôpital, F75651 Paris, France (C. Lalanne, J. Lorenceau). Tel.: +33-1-42-16-11-83; fax: +33-1-45-86-25-87.

E-mail address: lalanne@chups.jussieu.fr (C. Lalanne).

0928-4257/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jphysparis.2004.06.001

the kitchen; he looks towards the object in question, and it quickly appears that the saucepan must be taken away from the cooker. Now, what has to be done is to reach for the saucepan: crucial information about the object to be grasped are conveyed by vision (water steam), audition (sound of the boiling water), and haptics (heat and vibration of the handle). All of these perceptual cues, gained through different sensory modalities, concern the same object and, together, should influence the actor to perform this action carefully.

While from a phenomenological point of view we seem to have a direct and effortless experience of our surroundings, complex mechanisms are at work between sensory processing and cognitive interpretation or goal-directed motor action. Studies on crossmodal integration are particularly relevant for understanding spatial cognition because orienting in space and detection of singular events are common human behaviors that are to a large extent dependent on multiple and simultaneous sensory information [102]. Furthermore, a precise representation of our body and external space is necessary for planning accurate movements and controlling action directed towards objects in the environment [75,126], as illustrated in the preceding example. Many other situations involve crossmodal processing of different sensory information such as when one tries to identify an object by means of sight and hand movements, or when we decide that a car is moving away by perceiving a congruent decrease in sound intensity and apparent size. We would thus define crossmodal integration as the process by which the same or different perceptual features carried by distinct sensory modalities are bound together, both in the temporal and spatial dimensions, in order to get a common coherent representation of space and physical objects. However, the criteria and the mechanisms by which the brain is able to bind these modality-specific cues in order to yield a coherent percept remains an open issue.

In fact, several behavioral, electrophysiological and neuroimaging studies have shown that spatial cognition is not the monopoly of vision as traditionally advocated for many decades, and that sensory modalities do not work in complete isolation from each other, but rather share potentially common spatial representations and dedicated mechanisms of perceptual analysis that lead to fine interactions in both the spatial and temporal domains (e.g. reviews in [15,26,75,93,102]).

In this review, we shall discuss the following issues: At what level of perceptual organization do crossmodal interactions occur? How are crossmodal cues integrated into a unified and coherent multimodal percept? As in the introductory example, we shall take a progressive approach, starting from the more perceptual related processes (e.g. passive detection/identification) and leading to more elaborate sensorimotor behavior (e.g. pointing, tracking, exploration of objects). The review will thus be organized as follows: first, we shall deal with space representation through the description of visual 'dominance' effects commonly found in spatial localization tasks and the effect of voluntary or automatic orientation of attention in the detection of a spatial target (Section 2). We will then examine how congruent or non-related stimulation in one modality can affect sensory processing in another modality whether it be to identify dynamic external events, or more generally to analyze a perceptual scene (Section 3). Finally, more complex behaviors will be reviewed and in particular, we shall briefly discuss how sensory modalities interact for object recognition and sensorimotor control (Section 4). The crossmodal interactions reported in the above field of investigations, as well as experimental data collected in electrophysiological and neuroimaging studies, will be discussed in the light of recent theoretical models of crossmodal integration (Section 5).

2. Space representation: localization of events

Identifying the location of an object relies on a precise spatial representation, although spatial cues are often gathered through different sensory modalities. Location of a distal visual object emitting a sound can be recovered from both vision and audition, for instance. In this way, different modalities provide redundant, or convergent, spatial cues, which must be integrated into a coherent and unified percept in order to yield a perceptual decision or an accurate reaching movement. Furthermore, it is clear that from the huge amount of information received by our different sensors, and because our processing capabilities are limited, we have to select relevant information for our ongoing task and discard other irrelevant cues. On the one hand, we must deal with the selective filtering processes that are at work in spatial attention, and on the other hand, with the fusion of concurrent information perceived through different sensory modalities. Therefore, an outstanding question that remains is to understand how the brain combines these different signals to yield a coherent interpretation of surroundings and external events.

In the following section, we shall examine these two aspects of spatial perception, by reviewing the influence of visual cues on crossmodal spatial perception of external events and body parts, and the influence of orienting of attention in one modality on spatial processing in another modality.

2.1. Crossmodal localization of external events

2.1.1. Immediate and consecutive responses to intersensory discrepancy

In line with the original work of Stratton [105] on spatial localization of auditory or visual stimuli after a horizontal reversal of visual field (see also [30]), most studies have pointed to the 'dominance' of visual inputs over the other sensory signals, also referred as to 'visual capture' when it concerns body-parts [77,124]. This has been widely illustrated by the 'ventriloquism effect' [8,50], in which the perceived spatial position of an auditory source is biased toward the spatial position of a simultaneously presented visual stimulus. The magnitude of this effect depends on spatial disparity and temporal synchronization between the two sources, both of these determinants being involved in the constitution of unitary spatial entity. Such spatial mislocalization effects have also been reported with other combinations of sensory inputs. A significant biasing effect of proprioception on the perceived position of auditory stimuli has also been reported [14,125]. However, inducing visual biases with auditory distractors has never been as conclusive [7], and concurrent auditory stimuli have only little effect on the proprioceptively-perceived hand position [14,77,121]. Moreover, it has been shown that the 'ventriloquism effect' occurs even if the subject's attention is not voluntary or automatically directed to the visual stimulus [9,114], hence dismissing an explanation based on a strict attentional effect. Taken together, the preceding observations indicate that localization performance is not affected in a fully symmetrical way: auditory and proprioceptive targets give rise to the most salient mislocalization effect when a concurrent and synchronized visual stimulus is present, so long as this is not too distant, while the reverse—visual position estimation with concurrent auditory or proprioceptive targets—leads to perceived spatial discrepancy of a lesser extent, if any. Therefore, in the presence of conflicting spatial cues, human observers seem to rely more on the available visual information, keeping conflicting auditory and proprioceptive cues apart from the final perceptual decision.

In addition to verbal estimations or pointing movements, another method that has been widely used to demonstrate the dominance of visual inputs in crossmodal integration of discrepant spatial cues requires location judgements at the time of a distortion of the whole visual field with prismatic lenses or goggles (e.g. [47,120]; review in [123]). In these situations, subjects are asked to report the location of their hand or of a visual target, while the matching between visual and proprioceptive information is altered. Prism adaptation studies have pointed equally to the dominance of vision in spatial localization, as well as to the fast re-mapping of the reference frame used for pointing or judging hand location in the presence of conflicting spatial signals. In a typical task, subjects are wearing prismatic goggles that deviate the visual field by several degrees with respect to the sagittal plane. When subjects are asked to localize a target by means of an arm movement, it is generally observed that their pointing movement deviates accordingly, giving rise to significant spatial errors in the final pointing position ('visual capture', [77]). In a post-adaptation test, when the goggles are withdrawn, a strong sensorimotor after-effect is observed and subjects make systematic errors in the direction opposite to that found in the pre-adaptation test, which also reflects the biased recalibration of the matching between proprioceptive and visual spatial signals. Similarly, immediate estimation of hand location during conflicting visual information is biased toward perceived visual position, provided that the induced conflict is not too large [47,77]. Such spatial mislocalization effects have also been reported with visual distractors and tactile targets [76].

2.1.2. A common adaptive crossmodal fusion principle

A commonly accepted explanation for such dominance of visual inputs over the other sensory signals is that the perceptual system treats temporally coincident bimodal events, provided that they are not too distant, as a single unitary event on the basis of a 'unity assumption' [124]. Furthermore, it has been proposed that perceptual interpretation relies on the most reliable and precise information, both in the temporal and spatial domains, which is known as the 'modality appropriateness hypothesis' [124,125]. In spatial tasks, the perceptual system should therefore rely more on vision, which is the more accurate source of information due to its greater spatial resolution. Auditory and proprioceptive information are taken into account to a lesser extent in the final perceptual decision, as they are less precise than vision in the spatial domain.

As we have seen, perceptual responses to auditory or proprioceptive targets are located closer to a simultaneously perceived visual cue; nevertheless spatial mislocalization effects occur even without the involvement of the visual modality, and auditory cues do not give rise to significant spatial biases. Together with the fact that the 'ventriloquism' effect can be observed without the involvement of attention [9,114], it is unlikely that perceptual judgements regarding spatial location are based on the output of a bimodal competition driven by selective allocation of attentional resources, whereby the visual modality strictly dominates over another modality, as has been proposed earlier by some authors (e.g. [78]; see also [125]). Rather, human observers seem to process both sources of information and the direction of crossmodal bias depends on some automatic perceptual processes. Perceptual responses appear merely to be the result of a linear weighting mechanism between the contribution of each modality, the weights being determined according to the intrinsic reliability of each modality, i.e. its spatial acuity.

This is exactly what was observed in the preceding studies with discrepant spatial cues: immediate perceptual responses are biased in the direction of the more reliable of the two sensory modalities, i.e. vision in the case of auditory–visual or proprioceptive–visual localization, or proprioception in the case of auditory–proprioceptive localization. In the case of prism adaptation studies, a similar principle could explain the observed results, since visual acuity is better than proprioceptive acuity (e.g. [74,119]). After continuous exposure to an optical deviation, the re-mapping between visual and proprioceptive signals can lead to a re-weighting of these conflicting spatial signals, resulting in an enhancement of the proprioceptive contribution to the guidance of pointing movements.

Additional evidence in support of an adaptive combination of proprioceptive and visual inputs comes from the work of van Beers and colleagues [106–108]. Comparing unimodal and bimodal performance when subjects are asked to match the unseen position of one of their fingers with either proprioceptive information, visual information or both, van Beers et al. [106] have shown that subjects' responses are not based on the contribution of independent position signals. On the contrary, they result from an efficient weighting of the contribution of the two modalities, the weighting being related to the precision of each unimodal information. More precisely, the analysis of variable error reveals that, in the bimodal condition, subjects' responses are not distributed on a straight line between the two unimodal response distributions, as would have been expected if the perceptual response was the simple arithmetic mean of the two unimodal responses. In another experiment of visual–proprioceptive localization of the hand, van Beers et al. [107] examined the relative precision of the two sensory systems and found that visual and proprioceptive localization performance is affected in a direction-dependent manner. In the proprioceptive condition, hand positions are better localized in the radial direction, especially when the hand is closer to the shoulder, while in the visual condition, spatial judgments are more accurate in the azimuthal direction than in the radial direction. Similar results were obtained by van Beers et al. [109] in a sensorimotor adaptation study: the integration of visual and proprioceptive position information is dependent upon the direction of the pointing movements with respect to the observer. In particular, observers rely more on proprioceptive than on visual information when they have to point at visual and proprioceptive targets located in the radial direction, as compared to situations in which targets are located in the azimuthal direction (see also [108]). All these results highlight a finer interaction between proprioception and vision than has previously been reported.

2.2. Crossmodal spatial attention

Recent experimental work based on attentional paradigms provides new insights toward a better understanding of the internal linkage of bimodal signals, especially the mapping between different modality-specific reference frames. It has been shown for instance that previous and concurrent sensory information can facilitate the detection and location discrimination of a target delivered in another modality (e.g. [13]). Recent work on covert attention with endogenous and exogenous cues demonstrates strong crossmodal links and also suggests that spatial attention can be driven by a common multimodal representation (review in [24]).

In a series of elegant experiments based on an orthogonal cueing paradigm, Driver and collaborators [25,98,100] have demonstrated strong crossmodal links in covert spatial attention to tactile, visual and auditory stimuli. Observers were asked to judge the position of a visual, auditory or tactile stimuli having a high or a low probability of apparition at the top or bottom of a screen, while a preliminary cue indicated the likely target side for one modality only. For instance, a top (vs. bottom) target presented in the primary modality is cued by a left–right stimulation in the other modality. With this design, the two sensory modalities are purely orthogonal, ensuring that there is no priming artifact that could bias the upcoming sensory event: that is, tactile cueing on the left side does not give information about the azimuthal location of the visual stimulus. Spatial cueing effects were found in both auditory–visual [98] and visuo–tactile [100] tasks: reaction time analysis revealed that responses are shorter when high-probability stimuli are delivered to the same side of the visual field as the preceding cue, whether the cueing modality is the same as the test modality or a different one. However, cueing effects were always larger for the primary modality. Such spatial cueing effects suggest a tendency for auditory, tactile and visual endogenous attention to be directed together to the same spatial location.

The question of how the brain keeps an accurate mapping of space while spatial attention is oriented toward different locations in the visual field, has been addressed in another cueing paradigm. In this experiment [25], participants engaged in a crossed or uncrossed hands posture, were asked to judge the spatial location of a visual stimulus appearing in the right or left side of the visual field, after a tactile vibration had been delivered on the right or left finger. In the uncrossed condition, i.e. when the correspondence between somesthetic and visual spatial representations is preserved, it was observed that visual judgements on the left hemifield were faster when preceded by a tactile cue on the left finger. In the crossed condition, where the left/right sagittal relation between somesthesic and visual reference frames is reversed, the opposite pattern of results is observed: a vibration on the left finger results in faster visual judgements on the right hemifield. So, these results do not correspond completely to the classical view of an increased activation in the contralateral hemisphere due to crossed unimodal cortical projections spreading in turn to the other sensory areas located within the same hemisphere [53]. The authors suggest that judgements are made on the basis of an external, modality-free, spatial coordinate system. Thus, according to the authors, the brain maintains an unbiased bimodal representation of space by appropriately linking together unimodal spatial information onto a common frame of reference. Taken together, these results suggest that an unitary multimodal percept, elaborated on the basis of multiple spatially coincident sensory events, drives crossmodal spatial attention [25].

These experimental findings have also been replicated while recording evoked response potentials (ERP) in an odd-ball detection task, where subjects had to detect infrequent tactile targets among tactile, visual (Experiment 1) and auditory (Experiment 2) stimuli, while hand posture was varied between blocks of trials or across trials [27]. The initial hypothesis was that if crossmodal linkage is based on an external frame of reference, the effect of tactile attention on ERPs elicited by the task-irrelevant visual or auditory stimuli should lead to larger ERP amplitudes for these stimuli when close to the attended hand (i.e. on the same side of external space), regardless of hand posture. On the other hand, according to the hemispheric projections hypothesis [53], orientation of endogenous spatial attention toward one or the other side of the visual hemifield is determined by the relative levels of activation of the two hemispheres. Thereby, an increase of activation in the left hemisphere following a sensory cue leads to a rightward shift of spatial attention. This hypothesis would thus predict enhancement of activity for spatially congruent tactile and visual or auditory stimuli in the uncrossed condition only (or conversely, with spatially inverted tactile and task-irrelevant stimuli in the crossed condition). Indeed, the observed results are consistent with the former hypothesis, since a systematic increase was observed in ERP amplitude of visual and auditory components when visual and auditory stimuli were located on the same side of external space as the attended hand, for both crossed and uncrossed hand conditions. However, attentional modulations of somatosensory ERPs were affected by hand posture, suggesting that crossed posture should lead to a possible disruption of tactile spatial attention without affecting the crossmodal effect upon vision or audition [27].

In conclusion, the results suggest that crossmodal linkage involved in spatial attention is not determined by hemispheric projections but appears to be anchored to a common external coordinate system, with the exception that the tactile modality seems to be driven by both specific anatomical pathways and such an external reference frame ([27]; see also, [99]). Together with findings on spatial localization of bimodal events, all these results point to the dynamic elaboration of a multimodal percept, that is a global percept resulting from the adaptive integration of unimodal inputs. Such adaptive bimodal integration in spatial localization and the tight linkage between congruent signals in orientation of spatial attention suggest that human observers are able to build an unified multimodal representation of space. Since more elaborate spatial analyses generally rely on such visuo-spatial skills, a coherent crossmodal representation of sensory events that operates at early levels of perceptual processing would obviously be of relevant interest in spatial cognition.

3. Crossmodal interactions in perceptual scene analysis

We have seen how the brain solves the discrepancy between simultaneously presented bimodal signals for spatial localization, and how it takes advantage of sequential convergent bimodal information regarding stimulus location when orienting spatial attention. Crossmodal effects have also been extensively described in the more general framework of perceptual scene analysis. Indeed, as we will see in the next section, additional information delivered in a secondary modality can help to solve perceptual ambiguities, or facilitate the segmentation of a perceptual scene.

3.1. Auditory–visual interactions in motion analysis

Several experiments have described reciprocal auditory–visual interactions in the perception of dynamic displays. Using an auditory–visual adaptation task, Kitagawa and Ichihara [54] have shown that adaptation to the perceived motion in depth of a visual stimulus (induced by the contraction and expansion of its apparent size) can alter the perceived intensity of a sound paired with the visual stimulus. Likewise, Mateeff et al. [67] have demonstrated that a static auditory stimulus, when presented simultaneously with a moving visual display, can yield a sensation of displacement of the auditory source. Just as with a static auditory source, visual motion can also influence induced auditory motion after-effect [113]. Similarly, visual motion processing is modulated by concurrent auditory motion signals, though congruent auditory cues seem to facilitate visual motion identification only when visual motion cues are not reliable [71].

The observation that a consistent moving visual stimulus can influence the perception of static or moving auditory stimuli adds further support to the hypothesis that spatio–temporal correlation between sensory inputs, here motion signals, can be used by perceptual systems, precluding the occurrence of segmentation between bimodal events. However, there seems to be a more effective influence of visual motion cues on auditory motion processing than the reverse (see also [97]).

This suggests that visual cues can not only be the preferentially used static spatial information, as observed in spatial localization, but also that visual motion signals play a significant role in scene analysis and can be given a heavier weight in the perceptual interpretation, due to their greater reliability in comparison to auditory motion signals.

Such auditory–visual interactions are not limited to the case of continuous auditory stimulation, but have also been described with discontinuous or sudden auditory events. It has recently been shown that transient auditory stimulation can influence visual processing of position and motion information. First, Alais and Burr [3] have shown that perceived temporal lags in an auditory–visual version of the classical visual 'flash-lag effect' [1] were found to be intermediate between those observed with single auditory and visual stimuli, in both bimodal conditions (visual flash/auditory motion and auditory flash/visual motion). However, the estimated latency to perceive the two events as physically aligned was not the same in the two bimodal conditions: results for two subjects show that this latency was higher in the auditory motion/visual flash condition. Second, Vroomen and de Gelder [111] describe a temporal 'ventriloquism effect', where an auditory stimulus paired to a visual flash presented in close temporal relationship with a moving visual stimulus can bias the perceived temporal dimension of the visual flash. In this study, also based on the 'flash-lag effect', subjects had to judge the position of a visual spot relative to a moving bar (i.e. positive or negative spatial lag), where the spot was flashed in varying temporal phase (−66.7 to +66.7 ms) relative to the moving stimulus. On half of the trials, an additional auditory click was presented in synchrony with the visual spot. Results show that the presence of an auditory stimulus sharpens the discrimination response curve for spatial lag, while it decreases the magnitude of the flash-lag effect. When the sound was not synchronized to the flashed visual spot but delayed by −100 to +100 ms, a modulation of the effect was observed: sounds delivered before the visual flash decreased the flash-lag effect, while the reverse pattern was observed for sounds delivered after the visual flash. As emphasized by the authors, these findings are of relevant interest since they demonstrate that audition can interact with visual processing not only when auditory information has a structured rhythmic pattern as previously shown (e.g. [94,115]), but also with an isolated auditory signal. Moreover, it appears that the time window during which an auditory event is able to modulate visual processing is rather narrow and presents asymmetrical boundaries, since auditory cues perceived after a visual event seem to be less effective. Therefore, integration of auditory and visual information rely on the spatial correlation between perceived events, but also depends strongly on the temporal dimension and stimulus duration.

The role of the temporal correlation between auditory and visual events is also strengthened by the study of Sekuler et al. [89] who asked subjects to qualitatively evaluate the motion of two small spots translating along crossed paths. When the two spots cross in the middle of their trajectories, subjects report perceiving either two spots bouncing in opposite directions or two spots sliding one under the other, this latter being the most frequently reported percept. This bi-stable stimulus, initially devised by Metzger [70], is thus well-suited for studying the effect of an auditory cue on visual motion processing. Indeed, by adding a brief tone at the moment of spatial coincidence, the proportion of 'bouncing' percepts increases significantly as compared to the proportion of 'streaming' percepts. A similar phenomenon of lesser magnitude is observed when sound is delivered 150 ms before spatial coincidence, but it appears to be less consistent when sound is delivered 150 ms after spatial coincidence. Again, this highlights an asymmetrical time window for auditory–visual interaction to occur.

One obvious explanation of this auditory effect on visual motion perception, as suggested by these authors, is that it may result from strong perceptual associations developed between auditory and visual properties of physical objects, such as the sound they make when they enter into contact. Other experiments have also emphasized the role of transient auditory signals in the perception of physical causality [36], and the critical time window (<200 ms) needed between two events for observing such 'causality percepts' [88]. Again, an internal assumption of perceptual 'unity' [124], anchored on temporal and spatial relationships between sensory events, seems to drive perceptual organization in scene analysis.

In summary, auditory motion cues appear to be less effective in comparison to static and transient cues to induce significant crossmodal interactions in visuo–auditory scene analysis. As in location judgements, the elaboration of a coherent percept seems to result from an efficient weighting of both auditory and visual contributions according to their spatial and temporal reliability. In the case of dynamic signals, visual motion cues appear to be used preferentially compared to auditory ones, but transient auditory signals could also participate in the segmentation of conflicting or ambiguous visual information. This selective linkage between audition and vision in perceptual analysis could obviously follow from crossmodal apprehension of space and physical objects during perceptual learning. However, this implies that relevant crossmodal associ-

[1] Whereby a briefly flashed visual stimulus that is physically aligned with a moving visual contour appears to lag behind this last one.

ations must have been differentiated from arbitrary relations between perceived events. The observation that perception of temporal properties of visuo–auditory display is effective in 8-month-old infants [64], and that 6- and 8-month-old infants show comparable modulation of perception in the 'bouncing/streaming' display [87] suggest that auditory–visual interactions develop early in life. Around this critical period which also matches that for the development of spatial attention, selective processes increasing the salience of redundant auditory and visual information could participate in perceptual learning (see also [63]).

3.2. Auditory–visual interactions and perceptual modifications

We have seen that information that is redundant in one modality can facilitate the detection and identification of a stimulus delivered in another modality (e.g. faster reaction times with redundant bimodal spatial cues). Crossmodal interactions are also observed when additional information is not related to the task at hand: for instance, a salient auditory cue can help to detect a structured visual target embedded in distractors, even when this additional cue is not linked on any physical dimension to the target [112]. Such auditory influence on visual perception also exerts an effect on low-level features, since it has been shown that the perceived intensity of a liminal visual stimulus is modulated by the release of a simultaneous sound [103].

Most surprising are the observations made by Shams et al. [91]; see also Shams et al. [92] in an experiment studying the effect of concurrent auditory tones delivered during the presentation of successive visual flashes on the estimation of seen events. The results show that auditory tones strongly bias the number of perceived flashes: if a single visual flash is presented on the screen and more than one brief tone is delivered during a trial, subjects report perceiving more than one visual flash. Manipulating the temporal pairing of the auditory–visual events highlights a critical time window of ~100 ms (−75 to +115 ms), comparable to that found by Sekuler et al. [89]. The observation that a single beep paired with two visual flashes does not elicit any additional illusory flash, and leads to a performance similar to that with two beeps paired with one flash, rules out an explanation of this 'sound-induced illusory flashing' effect in terms of a cognitive bias. Rather, this suggests that crossmodal interactions occur at early stages of perceptual processing [92]. According to the authors, these results could be best understood as the selective influence of a discontinuous stimulus in one modality, i.e. a sensory transient (here the auditory tone), on the perception of a continuous stimulus in another modality (here vision). This explanation based on the influence of a concurrent and transient sensory cue on perception is also compatible with the findings of Sekuler et al. [89], since it has been shown that when the synchronized sound is not salient enough (for instance, by embedding it in a series of identical sounds with the same pitch), the 'bouncing' illusion declines [122]. Furthermore, the effect of a transient on perceptual interpretation of 'streaming/bouncing' motion is not dependent on the sensory modality in which it is delivered, as has also been observed with visual (brief presentation of a ring) and somatosensory (vibration of the finger) cues given in synchrony with the crossing of the visual spots.

In order to study at what level of perceptual processing these crossmodal effects take place (i.e. low-level sensory stage vs. decision stage), Shams et al. [90] have replicated the illusory flash experiment while recording visual evoked potentials. When subtracting the responses evoked by the bimodal condition from the summed unimodal conditions, they observed two significant periods in the difference wave for the illusory flash condition (a single flash paired with two beeps is perceived as two flashes). An early visual evoked response was observed around 174 ms post-stimulus (or 103 ms if stimulus-onset-asynchrony between successive visual flashes is subtracted), and another period of significant activity between 262 and 360 ms. Considering both this short latency response and the similarity between the global activity pattern—a positive peak and dual positive peaks in the two significant periods of difference waves—evoked by the illusory flash condition and that evoked by a true physical second flash (control condition), the authors suggest that auditory–visual interactions occur early in sensory processing in the visual cortex, and are not part of a higher decision level.

As in the preceding case of auditory–visual processing of motion and static signals, we see that one important temporal clue used by the perceptual system is the transient nature of the concurrent auditory event, regardless of the fact that it occurs with a continuous or a discontinuous stimulation in the visual modality. This has led some authors to propose that transient events possess a special status in perceptual scene analysis [93]. However, this raises at least two questions at the developmental level: how are contingent auditory and visual stimulation associated during early life, and how can these relevant crossmodal associations spread over other arbitrary spatio–temporal relations, since we have already seen that irrelevant auditory cues also modulate perceptual interpretation. Finally, this further suggests that the visual modality, which has been viewed for a long time as the dominant sense in spatial tasks, shares strong relationships with auditory processing of salient signal, at the level of perceptual organization (see also [11]). Here again, spatio–temporal integration of auditory–visual dynamic and static cues appears to follow an adaptive combination rule, in close analogy with the one that is observed in location judgements, that is dependent on the nature of the concurrent auditory cue.

4. Crossmodal 'active' perception

Whereas, above, we have focused our review mainly on crossmodal effects in the analysis of spatial properties of distal events, we now review other experimental data showing how the brain combines different complementary local features conveyed by different modalities in order to yield a coherent interpretation of object properties.

Commonly used techniques found in the literature on crossmodal object cognition are recordings of free exploratory movements (e.g. [55,56,60]), unimodal vs. bimodal matching and identification tasks (e.g. [4,17,39,44,73]), and crossmodal transfer tasks [45,48]. We shall limit this review to crossmodal discrimination tasks and sensorimotor control, as these two fields of research allow to offer a valuable insight into crossmodal processes that are at work in these active behaviors. The interested reader should refer to the more extensive review of Calvert [15], which encompasses a broader range of experimental data collected in neuroimaging studies.

4.1. Crossmodal object and shape analysis

When one has to identify an object that can be seen and felt at the same time, both vision and touch provide important clues on object structure. Indeed, some of these clues can be accessed through both modalities, such as size and position, while other are modality-dependent, e.g. volume and temperature for the haptic sensem, [2] and color for vision [56]. According to Lederman and Klatzky [59,60], specific exploratory procedures are used for the extraction of these specific local features with the haptic modality. Thus, different complementary spatial attributes are conveyed by different modalities and the brain has to combine them appropriately. As proposed by Lederman and coworkers, since the haptic modality serves both a sensory and motor function, haptic sense can be considered as the cornerstone between perception and action. Studying interactions between haptic and visual functions is thus an interesting approach to bridging the gap between perception and action.

Early studies on crossmodal interactions with conflicting spatial cues (e.g. [84]) have shown that vision took precedence over touch. However, haptic sense could take a more prominent role in object recognition because it is specifically a contact modality, specialized for the analysis of object material properties [60]. Moreover, the touch modality is not limited to a frontal 'field of view' and can access the back of an object, unlike vision, hence perfectly complementing object features accessed by the visual modality (e.g. [73]). Thus, while it has been argued that vision dominates our perception of spatial properties, in the case of texture analysis, haptic information could be used preferentially depending on the spatial features being analyzed [49,62]. Indeed, crossmodal integration between touch and vision should be dependent upon the task at hand: in spatial localization or analysis of spatial properties of object, vision provides reliable sensory information and should contribute to a greater extent to the perceptual interpretation, whereas in the analysis of material properties, haptic information should be used preferentially. An important issue is to understand how the two sources of information are integrated together, given their own domain of expertise, to yield a multimodally determined percept.

Several experiments have tried to address the question of optimal combination of haptic and visual information in the processing of object or surface properties [58,61,62]. In a study with raised dot patterns, it has been shown that visual or haptic dominance effects could be induced by manipulating task instructions [62]. When subjects were asked to evaluate the degree of perceived roughness (i.e. a purely material property of the object) of a textural pattern, by means of qualitative unimodal judgments, touch dominated vision. In another experimental condition, subjects were asked to evaluate in a similar way the perceived spatial density (i.e. a purely spatial property) of the same raised dot patterns. In this case, a reverse dominance of vision on touch was found. According to the authors, these findings are another illustration of the optimal use of sensory-specific information, as proposed by Welch and Warren [124]. It would have been interesting to see if one can observe a gradual modification of the weight given to each modality when manipulating perceptual reliability of the textural information (for instance, by distorting the visual field or by manipulating haptic feedback).

These findings are in close agreement with another study bearing on crossmodal perception of surface properties [28,29]. After having demonstrated that haptic feedback can bias visual interpretation of 2-D surface slant [29], Ernst and Banks [28] have shown in a visuo–haptic discrimination task that, when estimating the height of a raised ridge, human observers combine haptic and visual information in a very efficient manner, depending on the amount of noise added to visual cues. Visual information thus dominates in perceptual decision only when visual cues are perceived as sufficiently reliable. To account for these results, the authors present

[2] Considered here as the modality analyzing both tactile and proprioceptive cues.

a theoretical model based on maximum-likelihood estimation where unimodal sensory estimates are weighted in inverse relation to their reciprocal variance. Thus, both visual and haptic cues appear to be taken into account according to their intrinsic reliability to yield an optimal bimodal percept, as proposed earlier in the case of bimodal localization of spatial events or body-parts [108].

Finally, in addition to these visuo–haptic interactions in the perception of surface properties, crossmodal interaction between touch and audition have also recently been investigated. By comparing unimodal and bimodal performance, Lederman et al. [61] have shown that when subjects have to judge surface roughness with a rigid probe providing auditory feedback at the contact point, both kinds of information are also taken into account, but tactile cues are used preferentially compared to auditory cues, although subjects appear to be more confident in their perceptual judgements in the bimodal condition. Here again, such findings highlight the fact that human observers take into account all of the available sensory cues, but preferentially use the more reliable ones.

This adds further support to the idea that our perceptuo-motor abilities follow from an optimal use of our different senses, both in our perception of extra-personal space and in relation to objects with which we can potentially interact.

4.2. Sensorimotor integration and motor control

Crossmodal integration is not only a mandatory perceptual mechanism, involved in processing redundant, congruent or orthogonal spatial signals, but also deals with perceptual and motor coordination. Indeed, accuracy of pointing movements or precise manipulation of an object imply that subjects are able to evaluate their own performance through the use of sensory feedback and to rapidly correct their movements in response to upcoming events. The sensory and motor systems should thus interact in an efficient way in order to ensure such a precise interaction.

This is underlined by several studies that have shown that subjects are able to adapt their movement trajectory on a short time scale when visual feedback is temporally or spatially delayed (e.g. [82,95,110]). Likewise, reaching trajectories are affected when a visual target is suddenly displaced after arm movement onset (e.g. [81]), even if target displacement is not consciously perceived [22,40]. In addition to prior visual information regarding target location, non-visual signals operate in the control of arm movement during its execution. Eye–hand coordination in response to a target displacement is also a typical example of the need for fine tuning between sensory processing of visual and non-visual inputs and control of oculomotor and hand motor systems. When subjects are asked to point at the extrapolated final position of a moving visual spot temporarily occluded by a band of moving random dots, hand pointing errors are correlated with final gaze errors, which result from a combination of saccadic and ocular following responses [96]. These results are not consistent with an hypothesis in which a common signal could be used by the saccadic and hand motor systems, as in such case hand pointing errors would not have been expected to be influenced by the ocular following response elicited by the moving random dots. According to the authors, this further suggests that a gaze position signal, which could be derived from extra-retinal signals such as efference copy of the input to ocular motoneurons, provides a relevant target signal for the guidance of hand pointing movements, although other studies do not favor such a model of hand movement control based on gaze information (e.g. [85]). Nevertheless, such extra-retinal signals could be selectively used, depending on the availability and reliability of retinal stimulation. For instance, it has been shown that the matching between gaze and hand pointing directions is affected in total darkness [10]. According to these authors, the wrong calibration of motor command when vision is not allowed is an indication that extra-retinal signals could be better used in the presence of a concurrent retinal stimulation. Finally, in addition to gaze information, proprioception participates to a large extent in the control of limb movements, as illustrated for example by the observation that movements made by subjects without proprioceptive inputs (either following deafferentation, or when executing task in weightlessness) show strong unnoticed directional errors [41]. Proprioceptive information could also influence the control of smooth pursuit eye movements: for instance, ocular pursuit of a moving target is improved when target trajectory is manually monitored by subjects [104]. Taken together, these observations suggest that control of eye and hand movements depends on multiple sensory cues (retinal, extra-retinal, proprioceptive) which interact from initial planning of movement up to on-line motor control.

Current computational models of motor control (e.g. [51,52,127]) assume that sensory signals can indeed interact with feed-forward motor commands in an efficient manner. Accurate control of movements can be achieved by anticipating the sensory consequences of motor action through the use of an internal model of the upcoming interaction between our body and the environment [127]. Such predictive mechanisms could therefore ensure the maintenance of correct performance despite slow feedback loops and variable gain in sensory systems. Furthermore, the coupling between multiple contextual controllers could ensure flexible responses depending on the sensory context and past experience [128]. In this perspective, sensorimotor adaptation, as observed in judgements of hand location under optical deviation, could reflect the adaptation or learning of a new model of interaction between motor output and afferent sensory feedback.

We have already seen how touch can dominate in our perceptual interpretation of specific object properties. Proprioceptive information can be used, together with vision, in an optimal way, in judging hand location or when controlling limb movements. Taken together with the preceding descriptions of adaptive crossmodal interactions in perceptual analysis, all these studies converge toward the idea that human perception relies on a tight interplay between our sensory organs and active behavior, that are to a large extent dependent on adaptive and contextual crossmodal representation of space (see also [75,126]).

5. Discussion

5.1. A bayesian solution to crossmodal perception

To summarize, we have reviewed experiments showing that human observers tend to bind together sensory signals and assign them to a common external source, provided that they are perceived in close spatio–temporal relationships. In order to get the best sensory estimate and a coherent representation of space, a general underlying mechanism that could be used by the brain is to rely on the different modalities according to their relative precision, that is in a statistically optimal manner. In other words, perceptual responses could be the result of a weighted arithmetic mean between the sensory signals, the weights being related to the intrinsic reliability of each signal, which is qualitatively equivalent to a maximum-likelihood estimation. Obviously, each modality possesses its own domain of expertise, which could lead to fine interactions both in the temporal and spatial dimensions, depending on the task at hand and on the available sensory cues. Auditory and tactile cues can help to rapidly shift attentional resources toward a new region in visual space, and they can facilitate segmentation or grouping of sensory events when analyzing a visual scene. Object recognition and motor control also involve adaptive integration of reliable perceptual cues gathered through several sensory modalities and interacting with motor command.

Theoretical models based on maximum-likelihood estimation provide simulated responses in close agreement with behavioral performance in the case of auditory–visual localization [5,38], visuo–proprioceptive localization of hand [108], visuo-manual reaching [57], or visuo–haptic surface discrimination [28]. Interestingly, Battaglia et al. [5] have suggested that, in addition to minimum variance estimation, perceptual decision in auditory–visual localization could rely on prior information regarding the reliability of visual cues, turning this kind of model into a bayesian model of crossmodal integration. This boils down to raising the visual contribution for some aspects of spatial processing, e.g. bimodal localization, hence providing an elegant refinement of the former 'modality appropriateness hypothesis' [124]. How this prior information could be obtained remains to be determined, but it has recently been proposed that during a reaching task with varying amount of visual feedback, human observers make use, in an optimal probabilistic manner, of both estimated sensory uncertainty and prior information about visual reliability depending on past trials [57]. Therefore, like short-term sensorimotor learning, a similar principle could be considered on a larger time scale in the case of bimodal localization or scene analysis: prior information regarding the precision of visual cues in locating a target or the salience of auditory and visual signals could have been gained during development, although the implication of attentional processes remains to be further clarified. This optimal integration principle between relative contributions of each sensory modality is not limited to the above mentioned situations but also applies to the case of crossmodal exploration of a 3-D object, as specific perceptual attributes (e.g. size, volume, texture, resonance) are carried by dedicated modalities. To our knowledge no such computational model has been proposed.

5.2. Neurophysiological correlates of crossmodal processing

However, this raises the question of how the brain performs such an adaptive combination of simultaneous sensory inputs, yielding the best sensory estimate and an unified representation of space. Neurophysiological evidence indicates that the brain is able to perform such multimodal computations through the coordinated activity of distributed population of neurons (reviews in [1,15,101,102]). It has been proposed that the brain relies on the spatio–temporal correlation between the bimodal signals, that could be analyzed by dedicated populations of neurons responding to this kind of correlated inputs. Indeed, multisensory integration has been extensively studied at the level of single neurons (see, e.g. [68,69, 117,118]). A 'response enhancement' was observed in multimodal neurons located in the deep layers of the superior colliculus that are selectively activated by spatially and temporally coincident stimuli—either visual, auditory and/or somesthetic. This response enhancement is characterized by a stronger activity in response to bimodal stimuli than the sum of the responses to unimodal stimulation, especially for weak unimodal stimuli. In contrast, spatial discrepancy leads to a 'response depression'. Furthermore, this pattern of activity is not confined to the superior colliculus but has been found in cortical areas as well [42,83,116]. Therefore, such bimodal neurons are well-suited to detect coincident inputs. Since auditory and visual maps are organized in register, together with premotor maps, this could ensure a fast coordinated behavior in response to spatial events.

At the cortical level, parietal areas, among others, have been considered as potential sites of multimodal convergence (e.g. [1]). These areas seem to be involved in spatial attention processes and coordination of different reference frames (e.g. [1,19]; see also [20]), but also in crossmodal dynamic information processing [12], control of movement [42], object recognition and manipulation [44,72,86]. They are thus well-suited candidates for computing a multimodal representation of space and objects. Reciprocal connections between parietal areas and motor, visual, auditory and vestibular areas, as well as the cerebellum and frontal cortex, suggest that the parietal cortex could be involved in the coordination of different processing stages, in direct relation with intentional motor behavior [2,18,83].

All these neurophysiological correlates of crossmodal processing of sensory inputs indicate that the brain is able to maintain a coherent representation of space, although the precise mechanisms likely to carry on such optimal multisensory estimation are less well understood. As proposed by Ernst and Banks [28], interactions among modality-specific populations of neurons are sufficient to provide a unitary bimodal response, without the need to compute explicitly the weights given in inverse relation to the variance of the sensory estimators (see also [5]). Though this explanation relies on the implicit assumption that the brain performs the multiplication of two probability distributions regarding spatial features (e.g. depth cue), several electrophysiological studies have shown that spatial features, such as position and movement, could be encoded in the activity of selective populations of cortical neurons (e.g. [37]). The use of prior probability distributions in computing maximum-likelihood estimation through population codes has received much attention in recent computational models [21,79,80], which provide valuable explanations of the way perceptual system could perform crossmodal integration. Similar approaches that directly exploit the amount of noise carried by the sensory signals to derive a relevant sensory estimate have been proposed in the case of motor planning and sensorimotor learning (e.g. [46,57]).

5.3. Toward a 'rewired' vista of brain functional organization

Several crossmodal interactions that have been reviewed here mostly seem to affect early perceptual processing stages, suggesting that sensory processing in one modality can interact with correlated activity in another sensory modality. The observation that the activity of some brain areas usually considered as purely unimodal can be modulated by inputs from another modality [16,31,33,39,66,90] has recently led some authors to propose that crossmodal integration could be achieved through feedback projections from multimodal areas in addition to lateral interactions at the level of sensory-specific areas [15,16,24,26]. Higher cortical levels could therefore provide feedback connections to primary and secondary sensory areas, whose main activity most probably reflects a modality-specific coding of different spatial attributes in an internal reference frame (e.g. eye-centered, ear-centered, body-centered). Driver and Spence [26] have proposed that such feedback from multimodal levels could also influence unimodal processing of other spatial attributes through the lateral connections found in sensory areas. For example, orienting of attention to spatial visual location could enhance depth processing. Furthermore, these lateral modulations could be effective between areas devoted to the same modality or between areas of different modalities.

In her extensive review on anatomical studies of crossmodal processing, Calvert [15] has reported several structures—the superior temporal sulcus (STS) and inferior parietal sulcus (IPS), the posterior insula, the claustrum and the frontal cortex, that could be specifically activated in the perception of audiovisual speech, spatial localization and attention, the detection of bimodal temporal coincidence, crossmodal matching and learned crossmodal associations, respectively. Therefore, a wide network of potential sites is able to support crossmodal binding processes, depending on the task at hand and the sensory modalities involved in processing specific combination of perceptual features. Furthermore, the possibility that different cortico-cortical and cortico-subcortical networks could be involved in crossmodal processing is strengthened by some electrophysiological studies that have demonstrated significant interaction between associated somatosensory and auditory inputs [34], as well as between visual and auditory inputs [32,33,39], at early levels of the cortical processing hierarchy. These crossmodal interactions are expressed as modulations of brain responses in sensory-specific cortices, which depend on the kind of paired stimuli (redundant [32,39] vs. non-redundant [33]) and the kind of task ('passive' stimulation [34,35], detection [32], recognition [33,39]). These observations further challenge the hypothesis of a single route for crossmodal processing through backprojections from multimodal associative areas. In addition, several authors have highlighted the role of premotor cortex in multimodal representation of limb position [65], arm and face [43]. Crossmodal computation in the brain is thus not only at work in perceptual processing but could also be the basis of sensory guidance of movement, as previously discussed. In summary, the neural operations underlying crossmodal integration seem to involve several dedicated cortical and subcortical structures [15,32], although the way they could interact together and with modality-specific processing remains to be determined more precisely through further investigations.

All these considerations call into question the classical view of the functional organization of the brain, which has been for a long time considered as a set of distinct modules carrying dedicated sensory- and motor-related functions. Therefore, an important matter that needs to be addressed in future research is to better characterize the cortical pathways involved in sensory processing of redundant and complementary spatial cues, and their possible interactions with motor control and high-level cognitive functions, since crossmodal integration appears to be involved in cortical reorganization following sensory deprivation [6], social behavior [129] or perception of emotion [23].

6. Conclusion

In this review, we have described several perceptual crossmodal interactions, and neurophysiological correlates of multimodal computation in the brain. It appears that our perceptuo-motor abilities follow from crossmodal processing of our surroundings, and that a coherent and unified percept can emerge from crossmodal binding of different sensory cues with varying amounts of reliability. In the light of all these converging behavioral and neurophysiological data, and with the support of challenging theoretical framework that could lead to several new investigations, the functional organization of the brain is gradually becoming regarded as a potentially more dynamic network than previously thought, where some modality-specific areas will better be described as modality-dominant areas. We would like to add that the coupling of all anatomical subsystems can perhaps be understood as one integrated dynamical network in which adaptive crossmodal binding of spatial and temporal features is achieved through recurrent activity between multimodal and sensory-specific areas, that not only modulates but also helps to define perceptual organization and motor control.

Acknowledgements

This work has received financial support from France Télécom R&D.

References

[1] R.A. Andersen, L.H. Snyder, D.C. Bradley, J. Xing, Multimodal representation of space in the posterior parietal cortex, Ann. Rev. Neurosci. 20 (1997) 303–330.

[2] R.A. Andersen, L.H. Snyder, A.P. Batista, C.A. Buneo, Y.E. Cohen, Posterior parietal areas specialized for eye movements (LIP) and reach (PRR) using a common coordinate frame, Novartis Foundation Symp. 218 (1998) 109–122.

[3] D. Alais, D. Burr, The flash-lag effect occurs in audition and cross-modally, Curr. Biol. 13 (2003) 59–63.

[4] R.B. Banati, G.W. Goerres, C. Tjoa, J.P. Aggleton, P. Grasby, The functional anatomy of visual–tactile integration in man: a study using positron emission tomography, Neuropsychologia 38 (2000) 115–124.

[5] P.W. Battaglia, R.A. Jacobs, R.N. Aslin, Bayesian integration of visual and auditory signals for spatial localization, J. Opt. Soc. 20 (2003) 1391–1397.

[6] D. Bavelier, H.J. Neville, Cross-modal plasticity: where and how?, Nature Rev. 3 (2002) 443–452.

[7] P. Bertelson, M. Radeau, Crossmodal biases and perceptual fusion with auditory–visual spatial discordance, Percept. Psychophys. 29 (1981) 578–584.

[8] P. Bertelson, Ventriloquism: a case of cross-modal perceptual grouping, in: G. Aschersleben, T. Bachmann, J. Müssler (Eds.), Cognitive Contributions to the Perception of Spatial and Temporal Events, Elsevier Science, Amsterdam, 1999, pp. 347–362.

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

[10] J. Blouin, N. Amade, J.-L. Vercher, N. Teasdale, G.M. Gauthier, Visual signals contribute to the coding of gaze direction, Exp. Brain Res. 144 (2002) 281–292.

[11] A.S. Bregman, Auditory Scene Analysis: The Perceptual Organization of Sound, MIT Press, Cambridge, 1990.

[12] F. Bremmer, A. Schlack, J.-R. Duhamel, W. Graf, G.R. Fink, Polymodal motion processing in posterior parietal and premotor cortex: a human fMRI study strongly implies equivalencies between humans and monkeys, Neuron 29 (2001) 287–296.

[13] C.M. Butter, H.A. Butchel, R. Santucci, Spatial attentional shifts: further evidence for the role of polysensory mechanisms using visual and tactile stimuli, Neuropsychologia 27 (1989) 1231–1240.

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

[15] G.A. Calvert, Crossmodal processing in the human brain: insights from functional neuroimaging studies, Cereb. Cortex 11 (2001) 1110–1123.

[16] G.A. Calvert, M. Brammer, E. Bullmore, R. Campbell, S.D. Iversen, A.S. David, Response amplification in sensory-specific cortices during crossmodal binding, NeuroReport 10 (1999) 2619–2623.

[17] G.A. Calvert, P.C. Hansen, S.D. Iversen, M.J. Brammer, Detection of multisensory integration sites by application of electrophysiological criteria to the BOLD response, NeuroImage 14 (2001) 427–438.

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

[19] C.L. Colby, M.E. Goldberg, Space and attention in parietal cortex, Ann. Rev. Neurosci. 22 (1999) 319–349.

[20] J.C. Culham, N.G. Kanwisher, Neuroimaging of cognitive functions in human parietal cortex, Curr. Opt. Neurobiol. 11 (2001) 157–163.

[21] S. Deneve, P.E. Latham, A. Pouget, Efficient computation and cue integration with noisy population codes, Nat. Neurosci. 4 (2001) 826–831.

[22] M. Desmurget, H. Gréa, J.S. Grethe, C. Prablanc, G.E. Alexander, S.T. Grafton, Functional anatomy of nonvisual feedback loops during reaching: a positron emission tomography study, J. Neurosci. 21 (2001) 2919–2928.

[23] R.J. Dolan, J.S. Morris, B. de Gelder, Crossmodal binding of fear in voice and face, Proc. Nat. Acad. Sci. USA 98 (2001) 10006–10010.

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

[25] J. Driver, C. Spence, Crossmodal links in spatial attention, Proc. Roy. Soc. Lond., Ser. B, Biol. Sci 353 (1998) 1–13.

[26] J. Driver, C. Spence, Multisensory perception: beyond modularity and convergence, Curr. Biol. 10 (2000) R731–R735.

[27] M. Eimer, D. Cockburn, B. Smedley, J. Driver, Cross-modal links in endogenous spatial attention are mediated by common external locations: evidence from event-related brain potentials, Exp. Brain Res. 139 (2001) 398–411.

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

[29] M.O. Ernst, M.S. Banks, H.H. Bülthoff, Touch can change visual slant perception, Nat. Neurosci. 3 (2000) 69–73.

[30] P. Ewert, A study of the effect of inverted retinal stimulation upon spatially coordinated behavior, Genetic Psychol. Monog. 7 (1930) 242–244.

[31] A. Falchier, L. Renaud, P. Barone, H. Kennedy, Extensive projections from the primary auditory cortex and polysensory area STP to peripheral area V1 in the macaque, Soc. Neurosci. Abs. 27 (2001) 511–521.

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

[33] A. Fort, C. Delpuech, J. Pernier, M.-H. Giard, Early auditory–visual interactions in human cortex during nonredundant target identification, Cogn. Brain Res. 14 (2002) 20–30.

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

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

[36] W. Gaver, What in the world do we hear? An ecological approach to auditory event perception, Ecol. Psychol. 5 (1993) 1–29.

[37] A.P. Georgopoulos, A.B. Schwartz, R.E. Kettner, Neuronal population coding of movement direction, Science 233 (1986) 1416–1419.

[38] Z. Gharamani, D.M. Wolpert, M.I. Jordan, Computational models of sensorimotor integration, in: P.G. Morasso, V. Sanguineti (Eds.), Self-Organization, Computational Maps, and Motor Control, North-Holland, Amsterdam, 1997.

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

[40] M.A. Goodale, D. Pélisson, C. Prablanc, Large adjustments in visually guided reaching do not depend on vision of the hand or perception of target displacement, Nature 320 (1986) 748–750.

[41] J. Gordon, M.F. Ghilardi, C. Ghez, Impairments of reaching movements in patients without proprioception. I. Spatial errors, J. Neurophysiol. 73 (1995) 347–360.

[42] M.S.A. Graziano, C.G. Gross, Spatial maps for the control of movement, Curr. Opt. Neurobiol. 8 (1998) 195–201.

[43] M.S.A. Graziano, S. Gandhi, Location of the polysensory zone in the precentral gyrus of anesthetized monkey, Exp. Brain Res. 135 (2000) 259–266.

[44] C. Grefkes, P.H. Weiss, K. Zilles, G.R. Fink, Crossmodal processing of object features in human anterior intraparietal cortex: an fMRI study implies equivalencies between human and monkeys, Neuron 35 (2002) 173–184.

[45] N. Hadjikhani, P.E. Roland, Cross-modal transfer of information between the tactile and the visual representations in the human brain: a positron emission tomographic study, J. Neurosci. 18 (1998) 1072–1084.

[46] C.M. Harris, D.M. Wolpert, Signal-dependent noise determines motor planning, Nature 394 (1998) 780–784.

[47] J. Hay, H. Pick, K. Ikeda, Visual capture produces by prism spectacles, Psychon. Sci. 2 (1965) 215–216.

[48] H. Hecht, S. Vogt, W. Prinz, Motor learning enhances perceptual judgment: a case for action–perception transfer, Psychol. Res. 65 (2001) 3–14.

[49] M.A. Heller, Visual and tactual texture perception: intersensory cooperation, Percept. Psychophys. 31 (1982) 339–344.

[50] I. Howard, W. Templeton, Human Spatial Orientation, Wiley, New York, 1966.

[51] M.I. Jordan, D.M. Wolpert, Computational motor control, in: M.S. Gazzaniga (Ed.), The Cognitive Neurosciences, MIT Press, Cambridge, 1999.

[52] M. Kawato, Internal models for motor control and trajectory planning, Curr. Opt. Neurobiol. 9 (1999) 718–727.

[53] M. Kinsbourne, The mechanisms of hemispheric control of the lateral gradient of attention, in: P.M.A. Rabbitt, S. Dronic (Eds.), Attention and Performance V, Academic Press, London, 1975, pp. 81–93.

[54] N. Kitagawa, S. Ichihara, Hearing visual motion in depth, Nature 416 (2002) 172–174.

[55] R. Klatzky, S.J. Lederman, D.E. Matula, Haptic exploration in the presence of vision, J. Exp. Psychol. Hum. Percep. Perf. 19 (1993) 726–743.

[56] R. Klatzky, S.J. Lederman, C. Reed, There's more to touch than meet the eye: the salience of object attributes for haptics with and without vision, J. Exp. Psychol. Gen. 116 (1987) 356–369.

[57] K.P. Körding, D. Wolpert, Bayesian integration in sensorimotor learning, Nature 427 (2004) 244–247.

[58] S.J. Lederman, S.G. Abbott, Texture perception: studies of intersensory organization using a discrepancy paradigm, and visual versus tactual psychophysics, J. Exp. Psychol. Hum. Percep. Perf. 7 (1981) 902–915.

[59] S.J. Lederman, R. Klatzky, Hand movements: a window into haptic object recognition, Cogn. Psychol. 19 (1987) 342–368.

[60] S.J. Lederman, R. Klatzky, Extracting object properties through haptic exploration, Acta Psychol. 84 (1993) 29–40.

[61] S.J. Lederman, R.L. Klatzky, T. Morgan, C. Hamilton, Integrating multimodal information about surface texture via a probe: relative contributions of haptic and touch-produced sound sources, in: Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems 2002, 2002, pp. 97–104.

[62] S.J. Lederman, G. Thorne, B. Jones, Perception of texture by vision and touch: multidimensionality and intersensory integration, J. Exp. Psychol. Hum. Percep. Perf. 12 (1986) 169–180.

[63] D.J. Lewkowicz, Heterogeneity and heterochrony in the development of intersensory perception, Cogn. Brain Res. 14 (2002) 41–63.

[64] D.J. Lewkowicz, Perception of serial order in infants, Develop. Sci. 7 (2004) 175–184.

[65] D.M. Lloyd, D.I. Shore, C. Spence, G.A. Calvert, Multisensory representation of limb position in human premotor cortex, Nat. Neurosci. Dec 16 (2002) (online).

[66] E. Macaluso, C. Frith, J. Driver, Modulation of human visual cortex by cross-modal spatial attention, Science 289 (2000) 1206–1208.

[67] S. Mateeff, J. Hohnsbein, T. Noack, Dynamic visual capture: apparent auditory motion induced by a moving visual target, Perception 14 (1985) 721–727.

[68] M.A. Meredith, B.E. Stein, Spatial determinants of multisensory integration in cat superior colliculus, J. Neurophysiol. 75 (1996) 1843–1857.

[69] M.A. Meredith, J.W. Nemitz, B.E. Stein, Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors, J. Neurosci. 10 (1987) 3215–3229.

[70] W. Metzger, Beobachtungen über phaenomenale Identität, Psychologische Forschung 19 (1934) 1–60.

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

[72] A. Murata, V. Gallese, G. Luppino, M. Kaseda, H. Sakata, Selectivity for the shape, size and orientation of objects for grasping in neurons of monkey parietal area AIP, J. Neurophysiol. 83 (2000) 2580–2601.

[73] F.N. Newell, M.O. Ernst, B.S. Tjan, H. Bulthöff, View-point dependence in visual and haptic obejct recognition, Psychol. Sci. 12 (2001) 37–42.

[74] J. Paillard, M. Bouchon, Active and passive movements in the calibration of position sense, in: S.J. Freeman (Ed.), The Neuropsychology of Spatially Oriented Behavior, Dorsey Press, Homewood, 1968.

[75] J. Paillard, Brain and Space, Oxford University Press, Oxford, 1991.

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

[77] H. Pick, D. Warren, J. Hay, Sensory conflict in judgemnets of spatial direction, Percept. Psychophys. 6 (1969) 203–205.

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

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

[80] A. Pouget, P. Dayan, R.S. Zemel, Computation and inference with population codes, Ann. Rev. Neurosci. 26 (2003) 381–410.

[81] C. Prablanc, D. Pélisson, M.A. Goodale, Visual control of reaching movements without vision of the limb. I. Role of extraretinal feedback of target position in guiding the hand, Exp. Brain Res. 62 (1986) 293–302.

[82] J. Pratt, R.A. Abrams, Practice and component submovements: the role of feedback in rapid aimed limb movements, J. Motor Behav. 28 (1996) 149–156.

[83] G. Rizzolatti, L. Fogassi, V. Gallese, Parietal cortex: from sight to action, Curr. Opt. Neurobiol. 7 (1997) 562–567.

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

[85] U. Sailer, T. Eggert, J. Ditterich, A. Straube, Spatial and temporal aspects of eye–hand coordination across different tasks, Exp. Brain Res. 134 (2000) 163–173.

[86] H. Sakata, M. Taira, Parietal control of hand action, Curr. Opt. Neurobiol. 4 (1994) 847–856.

[87] C. Scheier, D.J. Lewkowicz, S. Shimojo, Sound induces perceptual reorganization of an ambiguous motion display in human infants, Develop. Sci. 6 (2003) 233–244.

[88] B.J. Scholl, K. Nakayama, Causal capture: contextual effects on the perception of collision events, Psychol. Sci. 13 (2002) 493–498.

[89] R. Sekuler, A.B. Sekuler, R. Lau, Sound alters visual motion perception, Nature 385 (1997) 388.

[90] L. Shams, Y. Kamitani, S. Thompson, S. Shimojo, Sound alters visual evoked potentials in humans, NeuroReport 12 (2001) 3849–3852.

[91] L. Shams, Y. Kaminati, S. Shimojo, Illusions: what you see is what you ear, Nature 408 (2000) 788.

[92] L. Shams, Y. Kamitani, S. Shimojo, Visual illusion induced by sound, Cogn. Brain Res. 14 (2002) 147–152.

[93] S. Shimojo, L. Shams, Sensory modalities are not separate modalities: plasticity and interactions, Curr. Opt. Neurobiol. 11 (2001) 505–509.

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

[95] W.M. Smith, Feedback: real-time one's own tracking behavior, Science 176 (1972) 939–940.

[96] J.F. Soechting, K.C. Engel, M. Flanders, The duncker illusion and eye–hand coordination, J. Neurophysiol. 85 (2001) 843–854.

[97] S. Soto-Faraco, A. Kingstone, C. Spence, Multisensory contributions to the perception of motion, Neuropsychologia 41 (2003) 1847–1862.

[98] C. Spence, J. Driver, Audiovisual links in endogenous covert spatial attention, J. Exp. Psychol. Hum. Percep. Perf. 22 (1996) 1005–1030.

[99] C. Spence, A. Kingstone, D.I. Shore, M.S. Gazzaniga, Representation of visuotactile space in the split brain, Psychol. Sci. 12 (2001) 90–93.

[100] C. Spence, F. Pavani, J. Driver, Crossmodal links between vision and touch in covert endogenous spatial attention, J. Exp. Psychol. Hum. Percep. Perf. 26 (2000) 1298–1319.

[101] B.E. Stein, Neural mechanisms for synthesizing sensory information and producing adaptive behaviors, Exp. Brain Res. 123 (1998) 124–135.

[102] B.E. Stein, M.A. Meredith, The Merging of the Senses, MIT Press, Cambridge, 1993.

[103] B.E. Stein, N. London, L.K. Wilkinson, D.D. Price, Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis, J. Cogn. Neurosci. 8 (1996) 497–506.

[104] M.J. Steinbach, Eye tracking of self-moved targets: the role of efference, J. Experimental Psychol. 82 (1969) 366–376.

[105] G.M. Stratton, Upright vision and the retinal image, Psychol. Rev. 4 (1897) 182–187.

[106] R.J. van Beers, A.C. Sittig, J.J. Denier van der Gon, How humans combine simultaneous proprioceptive and visual position information, Exp. Brain Res. 111 (1996) 253–261.

[107] R.J. van Beers, A.C. Sittig, J.J. Denier van der Gon, The precision of proprioceptive position sense, Exp. Brain Res. 122 (1998) 367–377.

[108] R.J. van Beers, A.C. Sittig, J.J. Denier van der Gon, Integration of proprioceptive and visual position-information: an experimentally supported model, J. Neurophysiol. 81 (1999) 1355–1364.

[109] R.J. van Beers, D.M. Wolpert, P. Haggard, When feeling is more important than seeing in sensorimotor adaptation, Curr. Biol. 12 (2002) 834–837.

[110] J.L. Vercher, G.M. Gauthier, Oculo-manual coordination control: ocular and manual tracking of visual targets with delayed visual feedback of the hand motion, Exp. Brain Res. 90 (1992) 599–609.

[111] J. Vroomen, B. de Gelder, Temporal ventriloquism: sound modulates the flash-lag effect, J. Exp. Psychol. Hum. Percep. Perf. 30 (2004) 513–518.

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

[113] J. Vroomen, B. de Gelder, Visual motion influences the contingent auditory motion aftereffect, Psychol. Sci. 14 (2003) 357–361.

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

[115] J.T. Walker, K.J. Scott, Auditory–visual conflicts in the perceived duration of lights, tones and gaps, J. Exp. Psychol. Hum. Percep. Perf. 7 (1981) 1327–1339.

[116] M.T. Wallace, M.A. Meredith, B.E. Stein, Integration of multiple sensory modalities in cat cortex, Exp. Brain Res. 91 (1992) 484–488.

[117] M.T. Wallace, M.A. Meredith, B.E. Stein, Converging influences from visual, auditory and somatosensory cortices onto output neurons of the superior colliculus, J. Neurophysiol. 69 (1993) 1797–1809.

[118] M.T. Wallace, L.K. Wilkinson, B.E. Stein, Representation and integration of multiple sensory inputs in primate superior colliculus, J. Neurophysiol. 76 (1996) 1246–1266.

[119] J.P. Wann, S.F. Ibrahim, Does limb proprioception drift? Exp. Brain Res. 91 (1992) 162–166.

[120] D.H. Warren, W. Cleaves, Visual-proprioceptive interaction under large amounts of conflict, J. Exp. Psychol. 90 (1971) 206–214.

[121] D.H. Warren, H. Pick, Intermodality relations in blind and sighted people, Percept. Psychophys. 8 (1970) 430–432.

[122] K. Watanabe, S. Shimojo, When sound affects vision: effects of auditory grouping on visual motion perception, Psychol. Sci. 12 (2001) 109–116.

[123] R.B. Welch, Perceptual Modifications. Adapting to Altered Sensory Environments, Academic Press, New York, 1978.

[124] R.B. Welch, D.H. Warren, Immediate perceptual response to intersensory discrepancy, Psychol. Bull. 88 (1980) 638–667.

[125] R.B. Welch, D.H. Warren, Intersensory interactions, in: K.R. Boff, L. Kaufman, J.P. Thomas (Eds.), Handbook of Perception and Human Performance, vol. 1, Wiley, New York, 1986, pp. 25.1–25.36.

[126] M.A. Wing, P. Haggard, J.R. Flanagan, The Neurophysiology and Psychology of Hand Movements, Academic Press, San Diego, 1996.

[127] D.M. Wolpert, J.R. Flanagan, Motor prediction, Curr. Biol. 11 (2001) R729–R732.

[128] D.M. Wolpert, M. Kawato, Multiple paired forward and inverse models for motor control, Neural Netw. 11 (1998) 1317–1329.

[129] D.M. Wolpert, K. Doya, M. Kawato, A unifying computational framework for motor control and social interaction, Phil. Trans. Royal Soc. 358 (2003) 593–602.