Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Physiology - Paris 98 (2004) 1–18
Journal of Physiology Paris
www.elsevier.com/locate/jphysparis

Foreword

Brain encoding and representation of 3D-space using different senses, in different species

What are the strategies used by flies, fish, birds or mammals to navigate in their immediate environment? How do the nervous systems of living beings extract and analyze information from the surrounding 3D-space in order to build a coherent representation of the external world, which will allow or guide sensory–motor interaction within a spatial reference frame? Answers to these general questions involve the merging of inputs from various fields of expertise, ranging from neuroethology and neurophysiology to modelling and robotics. A comparative approach to these problems through evolution and different senses is crucial, not only to assess the specificity of strategies modelled by external and internal constraints, but also to extract general operational and computational rules characterizing the same perceptual integration process in different species. The meeting that we organized in Treilles in May 2002, entitled "*Brain representation of 3D space using different senses, in different species*", was motivated by our desire to compare the viewpoints of specialists in different fields of research, all concerned with the study of the representation of 3D space. The following text summarizes the major issues raised during the five days spent at the Fondation des Treilles.

1. Computational strategies

This first section reviews the neural basis for the encoding of various specific physical attributes that can be extracted in space and time in order to reconstruct a unified 3D-representation of the proximal environment. It will be shown that, in addition to vision and audition, a variety of senses guide innovative biological strategies for estimating distances (position-in-depth) and movement of objects. Performing these tasks, even in the dark or in aquatic environments, is of primary importance in navigation for locating prey or sexual partners, and for avoiding predators.

1.1. Motion detection

The first example deals with predictive coding using a dominant sense, vision. The presentation made by Fabrizio Gabbiani concerned giant motion detector neurons in the locust visual system specialized in the detection of approaching objects ("looming") and whose spike activation triggers a very effective avoidance reflex. The nervous system of insects is a particularly attractive experimental model because of the presence of giant identified neurons, whose morphology and function are invariant across members of the same species, and whose size is such that these cells are easily intra-cellularly recorded and stained. Furthermore, these cells are clustered within paucineuronal networks, the dynamics and functional organization of which can be quantitatively modelled. Another advantage of the insect model is its potential application to the field of robotics, due to the relative simplicity of the neural mechanisms involved, which allows for the neuromimetic implementation of these elementary mechanisms in autonomous robots, an issue further illustrated by Nicholas Franceschini.

The locust visual system contains two (one on each side) identified neurons (the so-called LGMDs or "Lobula Giant Movement Detectors"), that collect visual information simultaneously sampled from the multiple facets forming the composite eye. These sensory detectors, which respond almost as efficiently to a local (one facet only) or global (composite eye) stimulus, constitute the optomotor interface controlling the insect flight: they respond vigorously to visual objects approaching on a collision course and produce an activity burst which triggers in an all-or-none fashion an escape behaviour. Processing of visual information by LGMD neurons seems to rely on two computational principles, namely (1) the multiplication of two independent input signals and (2) the invariance of the time-course of the output spiking pattern for a broad

0928-4257/$ - see front matter
doi:10.1016/j.jphysparis.2004.09.001

range of contextual parameters, that are shared by numerous other sensory systems. The group of Gilles Laurent has shown that such a looming-sensitive neuron indeed computes an angular size threshold for the approaching object: the stimulus-locked firing of the LGMD neuron peaks precisely at a specific delay which encodes for the kinematics of object approach. This delay, which corresponds to the time value at which an object will reach a fixed angular size on the retina (threshold detection), thus encodes the likely time of collision [1].

The most remarkable aspect of the sensory response is that the peak firing time is independent of particular features of the stimulus (luminance, contrast, shape and texture) as well as of the general context associated with the stimulus presentation (angle of approach of the stimulus, body orientation). This invariance cannot be accounted for by summation of excitation and inhibition operating at an earlier stage of integration in the visual pathway and is likely to be implemented by combining different sources of inputs within the dendritic tree of the LGMD itself. It requires several non-linearities, the substrates of which are partly identified. The characteristic time-course of LGMD firing rate patterns in response to an approaching object has been quantitatively modelled by a multiplicative operation combining two distinct input features of the target, namely its angular velocity and its angular size. LGMDs indeed collect on separate dendritic fans three independent sources of inputs, one corresponding to a motion-dependent feedforward excitation and the two others to feedforward inhibition signaling size-dependency on the one hand and lateral pre-synaptic inhibition on the other hand. This last component plays a role only in the initial phase of the looming response and will be ignored here. LGMD might thus operate as a biophysical device multiplying post-synaptically the two feedforward inputs. The response ceases when the feedforward inhibition, which covaries exponentially with angular size reduction, overrides excitation at the latest phase just before collision. The biophysical substrate of this operation appears rather complex in its implementation since the synaptic inputs are first transduced into membrane potential changes expressed on a logarithmic scale. Multiplicative operations between the post-synaptic effects of each source of synaptic inputs are thus transformed into simple subtractive mechanisms at the membrane potential level before an inverse exponentiation associated with the frequency/voltage input characteristic of the cell and the sodium channel spike generation restores the conversion into the appropriate spike rate change.

Following Gibson's ethological approach to visual perception, Nicolas Franceschini presented biology-derived (particularly insect-inspired) robots, whose computational architecture was meant to test both the soundness and the robustness of biological principles and implement these life-like principles to guide the navigation of autonomous machines and vehicles. This approach has been successfully used to show how optical flow produced by the retinal slip of contrasted objects in the environment during body and eye-movements of the animal participates in the construction of an internal 3D representation of space in flies during navigation: It can indeed be used by visually guided terrestrial and aerial robots to detect, locate and avoid environmental features, even when limiting the artificial central processing unit to a parsimonious "brain representation".

Insects, flies as much as locusts, rely on the continuous processing of optical flow to orient themselves during their flight in complex environments. In the blowfly optical flow sensors or motion detecting neurons are located at a rather integrated stage of processing, in the lobula plate, a central relay two synaptic steps away from the retina (3rd visual ganglion): for instance, the H1 neuron is one of the 50 giant neurons collecting local information on motion along a given retinal motion axis. It can be activated by full field optical flow as well as by the successive lasergun activation, in a specific order and a specific delay, of two neighbouring photoreceptors (R1 and R6) located within the same ommatidia cartridge (elementary facet and light guide replicated across the composite eye of the insect). By cross-correlating the inputs fed at different time delays by neighbouring photoreceptors, the Elementary Motion Detector (EMD) computes the direction of motion. The functional diagram of the operation realized by each EMD has been mimicked and implemented by analog hardware on an electronic board (miniaturized in size and weight) and replicated in the form of distributed sensor arrays feeding an autonomous robot. This "robofly" is designed so as to self-orient in a 2D space on the basis of the translational component of optical flow only (the rotational flow component is ignored) and receives a continuous information update fed by a crown of regularly spaced biologically-based EMDs. On its way to a target, under the influence of a drive signal mimicking some kind of elementary phototaxic behaviour, the robofly continuously processes the optical flow during sequences of translational trajectories, and modifies its motion axis each time the perceived speed of an obstacle image passes a chosen threshold. The result is a jerky (fly-like) non-optimal exploratory trajectory where the robot succeeds in avoiding obstacles even at relatively high speed (50 cm/s).

The principle of this motion detector has been recently combined with an active scanning device supposed to mimic the fast saccadic tremor that is associated with ocular fixation in vertebrates and invertebrates. As discovered by Nicolas Franceschini, a specialized pair of extraocular muscles is located behind the ocular lens of the fly and is constantly engaged in

a periodic microscanning during natural vision. This nystagmus-like behaviour, demonstrated by chronic EMG recordings in a fly suspended from a rod, results in a spatial resolution performance 40 times higher than that predicted from the photoreceptor array geometry. This process, unique to the fly, performs a computation on the visual image which is functionally equivalent to the performance of hyperacuity long ago reported in human. The robofly has consequently been equipped with a dedicated OSCAR (Optical Scanner for Autonomous Robot) technology, allowing it to track a nearby target even at angular speeds exceeding 30 deg/s (which is the maximal tracking speed of the human eye during smooth pursuit) [2].

When mammals observe an object moving in space, or navigate in their proximal environment, they continually anchor their gaze fixation onto the object of interest in order to keep its representation within the foveal-like region of the retina. The initial pattern of oculomotor response, before it becomes enslaved by sensory–motor loops to the trajectory of the target, has been proposed to reflect a feedforward estimate of object motion. Guillaume Masson summarized a series of investigations in primates, showing that an instantaneous read-out of motion perception can indeed be derived from the pattern of eye-movements. When a single step displacement (between two fixed positions) is applied to a large random dot pattern, both an apparent motion percept (Phi motion, where the subject consciously reports a smooth shift of the dot pattern) and a short-delay pre-saccadic response (at around 75 ms latency in humans vs. 140 ms for classical saccades) are elicited. The kinetics of this early oculomotor activity are more reminiscent of an eye movement pursuit than of a saccade, in the sense that the induced oculomotor response is of gradual amplitude and its direction covaries with that experienced by the subject. If the contrast of the dot pattern is reversed concomitantly with the single step displacement, both the direction of the perceived apparent motion and the direction of the early-induced eye-movement are reversed [3]. Electrophysiological correlates of these effects have been reported at the level of V1 and in MT neurons. Moreover, peak oculomotor gain is observed for a step size of half the dot size and is reduced for step sizes larger than the dot size. Those features reveal the involvement of low-level motion detectors in the initiation of pursuit-like eye movements, faithfully reflecting the perceptual experience.

Because the elementary motion detection performed at the level of single cells such as in V1 cannot directly account for the disambiguation of 2D movement integration of a contrast edge or feature surface (the classical ''aperture problem''), further higher level composition processes are required. A way to disambiguate the motion ambiguities is to integrate simultaneously several elementary motion signals across the visual field. A plaid pattern is a composite stimulus well suited to study this problem because its apparent motion is dependent on sets of constraints that can be defined on the basis of the individual motion components relative to each of the individual gratings which constitute the plaid [4]. For the specific case of unikinetic plaids (in which one of the two gratings is static while the other one is drifting at constant velocity), the early onset phase of the eye-movement pursuit response is first initiated in the direction of the moving grating component and deviates 20 ms later towards the motion direction corresponding to that perceived for the global plaid pattern. This two-phase process may involve the sequential recruitment of the magno- and parvo retino-thalamo-cortical pathways and is taken as an indication that both linear and non-linear processes are involved, with different dynamics. Using ''barber-pole'' stimuli, where a moving grating is seen behind an elongated oriented aperture, the two components of the pursuit eye-movement have been shown to also have different temporal dynamics, correlated with the time-course of spike responses of MT neurons to such stimuli [5]. MT neurons are tuned to both direction and inter-ocular disparity, and the combinations of both properties could underlie motion segmentation in 3D-space.

1.2. Spatial localization

Localization in space, and more specifically estimation of distance in depth, is achieved in different species, ranging through insects, batrachia, fishes, birds, bats and higher mammals, using different sensory systems, on the basis of electrical, somatosensory, auditory and visual cues. Various computational solutions were illustrated during the symposium, which allow the reconstruction of 3D-space after an initial transduction step through one or several receptor sheets, corresponding to the same or different sensory modalities. In the case of a single receptor array, the 3D-reconstruction is inferred by sophisticated local gradient analysis from the extrapolation of ''projective'' properties of the environment on the 2D sensory map. In the case of simultaneous mapping of space through different sensory filters or spatial view-points, the neural system derives, from comparison of the multiple registers, the direct computation of stereo-cues which are processed by specialized inter-map phase detectors tuned either in the spatial or temporal domains. We detail below unusual methods to infer direction and distance in a reconstructed 3D-space from a single sensory 2D image of the environment.

Sound propagating in a liquid creates pressure waves with complex local patterns (devoid of directional bias) and distant wavefronts (which give a directional cue). The inner ear in fish is capable of detecting vibrations with amplitudes of a few Angströms in water, and

certain species are sensitive enough to detect the phase relationship between the pressure wave due to its own movement and the acoustic motion of particles in suspension in the otoliths of the vestibular system.

The fish lateral line, studied by Horst Bleckmann, is a specialized sensory system used to detect hydrodynamic stimulation along the body axis. Its efficiency, superior to that of classical aerial audition, comes from the facts that sound travels faster in water than in air, and that binaural delay and intensity cues are lessened. The lateral line is composed of superficial neuromasts located on the skin and deeper canal neuromasts recessed in fluid-filled subepidermal pores. Superficial neuromasts detect the velocity of water flow resulting from the fish's own movements and/or from the surrounding water itself, whereas canal neuromasts, because of the mechanical filter properties of lateral-line canals, are more sensitive to water flow acceleration. The orienting behaviour of the fish (body turn) towards its prey depends on the rostro-caudal position of the first neuromast to be activated, and the fish swimming time is correlated to specific frequency bands in the spectrum of the detected hydrodynamic signal. Interestingly, the frequency dependence of this biological detection system shares similarities with man-made algorithms developed independently in the field of oceanography.

In the absence of light, the mottled sculpin is still able to orient itself and to approach a vibrating stimulus using only the lateral line. It does this by swimming along iso-pressure domains in the environment. Rostral neuromasts are directly involved in the orienting response and a lesion of those neuromasts leads to a disturbance of the orientation skill (while a lesion in a more posterior part of the body does not). On the one hand, the azimuth of the target is computed using the information contained in the pressure gradient along the fish's body. On the other hand, the spread of the excitation along the lateral line carries information about the target's distance. In contrast with the visual system, where the size of an object's retinal image decreases as a function of its distance, the spatial spread of the excited zone in the lateral line increases with the target's distance. Lateral line excitation spread contains enough information to discriminate between a small target far away and a large target nearby.

For stationary objects (which are not hydrodynamic stimuli), the blind cave fish, *Astyanax mexicanus*, is able to use self-induced water motion to identify and localize very close objects which remain stationary. Both water velocity and pressure gradient permit the fish to estimate the object's size and distance. Some fishes can even use lateral line information to construct an inner map of their environment. The depth-related computation performed by the lateral line system is in fact rather similar to that in electric fish using the electrical sense, which is described in the next section.

Some fishes in search of food are attracted by wave stimuli evoked at the water surface by the landing of insects. These fishes use the time interval between the arrival of successive wavelets to determine the target angular direction. Distance discrimination can be performed using different wave cues (frequency spectrum, curvature, frequency modulation) that have greater or lesser weight according to their reliability. Recent evidence has shown the involvement of the lateral line system in the ability of fish to track hydrodynamic wakes. When a catfish (*Siluris glanis*) navigates in search of potential prey, it is able to detect 3D vortex trails persisting for up to one minute. It can successfully capture passing fish at distances up to 55 times prey-body length, even if the initiation of the tracking search was initiated more than 10 s after the prey's passage. The remarkable feature of the orienting chase behaviour is that the predator calculates the direction taken by the prey at the time when the predator intercepts the wake [6]. Similar predictive behaviour has been reported in seals, a species that can be trained to track submarines using the vibrissae even in the complete absence (or suppression) of visual and auditory cues.

Dancing in the dark requires other senses and computations. Weakly electric fishes are nocturnal predators, searching for small insect larvae in tropical freshwater rivers and streams. They produce weak electric currents in the water. Their electric organ emits brief electrical discharges (EODs), implicated in both intra-specific communication and object detection. Each discharge results in a 3D electric field around the fish, which is altered by the presence of objects of conductivity different from that of the water. These field distortions are detected by three types of epidermal electroreceptors situated on the whole surface of the body. The ability to use a self-generated electrical field in order to detect objects is called "active" electrolocation, in contrast with "passive" electrolocation which consists of receiving and detecting electrical fields emitted by other electrically active sources.

The electric fish orients itself in complete darkness by analyzing the electrical "shadow" of surrounding objects that the self-generated field "projects" onto its epidermal body envelope. The electosensory image of an object is defined as the area on the fish skin where the amplitude and/or the phase and polarity of the locally induced electrical field are modified because of the object's presence. The electroreceptor distribution is not uniform, and two high-density zones of electroreceptors play a functional role similar to that of two specialized "foveas": The more caudal one is linked with navigation and far field analysis, and the other one, positioned near the nose end, is involved in higher spatial acuity and object identification. The bending of the electric fish body is also an active process, slightly similar to "foveation" of gaze in mammals during visual exploration. In spite

of these heterogeneities, the position of the centre of gravity of the electrically induced image on the body tells the fish where the object is located. The polarity of the amplitude change in the local electrical field informs the fish about the object's impedance relative to that of the water. Phase shift and/or EOD waveform distortions convey information about the capacitive properties of the object. Other aspects, such as object size, shape, or distance, involve complex central neural processes of computation, based on the extraction and combination of several physical attributes of the electrical image.

In spite of the ambiguity raised by the combined effects of size, resistivity and distance in object identification, the weakly electric fish *Gnathonemus petersii* is able to discriminate the relative distance of two objects independent of their size or shape, as shown behaviourally by Gerhard von der Emde. Although the absolute distance of an object cannot be recovered directly from any of the electrical image features, forced two-choice discrimination experiments (where the fish, posted at a Y-gate, can target each object sequentially) show that the perceptual performance is based on the ratio of the maximum slope of the electrical image to its peak amplitude. This slope/amplitude ratio is proportional to the inverse of the object distance and represents how much the image is ''out of focus'', independent of its shape, size or conductivity. A peculiar ''distance offset illusion'' has been observed with metal spheres yielding smaller slope/amplitude calibration ratio than other objects at the same distance. The fish always misjudged the sphere distance compared to that of other non-spherical objects, with a systematic bias in the error suggesting a displacement by a constant distance of the slope/amplitude ratio characteristics. This illusion resulted on average in perceiving the sphere to be 1.5 cm further away than a cube, whatever the absolute distance to the fish [7]. Remarkably, this illusion can be corrected when submitting the fish to classical aversive conditioning. Other work has shown that it is possible to reinforce a bias for a specific shape, independent of conductivity differences, by operant conditioning using social communication (under the form of a playback signal) as a behavioural reward. In view of these findings, it is likely that perceptual learning can take place in fish as already demonstrated in higher mammals and humans and that the internal representation of space can be distorted according to the class of object and the past associative experience of the fish.

The demonstration of depth and shape perception in weakly electric fish provides a remarkable illustration that a single 2D receptive surface, even receiving only stationary input, is sufficient to extract or infer stereo information. This type of computation differs from visual or auditory processing, where spatial and temporal disparities between two receptor arrays (eye or ear) are generally required to form 3D-percepts. The analogy with the visual system would be to consider that the electrical system relies on the luminance gradient of the outer shadow of the object projected on the sensory surface of a ''cyclopean'' eye, when lit by a distant point-like source of light (a virtual ''sun''). At each EOD, the fish can thus extract an electric ''snapshot'' of the environment and determine the distances of objects. The neuronal implementation of this computation is only partly known.

This type of computation is somewhat unique, since in most species computation of depth requires the merging of not one but several perspectives of the environment generated either through channels of the same sensory modality (two ears, two eyes) or through different modalities (e.g. combining vision and extraocular proprioception, or combining mechano and electroreception).

Several examples can be found which illustrate in the auditory system of birds and mammals: (1) the precision of inter-aural auditory encoding and (2) the fusion of auditory information with non-auditory cues in spatial location identification. Precise temporal encoding appears as a pre-requisite for both sound localization and interpretation. In birds and mammals, parallel evolution of the ears relative to the head anatomy has led to different reorganizing effects upon the central structures in the brainstem receiving input from the auditory nerves. In particular, sensitivity to the high frequency content of airborne sound is a recent event in evolution. In spite of this possibly divergent evolution, it is of interest to understand how neural mechanisms and underlying circuits serving precise temporal encoding of sound may have been adapted to achieve the same computational principles in both birds and mammals, as presented by Catherine Carr.

Several pre-synaptic and post-synaptic features account for a precise temporal coding. At the first step in processing, auditory nerve fibres encode temporal information by phase locking the relayed activity to the waveform of the acoustic stimulus. In birds and mammals, preservation of the temporal information at the post-synaptic level is then ensured by the remarkable transmission efficacy of endbulb synapses formed by auditory nerve afferents encapsulating the soma of the target cell. Studies of endbulbs, such as the Held calyx, have shown several common features in both the avian nucleus magnocellularis (NM) and the mammalian medial nucleus of the trapezoid body (MNTB), whatever the species considered (lizard, alligator, owl, cat, ...). In presynaptic terminals, one observes a brief calcium influx, large pools of releasable vesicles, and a fast and modifiable rate of release. At the post-synaptic level, a specific potassium conductance, which reduces the membrane time constant, limits the spike duration to one or two hundred microseconds and specific types of AMPA

receptors exhibit very rapid desensitization rates, making the output signal still more transient than the presynaptic message. As a result, both processes increase the temporal "contrast" and precision in relaying the auditory signal to higher centres.

Inter-aural time difference (ITD) detection is a common feature of avian and mammalian auditory systems. Coincidence detectors in the auditory brainstem of birds and mammals (situated respectively in the nucleus laminaris, NL, and in the medial superior olive, MSO) are binaural neurons whose evoked spike discharge is the highest when they receive simultaneous inphase inputs from the two ears. This condition is met for a given ITD when the relative transmission delays corresponding to inputs from each ear exactly compensate for the coded inter-aural delay. The array formed by these ITD selective cells implements a place coding architecture since the preferred delays which are represented can be predicted from the position of the cells along a midline axis in the auditory nuclei. These coincidence detectors have specific potassium conductances leading to a single or a few precisely-timed spikes in response to a depolarizing stimulus. A notable morphological characteristic of these detectors is a bi-tufted feature (with bipolar dendrites), shared by both birds and mammals: the projection of inputs from each ear on a different dendrite probably improves the detection capacity. Thus, the essential features of temporal coding appear to be the same in birds and mammals, and comparative studies point to shared computational principles.

These observed similarities in functional specialization may also extend to other species. The mustache bat sonar represents an exquisitely elaborated analog-based computation, used to form in the brain a neural representation of the in-depth dimension. A biosonar network is entirely devoted to the computation of distance: the spatial parameter "target range" is mapped spatially within a specialized cortical region of the bat's auditory cortex, the FM–FM area. The computation relies on the temporal delay between the time of emission of the self-generated call and the time of reception of the echo bouncing back from the target—this delay being proportional to the target distance. The temporal tuning of neuronal responses in this biosonar receiver is highly selective, so that no overlap in cell activation exists between the detection of the call, characterized by the frequency of its fundamental harmonic (FM_1) and the detection of the echoes, the energy spectrum of which falls in distinct ripples (FM_x), with most energy confined to the second harmonic (FM_2). The combination of both these information sources, FM_1–FM_2, provides an accurate measurement of the distance and is represented in an orderly map of iso-delay bands in a specialized cortical area, the FM–FM area. The approach of another bat or of prey evokes an activity wave rolling from posterior to anterior in the neural structure. Doppler shifts need to be actively compensated by progressively changing the frequency of the primary call so that the echo frequency is kept around 60 kHz, where the input detection selectivity is the best. This system follows a general principle of neural substrate organization for encoding spatial representations: sensory space is chartered spatially and in an orderly manner in the neural tissue. The spatial dimension being represented here is distance, and the information is decoded using a SONAR principle. One interesting feature of the decoding schema (FM_1–FM_x) is that extraction of the signal relative to the external echo (FM_x) can be retrieved only if the primary component FM_1 is a priori known by the emitter–receiver and correctly filtered out by its own auditory system.

Howards Hughes addressed the possibility that humans may extract distance information using auditory cues. The human perceptual system is highly sensitive to visual cues that specify approach along the in-depth-axis, which suggests that the efficiency of processing "looming" auditory stimuli is worth investigating. Preliminary findings based on the recording of binocular vergence movements show that humans are not capable, using loudness cues alone, of tracking auditory motion along the in-depth-axis in the straight ahead direction. While the addition of cues such as Doppler shifts may be beneficial, the bandwidth of frequency-tuned channels in the human auditory system is not adapted to this type of computation—at least for moderate stimulus velocities. In the absence of specialized built-in features as exemplified in the bat sonar system, the human auditory system appears poorly suited to process motion cues, especially motion in depth. However, the context chosen in Howards Hughes' paradigm might not be the most sensitive, and adaptation protocols to repeated auditory stimuli that indicate a predictable direction in space are known to elicit visual after-effects when the subject is exposed to a visual test: visual objects are reported to move in the same direction as that previously experienced with sound, in spite of the fact that the visual image presented at the same rate is flashed in a fixed position in the visual field. Once again, as in the case of the protocols described by Elisabetta Ladavas, the exposure to one sensory modality evokes in the subject a perceptual interpretation in space which is transferred to another sensory modality and interpreted as a change in the input map.

Computational problems shared by stereo and sound localization are (1) the correspondence problem, i.e. that signals from the two sensors (the two eyes or the two ears) related to a single object have to be matched, and (2) the interpretation problem, related to the information content of small (spatial or temporal) differences between the two sensors' inputs. In the visual modality, position, colour, luminance and contrast are processed independently through each eye, but the spatial phase

is not registered. The left and right retinal positions of an object have to be compared (''spatial'' disparity) at a higher level of integration. For the hearing modality, frequency content and intensity are processed in parallel through each ear. It is the difference in the arrival time of a sound between the two ears which becomes the key feature of binaural cell activation (''temporal'' disparity). How similar are the underlying algorithms used to reconstruct space from each modality? Distance-in-depth has to be computed by both sensory systems, and the auditory system may need additional computation to extract spatial location.

Hermann Wagner compared the precision of the visual and auditory modalities for 3D-space localization in the owl. In the auditory pathway, the receptor cells are finely tuned to the frequency of the stimulus. In order to detect temporal disparities as small as 2 ms between the arrival times of the sound source echoes for the two ears, the temporal code has to be very precise. Specific potassium channels exist in this system that shorten the duration of both the post-synaptic potential and the action potential (about 100 ms in the owl), allowing tight phase-locking with the auditory signal. Inter-aural phase differences are integrated across frequencies to be transformed into the inter-aural time difference.

In the visual pathway, the receptive fields are spatially organized, with antagonist ON and OFF subregions, and their spatial profile can be approximated by Gabor functions (cosine $\times$ Gaussian) of different spatial phases. The first stage of binocular integration in visual cortex takes place at the level of layer IV cells. Their ''simple'' receptive fields linearly sum the information from both eyes. Afferents having equivalent spatial receptive fields are combined. However the computation process would remain ambiguous since the response of binocular simple cells depends both on monocular and inter-ocular phase. The disparity energy model, where the receptive field of complex cells receives inputs from pairs of simple cell receptive cells in phase quadrature, could explain how phase independency is achieved at a latter stage of processing in the visual pathway. However, this process only leads to the computation of the binocular phase difference, not the absolute spatial disparity. How the visual system solves this last ambiguity remains an open question.

In contrast with the visual system, the first binaural interaction is already a non-linear process. The two phase-locked inputs are combined in a way that may well be described by a multiplication, although the details of the interaction are still a matter of debate and may differ from species to species. Nevertheless, this combination of the inputs from the two ears remains ambiguous, because of the band-pass nature of the inputs, and involves delay lines and coincidence detection. The delay line is implemented by a short portion of axon that shifts the signal in time for a certain amount that depends on the conduction velocity of the neural signal [8]. Thus, there is an anatomical delay that is superimposed on physiological delays due to phase locking. Consequently, the activity of cells at this level of integration represents the phase equivalent of inter-aural time difference, i.e. the inter-aural phase difference.

Because of the lateral separation between our two eyes and in spite of their frontal positioning in mammals, the left and right retinal images of the same visual scene in the binocular field are slightly different. Inter-ocular disparities are used to quantify differences between corresponding features (position, orientation) on the retinas corresponding to the optical projections arising from the same or different objects in space. Absolute measurements are done relative to the homologous projections of the same object situated in the binocular horopter fixation plane. Our ability to appreciate the relative distance between objects, and their 3D shape, on the basis of the horizontal relative disparity (HD), is known as stereoscopic vision. The HD information encodes only relative distances, which are independent of the binocular fixation point. However, in order to reconstruct 3D-space in a head-centred co-ordinate framework, absolute cues, i.e. information about geometrical features of the binocular fixation (gaze direction and viewing distance) are required. The position of the eyes in their orbit allows these viewing parameters to be untangled since the convergence angle is related to the viewing distance and the version angle gives the gaze direction.

As stated earlier, the primary visual cortex, V1, is the first cortical area where spatially co-registered information coming from each eye converges on the same target cells. Several studies have shown that positional and spatial phase disparities between the two monocular receptive fields and measured along the horizontal axis (horizontal disparity, HD) are encoded as early as in V1, see [9]. The issue raised by Yves Trotter was the elucidation of the neural mechanisms through which HD input (relative cue) and information related to the viewing condition (absolute cues) are integrated and possibly interact at the single cell level in V1. It has been shown in behaving monkeys that the tonic level of ongoing activity and the phasic amplitude of visual responses are modulated in 75% of V1 cells by changing the viewing distance. These modulatory effects can be reproduced by changing the vergence angle without changing the distance of optical fixation using prisms. This strongly suggests that the modulation effect is dependent on the state of ocular vergence. This interpretation is supported by the earlier demonstration of the presence of extra-retinal information (extraocular proprioception) at various levels of integration of the thalamo-cortical pathways [10]. In a second series of experiments, the authors have shown that modulations of the neural activity were also present for changes in gaze direction (asymmetrical vergence), and the

observed kinetics of the visual gain control are compatible with a feedforward integration of the eye position signal [11]. At larger retinal eccentricities, surprisingly, a high proportion of HD selective cells was found, which were also modulated by changes in gaze direction. However, at these large eccentricities, another signal could potentially contribute to the observed effects, i.e. vertical disparity (VD), which also conveys measurable information about the eye-viewing parameters. Whereas VD is small in the central visual field, it increases significantly with retinal eccentricity. A neural substrate for the encoding of VD has been found in peripheral V1 neurons (eccentricity larger than 10°). Both interact strongly, suggesting that VD is probably involved in depth perception as an absolute cue as well.

These different results, showing the combined integration of retinal and extra-retinal cues and of relative and absolute distance cues, suggest a strong implication of the primary visual cortex V1 in the neural process of localization in 3D space. Consequently, visual properties at the level of V1 should no longer be considered as coded in pure retinotopic co-ordinates, although the visuo-proprioceptive referential that seems to emerge from the experimental evidence is not fully deciphered and is probably not defined in purely head-centred co-ordinates.

In the reconstruction of 3D space from sensory information, comparison of two sensory arrays receiving slightly different information from the environment often occurs. Two examples of such systems were presented by Jack Pettigrew and concerned binocular vision in cats and owls and mechanoreception/electroreception in a more exotic species, the platypus.

The integration of the inputs forwarded by the two sensory arrays requires first a ''segregation'' between each map in order to perform a ''fusion'' at a higher level of processing. The segregation process is heavily constrained by topographic cues linking each projection as well as the association mapping to absolute space co-ordinates. The final architecture of the target area must be organized as an orderly topographical representation of space whatever sensor is used to acquire the positional information. In the mammalian primary visual cortex (V1), the ocular dominance pattern observed in the tangential plane of layer IV consists of alternating stripes, grouping thalamo-cortical afferents corresponding to each eye. However, the spatial period of the ocular dominance network is small enough to not disrupt the global retinotopy of the binocular map formed in cortex (outside layer IV). Interestingly, a stripe orientation bias for the horizontal is often observed. This bias is preserved across several species and could possibly result from the constraint of minimizing the length of intra-cortical wiring needed to recombine information from the two eyes in order to achieve depth perception.

In the owl visual system, there is a complete decussation of retinal projections at the optic chiasma level, whereas only half of the retinal fibres cross the midline in cats and primates. In this particular bird family, binocular projection from the monocular thalamic nuclei is realized at the level of a specifically identified relay, the Wulst. This binocular relay is the recipient of two precisely superposed spatial maps of the visual field and is involved in binocular matching and depth processing. Thus, Wulst neurons share many features with cat and monkey visual cortical neurons, including a precise topographic organization, a high degree of binocular interaction, and selectivity for orientation, direction of movement, and binocular disparity of straight-line contours. Output from the Wulst reaches, in a highly non-topographic fashion, the tectum, possibly providing a generalization across azimuthal space and a further specialization for encoding iso-depth preference. This high level of convergence is found to be confined to a few bird families. For instance, the night jar has a monocular contralaterally activated Wulst. Already, long ago, Jack Pettigrew made the highly debated claim that birds could be separated into two evolutionary branches: the more primitive one has a monocular Wulst and is mostly involved in aerial feeding, like a kind of ''flying rabbit''; the other branch, including the owl, is endowed with a predator brain characterized by a binocular Wulst, a sign of a higher rank in the evolutionary scale, and thus should be considered of the ''flying cat'' type! [12].

The platypus, species extensively studied by Jack Pettigrew, can be considered as the national animal ''anthem'' of Australia. The beak-nose (''bill'') of the platypus is a mixed sensory organ, which is formed by inter-woven arrays of pressure-rod mechanoreceptors and electroreceptors hidden in the depths of mucuous glands. The S1 somatosensory cortex of the platypus receives input from these two types of receptors in a spatial arrangement very similar to ocular dominance columns in visual cortex or artificial ocular dominance columns in the tectum of three-eyed *Xenopus*. However, in the present case, each domain codes for a distinct sensory modality, resulting in ''stripes'' alternately activated by mechanical or electrical stimulation. The mapping rule is nevertheless conserved since two neighbouring mechanoreceptor and electroreceptors project to neighbouring target regions in the S1 cortical map. When, for instance, a shrimp flicks its tail, concomitant electrical and mechanical signals are detected in spatial register. The electrosensory input results from detection of the nerve and EMG related activity while the mechanical input is induced by the water turbulence generated by the shrimp tail movement. The electrical signal is first detected and followed, with a specific delay, by the mechanical signal. From this delay between the two spatially congruent stimuli, the information required for prey localization

can be extracted, in a ''thunder and lightning'' kind of central computation.

1.3. Path integration and cognitive mapping of space

Can integration of distance information during active exploration be used to reconstruct an allocentric view of the surrounding environment? This question, interestingly, can be best addressed in insects. How do ants and bees represent familiar space? By linking landmarks to a global scene co-ordinate system? By keeping a continuous track of an action code validated by proprioceptive feedback? Is navigation instrumental in building a global representation that will become independent of the exploration path?

The presentation of Thomas Collet reviewed the various codes used by ants and bees to store information relative to their environments. Some of the encoding is global and referenced on a nest-centred co-ordinate system. Other codes are local and reflect specific landmarks acquired in the environment. A fundamental question for insect navigation is whether views acquired at a specific landmark can be labelled with absolute positional co-ordinates through associative links with global path integration. If the answer proved to be positive, the underlying implication would be that insects could possess some internal representation of space akin to a cognitive map. Thomas Collett asked therefore whether associations between path integration co-ordinates and landmark memories exist and are coded in an interdependent way.

It has been long known that naïve ants use path integration to navigate in unknown environments. During exploration of unfamiliar ground, where no optimization of length can be made, path integration along translational segments of apparently random directions is used to monitor the net distance that the ants travel and the direction linking from the nest to the feeding site. This process is achieved by various means, such as computing optic flow, reconstructing direction from a sun compass or continuously updating sensory–motor information about their displacement. If the ant near the feeding is site is hijacked and released at a distinct diversion site, the return path will be made according to the return ''home vector'' from the feeding site to the nest, as if its position had not been moved, thus resulting in a target error vector equal to the opposite of the initial displacement (''diversion'') vector. In contrast, the same experiment done in bees shows that these insects are able to construct a cognitive map of their environment and that the mean direction of the return flight points toward the hive independently of the release point, even if the return flight often results in some overshoot. Interestingly, neither species records co-ordinates on their way back to the nest or the hive.

In familiar environments, bees and ants can also navigate using landmark guidance (recognition of a particular view of a landmark). Insects not only monitor their co-ordinates with respect to their nest (global path integration) but they also monitor the distance they travel along separate segments of routes between landmarks (local path integration). After an insect has visited a food site on several occasions, it tends to follow a fixed path, having learnt the appearance of landmarks along the route, the distance and direction from one landmark to the next, and having linked these two memories together. When an insect unexpectedly encounters a landmark, this new site can be memorized and included in the route plan. Through this interaction between landmarks and local path integration, the insect knows what action to take next along the route when searching to reach a goal, but this interaction does not tell it its absolute co-ordinates relative to its nest. Path integration is not the only mechanism that allows insects to acquire relevant, such as the spatial co-ordinates of significant sites (e.g., food sources) that they visit, and store it over long periods; the insects also memorize the quality of the food, and, in bees, the time at which the food source was visited [21], which allows them to store multiple feeding sites in some kind of hierarchically organized memory.

Hence, a familiar environment seems to be represented in two rather different ways. In familiar surroundings, guidance by landmarks overrides guidance by path integration. When given conflicting cues from the two systems, insects follow the dictates of their route-landmark based memories rather than their global path integration system. The global path integration system is then subordinate to the landmark and local vector based representation, perhaps because the latter referential system appears to provide a more accurate solution, especially for long distance travel. The Path integration accumulator is however updated all the while so that this latter type of formation is available as a back-up strategy used when landmark references fail. Keeping the two systems independent may have drawbacks, but it means that the errors of one system do not propagate to the other, giving more chance to the insect to rejoin the nest.

1.4. 3D shape representation of objects

Another issue, addressed during the meeting, concerned the anatomical and functional identification of higher cognitive cortical areas involved in 3D-shape representation in the primate cortex. Nikos Logothetis presented advanced techniques, coupling high resolution functional magnetic resonance imaging (fMRI) with simultaneous local field and/or single or multi-unit electrophysiological recordings in the anaesthetized or behaving monkey. Single or simultaneous multiple unit

recordings of neural activity are very useful tools to investigate how sensory information is processed at the cortical cell level, and what are the codes used to handle spatial and temporal parameters that can be extracted from the 3D-environment. However, because of their invasive nature, these methods are rarely applied to humans. Coupling chronic electrophysiology with fMRI in higher vertebrates still remains the experimental paradigm which is the best suited to reach a more global and comprehensive view of the dynamic functional interplay between co-activated cortical areas. In that regard, experimental data acquired in monkeys can serve to extract homologies between humans and non-human primates. The BOLD signal used in fMRI is a measure of local magnetic susceptibility changes produced by changes in the relative concentration of deoxy-haemoglobin vs. oxy-haemoglobin in venous blood vessels. This technique has been used successfully in the study of human cognitive processes and of psychiatric and neurological disorders.

Nikos Logothetis and colleagues have developed new tools allowing high resolution fMRI combined with single or multi-unit recordings in the anaesthetized and awake monkey. Recently, a neuronal tract-tracing method, which is detectable using MRI in living animals, takes advantage of the anterograde transport of manganese ($Mn^{(2+)}$). Trans-synaptic tract tracing in living primates allows chronic studies of development and plasticity and provides valuable anatomical information for fMRI and electrophysiological experiments in primates.

An issue addressed by these methods, and as yet unsolved, is to know how the haemodynamic changes measured in fMRI are linked to neural activity. A recent work has shown that, within certain limiting assumptions, the BOLD response may be correlated to the local field response elicited by a stimulus convolved with a standard haemodynamic filter. Although this correlation holds only for a given stimulus condition (no real transfer function and inverse analysis are yet possible), it demonstrated for the first time that the BOLD signal reflects more the synaptic activity (dendro-somatic components of the input signal, measured by the local field potentials) than the spiking activity (output of the neural population, recorded in single-unit or multi-units). The activation of an area visualized using fMRI thus reflects the incoming input and the local processing rather than the spiking activity of this area. This could explain inconsistencies observed in some studies on binocular rivalry or visual attention between electrophysiological recordings in monkeys and fMRI on humans [13], because synaptic activity produced by lateral excitation or inhibition and feedback from higher cortical areas may be more easily integrated with imaging techniques but may not necessarily reach the spiking activation threshold with single unit extracellular recordings.

Nikos Logothetis and colleagues have investigated how the primate visual system constructs representations of 3D-shapes from a variety of cues, using fMRI in anaesthetized monkeys [14]. Computer-generated 3D-objects defined by shading, random dots, texture elements, or silhouettes were presented either statically or dynamically (rotating). Results suggest that 3D-shape representations are both highly localized, although widely distributed in multiple "hot" spots in occipital, temporal, parietal, and frontal cortices, and may involve common brain regions regardless of shape cue. This imaging of composite context-dependent activation across a distributed network of areas differs from the classical division between two streams processing "what" and "where" in parallel [15], possibly reflecting multiple uses for 3D-shape representation in recognition and action.

2. Multi-modal representations

In order to guide interaction with objects, information acquired from different sensory systems (vision, hearing, touch, proprioception, vestibular, etc.) must converge and be reformatted via various encoding and integration processes in order for a coherent percept to emerge. A consensus has not yet been reached concerning the nature of neural representations subserving multi-modal integration. A minority viewpoint, defended by Howards Hugues at the meeting, claims that a perceptual representation of peri-personal space needs only to contain a symbolic correspondence with objects in the outside world. Object representations under a variety of experimental conditions should be considered as metamers, in a way that it will become possible in a unified perceptual space to perceive an apple as "redder" than an orange, as well as to report lead as "heavier" than aluminium, although no strict isomorphism can be established. An important issue is to define conditions where a perceptual problem, such as shape recognition invariance during mental rotation, involves a holistic level of recognition or can still be explained on the basis of local cues. Nevertheless, the dominant view, strengthened by imaging techniques, supports the existence of distributed multi-modal representations in the brain. This second part of the report illustrates progress, made in recent years, in understanding how and where the brain integrates information from these sensory modalities and reconstructs coherent multi-modal representations of the peri-personal space, used for guiding actions efficiently, through eye- or limb-movements.

2.1. Cross-modal interactions and adaptability of brain representations of peri-personal space

The perceptual constancy of an object results from our ability to recognize it independently of the "point

of view'' (perspective) and of the sensory modality. Given that our exploration of the environment is generally multi-sensorial, object and scene constancy might be achieved by building a higher dimensional internal representation of peri-personal space. Our knowledge remains, however, still fragmentary about the way information from different modalities is combined and to what degree the compositionality of the different input streams results in the stabilization of a perceptually coherent framework.

For visuo-tactile integration, efficient cross-modal recognition of shapes appears to be based on equivalent forms of encoding across modalities. By investigating cross-modal recognition performance Fiona Newell directly compared the nature of spatial encoding within each modality and tested how this information is shared across modalities for recognition. She summarized, at the meeting, recent investigations of visuo-tactile cross modal recognition of objects and scenes and the dependency of intra- and inter-modal recognition on changes in the orientation and perspective.

A classical observation within each modality is that recognition performance is degraded when the orientation of a learned object or a scene is changed. Object recognition is found to be improved when the subject is able to integrate complementary information across visual and haptic modalities, i.e., for rotations that involved an exchange between the front and back views of an object. This suggests that both sensory systems code view-specific representations of objects, but that each one has its own preferential view for the visual system, it is the surface of the object facing the observer, whereas, for the haptic system, it may be the surface of the object that the fingers explore preferentially during handling, namely the back side of the object. Fiona Newell concludes that efficient cross-modal recognition is based on equivalent encoding of shapes across modalities (mediated by surface-dependent descriptions). She suggests that cross-modal recognition relies on a common code shared between modalities (i.e. surface related), coupled with common processing (i.e. orientation dependence). When a possible discrepancy in encoding exists across modalities, as in scene perception (e.g. spatial encoding for vision vs. serial encoding for haptics), cross-modal recognition is less efficient. For scene recognition, a cost of cross-modal transfer is indeed observed. This reduction in efficiency may be due to a re-coding of the input provided by one modality into a format congruent with the other modality.

Multi-sensory representations exist in the brain for both peri-personal space and more distant space. This kind of representation allows (1) a spatial co-registration of stimuli of different modalities, (2) a dynamic linkage to the body position and (3) a possible generalization to build a virtual peri-personal space.

By studying right brain damaged (RBD) patients with a left tactile extinction syndrome, Elisabetta Ladavas demonstrated the existence of a multi-sensory integrative system representing space by combining vision and the haptic sense. Visuo-tactile integration seems to be processed in a privileged manner within a limited sector of near-peri-personal space. The extinction syndrome is taken as a sign of the existence of a brain lesion, whereby patients fail to report a stimulus shown on the contralateral side of the lesion when a concurrent stimulus is presented simultaneously on the ipsilateral side. In spite of this deficit, the patients are still able to detect a single stimulus presented either to the ipsi- or contra-lesional side of the body. In those patients, Elisabetta Ladavas showed that a cross-modal (visuo-tactile or audio-tactile) stimulation paradigm can cause a cross-modal (visuo-tactile or audio-tactile) extinction for near space to the same extent as an ipsilesional tactile stimulation (unimodal tactile extinction) did [16]. The generalization of the extinction syndrome across different modalities argues for the existence of multimodal representations specific to the near-peri-personal space.

It is well established that, when we move the hand, the visual peri-personal space remains anchored to the hand and therefore the spatial co-ordinate system required to merge visual and tactile information moves with the hand. Elisabetta Ladavas' studies suggest the existence of central relays that integrate both visual and tactile inputs within peri-personal space around the face and the hand. In the case of hand-centred visual peri-personal space, visual information about a given body part appears more relevant than proprioceptive information. This specialized system is functionally separated from other systems processing visual information further away in the extrapersonal space.

The question of how the combined integration of multiple sensory modalities may constrain the deployment of attention through space was addressed by Charles Spence. His latest findings in human psychophysics and cognitive neuroscience on the nature of cross-modal links in spatial attention (primarily involving vision, audition, touch, and proprioception) illustrate some of these constraints, identified in both normal people and in brain-damaged patients. When attention is directed to a particular location within one modality, attention for the other modalities tends to follow to the same spatial location [17]. For instance, the detection performance in auditory speech is improved when the auditory source has its origin in the same location where the subject concomitantly deciphers lip reading. Conversely, it is difficult to ignore sounds originating from the location to which the observer's gaze is directed. It is also difficult to divide attention between different locations through the separate activation of sensory channels. In other words, multi-sensory inte-

gration produces implicit reference to a multi-modal spatial representation within which attention is directed. A remapping of this multi-sensory ''working'' space can be triggered following a variety of postural changes, such as when deviating the eyes with respect to the head, crossing the hands with respect to the body, or interleaving the fingers of the two hands. Exposure to a constant spatial mismatch between auditory and visual space, induced for instance by optically deviating the observer's gaze, leads to a remapping of visual space such that attention is directed to the same location whatever modality is put into play. So, these cross-modal shifts of attention are directed on the basis of representations of space that are updated in a system of multiple, body-referred co-ordinates which allows the spatial co-registration of stimuli sampled by the different sensory modalities.

A remarkable feature is the high level of plasticity of near-peri-personal space representation. It can be dynamically extended to form a virtual personal space by the ''incorporation'', into the inner view that the brain has of the body, of any given object/tool that can be manipulated instead of the arm or actively used as an extension of the arm. By using a variety of tasks that involve manipulating the congruency across sensory modalities, it can be shown that the representation of visuo-tactile space can also rapidly adapt to a variety of ''virtual'' spatial transformations (visual exposure to artificial limbs or to the image of body parts reflected by a mirror). One striking illustration in humans is the illusion of ''being touched'' by a virtual object: this occurs in subjects who are prevented from seeing their arms, but are exposed through mirrors to the image of an artificial arm (shown on a video monitor) positioned approximately where the hidden observer's arm should be in space. When the observer watches the animated movie of a paint brush stroking the artificial limb, he reports that some object strikes his hidden arm. To be induced, such an illusion requires contextual congruency in space and orientation, since no effect is reported for mispositioned or misaligned arms and hands. The remarkable result of this paradigm of exposure to a virtual environment is that a high-level interpretation of the context of a scene processed through one modality (requiring segmentation, identification and explicit decoding of the relations between objects) generates a self-induced activation of a sensory analyser driven by another modality in a way which is congruent with the context of the original scene. These behavioural reports converge with recent single-cell findings reported in area 5 of monkey cortex and with neurological and behavioural data emerging from the study of patients suffering from the attentional deficit of cross-modal extinction (see above).

In spite of this plasticity, notable failures to remap near-peri-personal space can occur in both brain-damaged patients and normal subjects under a variety of conditions: for a split brain patient, failures are expected since inter-hemispheric connections are required to remap visual peri-personal space according to the current hand position, at least when the hands cross the midline; in normal subjects, a similar failure to integrate multi-modality information in the peri-personal space is observed when the subject is confronted with unusual cases of competition between different frames of reference in which stimuli can be coded—such as body-centred, retinotopic, allocentric, etc. For instance, failure to remap the world seen in a mirror could be specific for stimuli presented beyond the peri-personal space. These multi-sensory representations of space can therefore break down, resulting in the loss of spatial co-registration of stimuli across the different sensory modalities.

It is often assumed that each sense gives a view or interpretation of our environment which is subjectively unique, and that the recruitment of higher cognitive processes is required to produce an amodal representation of space. The Molyneux question raises the issue of whether the high level percept of 3D-shape acquired through one sense (let us say the concept of a ''sphere'' or a ''cube'') can be transferred and merged with that acquired through another sense (in the case where a person blind from birth recovers vision and has to recognize the differences between the two shapes by using vision only). There are of course many striking examples of the congruency constraints between sensory maps, which support the existence of a higher level of ''percept'' representation which influences in a top-down way the read-out made by each of our senses. For instance, the McGurk effect shows that the association of a ''ba'' auditory cue with the visual lip reading of a ''ga'' is interpreted by a naïve observer as a ''da'' auditory cue. Another classical illusion is that of ventriloquism where the auditory cues provided by the ventriloquist are ''captured'' and associated with the viewed puppet movements as long as they remain interpretable as parts of a ''whole'', congruent percept. It still remains to be seen if the McGurk effect can be reproduced when using sensory inputs unrelated to linguistic compositionality and check if the errors in sound identification reflect a misclassification in a higher level multi-modal relational representation, or simply result from a change in the global sensitivity of a specific sensory channel.

2.2. Cortical sites for multi-modal integration

Which cortical areas are involved in multi-sensory integration? The role of both ''unimodal'' and multi-sensory cortices in multi-modal integrations has been addressed by using complementary experimental paradigms focusing on the non-linearity of multi-sensory integration. The precise localization of the activated cortical areas, identified by a combination of brain imaging

techniques, is shown to depend on the modalities, the attention level and specific constraints defining the task. Multi-modality, to a certain extent, is also built-in in so-called primary sensory-specific cortical areas. We will illustrate this point later in the text by showing that the concepts ''right'' and ''non-right'' in peri-personal space are linked to the merging of both visual and proprioceptive information and that visuo-tactile integration, which could take place in the posterior parietal cortex, is already present in the primary occipital visual cortex.

The cerebral site(s) where the integration of cues initially processed by anatomically distinct sensory pathways takes place are still largely unknown in humans. Gemma Calvert described activation patterns that appear to be distributed across distinct neuronal networks and whose identity varies depending on the nature of the shared information between different sensory cues. Higher spatial and temporal resolution can be gained by combining respectively fMRI and electromagneto-encephalography (MEG). These techniques provide converging evidence of multi-sensory interactions at both ''early'' and ''late'' stages of neural processing, which suggests a cascade of synergistic processes operating in parallel at different levels of the cortical hierarchy. Sensitivity to shared temporal onset across different sensory cues has been shown in the superior colliculus and insula-claustrum. Several regions of the inferior and superior parietal lobe, more specifically the intra-parietal sulcus, appear to be involved in the detection and integration of multi-sensory cues based on the detection of shared phonetic features. In addition to these regions of heteromodal cortex, a number of recent studies now suggest that multi-sensory interactions also occur at early stages of the processing hierarchy, in primary sensory-specific areas [17,18].

Electrophysiological and neuroanatomical studies in the animal (non-human primate) have shown the existence of numerous areas in the association cortex and subcortical structures (claustrum, superior colliculus, thalamus) that receive convergent afferents from different sensory modalities. Furthermore, many neurons in these multi-modal areas are multi-sensory—i.e. they respond to information from more than one modality. The principle of multi-sensory integration calls for a diversity of possible computation schemas for its implementation: if the sensory-specific information channels act independently, a linear summation is to be expected; if this is not the case, interaction may be viewed as a normative combination of inputs with multiple partial gains, each specifying the relative dominance of one sense in the interaction effect. True interactive effects call for correlation terms, e.g. a multiplicative operation between the two afferents. Pragmatically, experimenters most often limit their study to a rough evaluation of the degree of non-linear interaction, without characterizing the nature of the computation (but see Fabbrizio Gaziani, this volume). A ''super-additive'' interaction is often observed when the firing rate of a multi-sensory integrative neuron to a stimulus presented in one modality is enhanced in the presence of a stimulus from another modality originating from the same location in space and at the same time; in contrast, a ''sub-linear'' behaviour is reported when the neural response is depressed by the association of the text input with a spatially disparate cross-modal cue. These phenomenological features of multi-sensory integration at the cellular level have also been found in behavioural studies of cross-modal processing in humans: congruent stimuli from two modalities improve performance superadditively and the reverse is true for incongruent stimuli.

Recent imaging studies show that performance in humans in cross-modal tasks may rely on the activation of what are regarded as ''purely'' unimodal cortical areas (with reciprocal amplification) as well as of association cortices. The discrepancy between some of these studies can result from the differences between the experimental paradigms and analyses used. In order to unify the identification of multi-sensory integration sites, it appears more important to focus on the non-linearity of the multi-sensory integration (in relation to congruency between the stimuli) than to try to identify them from the cortical topology of the overlap zone between two unimodal maps of activation. Indeed, considering the crudeness of spatial sampling by the imaging techniques, an overlap analysis could not distinguish true multi-modal integration from the activation of two distinct unimodal populations within the same voxel box.

Gemma Calvert showed in humans that convergence onto multi-sensory cells occurs and that non-linear integration is a general property of multi-modal interaction. Some sites of integration could be clearly identified, mainly in the ''association'' multi-sensory cortex. For example, the left superior temporal sulcus appears to support bimodal integration during audio–visual speech perception. This comes in addition to the amplification of activity in unimodal cortices (visual and auditory) during conjunction stimulation. More data are expected concerning the identification of possible cellular mechanisms implied in the binding of multiple sensory sources. Features such as proximity in space and time could be relevant cues for this binding. The superior colliculus activation could mediate an integration of the spatial and temporal relationship between non-speech audio–visual stimuli. Therefore, identification of the multi-sensory integration site depends on the modalities combined (visuo-tactile integration takes place in parietal regions) but also on the point of correspondence used to bind the information: time, space and meaning. Furthermore, attention and task can modulate the

importance of each of these factors in the performance and can also change the site of multi-modal integration.

2.3. Deciding what is "right" or not, in near-peri-personal space

Early processing of cross-modal information in a spatial representation has also been reported in humans by Satoru Miyauchi, by addressing the question of what is "Left" and what is "Right" in peri-personal space, using both fMRI and behavioural experiments.

The high level of adult brain plasticity in sensory–motor co-ordination was investigated by forcing (and rewarding) human subjects to adapt to a visually left–right transposed world. The use of left–right reversing goggles causes strong dissociation between visual, tactile and proprioceptive inputs. When wearing these goggles, the stabilization of the adaptation process at the behavioural level (postural control and walk) requires more than one week. However, visual function adaptation is already reached within a few days: at this delay one starts to observe the metabolic or haemodynamic activation of the ipsilateral occipital lobe. The ipsilateral primary visual cortex is likely to be activated by backward and heterotopic commissural connections from extrastriate and/or association cortices. The appearance of the ipsilateral activation is thought to relate to the left–right reversal of the internal representation of personal space.

Application of left–right reversing goggles to patients suffering from phantom limb illusions and the associated pain gives more insight into the visuo-tactile links formed in the brain: i.e. the concept of "left" and "right" in spatial perception is not entirely visually driven and there is some kind of structural association with our own awareness of "left hand" vs. "right hand". Indeed, when phantom limb patients put on left–right reversing goggles, they experienced no problem in superimposing the phantom representation of the missing hand with the existing representation of the intact hand. This leads to the perceptual illusion that the phantom hand is anchored, i.e. moves, with their intact hand. Moreover, after a two-week training, the subject was able to trigger the illusion of motility in his phantom hand by voluntarily moving his intact hand, and thus to relieve the pain. Changes in activation in the parietal cortex can be correlated with the perceived movement of the phantom hand: in addition to the contralateral somatomotor area of the intact hand, activation was found in bilateral supplementary motor areas (SMA) and the ipsilateral somatomotor area, which is contralateral to the phantom hand. Strong activation areas around the post-central sulcus (PCS) also appeared. It can be concluded that the posterior parietal cortex and perhaps the SMA are implicated in the integration of tactile proprioceptive and visual inputs.

To further test the hypothesis suggested previously that brain activity for "left–right" cognition may be affected by proprioceptive inputs from hands, fMRI imaging during a visuo-tactile matching task was conducted with crossed hands. So doing, dorsal occipital cortex (Areas 18/19) was found to be more activated by changes in proprioceptive input. Right inferior frontal cortex (IFC) was activated by a tactile stimulus when it was applied to the "right", whether this context was defined by the right hand itself or by the right side of the visual field where stimulation took place. This observation is reminiscent of the recording of bimodal visuo-tactile neurons in monkey pre-motor cortex that code stimulus position in body-centred co-ordinates. No equivalent of a contextually defined "left" area has been shown.

Not only visual, but also proprioceptive information relative to the body image is profoundly associated with the definition of "left vs. right" context. This opens the possibility that a representation of visuo-tactile spatial correspondence depending on proprioceptive inputs exist in the PCS and an internal representation of space exist in occipital cortex. It remains thus likely that the space reconstructed by our brain may be divided not as "left vs. right", but as "non-right vs. right" words.

2.4. Cross-modal effects in primary sensory neocortex

The previous results raise the issue that the integrative function of primary visual cortex should not be defined as processing retinal information within a visual-only referential, but could also be used to chart some form of amodal representation of space. The potential role of primary sensory cortex in early stages of multi-modal convergence was already mentioned in Section 1. Despite this, extraretinal influence is generally considered as subliminal, each primary sensory cortex being driven by definition by a unique modality. Cross-modal cooperation could have another function, namely to compensate for a lack of, or suboptimal, stimulation, allowing the extraction of "equivalent" information by processing input from another sensory modality. For example, auditory cues could help visual attention (in the peripheral visual field) but there is no equivalence between these two senses for spatial location: for instance, audition cues are not used for looming stimuli in the fovea field (see Section 1).

Hugo Theoret reconsidered the supposed modal organization of the brain: Are there really areas devoted to the treatment of information from one sensory modality (audition, vision, tactile,...) or shall we consider a metamodal organization where the areas are devoted to one modality just as a consequence of their specialization in processing information of one kind (spatial for vision vs. temporal for audition for example)? If the second

hypothesis is true, we should be able to observe an unmasking of the totipotency of these cortical areas when their major sensory input source is impaired. The study of blind individuals by Hugo Theoret and Alvaro Pascual-Leone provides insight into the brain reorganization and behavioural compensations occurring after long-term sensory deprivation. Loss of vision leads to behavioural enhancements in other modalities, i.e. the auditory and tactile capabilities. This effect is associated with metabolic reorganizations at the cortical level such as activation of the occipital cortex (considered visual) by non-visual stimuli. The primary visual cortex is also activated during Braille reading. This invasion of the occipital lobe, seen in the imaging of the cortical activation pattern following early blindness, appears as the necessary substrate for cross-modal plasticity.

Is this effect induced de novo by the initial period of sensory deprivation? Or is the neuroanatomical substrate of cross-modal interactions between primary sensory areas present during normal development but functionally dormant, unless a sensory deprivation demands their recruitment? In this case, we could expect that a period of blindfolding in adult sighted subjects will induce the same kind of cross-modal plasticity. Experiments reveal that blindfolding enhances tactile discrimination performance, and that this effect is correlated with the imaging of activity recruitment in visual cortical areas (bilateral primary visual cortex) during tactile perception and auditory stimulation It is also shown that occipital cortex activation is necessary for tactile processing. Thus, prolonged (5 days) sensory deprivation in sighted individuals mimics the effects of blindness.

We conclude that visual cortex function is not only restricted to processing ''visual'' information, but is able to integrate auditory and tactile information by the unmasking of pre-existing cross-modal connections between primary areas or by feedback from multi-sensory association areas. However, this capability is ''suppressed'' during normal visual processing.

2.5. Transformation of co-ordinates

We have seen in Section 1 that extra-ocular proprioceptive information modulates the response of primary visual cortical neurons [11]. These observed modulations plead for a recalibration process realizing the transition from a retinotopic to a head-centered spatial reference frame. Such data support the hypothesis that associative networks should have the built-in capacity to generate new referentials obtained by combinations of multiple sensory co-ordinates, as proposed on theoretical grounds by Yves Burnod and Alex Pouget in the sensory–motor domain.

The ''sensory–motor transformation'' is the computation from a basic set of sensory inputs leading to the programming of motor commands. It is often described as a succession of transformations of the spatial frame of reference. If one considers a simple action, such as grasping a target under view, most models of brain function assume that eye-centered co-ordinates should be transformed into head-centered co-ordinates by integrating the eye position signal, and then into body-centered co-ordinates by integrating the head position relative to the body and so on. In the auditory modality, sound is encoded in a head-centered co-ordinate frame and, in order to make a visual saccade toward an auditory stimulus, the sound has to be somehow encoded in an eye-centred co-ordinate frame.

The model proposed by Alex Pouget and Sophie Deneve assumes the existence of an intermediate frame of reference, permitting co-ordinate transformation in both directions (sensory–motor and motor–sensory). It is hypothesized that sensory inputs of different modalities (vision, audition, proprioception), each having its own spatial reference frame, are merged to form a common spatial reference frame. The multi-sensory spatial representation is encoded in an intermediate layer by implementing basis functions which compute correlations between inputs and outputs at the population level [19]. This kind of architecture has been shown to achieve adequately: (1) sensory–motor transformation, (2) sensory prediction of motor command, (3) sensory expectation and (4) multi-sensory integration. The basis function model makes testable predictions about characteristics of the units to be recorded in the intermediate layer. In a pure feedforward architecture (sensory–motor transform), units of the intermediate layer should display gain fields. In a feedforward and feedback architecture, units should display both gain fields and partially shifting receptive fields, a prediction that has been actually observed in the parietal area VIP [20]. The model has also a robust noise-invariant behaviour. Using maximum likelihood algorithms, the reliability of mapping improves with iterations of associations.

Neuronal populations involved in 3D space representation or sensory–motor transformations are distributed over different areas, linked by reciprocal connections, and are not confined to individual areas. The model proposed by Yves Burnod integrates these cortical properties in a network devoted to the processing of 3D space. Its organization is that of a meta-network, i.e., a network of computational nodes (corresponding to subnetworks of identified neuronal populations) with combinatorial properties distributed across large gradients. Properties of units depend upon their location relative to three main axes defined within the anatomy of the cortex.

(1) The ''visual-somatic'' axis is defined by symmetrical gradients relative to the central sulcus. From frontal and occipital regions towards the central sulcus,

combinatorial domains pass from retinal/gaze, to gaze/arm, to arm/muscles output.

(2) The "position-direction" axis. Positional and directional information are encoded in a parallel fashion by neurons within the same combinatorial domain. These combinations can be viewed as representing 3D pathways for gaze or hand.

(3) The "sensory–motor" axis. Neurons sharing the same combinatorial domain and similar positional or directional tuning may have different temporal relationships with the signals relevant to reaching. Matching units serve to learn different neural representations of virtual 3D pathways for gaze and hand, permitting predictions of possible motor commands and sensory consequences. Specialized units fulfill the task of linking the matching operation to reinforcement contingencies. They are suited for the recruitment of neuronal populations along the three axes of the network.

Yves Burnod proposed an extension of the model to human cortical organization by adding two "man-specific" axes in the network: the "parieto-temporal" axis, along which a new kind of matching unit can store 3D relationships between different parts of the body and the different elements of the scene; and the "left–right" axis (see Section 2.3), taking into account a differential distribution of properties between hemispheres.

3. Conclusion

This review has illustrated various strategies used by living organisms in order to compute the spatial location or distance-in-depth, or to recognize the 3D shape of an object. Different species rely dominantly on the peculiar sensitivity of the specialized sensory systems which have been selected through evolution for optimizing survival in their preferred environment. These environments are not only diverse (aerial for birds and mammals, liquid for fish, underground for the star-nose mole,...) but highly constrained by the diurnal or nocturnal behaviours engaged in eating, mating and communicating. Vision and audition dominate most often, but other senses emerge or are kept in a vestigial form when some features of the environment are susceptible to vary greatly and hence impede the most likely sources of sensory input (e.g. increase in the turbidity of water, masking of the sun during navigation,...).

The neural representation of our environment seems to be predominantly decomposed along two principal component axes, time and space. The question that the central nervous system has to solve under the continuous bombardment of multiple sensory streams is not always to know: "where is the target?", but rather, "where will it be?" when I will capture it, or "when should it hit me?" before it is too late. Thus, the inner representation of space should not be purely static, but should also depend on dynamic variables that are computed by the brain on the basis of gradients of input flow in space and time.

- Time-dependent selectivity is achieved in multiple ways. The auditory system illustrates a diversity of strategies which allow the computation of direction in space as a function of inter-aural disparity, as shown in the owl, or the extraction of prey distance from frequency change in echo-location, as shown in the bat. Similar temporal selectivity is also engineered in the visual system to extract direction preference based on the precise asynchronous sampling of target position through elementary motion detectors.
- Spatial localization can be gained almost instantly by the high level of binocular convergence observed in cortical-like areas (cortex for mammals, Wulst for birds) and of the topographic precision (retinotopy) maintained in the binocular map. Reconstruction of in-depth information does not obligatorily require a double mapping of space. In fish, for instance, a remarkable strategy appears to be based on local computation of the slope/amplitude ratio of the sensory evoked excitation, transduced by the lateral line system in response to a pressure gradient or received by the electrosensory system in response to a self-generated electrical field.

3D-reconstruction is done under the assumption that the spatial representation is computed by unfolding in a higher dimension space the sensory projection of the environment on one or several receptor sheets. The sensory image can be produced by an external source (sound pressure, light source) or by a self-generated action (eye-movement, electrical organ discharge) creating a virtual "lightning and thunder". The fact that this snapshot takes its origin externally or internally makes no difference, and the neural mechanisms involved in the stereo-interpretation remain the same. This view is reminiscent of the Aristotle metaphor, where humans are kept inside a cavern and perceive only the 2D print on their retina of shadows of a higher dimensional space produced by the light of the fire. In a few species, only one snapshot is necessary, such as in the fish, during active electrolocation or passive mechanolocation using the lateral line. In other species, several snapshots are required, from a different spatial referential "perspective" when using the same sensory modality. In the latter case, the acquisition of information is done simultaneously

and disparity between maps is processed using space and time difference detectors. When using a time code, natural delay lines in the neural system are used to reconstruct a spatial representation along specialized morphological dimensions of sensory nuclei. In some instances, perceptual abilities are improved by merging several sensory modalities, for instance mixing mechanoreception with electrolocation, as demonstrated in fish and platypus.

There is no unique way of computing distance or identifying the 3D shape of an object in the environment, but rather a battery of potential strategies based on specialized sensory systems to be found unequally distributed across species: echo-location of the bat, mechano and/or electrolocation in fish and platypus, location using visual, tactile, or hearing cues in birds and mammals. Despite the existing differences between all these sensory systems, they operate under similar constraints, suggesting the existence of general species-independent principles for neural computation and the spatial organization of neural representations of the environment. Moreover, at least in mammals, the representation of the external world is built by using inputs from different modalities. These cross-modal interactions, generally thought to be performed in "associative" cortices, have been shown to also involve primary sensory areas. The classical hierarchical view of the general functional architecture of the cortex, starting from primary areas devoted to the processing of a specific sensory modality and progressing towards higher-order convergence integration sites, seems somewhat obsolete. Experimental results, as well as computational models, highlight the fact that an intrinsic potential for and versatility with multi-modality processing is a general rule in the brain and that plasticity permits a reorganization even of primary sensory areas when these are deprived of specific sensory input. Although input of the primary modality (vision for V1, haptic for S1, auditory for A1) is sufficient per se to trigger specific processing in primary sensory cortical areas, the multi-modal nature of such networks should not be understated: they constantly integrate multiplexed inputs from other areas, combining multiple perspectives on the same environment filtered through various modalities/senses, to which different weights are given according to the contextual nature of the environment.

Acknowledgments

We thank Andrew Davison for helpful comments. This review work was supported by a grant from EC to Y.F. (Life-Like Perception: IST-2001-34712).

References

[1] F. Gabbiani, H.G. Krapp, C. Koch, G. Laurent, Multiplicative computation in a visual neuron sensitive to looming, Nature 420 (2002) 320–324.
[2] N. Franceschini, From fly vision to robot vision: re-construction as a mode of discovery, in: Barth, Humphrey, Secomb (Eds.), Springer-Verlag, Wien/New York, 2003, pp. 223–236.
[3] G.S. Masson, E. Castet, Parallel motion processing for the initiation of short-latency ocular following in humans, Journal of Neuroscience 22 (2002) 5149–5163.
[4] E.H. Adelson, J.A. Movshon, Phenomenal coherence of moving visual patterns, Nature 300 (1982) 523–525.
[5] C.C. Pack, R.T. Born, Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain, Nature 409 (2001) 1040–1042.
[6] K. Pohlmann, F.W. Grasso, T. Breithaupt, Tracking wakes: the nocturnal predatory strategy of piscivorous catfish, Proceedings of the National Academic Science, USA 98 (2001) 7371–7374.
[7] G. von der Emde, S. Schwarz, Three-dimensional analysis of object properties during active electrolocation in mormyrid weakly electric fishes (*Gnathonemus petersii*), Philosophical Transactions of Royal Society London B, Biolgical Science 355 (2000) 1143–1146.
[8] J.M. Goldberg, P.B. Brown, Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization, Journal of Neurophysiology 32 (1969) 613–636.
[9] D.Y. Tsao, B.R. Conway, M.S. Livingstone, Receptive fields of disparity-tuned simple cells in macaque V1, Neuron 38 (2003) 103–114.
[10] P. Buisseret, Influence of extraocular muscle proprioception on vision, Physiological Reviews 75 (1995) 323–338.
[11] Y. Trotter, S. Celebrini, Gaze direction controls response gain in primary visual-cortex neurons, Nature 398 (1999) 239–242.
[12] J.D. Pettigrew, M. Konishi, Neurons selective for orientation and binocular disparity in the visual Wulst of the barn owl (*Tyto alba*), Science 193 (1976) 675–678.
[13] R. Blake, N.K. Logothetis, Visual competition, Nature Reviews Neuroscience 3 (2002) 13–21.
[14] M.E. Sereno, T. Trinath, M. Augath, N.K. Logothetis, Three-dimensional shape representation in monkey cortex, Neuron 33 (2002) 635–652.
[15] M. Mishkin, L.G. Ungerleider, K.A. Macko, Object vision and spatial vision: two cortical pathways, Trends in Neurosciences (October) (1983) 414–417.
[16] E. Ladavas, F. Pavani, A. Farne, Auditory peri-personal space in humans: a case of auditory-tactile extinction, Neurocase 7 (2001) 97–103.
[17] E. Macaluso, C.D. Frith, J. Driver, Modulation of human visual cortex by crossmodal spatial attention, Science 289 (2000) 1206–1208.
[18] M.H. Giard, F. Peronnet, Auditory–visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study, Journal of Cognitive Neuroscience 11 (1999) 473–490.
[19] S. Deneve, P.E. Latham, A. Pouget, Efficient computation and cue integration with noisy population codes, Nature Neuroscience 4 (2001) 826–831.
[20] J.R. Duhamel, F. Bremmer, S. BenHamed, W. Graf, Spatial invariance of visual receptive fields in parietal cortex neurons, Nature 389 (1997) 845–848.

Yves Frégnac
Editor-in-Chief of the Journal of Physiology (Paris)
UPR-CNRS 2191 (UNIC)
Unite de Neurosciences Integratives et Computationnelles
Bat. 33, 1 Ave de la Terrasse
91 198, Gif sur Yvette, France
Tel.: +33 1 69 82 34 15; fax: +33 1 69 82 34 27
E-mail address: yves.fregnac@iaf.cnrs-gif.fr

Alice René
UPR-CNRS 2191 (UNIC), Gif sur Yvette, France

Jean Baptiste Durand
UMR 5549 (CERCO), Toulouse, France

Yves Trotter
Guest-Editor
UMR 5549 (CERCO), Toulouse, France