Compensating time delays with neural predictions: are predictions sensory or motor?

BY ROMI NIJHAWAN1,* AND SI WU2

1Department of Psychology, University of Sussex, Falmer, East Sussex BN1 9QH, UK
2Institute of Neuroscience, Shanghai Institutes for Biological Sciences, State Key Laboratory of Neurobiolo, Chinese Academy of Sciences, Shanghai 200031, China

Neural delays are a general property of computations carried out by neural circuits. Delays are a natural consequence of temporal summation and coding used by the nervous system to integrate information from multiple resources. For adaptive behaviour, however, these delays must be compensated. In order to sense and interact with moving objects, for example, the visual system must predict the future position of the object to compensate for delays. In this paper, we address two critical questions concerning the implementation of the compensation mechanisms in the brain, namely, where does compensation occur and how is it realized. We present evidence showing that compensation can happen in both the motor and sensory systems, and that compensation using ‘diagonal neural pathways’ is a suitable strategy for implementing compensation in the visual system. In this strategy, neural signals in the early stage of information processing are sent to the future cortical positions that correspond to the distance the object will travel in the period of transmission delay. We propose a computational model to elucidate this using the retinal visual information pathway.

Keywords: neural processing; brain dynamics; motion extrapolation; flash-lag effect; diagonal neural pathway

1. Introduction

Animals sense the environment and produce adaptive behaviours to obtain nourishment, avoid dangerous objects and procreate. The ease with which animals are able to execute these tasks hides some complex neurophysiological issues. Among these, one of the most intriguing and pervasive ones is related to the speed of neural processing—in the event chain from sensory transduction to muscular contraction, information flows relatively slowly (Maunsell & Gibson 1992; Schmolesky et al. 1998; DiCarlo & Maunsell 2005). In 1850, Helmholtz demonstrated that the speed of neural signals travelling along a vertebrate motor nerve is an order of magnitude slower than the speed of sound through air. There is a large range of neural conduction speeds. The slowest of these speeds is a few orders of magnitude slower than the speed of sound.

* Author for correspondence (romin@sussex.ac.uk).

One contribution of 10 to a Theme Issue ‘Delay effects in brain dynamics’.
Neural processing delays are of interest for several reasons. First, consider the change in the state of some environmental variable over time. During visual motion, for example, moving objects can cover a significant distance over the retinal surface in the time it takes the retina to produce an output and for the brain to register this output. The initial stages of visual processing alone can incur a delay of 20–30 ms. The engagement of the primary visual cortex and the areas of the association cortex, where multiple pieces of information from different sensory modalities come together, and where non-sensory information from the brain centres concerned with the animal’s current needs, goals and arousal level are integrated, inevitably incur further delays. Both in nature and in modern sports there are plenty of examples showing the impact of these delays (figure 1). How then are animals able to interact with quickly changing, e.g. fast moving, objects?

The receptors for sensory transduction are exquisitely sensitive (Kirschfeld 1983)—chemoreceptors can detect single molecules, some mechanoreceptors can respond to displacements close to the diameter of a hydrogen atom and photoreceptors can detect single photons. In the domain of temporal processing, behavioural studies have revealed that animals can use extremely small time intervals in computations (Simmons 1979; Rose & Heiligenberg 1985; Carr 1993). Thus, the relative slowness of neural processing does not seem to be a necessary consequence of a general limitation of neurobiological processes. So, why then are these delays there?

2. Are neural delays inevitable?

We suggest that neural delays, rather than being a passive consequence of the anatomy and physiology of the nervous system, are a result of a basic feature and active strategy used by the nervous system for information processing—namely,
neural integration and temporal coding. Thus, time (delay) is part and parcel of neural computations. The use of time, however, raises the issue of neural representations, both sensory and motor, that must track time-dependent stimulus events. This leads to a second type of computational feature, that of compensating for time delays that neural integration and temporal codes incur. We suggest that the mechanisms compensating for neural time delays have co-evolved with mechanisms that use time as a medium in which to carry out computations.

3. Neural integration and temporal coding

At the cellular level, the neurons are performing what Charles Sherrington regarded as the fundamental task of the nervous system: decision making. Decision making roughly corresponds to signal integration over time and space, which is a basic feature of neural computation.

Say a presynaptic neuron (neuron A) communicates with a postsynaptic neuron (neuron B). Whether neuron B will produce an action potential or not depends on whether depolarization of neuron B reaches a threshold value. Typically, neuron B will receive inputs from many neurons, and whether neuron B produces its own action potential will depend on the status of these inputs and the instantaneous membrane potential. The instantaneous membrane potential will, in turn, depend on the neuron’s time constant, which for different neurons ranges from 20 to 50 ms. In order to produce an overt action or a sensory percept, neurons integrate inputs from diverse sources, a process that takes time.

One function of integration is signal amplification. At the absolute threshold of ganglion cell response in the toad, for example, the light level is such that there are only one or two photoisomerizations per second per 100 rods. At this level of illumination, ganglion cells integrate signals from 400 to 1000 rods over a period of 1–2.5 s (Aho et al. 1993), which necessarily makes the ganglion cell output to the brain slow. Integration is necessary for the detection of extremely small differences in position or time intervals. Human observers, for example, can perceive a separation between line stimuli that is a fraction of the distance between neighbouring photoreceptors in the fovea—a phenomenon known as hyperacuity (Westheimer 1979). In the domain of time, electric fish can detect the modulation in the timing (phase) of an electrical field as small as 400 ns. Such hyperacuity performance is possible only on the basis integration of the data from many afferent fibres (Rose & Heiligenberg 1985). Thus, neural integration, which allows detection of some stimuli and efficient processing of others, contributes significantly to neural processing delays.

Following the initial neurophysiological studies, the average rate of action potentials was thought to be the main carrier of sensory information (Adrian 1928). However, over the past many decades, alternative schemes have been developed in which time is the medium in which coding of information takes place (Bialek & Rieke 1992; Hopfield 1995; Singer & Gray 1995; Johansson & Birznieks 2004; VanRullen et al. 2005; Tiesinga et al. 2008). A main feature of the temporal coding schemes is time delays, which typically arise owing to the natural time dependence of the stimuli, but can be actively introduced by the encoding scheme (Hopfield 1995). For example, a postsynaptic neuron with
a short integration time constant can serve as a ‘coincidence detector’ for inputs arriving from two presynaptic neurons, one of which is slightly delayed relative to the other.

4. Neural prediction and forward shifts

Among the most intuitive examples of how time delays are used in neural computations are models of sound localization (Knudsen & Konishi 1979) and direction selectivity in motion perception (Borg-Graham 2001). We suggest that the biophysical principles that establish the circuitry and mechanisms for time delays are also responsible for mechanisms that compensate for the delays, thus assuaging harmful effects of delays. In figure 2, we outline a very simple mechanism compensating for delays arising from computation of direction selectivity.

No matter which coding scheme one considers, neural processes that feed into the computations are slow in relation to animal behaviour. So, an inescapable conclusion is that there must be mechanisms compensating for the delays. Where are the compensatory mechanisms located and what are their properties? Are these mechanisms sensory (Nijhawan 1997), motor (Wolpert & Flanagan 2001) or intermediate? In other words, are these mechanisms located in the pathway between the receptors and the primary sensory cortex, in the pathway from the primary motor cortex to the muscles, or between the primary sensory and the primary motor cortex? In §5, we attempt to address these questions.

5. Neural prediction: sensory or motor?

(a) Evidence from head-direction system and motor flash-lag

Anticipatory behaviour is seen in many situations, in various species and at multiple levels within the nervous system. Anticipatory control is seen in diving birds that retract their wings before entering water (Lee & Reddish 1981) and in

Figure 2. (a) The standard Reichardt detector. The detector (circle with arrow) is a coincidence detector—only if the excitatory (+) inputs from the receptors (semicircles) arrive nearly simultaneously will the detector respond. This detector circuit uses a delay in the fibre on the left, which allows it to be directionally selective—the detector fires only when motion is rightward. (b) A similar circuit in which the detector is shifted to the right—the shift is in the detector’s preferred direction of motion. The rightward shift compensates for the spatial lag that occurs, as the circuit requires a delay in the processing of motion. The retinotopic position shift shown is arbitrary, and can be adjusted (dots) with the delay used and the required magnitude of compensation. This scheme predicts that the moving objects should appear shifted in the direction of motion, for which there is a large body of evidence (see below).
houseflies as they land (Wagner 1982). When a visual object heads directly on a collision course towards a pigeon, the firing rate of ‘looming-sensitive’ neurons in the nucleus rotundus of the pigeon brain increases approximately 1 s before the expected collision. Following this increase in firing rate, there is an increase in wing muscle activity and heart rate (Wang & Frost 1992). There is evidence that even homeostatic regulation is anticipatory, and the mechanisms are turned on or off before tissue deficit, and sometimes even prior to felt need (Kupfermann et al. 2000).

Catching a ball requires both anticipatory and reflexive (feedback) mechanisms (Lacquaniti & Maioli 1989; Ghez & Krakauer 2000). The anticipatory responses are revealed by the activity of the arm muscles (biceps and triceps) prior to the ball impacting on the hand (figure 3). Following contact, the reflexive feedback mechanisms are engaged by the receptors in the hand and the arm.

A remarkable anticipatory system, known as the head-direction (HD) system, exists in rodents. The HD neurons were originally discovered in the postsubicular cortex (PSC) of the rat hippocampus (Ranck 1984) and are present in other parts of the rat brain. These neurons are activated when a freely moving rat’s head points in a specific direction in the horizontal plane relative to a static environment (Taube et al. 1990). The HD cell response drops off as the head direction moves away from the centre of a narrow range of head directions. Different neurons have different tunings for head direction, so the population of cells covers the different directions of the head in the horizontal plane. Although these neurons use continuous visual input from the environment (in addition to vestibular, proprioceptive and other inputs), the response persists even in the dark (Mizumori & Williams 1993).

The input from various sources arriving at HD neurons is delayed, which implies that, if a rat’s head is moving continuously to the right (say), the sensed head direction should trail behind (shifted to the left) in relation to the actual instantaneous head direction (van der Meer et al. 2007). It has been found, however, that neurons in the PSC represent the current (as opposed to a lagging) position of the head with close to 0-lag. Most interestingly, however, neurons in another part of the brain (the anterior dorsal thalamic nucleus, ADN) anticipate the future position of the head (Blair & Sharp 1995). The anticipatory response of the ADN neurons may be due to the integration of motor efference copy resulting from the motor command instructions to move the head and sensory feedback (Taube & Muller 1998).

A large volume of mathematical work has been devoted to understanding how the HD information is acquired and maintained in a neural network system. Zhang (1996) proposed an interesting model in which the neuronal interactions were structured to have both symmetric and asymmetric components. The symmetric component was static enabling the neural system to hold a bump-shaped persistent activity even when the visual cues are absent (e.g. in a dark room). This bump population activity is the internal representation of the HD system (Taube et al. 1990). The asymmetric neural interactions, on the other hand, were dynamic and conveyed the information of head rotation, e.g. owing to the self-motion of the body. This asymmetric component drove the bump to catch up with the changes in head direction, so that a world-centred, rather than a body-centred, representation of head direction was achieved.
Helmholtz first outlined the notion of an efference copy. When humans make a saccadic eye movement, the visual image transiently shifts over the retina. Yet, we perceive the world as remaining stable. Helmholtz suggested that a copy of

![Diagram of eye movement](image)

Figure 3. (a) The ball is dropped from a certain height (0.8 m in the example) in full view of the observer ((i) feed-forward and (ii) feedback). (b) There are anticipatory responses of muscles 130 ms after the release of the ball, as well as 100 ms before the ball makes contact with the hand (Lacquaniti & Maioli 1989). These muscular responses occur irrespective of the height from which the ball is released, but the second response is stronger if the ball is dropped from a greater height, so it has greater momentum. The response that occurs at a fixed interval after the ball’s release is weak and cannot be discerned in the figure. The response that appears 100 ms before contact is strong and easily discernable (marked by arrowheads) in the figure for both the biceps and the triceps muscles. Such anticipatory responses are based on the visual input and memory based on past experience with falling objects. Feedback control comes into play after the ball makes contact with the hand. Four line traces are shown as a function of time, with elbow and wrist angles in the top line traces, and muscular activity (electromyographs) at the bottom. The preparatory response (as seen in the muscle activity before impact) is essential for catching the ball. Adapted from Ghez & Krakauer (2000).

Helmholtz first outlined the notion of an efference copy. When humans make a saccadic eye movement, the visual image transiently shifts over the retina. Yet, we perceive the world as remaining stable. Helmholtz suggested that a copy of
the motor command responsible for the saccade is integrated with the visual motion input from the retina causing us to perceive the visual scene as motionless (Sperry 1950; von Holst & Mittelstaedt 1950). The following demonstration supports this notion. If we move our eyeball with the gentle pressure of a finger, then we do see the visual scene as shifting. When the motor command to the eye muscles is missing, so is the efference copy. In this case, image motion translates into perceptual motion.

The suggested explanation of the anticipatory response in the rat HD system is conceptually similar to Helmholtz’s explanation of perceptual stability during a saccade. The size and direction of a saccade is known in advance of the actual movement of the eye, so the shift in the retinal image is predictable. According to a recent proposal, the pathway carrying the motor command information from the superior colliculus via the thalamus to the frontal cortical areas (frontal eye field) contains the mechanism responsible for maintaining visual stability during saccades (Sommer & Wurtz 2006). Interestingly, the involvement of the thalamus has also been suspected in the anticipatory response of HD neurons in rats (Bassett et al. 2005).

Predictive responses in single cells were discovered in the lateral intraparietal cortical area of rhesus monkeys when the animals made a saccadic eye movement to a visual target (Duhamel et al. 1992). The finding was that, when a visual stimulus was switched on within the parietal cell’s receptive field, the cell responded with a delay of 70 ms. If, however, the same stimulus, which was initially outside the cell’s receptive field, was brought into the cell’s receptive field by a saccade made by the monkey, then the response of the same cell appeared 80 ms before the onset of the saccade. Thus, these neurons are capable of responding to stimuli located at two different regions of the retina. First, the region that is the classically defined receptive determined when the eye is stationary. Second, the region that is spatially distinct from the first, to which these neurons respond transitorily around the time of the saccade. We suggest that transitorily the information from the second region of the retina travels to the neurons in the parietal cortex along a diagonal pathway, which is a pathway that is momentarily activated around the time of the saccade (see below).

A mechanism similar to the extensively studied HD system in rats is likely to be present in humans (Zhang 1996). Indeed, localization mechanisms for other end effectors, such as the arm or the hand, are known to exist in the premotor and posterior parietal cortex of monkeys (Taira et al. 1990; Graziano 1999). But are these localization mechanisms anticipatory? Human psychophysics experiments have shown that this is the case. In one experiment (Nijhawan & Kirschfeld 2003) demonstrating anticipatory arm localization, the subjects actively moved a hand-held metal rod (220 g) to which a light-emitting diode (LED) was attached. There was no visual feedback from the arm/rod. The LED was flashed (for 1 ms) once during the movement. The subject judged whether the flashed LED appeared to lead or lag the kinaesthetically sensed position of the rod. The finding was that when the LED was physically aligned with the rod it appeared in a position that trailed the rod by approximately 128 ms.

Although the neural mechanism responsible for this ‘motor flash-lag’ effect is keenly debated, the following explanation is plausible: during movement of a limb, a human subject must monitor the limb’s instantaneous and immediate future positions precisely via anticipatory control (Georgopoulos 1986). Without
this anticipation, and, in particular, if there were uncompensated delays in the sensorimotor pathway controlling the limb, the subject’s kinaesthetic sense of the arm will lag behind. Thus, the subject will be unable to perform actions such as reaching and grasping an object. So, just as the HD neurons are predictive, there are mechanisms in humans devoted to representing the kinaesthetic position of the moving arm after the delays have been taken into account. Thus, the representation of the arm position during movement is the predicted (current or future) position. By contrast, a flash is an unpredictable event, so its visual representation constructed from bottom-up input is delayed by a significant fraction of a second. So, the flash is seen in a position, which the moving rod has already crossed, which causes the flash-lag effect.

What is responsible for a predictive response of HD neurons? Is motor outflow necessary? In a further study, Bassett et al. (2005) experimentally tested the hypothesis that the predictive response of the HD neurons in the ADN was linked to the corollary discharge of motor neurons responsible for moving the head. Clearly, a motor command for moving the head and the accompanying corollary discharge occur only for actively produced movements. Thus, the authors asked if active movement of the head was necessary for observing the predictive response of HD neurons. In the study, the rats were moved passively, either by hand or by a cart, while recordings were made from the HD neurons. To the surprise of the authors, not only did the anticipation in the ADN neurons persist during passive movement it actually increased! Thus, under passive movement conditions, the HD neurons predicted the head direction farther into the future. This study clearly shows that there are non-motor contributions to the predictive response of HD neurons. This suggestion is consistent with a previous proposal of predictive mechanisms within the sensory pathways. This proposal (Nijhawan 1994, 1997, 2002, 2008) suggests that human perceptions are built on the basis of expected sensory inputs (not actual instantaneous inputs) when the information regarding how the input is changing over time is available to the nervous system.

6. Sensory prediction: motion extrapolation in diagonal pathways

When a human subject views a moving object, and a flash is presented in alignment with the moving object, then the flash is seen in a position that trails behind the moving object. This visual flash-lag effect has been repeatedly observed (figure 4). Since there is no explicit involvement of the motor system, a likely cause of this effect is the prediction in the sensory pathways delivering an anticipated position of a smoothly moving object.

Another version of the flash-lag effect leads to the so-called colour decomposition effect. It is well known that a region where red and green stimuli overlap, with appropriate intensities of the red and green stimuli, will be perceived as yellow (additive colour mixture). Using the flash-lag display one can ask: will red and green stimuli ‘mix’ to yield yellow in a flash-lag display in which a green moving stimulus is presented with an overlapping red flashed stimulus? In an experiment, observers viewed a moving green bar with an overlapping flashed red line such that the mixture of the two produced a ‘yellow’ stimulus at the retina. Nonetheless, observers perceived the colour of the flashed line, which
was perceived as lagging behind the green bar, as red (Nijhawan 1997). This finding suggests that the predictive mechanisms may be located very early in the vision pathway, for which there is evidence (Berry et al. 1999).

These findings suggest the notion of a diagonal pathway that we mentioned above and will now outline. We will first describe the concept of the diagonal pathway by drawing a contrast between the differential latency (DL; Whitney & Murakami 1998) and the motion extrapolation (ME) accounts of the flash-lag effect. The DL model suggests that the moving object appears ahead of the flash because moving objects are processed more quickly than flashes. One possible basis for faster processing of moving objects could be ‘neural priming’ (Jancke et al. 2004). When a given neuron is stimulated, there is a ‘spread of activation’ via lateral connections (shown by horizontal lines in figure 1a) to neighbouring neurons. New incoming retinal signals could then more quickly stimulate these ‘primed’ cortical neurons in the path of motion, thus effectively reducing the latency with which these neurons respond to external input.

Consider a network of neurons A, B, C, ..., F of the retina connected to neurons in the cortex $A'$, $B'$, $C'$, ..., $F'$ (figure 5a). Once a retinal neuron receives a signal from the world (a stimulus) it sends action potentials to its counterpart in the cortex. Thus, neuron A communicates with $A'$, neuron B with $B'$ and so on. In addition, $A'$ can communicate with $B'$, and $B'$ can communicate with $C'$ and so on. As an object, in the present example a car, moves in the world it causes A, B and C to send action potentials to $A'$, $B'$ and $C'$ in succession. Suppose a moving car arrives at C and then D. Once the action potentials generated at C have arrived at $C'$, $C'$ communicates with $D'$ and ‘primes’ it for incoming action potentials. Thus, a primed $D'$ reaches its threshold of activation more quickly. In our example, the activation of $D'$ constitutes the perception of the car at D.

Figure 4. In this version of the flash-lag display, an observer holds his/her eyes still and views a black ring moving on a circular path. The figure shows three discrete positions of the ring moving in the direction of the arrow. A white disc is flashed in the centre of the ring at the 3 o’clock position when the ring arrives at that position. The observers perceive the white disc in a lagging position, as depicted by the ‘percept’ (Nijhawan 2001). The percept shows some additional details of what the observers see: a grey crescent-shaped space (called the ‘perceived void’), and illusory white/grey and black/grey edges that are not present in the world (or the retinal image).
The main weakness in the DL account is that a relatively large latency difference of approximately 80 ms is necessary to explain the flash-lag effect. As flashes are extremely effective in stimulating the visual system, it is difficult to imagine that flashes would take more than a few milliseconds longer to be processed than moving objects. Furthermore, 80 ms is a large time difference that should be relatively easy to measure neurophysiologically. Yet, a majority of neurophysiological studies do not show more than a few millisecond latency differences between the moving and flashed objects. Furthermore, the sign of the difference is not consistently one of the other. So, one must look for another explanation of the flash-lag effect.

The ME model using a diagonal pathway is quite different. Consider the two rows of neurons again. The crucial difference between the two models can be appreciated by considering what happens when the moving car arrives at position C. One effective way to compensate for neural processing delays for moving objects would be for neuron C to send the action potentials diagonally to D’ (figure 5b). Compensation would occur when the action potentials arrive at D’ precisely when the moving car arrives at D. The model containing two rows of neurons is overly simplistic. In a more realistic model, there would be many rows consisting of many neurons and connections between the neurons. However, the only requirement for the ME model would be that there are diagonal pathways,

Figure 5. A model to explain the difference between the DL and the diagonal pathway accounts. Neurons in two rows have a one to one mapping. The neurons A’, B’, C’, ..., F’ (the cortical row) receive action potentials from neurons A, B, C, ..., F (the retinal row). An object moving in the world (a car) arrives in position D at the instant depicted. (a) DL model. Owing to the communication between C’ and D’ via lateral interaction (horizontal line), D’ is primed, so its potential crosses the threshold more quickly (in less than one-twentieth of a second) as the signal from D arrives. If neuron D’ were not primed the delay would be one-twentieth of a second. (b) Diagonal pathway model. Owing to the communication between B and C (horizontal line), C sends the action potentials to D’ along a ‘diagonal’ pathway. Compensation is achieved when the action potentials arrive at D’ when the car arrives at D.
and the neural signal speed along the horizontal component of the diagonal is high compared with the speed of the moving object. Both these requirements are easily fulfilled. Segments of horizontal and vertical pathways, well known to exist in the visual system, form the ‘diagonal’ pathways. There are previous models for horizontal routing of inputs in the vision pathway (Anderson & Van Essen 1987).

Can the diagonal pathway be implemented more realistically in the nervous system? Below we combine previous findings (Berry et al. 1999) and the special properties of starburst amacrine cells in the retina, to outline a circuit with diagonal pathway in the visual system. A starburst cell has a very unique structure: its soma is very small and it does not generate spikes; its dendrites, on the other hand, very rich and show radial symmetry of 360° in space (Euler et al. 2002). The response of a starburst cell to stimuli mainly occurs at the dendrites; in particular, on the distal parts of the dendrites at some distance from the soma. When light signals move across a starburst cell, a [Ca²⁺] signal is generated at the distal part of the dendrites, which subsequently generate inhibitory currents in the ganglion cells. Different sectors of the 360° spread of amacrine cell dendrites work rather independently of each other, and each of these sectors has a different preferred motion direction. This unique structure of dendrites enables a starburst cell to sense the motion direction of an input for all possible directions (from 0 to 360°). In addition, the responses of dendrites are distinguished on the basis of whether light signals are moving towards or away from the cell body—motion away from the cell body elicits a larger response. This property of asymmetric responses is critical for starburst cells to selectively modulate the population activity of ganglion cells, and hence to dynamically route the visual information.

To elucidate more clearly how this scheme works, we consider a network of ganglion cells. The real network is two dimensional, but, for simplicity, here we consider a stimulus moving in a straight line and only study the one-dimensional case. The network dynamics is given by

\[
U(x, t) = g(x, t) \int_{-\infty}^{x} dx' \int_{-\infty}^{t} dt' K_E(x-x', t-t') E(x', t'),
\]

\[
g(x, t) = \exp \left[-A \int_{-\infty}^{x} dx' \int_{-\infty}^{t} dt' K_I(x-x', t-t') I(x', t') \right],
\]

\[
r(x, t) = F[U(x, t)],
\]

where \(U(x, t)\) represents the synaptic input at time \(t\) to the ganglion cells located at the retinotopic position \(x\), whose value is determined by the following two factors: (i) the excitatory input, \(E(x, t)\), which contains the spatial information of the stimulus and (ii) the gain control function, \(g(x, t)\), which represents the modulation contribution from the inputs of starburst cells. A previous model used similar dynamical equations (Berry et al. 1999). Here, the difference is that we link the activities of starburst cells to the gain function.

We choose the following network parameters. The final results are, however, qualitatively independent of these details.

The firing rate \(r(t)\) is a threshold-nonlinear function of \(U(x, t)\), i.e.

\[
r(t) = \begin{cases} 
0, & \text{if } U < 0, \\
U, & \text{if } U \geq 0.
\end{cases}
\]
The spatio-temporal filter for excitatory inputs is

\[ K_E(x, t) = A(x) e^{-at}, \]

where \( A(x) \) is a constant for \(-d < x < d\) and zero otherwise, reflecting the receptive field of ganglion cells. \( K_E(x, t) \) at first increases with time and then decreases to zero, reflecting the adaptation effect.

The excitatory input, \( E(x, t) \), is determined by the stimulus location, \( z(t) \). We choose it to be of a small Gaussian bump centred at \( z(t) \),

\[ E(x, t) = \alpha \exp \left[ -\frac{(x - z(t))^2}{2a^2} \right]. \]

The spatio-temporal filter for inhibitory inputs is given by

\[ K_I(x, t) = A(x) e^{-bt}. \]

Without loss of generality, we consider the stimulus moving in the direction in which the value of \( x \) increases with time (say, from left to right). Suppose the stimulus is at the position \( z \) at time \( t \), then the inhibitory input it generates at the location \( x \) is given by

\[
I(x, t) = \begin{cases} 
  b_1 |x - z + R| e^{-(x-z-R)^2/2b^2}, & \text{if } x < z, \\
  b_2 |x - z - R| e^{-(x-z-R)^2/2b^2}, & \text{if } x \geq z.
\end{cases}
\]

The form of \( I(x, t) \) is chosen to satisfy the following response nature of starburst cells.

(i) \( b_1 > b_2 \), which implies that the inhibitory current is larger when the light signal moves outwards from the soma \((x < z)\) than that for the light signal moving towards the soma \((x > z)\).

(ii) \( I(x, t) \) has the largest values at \( z = x \pm R \), i.e. when the light signal is at the distal parts of the dendrites (where \( R \) is the radius of the cell), and has small values at \( z = x \), i.e. when the light signal is at the soma (figure 6b).

Figure 6 illustrates how the scheme using the diagonal pathway works. Without the modulation from starburst cells, the population activity of ganglion cells has a symmetric bell shape (figure 6a), as previously noted (Berry et al. 1999). Starburst cells induce asymmetric inhibitory currents to ganglion cells, and they have larger inhibition for those ganglion cells located on the left-hand side of the instantaneous position of the stimulus than that on the right-hand side of the stimulus (for a stimulus moving from left to right). Consequently, the gain function has smaller values for those ganglion cells on the left than on the right-hand side of the stimulus. This shifts the spatial profile of the population activity of the ganglion cells to the right (figure 6c). As a result, the retinotopic representation of the stimulus position leads its actual position. Considering that there is retinotopic connectivity between the retinal ganglion cells and cortical neurons, the above scheme achieves the diagonal pathway for information transmission.
The size of skewness in the population activity of ganglion cells, i.e. the amount of compensation, increases with the motion speed of the stimulus. This is plausible, since a faster moving light signal should recruit more starburst cells to inhibit the responses of ganglion cells lying on the left-hand side of the instantaneous position of the stimulus. If the network parameters are properly chosen, we can achieve the amount of position shift of the bump matching the speed of moving images, so that a perfect compensation can be achieved. The amount of position led in the retina can be adjusted to compensate a fixed transmission delay to the visual cortex.

7. Conclusions

Neural delays are pervasive in the computations associated with the brain function. If these delays were not compensated then the action of animals would not be adaptive. Similarly, if compensation did not occur in the sensory pathways, e.g. the vision pathway, then our perception of time-dependent stimuli, such as a moving stimulus, will trail significantly behind the stimulus’s true position. Thus, the neural systems require mechanisms compensating for time delays. A plausible strategy to achieve compensation is that the neural systems use prior information of a moving stimulus to predict its future position, so that the transmission delay is cancelled out.

In sensory systems, the diagonal pathway approach is likely to be adopted. Based on the prediction, the neural system dynamically selects a route, which sends the neural signal directly to the future position of the moving stimulus encoded in the cortex, compensating for the distance the stimulus will travel in
the delay period. The details of the compensation mechanisms may vary with the part of the nervous system in which they are implemented and the function that part serves (Nijhawan 2008).

On the surface, neural delays appear to be due to the large amount of time consumed in signal transmission. We suggest that these delays are part and parcel of neural integration and temporal coding, two well-known strategies adopted by the nervous system for performing computations. We suggest, furthermore, that these strategies could not have been adopted by the nervous systems if the delays that these strategies incur could not be compensated. So, there is an intimate relationship between the computations that involve time delays and computations that compensate for those delays. In conclusion, we suggest that the neural delays and the associated compensation mechanisms are not a simple by-product of neural anatomy and physiology, but belong to a general principle of how the brain carries out computation efficiently in the spatio-temporal domain. Understanding the relationship between the neural delays and the neural compensation will help us to establish a complete picture of how information is processed in the neural systems.

References


*Phil. Trans. R. Soc. A* (2009)


