(Appeared in The Austrian Society for Cognitive Science Technical Report 97-01. 176-183)

Response Selectivity, Neuron Doctrine, and Mach's Principle in Perception.

Ken Mogi
Physiological Laboratory
University of Cambridge
Downing Street
CB2 3EG U.K.

Present Address
Sony Computer Science Laboratory Inc.
Takanawa Muse Bldg.
3-14-13, Higashigotanda
Shinagawa-ku, Tokyo, 141-0022 Japan
Tel +81-3-5448-4380
Fax +81-3-5448-4273

Abstract

I discuss the principle that bridges neural firing and perception. I start from the assumption that in order to understand perception, the state of neural firing in the brain is necessary and sufficient (the neuron doctrine in perception). I argue that the concept of response selectivity, currently the *de facto* central dogma in explaining the relation between neural firing and the brain, is incompatible with the neuron doctrine. I suggest that we start instead from Mach's principle as applied to the neural correlates of perception. I propose to define a percept as an interaction-connected firings of neurons, not as a single (or an ensemble of) neuron(s) which selectively respond(s) to a particular set of stimulus, as is assumed under the paradigm of response selectivity. This definition of percept by necessity leads to an interesting argument about the neural basis of psychological time, namely the principle of interaction simultaneity. Finally, I discuss the relevance of the twistor formalism to the foundations of neuropsychology.

1. Introduction

Perception can be approached from two points of view. One perspective is concerned with the subjective nature of perception, including ultimately such question as qualia (Chalmers 1995). In another, perception can be regarded as the computational process in the brain. For example, the currently much debated "binding problem" (e.g. Singer & Gray 1995) can be approached alternatively as a problem of the subjective integrity of perception, or as one of the computational process which integrates information represented in the various areas of the cortex. In order to study perception as an empirical science, the computational viewpoint is crucial. On the other hand, our ultimate interest in perception is propelled by its subjective nature, which lies at the core of the so-called mind-body problem.

In this paper, I suggest that the two approaches to perception can be successfully integrated by considering the following question; what is the natural framework for describing the dynamics of the neural network in the brain? Here, by "natural framework" I mean one in terms of which the dynamical evolution of the neural network can be described in a *causal* manner. The construction of the space-time structure that describes the dynamics of the neural network in a causal manner is a non-trivial problem. I critically review the idea of response selectivity as is applied to

neuropsychology. I suggest to adopt "Mach's principle in perception" as the basic principle that bridges neural firing and perception. I then go on to suggest how the neural correlate of a percept (element of perception) should be defined. This definition will lead to an interesting relation between perceptual time and the dynamics of neural network. Finally, I put forward the conjecture that the perceptual space-time actually corresponds to a twistor-like space constructed from the causal relation between neural firings.

2. The Neuron Doctrine in Perception

Barlow (1972) applied the neuron doctrine (e.g. Sherrington 1941) to the problem of perception, and proposed the neuron doctrine in perception. His proposal consisted of 5 dogmas. Of special interest here is the first and fourth dogma. Namely, that

a description of that activity of a single nerve cell which is transmitted to and influences other nerve cells, and of a nerve cell's response to such influences from other cells, is a complete enough description for functional understanding of the nervous system

and that

just as physical stimuli directly cause receptors to initiate neural activity, so the active high-level neurons directly and simply cause the elements of our perception.

Barlow's neuron doctrine has often been associated with the idea of a "grandmother" cell, especially in the context of the fourth dogma. Recent evidence of "sparse coding" in, for example, the temporal cortex of monkeys (Rolls & Tovee 1995) has been cited as an argument against the "grandmother" cell idea. However, the neuron doctrine itself does not necessarily imply a grandmother cell type coding. I propose to reformulate the neuron doctrine as follows.

Our perception is directly invoked by the neural firings in the brain. A non-firing neuron is as good as non-existent as far as perception is concerned. The characteristics of our perception should be explained by the nature of neural firings only.

The neuron doctrine, when phrased in this form, appears to be a reasonable starting point for neuropsychology at present. There are ideas, for example, that the molecular level processing at the microtubules participate in the conscious process (e.g. Hammerof and Penrose 1996). However, the available data, especially those from single unit recordings (e.g. Newsome et al. 1989) seems to be compatible with the idea that neural firings are necessary and sufficient to invoke perception, and the cellular processes, such as the release and binding of neurotransmitters at the synaptic cleft and the subsequent postsynaptic flow of ions across membranes, influence perception only as far as they affect the neural firing.

Although the neuron doctrine as rephrased above does seem to give a correct starting point for neuropsychology today, the most profound question, namely why the neural firing plays such a special role in our perception, and indeed our consciousness, remains unanswered. From the dynamical point of view, the non-linearity and the all-or-none character involved in the action potential generation is likely to be at the basis of the crucial importance of neural firing in perception. In this view, it is of interest to note that there is no sub-neural processes known at present which demonstrates the same degree of non-linearity or all-or-none character as the action potential generation.

3. Response Selectivity

"Response selectivity" is a concept of a central importance in neurophysiology today. For example, in the primary visual cortex (V1), we find neurons that selectively respond to a bar with a certain orientation (Hubel & Wiesel 1962) (Fig.1(a)). In areas MT, V4, IT, we find neurons that respond to motion, color (in the context of "color constancy" see Land (1983)), and form, respectively (Newsome et al. 1989, Zeki 1980, Tanaka 1993). As we go to the higher visual areas, we find neurons with more complex response selectivities, and larger receptive fields. One idea that emerges is the assumption that when a neuron with a response selectivity to a particular visual feature fires, the perception of that feature occurs. For example, when a neuron selectively responsive to a bar slanted by 45 degrees to the right fires, the perception of the slanted bar would be invoked. When a neuron selectively responsive to a "face" fires in area IT, the perception of a "face" is invoked, and so on.

There is, however, a fundamental flaw in this line of argument, which becomes apparent when one tries to answer the following question. "When a neuron selectively responsive to a feature A fires, how does the brain (or the subject) know that it is selectively responsive to feature A?" The visual feature space is vast and complex. The fact that a neuron fires vigorously to a particular feature A does not necessarily mean that the neuron has the response selectivity to feature A only. In fact, in order to establish the response selectivity of a neuron, every possible visual feature should, in principle, be presented to the neuron. Of course, this is impossible both in practicality and in principle. Moreover, the neuron doctrine dictates that our perception is constructed based on the neural firings at a particular psychological moment. It is impossible, just based on the firing of the neurons at a particular time, to establish the response selectivity of a particular neuron. The reason for this being that response selectivity is based on the idea of an statistical "ensemble", the set of all possible stimulus-response relations. If our perception is based on response selectivity, the brain must somehow have an instantaneous access to every element within the ensemble when only a sample from it is presented as stimulus. Such an assumption seems to be implausible.

In general, as we go higher up the visual system, the more difficult it becomes to define the response selectivity of a particular neuron in an operational manner (see Tanaka 1993 for example). Even if a neuron seems to fire rigorously only when a frog is presented, for example, it is virtually impossible to establish that the neuron is selective to a frog only, as the visual stimulus space is vast and complex (Fig.1(b)). This is in contradiction with the idea that the neurons in the higher visual areas play a crucial role in our perception through their response selectivity. Even in the case that our perception is evoked by the successive neural firings from the lower visual areas to the higher visual areas, the significance of the neurons in higher visual areas becomes obscure, if indeed the response selectivity plays an essential role in perception.

From these considerations, I conclude that response selectivity cannot be the foundation for the relation between neural firing and perception. Some other bridging principle(s) should come into the picture.

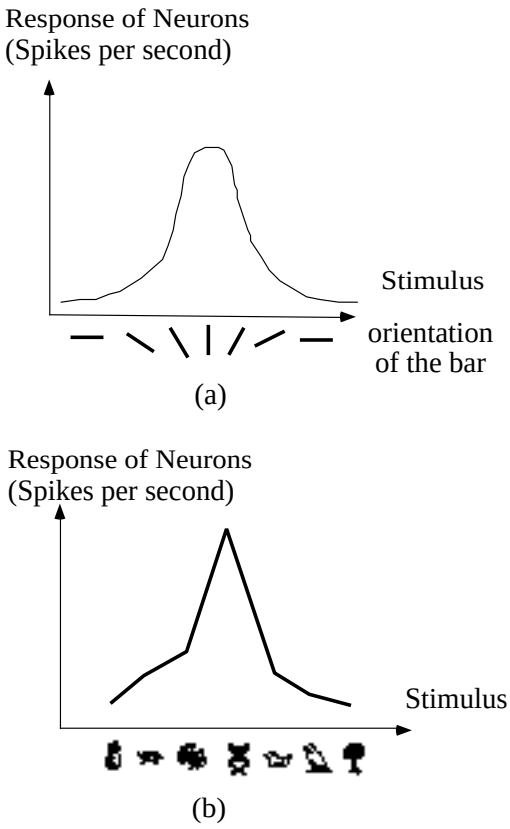


Fig.1 Response Selectivity

4. Mach's Principle in Perception

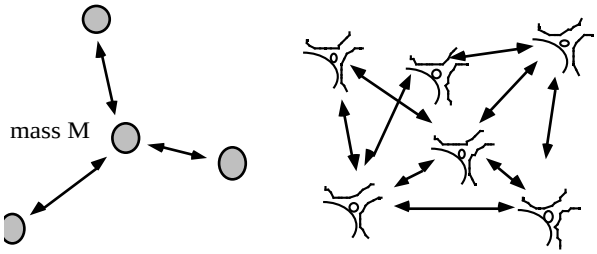
Ernst Mach (1838-1916) was a physicist, philosopher, psychologist who had a major influence on Albert Einstein in his development of the theory of relativity. "Mach's principle" states that the mass of a particle is determined by its relation to all the other particles in the universe (Fig.2 (a)). If there was only one particle in the universe, it is meaningless to question how large its mass is. In a nutshell, the idea behind Mach's principle is that the properties of an individual is determined by its relation to other individuals in the system.

A similar line of thought is relevant when we consider the neural correlates of perception. Namely, a neural firing plays a particular role in our perception, not because it is selectively responsive to a visual feature (the idea behind the neuropsychological

application of response selectivity), but because the neural firing is related to other neural firings in the brain in such a way that the particular role in perception is endowed on the neural firing in question. Namely, I propose to formulate the following principle.

In perception, the significance of a firing neuron is determined by its relation to other firing neurons at that psychological moment.

We shall call this idea "Mach's principle in perception" (Fig.2 (b)). For example, suppose a neuron in area IT fired, and a perception of "rose" is invoked in our mind. In this case, the perception of "rose" is invoked not because the neuron selectively responds to the presentation of a rose, but because the neural firing in question is endowed with the property of "rose" through its relation to other neural firings in the brain. Specifically, the cluster of neural firings connected by



(a) Mach's Principle

(b) Mach's Principle in Perception

Fig.2 Mach's principle and
Mach's principle in Perception

interaction through action potentials that is initiated in area V1 and leads up to the neural firing in area IT codes the perception of "rose". It is meaningless to consider a single neural firing in isolation and ask its significance in perception, even if the response selectivity of that neuron could be established unequivocally.

Under the scheme that I have put forward above, a percept is coded not by a single neural firing, but by a cluster of interaction-connected neural firings. This is the bridging principle that is consistent with the neuron doctrine in perception, and Mach's principle in perception. Namely, a percept is coded non-locally. Neural firings in spatially distant areas of the brain are integrated into a cluster through their mutual interactions, and form a percept.

That a percept is defined not as a single (or an ensemble of) neuron(s) which selectively respond(s) to a particular set of stimulus, but as a cluster of interaction-connected neural firings, has an immediate impact on some important issues in perception, such as the binding problem (Damasio 1989;Gray et al. 1989;Malsburg 1981;Singer & Gray 1995). The various visual features are represented by the neural firings in spatially separate areas of the brain. For example, color is coded in area V4, motion in area MT, form in area IT, and so on. However, our visual perception is not a collection of fragmentary features, but a coherent world view. The binding problem questions how the brain integrates the visual features into a coherent picture of the world.

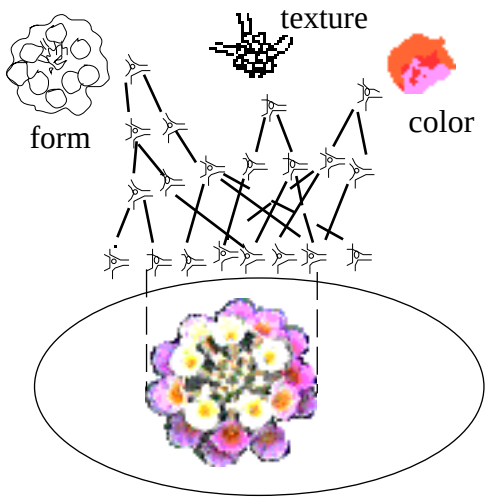


Fig.3 The Binding Problem

The binding problem arises from the view that perception is coded locally, by the firing of a neuron(s) that respond(s) selectively to a particular subset of features. As the features is coded locally, it becomes necessary then to question how the brain integrates the locally coded representation of features.

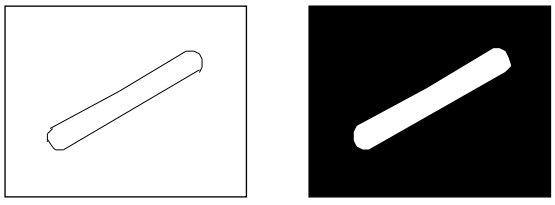
Under the view that a percept is coded non-locally, as a cluster of neural firings arising in area V1 and leading up to higher cortical areas such as V4, MT, and IT, the binding problem is not such an acute one any more. What is likely to be happening is that the neural firings in V1 function as a kind of "address" for the visual features to be integrated (Fig.3). As the percepts now have "tags" of interaction-connected neural firings

originating from area V1, they are embedded with the necessary information to be organized into the basically retinotopic visual field, which functions as the frame of reference for our coherent world view. Although the details are still to be worked out, and the binding problem does remain a difficult one, it is important that the very formulation of the binding problem assumes a local coding, which for reasons discussed above seems not to be the case.

The distinction between the excitatory (e.g. glutamate) and inhibitory (e.g. GABA) connections now becomes important. Specifically, it appears that only excitatory connections are included explicitly in the cluster of neural firings that forms a percept. Inhibitory connections affect the formation of percepts only indirectly. For example, in the color constancy mechanism (e.g. Land 1983), inhibitory inputs from surrounds will lead to a non-formation of a percept of a color in the center. Inhibitory connections have significance in that they can "veto" the formation of a percept. However, inhibitory connections are not included explicitly in the cluster of neural firings that forms a percept.

In order to see the intuitive meaning of this arrangement, consider a white bar in a black surround. In order that the white bar is a bar, it is necessary that the area surrounding the bar is black, rather than white. If the surround was white, then the bar would not be a bar (Fig.4(a)). So the surround contributes to the formation of a white bar by not being white. However, the black surround does not constitute an explicit part of the percept "white bar" (Fig.4(b)).

The idea that excitatory connections and inhibitory connections contribute differently in the formation of a percept should be grounded in the mathematical description of the dynamics of the neural network, as is outlined in section 6.



(a) (b)

Fig.4 White Bar in a Black Surround

5. Principle of Interaction Simultaneity

The definition of the neural correlate of a percept in the previous section lead to an interesting relation between the dynamics of neural networks and the time in our perception.

Albert Einstein, in his first paper of relativity theory published in 1905, stated thus.

One thing should be remarked here. Such a mathematical description is physically meaningless unless the way we construct time is made clear. All our judgements about time is one about events that occur simultaneously.

We see below that the argument that led us to the definition of a percept as an interaction-connected cluster of neural firings leads to an operational definition of simultaneity in perception, namely the psychological "now".

Let us start from the neuron doctrine in perception. Namely, we assume that knowledge about the firing neurons is necessary and sufficient to determine the content of perception. The concept of "interaction simultaneity" dictates how to determine the nature of psychological time in a way consistent with the neuron doctrine. Under the principle of interaction simultaneity, when a neural firing and another neural firing are connected by interaction (i.e., action potential propagation and subsequent synaptic interaction), these are considered to be simultaneous events (Fig.5). It takes a finite length of time (say 5 milliseconds) for the effect of a neural firing to propagate to a postsynaptic neuron. However, under the principle of interaction simultaneity, these firing events should be regarded as "simultaneous". We call the time parameter thus constructed "proper time" and write it as τ. When the presynaptic neuron fires at time t, and the postsynaptic neuron fires at time $t + \Delta t$ (as a result in part of the EPSP caused by the presynaptic firing), we assign the same proper time τ to the pre- and postsynaptic events.

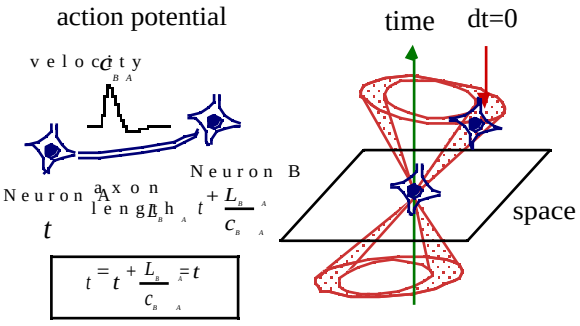


Fig. 5 Principle of Interaction Simultaneity in a Neural Network.

Interaction simultaneity is derived from a more fundamental principle, that of "causality". Here, "causality" is taken to mean that given the state of the system at proper time τ, we are able to derive the state of the system at a slightly later time $\tau + \Delta\tau$. Schematically,

$$\Omega(\tau) \rightarrow \Omega(\tau + \Delta\tau)$$

where $\Omega(\tau)$ is the state of the neural system at proper time τ. We need to use the proper time τ in order to describe the dynamical evolution of the neural network in a causal way. Note that a system described by a differential equation in time satisfies the above definition of causality. Also, this concept of causality encompasses both the deterministic (e.g. Newtonian) and stochastic (e.g. quantum mechanical) dynamics.

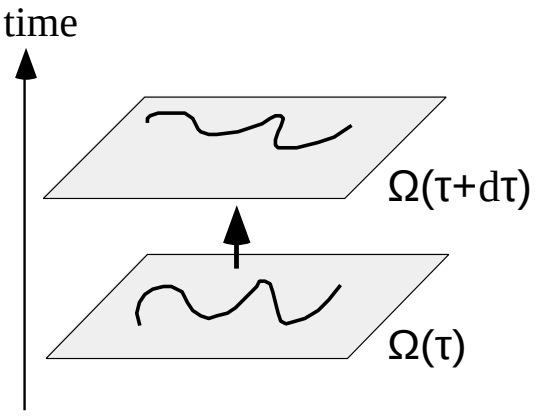


Fig. 6 Causality

Interaction simultaneity is based on the idea that if we are to derive the properties of our perception from neural firings, we should not adopt the position where we "observe" the neural firings from "outside the brain". If we observe the brain from the outside, we can describe the dynamical evolution of the neural network with any desired temporal accuracy. We may, for example, describe the release and diffusion of neurotransmitters at synapses with submillisecond temporal resolution. However, under the neuron doctrine in perception, only the neural firing enter explicitly in our perception. Therefore, properties of our perception should be described without resorting to the idea of an outside observer. Thus, the adoption of the principle of interaction simultaneity is justified.

Note that interaction simultaneity is necessary in order for the definition of a percept as an interaction-connected cluster of neural firings, as is required by Mach's principle, to remain consistent. Suppose that a percept, for example that of a "rose" is formed by the cluster of interaction-connected neural firings from area V1 up to area IT. As it takes a finite length of time for the presynaptic activity of neuron to be transmitted to the postsynaptic activity of neuron and result in its firing, there is a time delay (referred to as *latency* in the literature) involved. However, we subjectively assign a percept to a particular psychological moment. There is no *inherent* passage of psychological time involved in the percept of a "rose". If we consider the percept of a color (*e.g.* red), which is invoked by neural activities leading up to area V4, the assignment of that percept to a psychological moment, in other words the absence of any necessary passage of time *per se*, becomes more evident. The proper time τ constructed from the principle of interaction simultaneity fits such temporal characteristics of the percepts well. Namely, there is no passage of proper time, and therefore of psychological time, accompanying the formation of clusters of neural firings within the brain which underlie our perception (Fig.7)

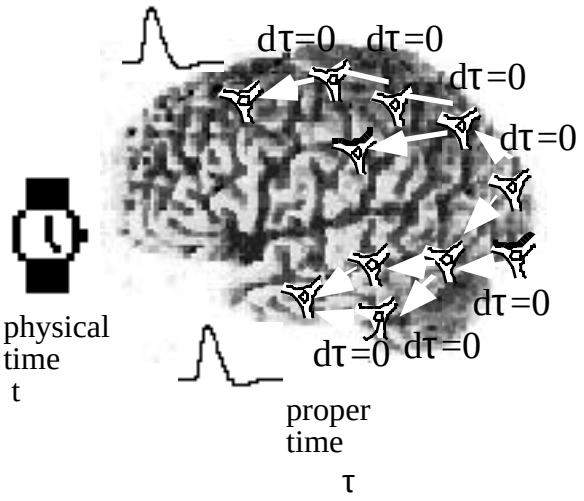


Fig.7 Element of Perception

We can obtain some interesting conclusions about the nature of psychological time. Firstly, the psychological "present" has a finite duration, when measured by the physical time t. The duration corresponds to the transmission delay present when the cluster of interaction-connected neural firings is formed. This would be of the order of ~ 50 ms. This time gives the measure of transmission delay necessary for neural excitation to travel across the cluster of neurons involved in the formation of a percept. In other words, there would be a minimum "unit" of the psychological time, with a duration of ~50ms. Despite the existence of such a finite duration of the psychological "moment", the flow of psychological time is shown to be smooth. Specifically, the displacement between the adjacent "moment" can be made arbitrarily small. This in turn means that there is an "overlap" between adjacent psychological moments. A particular neural firing is shared by the neighboring moments. Schematically, we can summarize the property of the neural system parametrized by the proper time τ as follows (Fig.8).

$$\Omega(\tau) \cap \Omega(\tau') \neq \quad (|\tau - \tau| < h)$$
$$\varepsilon > 0, \ \tau \quad \tau' \text{ s.t.} (|\tau - \tau'| < \varepsilon)$$

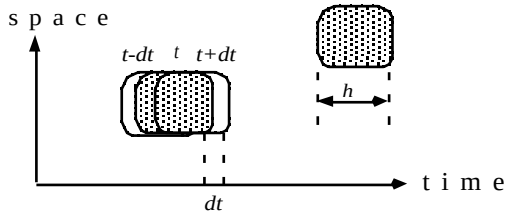


Fig.8 Nature of Time under Interaction Simultaneity

Such a picture of time seems to coincide with the subjective "feeling" of time flow.. Lockwood (1989) refers to a similar model of psychological time. It is of interest that Libet (1985) reports that in order that neural activities enter consciousness, they need to be maintained at least 500 milliseconds. At present it is not possible to make detailed arguments about the neural basis of psychological time, but I believe that the idea of interaction simultaneity is to play a significant role in the construction of future models.

6. Causality and Twistor formalism

Interaction simultaneity is concerned with the construction of the psychological time. This in turn is part of a more general problem, namely, how our perceptual space-time structure emerges from the neural firings in our brain. "Causality" is conjectured to be a leading principle in the construction of the perceptual space-time. Note that the perceptual space-time does not necessarily coincide with the physical space-time in which the neurons are embedded, as is evident from the marked difference in space-time structure of different sensory modalities.

The construction of the space-time in perception is a two-sided problem. One aspect is the nature of the space-time structure in our perception from a subjective point of view. The other is concerned with how to construct our perceptual space-time structure in such a way that within that framework, it becomes possible to describe the dynamical evolution of the neural network in a causal way. The principle of interaction simultaneity gives such a construction scheme.

What then is the mathematical language that we should adapt to describe the neural network dynamics in a causal way? It should be noted that to construct a system of variables and space-time structure that satisfies causality is a highly non-trivial problem, especially when we consider that (1) there is a finite delay in the transmission of signals between the neurons (2) only the neural firing are to enter explicitly in the description of the dynamics. For example, there

is an ambiguity in the meaning of the state of a neuron at proper time τ. If the firings of mutually connected neurons A and B at the physical time t reaches the postsynaptic neurons at the time $t+\Delta t$, the state of the two neurons $A(t+\Delta t)$ and $B(t+\Delta t)$ are simultaneous to the states $A(t)$ and $B(t)$ by the principle of interaction simultaneity. (Recall interaction simultaneity was a consequence of causality). Therefore, schematically we construct

$$(A(t),B(t),A(t+\Delta t),B(t+\Delta t)) \rightarrow (A(\tau),B(\tau))$$

where τ is the proper time. Here, $A(\tau)$ could either mean $A(t)$ or $A(t+\Delta t)$. The same ambiguity exists for $B(\tau)$ One way to avoid this ambiguity is to adopt a convention such that in the product $A(\tau)B(\tau)$, the term on the right side specifies the state at the time t, and the left term specifies the state at the time $t+\Delta t$ Namely, in general we have

$$A(\tau)B(\tau) - B(\tau)A(\tau) \neq 0$$

There are further complications owing to the fact that the time required for the transmission of signals between neurons varies depending on the length and diameter of axons, the duration of synaptic and dendritic delays, and the fact there are excitatory, inhibitory, and modulatory synapses. Specifically, if we look at the post synaptic potential φ_{BA} exerted by neuron B on neuron A, the rise time t_a, decay time t_m, synaptic delay t_s, dendritic delay t_d, and the action potential transmission time $\frac{l_{BA}}{c_{BA}}$, where l_{BA} is the length of the axon of neuron A, and c_{BA} is the transmission velocity, should all be taken into account (Fig.9). These parameters typically take the values

$$t_s <\approx 1ms$$
$$t_d <\approx 1 \sim 2ms$$
$$t_a <\approx 1ms$$
$$t_m \approx 10ms$$
$$\frac{l_{BA}}{c_{BA}} \approx 1ms$$

The variables parametrized by the proper time τ, and the space-time structure constructed upon them, should reflect all these properties.

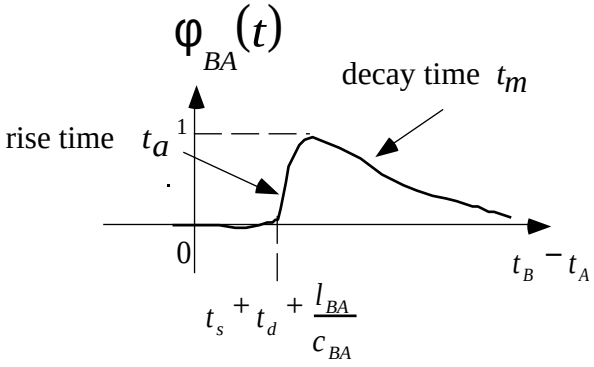


Fig.9 Time Constants in Interaction Simultaneity

Although the idea of interaction simultaneity strikes some similarity with the relativity theory, care should be taken as the parallelism is sometimes misleading. For example, in relativity theory the constancy of the velocity of light is one of the starting assumptions, indeed an axiom. In the case of neural networks the conduction velocity varies, as do other time parameters that describe the synaptic interaction between neurons. It is customary in relativity to define the "simultaneity" between events by the co-ordinate time t. Here, we defined the simultaneity by the proper time τ. So there is no "relativity of simultaneity" involved in the present scheme. In addition, there is no mathematical structure corresponding to the Lorentz transformation in the case of neural networks. Despite these differences, the relation between causality and proper time is basically the same in relativity theory and interaction simultaneity, and this could lead to some interesting consequences.

In order to develop the model of neural basis of perception put forward above any further, we need to have a solid mathematical background. At present, one is yet to be found. Penrose's "twistor" is a hint for the mathematical structure to be developed from such an approach. Penrose constructs a "twistor space" separately from the physical space-time. We consider a spinor field Ω which satisfies the twistor equation

$$\Omega^{(A\ B)}_{A'} = 0$$

Ω can be written in terms of some other spinor field π and a constant spinor ω as

$$\Omega^A = \omega^A - ix^{AA'}\pi_{A'}$$

where x represents the Minkowski coordinates.

Twistor is then expressed by a pair of spinors

$$Z^\alpha = (\omega^A, \pi_{A'}); \alpha = 0,1,2,3$$

A spinor, in turn, is a geometrical object closely related to the metric of the space-time. A spinor can be returned to itself only when it is rotated by 4π. A rotation of 2π sends a spinor into its negative. In relativistic picture, the trajectory of light (null vector) is represented as a straight line in physical space-time (assuming it is a Minkowski, i.e. flat space-time). In twistor space, it is mapped to a point. This can be interpreted in terms of causality structure as follows. The trajectory of light represents the world-line along which causal interaction propagates. In twistor space, the set of points in physical space-time that are connected by the interaction (which is represented by the trajectory of light) is mapped into a point. Therefore, twistor space can be considered to be based on the causal relations between world points in physical space-time (Fig.10 above).

In a nutshell, the idea behind the twistor formalism is to regard the causal connection between points in physical space-time as more fundamental than the space-time itself. The twistor space is in a sense a more fundamental framework for natural law than physical space-time itself. Penrose writes in Penrose & Rindler (1984)

We should think of twistor space as the space in terms of which we should describe physics.

We cannot apply the twistor formalism directly to neural network. However, the basic idea of the twistor approach, namely to regard the causal relation between the individuals as more fundamental than the individuals themselves, corresponds to our arrangement of regarding the cluster of interaction-connected neural firings as the percept, that is, the elementary unit in perception. The nature of our perception is ultimately determined by the dynamics of the neural network. Accordingly, the cluster of interaction-connected neural firings functions not only as an element of perception, but also as an element in the dynamics of neural network. In order to describe such a dynamics, we would need a mathematical structure similar to that of twistor space.

An intriguing possibility is that a twistor-like space can be constructed to describe the dynamics of a neural network, and the space thus constructed corresponds to our perceptual space-time (Fig.10 below)). The differential contributions of the excitatory and inhibitory connections in the formation of a percept is expected to be of importance here. If such a picture is found to be the case, our mind would inhabit the twistor-like space that describes the dynamics of neural networks in the brain. Of course, at present this is merely a conjecture.

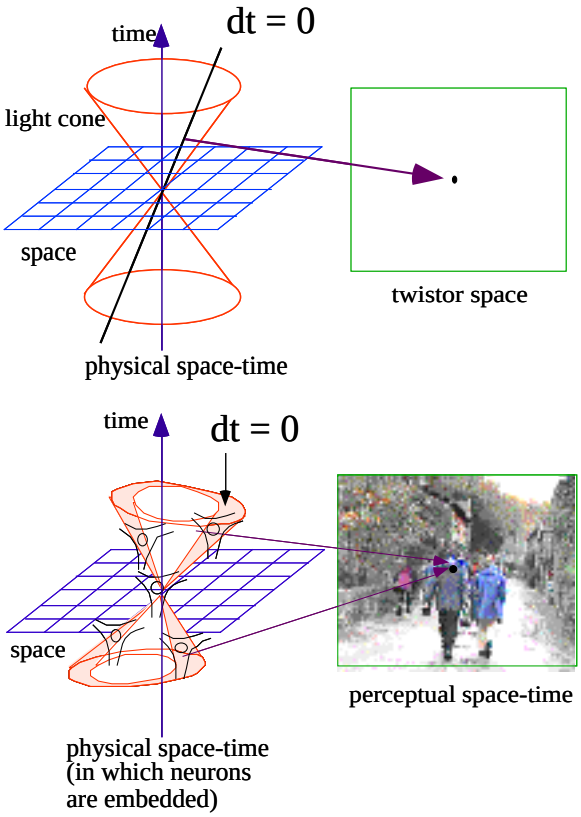


Fig.10 Twistor space and Perceptual space-time

Malsburg, C. von der (1981) The correlation theory of brain function (Internal Report 81-2, Max-Planck-Institute for Biophysical Chemistry).
Newsome W. T., Britten K. H., Movshon J. A. (1989) Neuronal correlates of a perceptual decision. Nature 341 52-54.
Penrose, R. & Rindler, W. (1984) Spinors and space-time vol I ;(1986) Spinors and space-time vol II Cambridge University Press.
Rolls, E. T. & Tovee, M. J. (1995) Sparseness of the neuronal representation of stimuli in the primate visual-cortex. J of Neurophysiol. 73, 713-726.
Sherrington (1941) in Man on his nature. Cambridge University Press.
Singer, W. & Gray, C.M. Visual feature integration and the temporal correlation hypothesis. (1995) Annu. Rev. Neurosci. 18, 555-588.
Tanaka, K. (1993) Neuronal mechanisms of object recognition. Science 262, 685-688.
Zeki, S. (1980) The representation of colours in the cerebral cortex. Nature 284, 412-418.