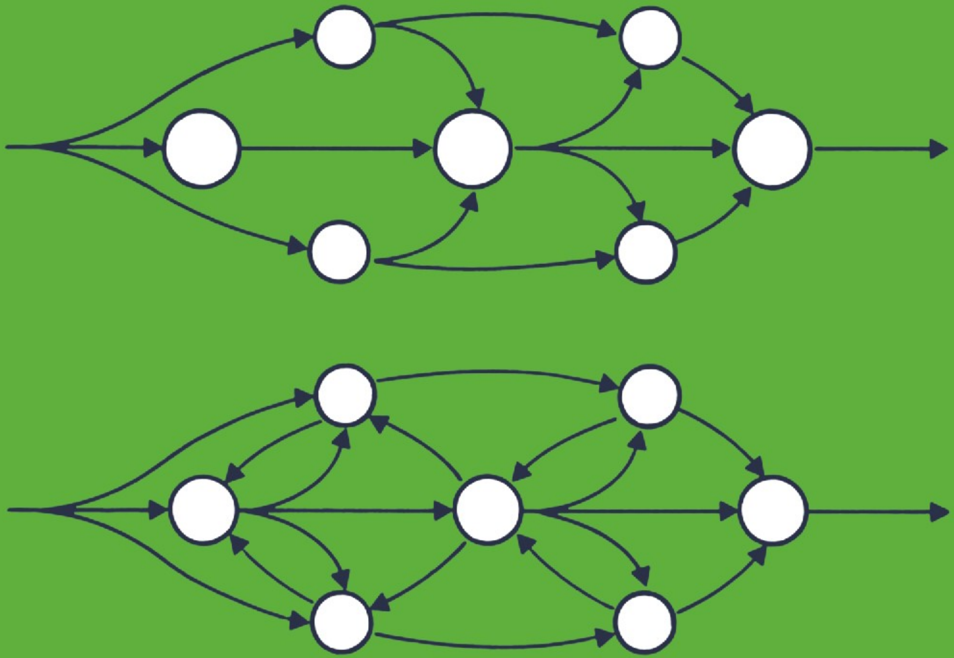
Brain Theory

Edited by
Günther Palm and Ad Aertsen



Springer-Verlag Berlin Heidelberg GmbH

Brain Theory

Proceedings of the First Trieste Meeting on Brain Theory, October 1–4, 1984

Edited by
Günther Palm and Ad Aertsen

With 75 Figures

Springer-Verlag
Berlin Heidelberg New York Tokyo

Dr. GÜNTHER PALM
Dr. AD AERTSEN

Max-Planck-Institut für
Biologische Kybernetik
Spemannstraße 38
7400 Tübingen, FRG

Cover illustration:
Reactive and creative system differing in absence or presence of closed loops. From P. Johannesma, this Volume, page 34, Figure 6.

ISBN-13: 978-3-642-70913-5 e-ISBN-13: 978-3-642-70911-1
DOI: 10.1007/978-3-642-70911-1

Library of Congress Cataloging in Publication Data. Trieste Meeting on Brain Theory (1st : 1984) Brain theory. 1. Brain–Congresses. 2. Intellect–Congresses. 3. Neurology–Philosophy–Congresses. I. Palm, Günther, 1949– . II. Aertsen, Adrianus, 1948– III. Title. IV. Title: Brain theory. QP376.T73 1984 153 85-27664

Softcover reprint of the hardcover 1st edition 1986

Offsetprinting: Beltz Offsetdruck, Hemsbach/Bergstraße.

2131/3130-543210

From Neuron to Assembly: Neuronal Organization and Stimulus Representation

A. AERTSEN[1], G. GERSTEIN[2], and P. JOHANNESMA[3]

1 Introduction

The study of information processing in the sensory nervous system may be viewed as an investigation of images. Let us consider, for instance, the auditory nervous system. Throughout the auditory system, starting at the hair cells in the cochlea and the auditory nerve fibres, through the various stages of the auditory processor, composed of the numerous individual neurons with their different patterns of interconnections, we have what might be called "the neural image of sound" in its different realizations. The external world is paralleled by an internal representation (e.g., Craik 1943, McCulloch 1965). The acoustic environment of an animal, consisting of patterns of air pressure variations at the external ears, is represented and transformed internally by a network of neurons which communicate by complex spatio-temporal patterns of action potentials, the all-or-none events generated by the individual neurons.

1.2 Single Unit Analysis

A central paradigm in the study of the sensory nervous system is that meaningful information regarding its principles of operation can be obtained from experimental investigation of the functional characteristics of its elementary components, i.e., the single neurons. This is reflected in the vast amount of literature on single unit experiments. In the periphery, the single neurons can quite adequately be described by characteristics such as frequency tuning and temporal pattern of firing to stimulation with tones. This type of characteristics, grossly speaking, serves to cover all relevant aspects of the various possible sounds occurring in the external world (e.g., Evans 1974). Moreover the periph-

---

[1] Max-Planck-Institut für Biologische Kybernetik, Spemannstraße 38, 7400 Tübingen, FRG

[2] Department of Physiology, School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

[3] Department of Medical Physics and Biophysics, University of Nijmegen, Nijmegen, The Netherlands

Brain Theory
Edited by G. Palm and A. Aertsen

eral neural image can be characterized by simple organization principles like tonotopy, i.e., an orderly spatial representation of the spectral composition of the acoustic environment. The picture becomes increasingly more complex and shows more and more blanks when progressing towards the more central parts of the auditory system. Classical experiments using "simple" stimuli such as tones and clicks appear to lose their claim to completeness.

A theoretical description and related experimental and data-analysis procedures developed by the Nijmegen group provide for single neurons in more central parts of the auditory nervous system, a functional representation of the neuron as an element of the auditory processor (for a review see Eggermont et al. 1983b). This approach is based on statistical analysis of the relation between the extra-cellularly recorded single unit activity and the presented ensemble of sound stimuli. A question of special interest has been to what extent such a description covers the neuron's behaviour under a variety of different acoustic stimulus conditions (e.g., tones, noise, natural sounds).

The analysis is focused on the evaluation of the spectro-temporal properties of those stimuli which precede the occurrence of action potentials, the neural events, as compared to the characteristics of the complete ensemble of stimuli presented to the animal. This approach, formally related to the evaluation of the second order stimulus-event cross correlation function, results in the "spectro-temporal sensitivity" (STS) as a characteristic of the neuron's stimulus-event relation. In order to get an impression of the extent of stimulus-invariance of this neuron characteristic, the outcome of the correlation analysis has to be normalized with respect to the a priori known spectro-temporal structure of the stimulus ensemble, used to measure it. Under certain conditions, regarding both the neuron's system function and the stimulus ensemble, a stimulus-normalization procedure can indeed be derived using the formalism of nonlinear systems theory. If the spectro-temporal sensitivity of the neuron can be normalized in this way, a "spectro-temporal receptive field" (STRF) can be derived. This STRF forms the functional characteristic of the contribution of this particular neuron to the auditory processor.

Experimental investigations have been made of the receptive field properties of single neurons in the auditory midbrain of the grassfrog (Rana temporaria L.) under a variety of stimulus conditions (tones, noise, species specific vocalizations). It appeared that, even after elaborate stimulus normalization procedures in a considerable number of cases the result cannot be reconciled into a single model of the sinle unit receptive field. This is illustrated in Fig. 1 which shows the spectro-temporal sensitivity for seven neurons from the frog's midbrain, determined with three different types of stimulus ensembles. As a general conclusion it can be stated that in a considerable number of cases the single-unit receptive field turns out to be stimulus-variant, possibly context-sensitive. This, obviously, poses interesting problems when using the receptive field as a conceptual tool in modelling brain function.

Theoretical considerations at the same time have shown that for neuron models which can be characterized as not having any feedback connections

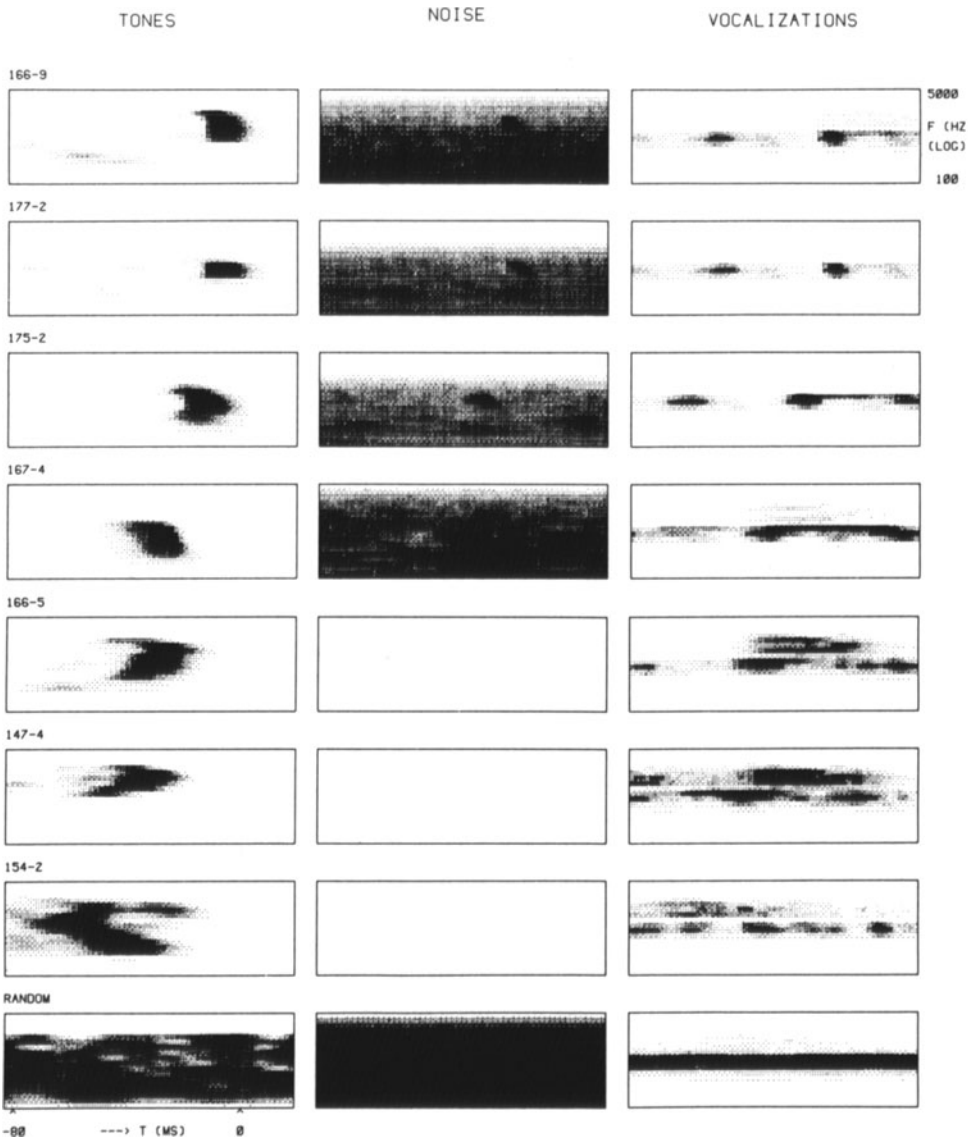


Fig. 1. Spectro-temporal sensitivity of 7 neurons (*upper 7 rows*) in the auditory midbrain of the grassfrog (*Rana temporaria L.*), determined with different types of acoustic stimuli (*3 columns*: tone pips, stationary Gaussian wideband noise and species-specific vocalizations). The average spectro-temporal structure of each of the three stimulus ensembles is shown in the *bottom row*. Further explanation in text. (From Johannesma and Eggermont 1983)

across nonlinear elements, the receptive field can be defined and, in principle be measured in a neat way (Van den Boogaard et al. 1985a). This receptive field then should be a stimulus invariant neuron characteristic. As soon as feedback loops (e.g., through circuits of neurons) enter into the description, a formal definition of a stimulus invariant receptive field becomes much more cumbersome, if not impossible (Van den Boogaard 1985b). The terms "reactive" versus "creative" systems have been coined in this context (Johannesma et al. this Vol.).

Both for experimental and theoretical reasons we thus appear to reach a boundary where single unit analysis, using concepts like the receptive field, ceases to be the single explanatory vehicle it is (implicitly) assumed to be.

1.3 Multi-Unit Analyis

Recent years have shown an increasing interest in going beyond the single unit paradigm. In several laboratories techniques have been developed to record the activity from different individual neurons simultaneously (for a review see Gerstein et al. 1983). Multi-unit recordings from some 10 to 20 neurons simultaneously have become feasible; the possibility of going to higher numbers is coming within reach (e.g., Grinvald 1984). Behind these experiments is the idea that essential properties of single neuron behaviour can only be understood when considering the neuron in its natural context: being an element in a massively connected network of similar elements. A further assumption is that meaningful information can be obtained from observation of the activity of a relatively small fraction of this network (e.g., 10 neurons from a connected total in the order of thousands).

The classical approach to the analysis of multi-unit data is the evaluation of the cross correlation function of pairs of simultaneously recorded spike trains (Perkel et al. 1967), or procedures derived from that (e.g., Gerstein and Perkel 1972, Perkel et al. 1975). This cross correlation gives an indication on the amount of (near) synchrony in firing (or silence) between the observed neurons. Since synchrony could also be induced by both neurons being driven by the same stimulus, a control measurement is performed. The stimulus ensemble is presented at least twice and one evaluates the cross correlation between spike trains from different units, now, however, with the time shift in the correlogram around the stimulus period ("shift predictor"). This correlogram measures the direct stimulus influence on the firing synchrony. The "difference" (not necessarily the algebraic difference) between these two measures of coincidence in firing is interpreted as the sign of true connectivity between the neurons.

Recent results using this type of approach (Eggermont et al. 1983a, Frostig et al. 1983, Bloom and Gerstein 1984, Espinosa and Gerstein 1984) have suggested that a similar problem exists as described above for the single unit receptive field: in a considerable number of cases the neural correlation, after application of a stimulus normalization procedure, may be stimulus-dependent. An example of this is given in Fig. 2, which shows the simultaneous and the non-simultaneous cross correlation of firing for a pair of neurons from the grassfrog's auditory midbrain, determined with different types of stimuli.

This observation gives rise to the hypothesis that the stimulus-variance of both the single unit receptive field and the neural correlation are intrinsically interconnected, in fact they are both manifestations of the same underlying mechanism: the neurons being elements in an interacting population, possibly giving rise to neuronal assemblies or related concepts (Hebb 1949, Braitenberg 1977, Palm 1982). This hypothesis can only be addressed by analysis methods which transcend the usual approach of addressing a group of neurons as a collection of all possible pairs or triplets, but instead, stress the properties of the entire group of neurons as an entity.

Also for purely pragmatic reasons such an approach seems inevitable. The very proliferation of correlation functions resulting from a multi-unit experiment clearly points to the necessity of a more integrated representation of the relevant interactions. A 10-unit recording, for example, results in 45 different cross correlograms to be calculated and, what is more, to be interpreted; a 20-unit recording leaves the poor experimenter with 190 correlograms. Fur-

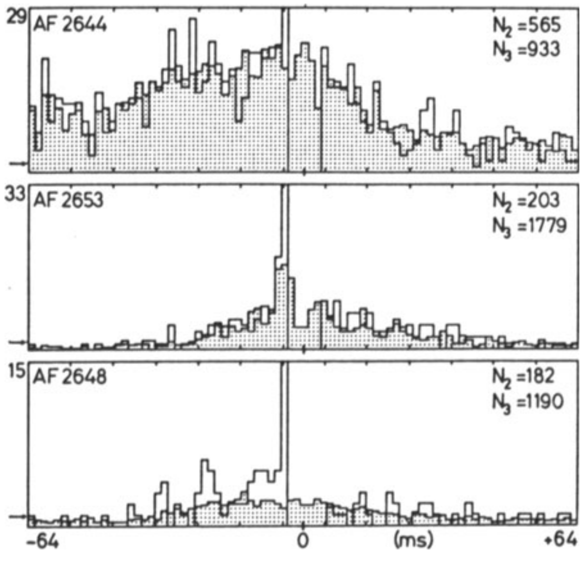


Fig. 2. Simultaneous and non-simultaneous cross coincidence histograms for two neurons from the auditory midbrain of the grassfrog (*Rana temporaria L.*), under tonal and noise stimulation. The non-simultaneous coincidence histograms are shaded. The difference between both histograms is interpreted as a measure for the strength of neural interaction. It is observed that this difference is negligible for stimulation with 48 ms tonepips presented once per s (*upper*); is confined to a few bins in case stimulation was done with 16 ms tonepips presented once per 128 ms (*middle*); and is quite clear and more extended for stimulation with stationary wideband noise (*lower*). (From Eggermont et al. 1983a)

thermore, these numbers have to be multiplied with the numbers of different stimulus conditions and possible other parameters of interest, such as time resolution of the correlogram. A more global and integrative method of analysis seems imperative, if only for logistics reasons.

A new conceptual representation of cooperative behaviour in a population of observed neurons, evolved from earlier ideas in statistical pattern recognition (Wright 1977), has recently been described (Gerstein et al. 1985, Gerstein and Aertsen 1985). This representation leads to a new technique for detecting and studying functional assemblies: "gravitational clustering". The multi-unit problem is mapped into an N-body problem; functional similarity between spike trains is translated into metrical distances. This representation being purely a mental construct, the rules governing the transformation can be defined at will, preferably such that the behaviour of the "model" system in a natural way induces inference regarding the relevant relations in the original problem.

2 Gravitational Representation of Multi-Unit Activity

2.1 Particles, Space, and Charges

As a first step we set up a geometrical representation of the problem at hand. With each one of the N units recorded from, we associate a fictitious point particle; this particle is located in an abstract Euclidean space. According to dynamic equations which will be specified in the following, the particles will travel through space. The rules governing these movements will be defined such that: higher than average coincidence in firing of different units leads to a higher degree of proximity of the corresponding particles.

The proximity is measured by the normal Euclidean distance. Before the particles are set into motion, a starting configuration has to be chosen. In the absence of any prior information regarding the association of neuronal firings, the most natural starting configuration is to require every particle to have equal distance to all the other ones. This necessarily leads to the representation space being N-dimensional: at time $t = 0$ the particles are positioned at the vertices of an N-dimensional hypercube, all initial mutual distances are equal to some starting value d_0. After the system starts to evolve the position of the particle corresponding to neuron i is at any time t given by the N-dimensional vector $\mathbf{r}_i(t)$.

We now have to specify the rules according to which the particles will move. To this end we associate with each particle an "electric charge" $Q_i(t)$, and have this charge be determined by the spike sequence $z_i(t)$ of the corresponding neuron i. In the present work we chose for the charge function a low-pass filtered version of the actual spike sequence:

$$Q_i(t) = \int ds\, q(s) z_i(t - s) \tag{1}$$

with the impulse response $q(s)$ given by

$$q(s) = q_0 \exp(-s/\tau) \ . \tag{2}$$

The time constant τ is a parameter, the value of which is chosen by the investigator. The charge history for a segment of a typical spike sequence is shown in Fig. 3.

Due to the electric charges the particles will exert forces onto each other, and, as a consequence, will start to move. In the present formulation we define the force between any two particles to be proportional to the product of their charges. The direction of the force is along the line conjoining the instantaneous position of the two particles involved, being attractive for equal signs of charge. Analogous to the Coulomb interaction in physics we may, in addition, have the force be distance dependent. The total force acting on a particle is obtained by vectorial addition of the individual contributions from all pair interactions involved.

Note that, unlike the normal case in physics, the electric charges are fluctuating functions of time. Through this time dependence the force field between particles is modulated in time by the activity patterns of the corresponding neurons. As a result those particles that correspond to neurons that tend to fire in synchrony will exert mutual attractive forces, and, consequently those particles will start to aggregate. As time proceeds their mutual distances will keep decreasing; after a sufficiently long interval has passed the sets of synchronously firing neurons will have collapsed to clusters in N-space, with every cluster corresponding to a different coherently firing group of neurons among the population that was recorded from. These clusters can be identified and characterized by standard methods of cluster analysis and pattern recognition (e.g., Fukunaga 1972).

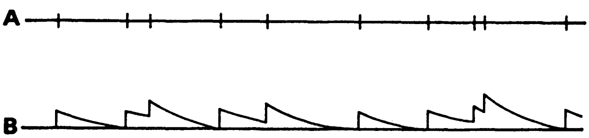


Fig. 3. Conversion of spike train to charge used in the gravitational representation. At the time of each neural impulse, the existing charge of the particle corresponding to that neuron is incremented by a fixed amount. Between neural impulses the charge decays with a fixed time constant. In the "rate normalization" used in the present paper the magnitude of the charge increment is different for each neuron, and is inversely related to its mean firing rate. (From Gerstein et al. 1985)

2.2 Dynamic Equations

The propulsive field $\boldsymbol{E}_{ij}$ at position $\boldsymbol{r}_i$ generated by the particle j at position $\boldsymbol{r}_j$ is given by

$$\boldsymbol{E}_{ij} = Q_j A(r_{ij}) \hat{r}_{ij} \tag{3}$$

with the unit vector $\hat{r}_{ij}$ given by

$$\hat{r}_{ij} = \frac{\boldsymbol{r}_{ij}}{r_{ij}} = \frac{\boldsymbol{r}_j - \boldsymbol{r}_i}{|\boldsymbol{r}_j - \boldsymbol{r}_i|} \ . \tag{4}$$

Since we are not really interested in intricate dynamic behaviour the distance dependence of the field has been eliminated by setting $A(r) = 1$. The total field $\boldsymbol{E}_i$ at position i is obtained by vectorial addition:

$$\boldsymbol{E}_i = \sum_{j \neq i} Q_j \hat{r}_{ij} \ . \tag{5}$$

The force $\boldsymbol{F}_i$ acting on particle i at position $\boldsymbol{r}_i$ is then given by

$$\boldsymbol{F}_i = Q_i \boldsymbol{E}_i \ . \tag{6}$$

As a consequence of this force the particle will move. For the equation of motion we define

$$\mu \dot{\boldsymbol{r}}_i = \boldsymbol{F}_i \ . \tag{7}$$

The acceleration term has been omitted because of computational convenience since, again, we are not really interested in detailed dynamics. Physically, it means that in our representation the particles move in a medium with high viscosity μ: the velocity of the particle is proportional to the force acting upon it.

The resulting displacement, finally, is obtained by simple numeric integration (Euler) using a time step δ:

$$\boldsymbol{r}_i(t + \delta) = \boldsymbol{r}_i(t) + (\delta / \mu) \boldsymbol{F}_i(t) \ . \tag{8}$$

This computational scheme is applied to all particles repeatedly as often as necessary to cover the duration of the multi-unit recording. The result is a collection of trajectories of the N particles in N-space.

Note that actually the movement is restricted to $N - 1$ dimensions since no external forces are involved. Furthermore, since all interactions are symmetrical, the center of mass of the system of particles is preserved.

2.3 Attraction and Repulsion

The gravitational rules given so far will necessarily lead to a collapsing universe. This is caused by the charges being definite non-negative [Eqs. (1–2)]: all particles attract and will eventually aggregate into the system's center of mass. Although it is certainly possible to infer functional proximity from studying the dynamics of evolving clusters, i.e., the time it takes the particles to aggregate (Wright 1977), we chose a slightly different approach. For our purpose we require particles which correspond to independently firing neurons to neither attract nor repel one another when studied over sufficiently long time. In this way only those particles will cluster which represent neurons that fire in a truly coherent fashion. The other ones will be subject to purely random influences; no systematic movement will occur: random walk with expected displacement equal to zero.

One way to obtain this behaviour is to impose an overall force on the particles which points outwards from the center of mass. This outward "drift" can be accomplished in several ways. In the present formulation of the gravitational representation we modify the charge rule [Eqs. (1–2)] to this effect: the original charge $Q(t)$ is replaced by shifted version $Q'(t)$, such that the time averaged value of $Q'(t)$ equals zero:

$$Q'(t) = Q(t) - \overline{Q(t)}. \tag{9}$$

The bar denotes taking the time average. A more general way to obtain zero mean charge would be to modify the impulse response $q(s)$ [Eq. (2)] to have no DC-component: the lowpass filter should be made bandpass to reject extremely low frequencies.

As a result of this modified charge rule the time averaged force for a system of two particles corresponding to two independently firing neurons, which is proportional to $\overline{Q'_i(t)Q'_j(t)}$, will be zero. When both neurons have fired recently, both charges Q' are positive: the particles will attract. When one neuron has fired recently and the other one remained silent, the signs of the charges differ: the particles will repel, the magnitude of the force being smaller than in the former case. Finally, when neither of the neurons fired recently, both charges will be negative: the particles will attract, the magnitude of the force dropping once more. These are precisely the properties needed to insure aggregation of those particles whose neurons either tend to fire in synchrony, or (more weakly) tend to be silent in synchrony. Particles that correspond to neurons which fire in an uncorrelated way will, on the average, exert no net force onto one another. Finally, particles whose neurons fire in anti-synchrony (antagonist type firing) will show a net divergent force.

2.4 Rate Effects

The original charge rule [Eqs. (1–2)] induces another peculiar property of the charge $Q(t)$. Due to the fact that for every particle the charge increases with an equal amount q_0 whenever the corresponding neuron fires, the propulsive influence of the charge would be proportional to the neuron's mean firing rate. Such a rate dependence may not be desirable. Therefore, in the original description of the gravitational representation (Gerstein et al. 1985), a "rate normalization" was applied: the increment of charge added at the time of each action potential was made proportional to the mean interspike interval for the neuron represented by that particular particle. In this way the mean value of charge $Q(t)$, considered over the whole recording (or a sufficiently long interval in the case of a sliding normalization) is the same for all particles; consequently the possibly undesirable effect of rate differences between neurons on the time averaged propulsive influence of the corresponding charges is canceled.

The more fundamental modification of the charge rule discussed earlier in connection with attraction and repulsion, i.e. going from $Q(t)$ to $Q'(t)$, causes a "rate normalization" in the above mentioned sense to become superfluous. Since the time averaged charge $Q'(t)$ by definition is zero, the overall effect of mean firing rate on propulsive "potential" no longer exists. More precisely: with equal increments in charge q_0 for every action potential clearly the magnitude of fluctuations in the charge $Q'(t)$ (the deviation from the expected value zero) will be proportional to the neuron's mean firing rate; however, the time avearage of the charge, and hence its global propulsive influence, is insensitive to it. As a consequence, when considering the force between any two particles it is not the rate of firing as such but truly the rate of "near-coincident" firing from both neurons which determines the net movement that will result. The only remaining effect of firing rate per se is on the "noisiness" of the trajectories, not on the systematic trends.

Not only is an additional "rate normalization" thus made superfluous, what is more, it even becomes counterproductive in the case of genuine correlated firing. The combination of the "rate normalization" and the transformation from Q to Q' [Eq. (9)] effectively amplifies the propulsive influence of "near-coincident" firings when one or more neurons with low mean firing rate are involved, while reducing it for neuron(s) with higher mean rate(s). This thus creates an undesired effect on the velocity of particle aggregation.

From the foregoing reasoning we conclude that, as far as the rate of coalescence is concerned, a charge rule without any specific "firing rate normalization" is probably the best choice for unbiased analysis of unknown data. This same conclusion was arrived at recently by a somewhat different reasoning (Gerstein and Aertsen 1985). It was pointed out there also, however, that the choice for a normalization rule must take into account additional factors, such as different selectivity for specific spike patterns. Results shown in the present paper were obtained while using a charge increment normalization as described by Gerstein et al. (1985).

3 Results

In this Section we show some results of applying the gravitational representation to simulated spike trains, generated by a simple neural network simulator (Gerstein et al. 1985, Aertsen and Gerstein 1985). More results can be found elsewhere (Gerstein et al. 1985, Gerstein and Aertsen 1985).

The aggregation process in the gravitational representation can be evaluated by studying the time evolution of the distance between pairs of particles. This should provide an indication about the functional connectivity of the corresponding neurons. Figure 4 shows the results for a simple neuronal circuit consisting of a pair of synaptically connected neurons together with a set of independently firing control neurons. The strength of the excitatory synaps was varied in the simulation, using values of 0.25, 0.50, and 0.99 (on a linear scale between 0 and 1) for different runs. Mean firing rates were in the order of 10 events per second. The figure shows, for each case, the time course of the distance between the points corresponding to pre- and postsynaptic neurons as well as those of an unrelated pair. The particles corresponding to unrelated neurons move about in a random fashion, their mutual distance remaining approximately constant. The particles corresponding to the connected pair of neurons, however, clearly show a systematic, albeit irregular, tendency

Fig. 4. The time to gravitational aggregation decreases as the strength of synaptic connection increases. Distance between selected pairs of points is shown as a function of step number (time) in the gravitational calculation. All pairs start at the distance of 100 arbitrary units; each time step corresponds to 2 ms of real time. The time constant for the charge decay was set at 10 ms, the distance moved per unit force was 3.5×10^{-5} units. The *three descending curves* show distances between particles that represent neurons connected with (*left to right*) decreasing excitatory synaptic strength of 0.99, 0.50 and 0.25, respectively. Distances between typcial particles representing independently firing neurons are at the *top* of the figure and remain essentially unchanged as the computation progresses. (From Gerstein et al. 1985)

to approach each other. The speed of aggregation appears to be monotonically related to the strength of the connection: the stronger the connection, the faster the coalescence. Note that coalescence was obtained already after about 40 spikes (4 s) in the case of strongest connectivity; the weaker interactions need approximately 7 and 10 s, respectively.

A more complex and interesting network of (simulated) neurons is shown in Fig. 5. The network consists of (1) two independently firing presynaptic neurons, (2) a set of four postsynaptic neurons for each presynaptic one, two neurons in each of these sets being driven by both presynaptic neurons, and (3) two independently firing control neurons. This example provides direct synaptic interaction, as well as two degrees of shared input. Firing rates of the individual neurons again were in the order of 10 spikes per second, connections were excitatory with a strength of 0.35, which puts them into the weak to moderate category encountered in physiological recordings. The time evolution of the clustering process is illustrated in Fig. 6, where, once again, pairwise distances are plotted as a function of time. The identities of the particles involved are indicated at the right hand side. Going from the top of Fig. 6 to the bottom, we observe the following: The distance between the particles 9 and 10 (unrelated control neurons) remains essentially constant throughout the entire run, which covers 16 s of "recording". The distances for the pairs (9,3) and (9,4) (one control neuron and in each case one postsynaptic neuron) initially decrease, later increase. The distances for the pairs (10,3) and (10,4) (the second control

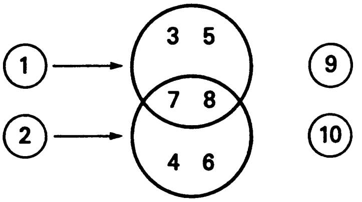


Fig. 5. A Venn diagram of more complicated neuronal connectivity that was simulated in a set of 10 spike trains. These data are analyzed in Figs. 6 and 7. Neurons *1* and *2* are drivers. Neurons *7* and *8* receive excitatory input from both drivers, while neurons *3,5* and *4,6* receive excitatory input only from one driver. Neurons *9* and *10* are independent control neurons. (From Gerstein et al. 1985)

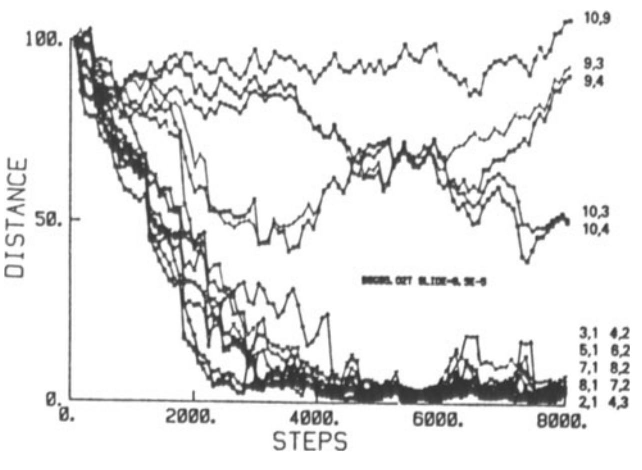


Fig. 6. Gravitational computation for the 10 spike train data set generated by the simulated neuronal circuit in Fig. 5. Ordinate: distance between particles representing selected neuron pairs; abscissa: time steps in the computation (2 ms step^{-1}). Further details in text. (From Gerstein et al. 1985)

neuron and one postsynaptic neuron) slowly decrease through the calculation. Finally, the pair distances for various transsynaptic pairs, for the two drivers (1,2) and for an independent postsynaptic pair (4,3) all show similar rapid coalescence. Full aggregation is obtained after some 8 s. Analysis of the same data, but excluding the spike sequences from the driver neurons 1 and 2 leads to essentially the same result (not shown here): curves look very similar, the main difference being that aggregation is somewhat slower. In the latter case full aggregation is reached after about 12 s (Gerstein et al. 1985). This shows that also in case the multi-unit recording fails to include the presynaptic driver neurons the interacting neurons can clearly be distinguished from the control neurons.

All synaptic connections in the spike generating network (Fig. 5) have been set to be equally strong. Thus each driver attracts equally (and is attracted by) each of its postsynaptic partners. Indirectly this will bring all of the postsynaptic partners of one driver neuron together in one small cluster, quite independent from the degree of direct interaction between the postsynaptic neurons. For similar reasons particles 7 and 8, which both are connected to both drivers 1 and 2, through the symmetry of the gravitational interaction will act as attractors for the clusters around the drivers. This eventually leads to clustering of all particles 1 to 8, although, for instance, the pairs (1,2) (the drivers) and (3,4) (unrelated postsynaptic neurons) represent neurons which, as such, are not in any way functionally related. Apparently the analysis of pair distances can identify a subset of interacting neurons; in itself it is insufficient to infer the circuit or the strength of connections in it.

The investigation of pair distances is a rather primitive tool to study the information that is contained in an ensemble of trajectories in N-space. More insight may be gained by adding one dimension in the representation of results: visualization of the trajectories by projection on an appropriately chosen plane. Figure 7 shows the results of a projection where information about the circuitry was used to define an adequate projection plane within the N-space: the plane determined by the instantaneous positions of the particles 1 and 2 (the drivers) and the center of mass of the particles 9 and 10 (control neurons). Note that, since these points are travelling through N-space, also the projection plane itself will be moving. Figure 7a–f shows "snapshots" of the projected positions of the 10 particles at regular intervals of 1.6 s. From the definition of the projection plane it follows that the particles 1 and 2 will move inwards in a straight line. The particles 9 and 10 (control neurons) can be observed to move rather randomly in the lower half of the pictures. All other particles (the driven ones) start at a point near the middle of the screen, and are rapidly dragged towards the aggregating drivers 1 and 2. The sensitivity of the gravitational representation is quite dramatically demonstrated in Fig. 7c: already after 4.8 s (i.e., after a mere 50 spikes from each neuron) the projection shows a configuration of particles that is reminiscent of the Venn diagram describing the network in Fig. 5. Each driver has collected its own exclusively driven particles, while the jointly driven particles 7 and 8 form a "bridge" between the

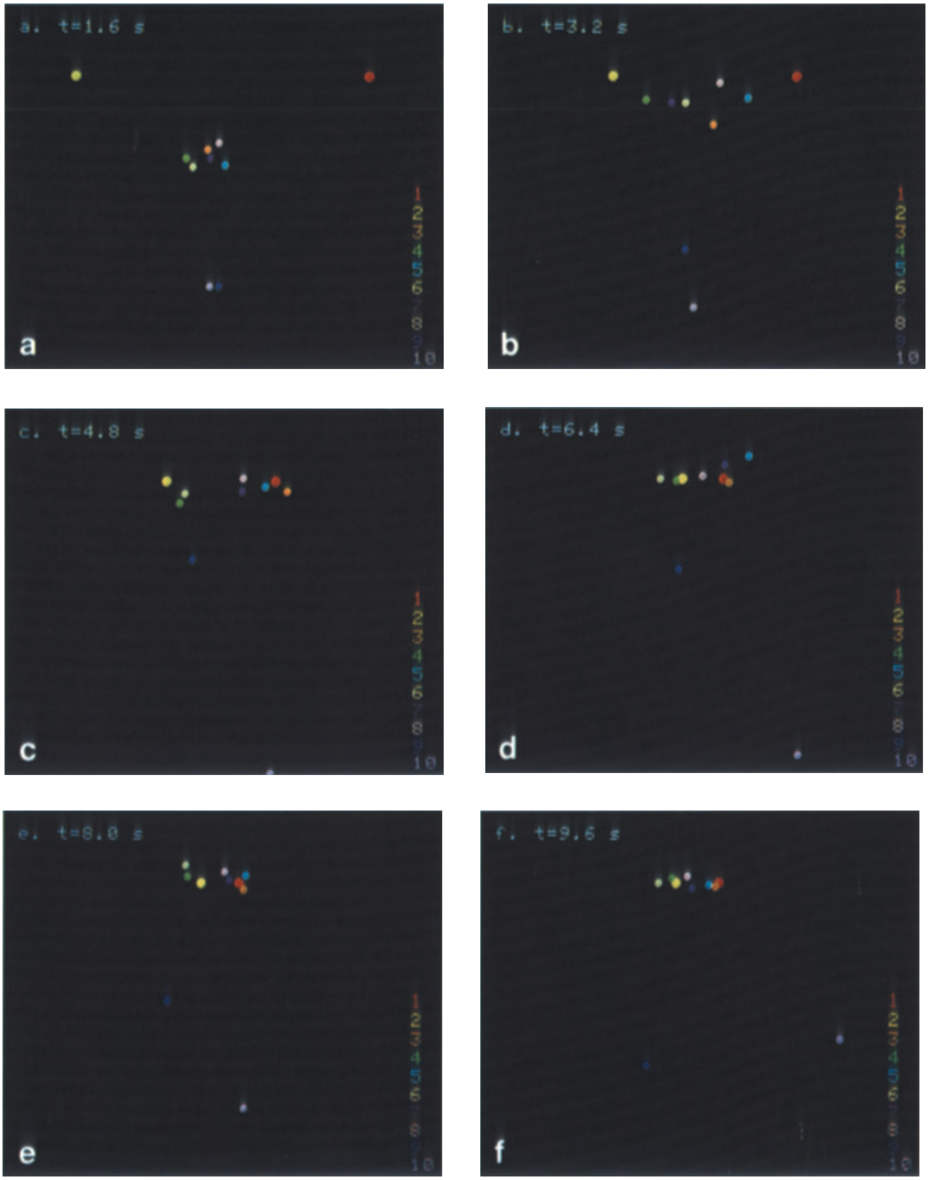


Fig. 7. Visualization of gravitational clustering by two-dimensional projection. Spike trains were obtained from the simulated network of 10 neurons, shown in Fig. 5. The projection of the *10 points* (representing the 10 neurons) from the 10-space onto an appropriate plane is shown in the form of "snapshots", taken at different moments in time with regular intervals of 1.6 s of "real" time. Particles are identified by their colour and the corresponding neuron number at the *right hand side* of each figure. Already after 4.8 s (i.e., after roughly 50 spikes from each neuron) the projection in Fig. 7c shows the Venn-diagram of the original network (cf. Fig. 5). As time progresses the configuration continues to aggregate, however, maintaining the basic character of the Venn-diagram (Figs. 7d–f). Further details in text

primary clusters. As time moves on, the configuration "boils" and continues to aggregate (Fig. 7d–f) until one single cluster remains, which still shows the characteristic structure of the Venn diagram, already clear in Fig. 7c.

A very good impression of the dynamics of the aggregation process can be obtained by combining a sequence of such snapshots into an animation movie. In such a movie especially the role of the "attractors" and the further evolution of the separate clusters becomes quite manifest. A more elaborate discussion on the possibilities of visualization by projection, also in the absence of prior information about the underlying network (used here to define the projection plane) can be found elsewhere (Gerstein and Aertsen 1985).

4 Discussion

4.1 Gravitational Clustering

We have described a new conceptual approach to representing and analysing spike train data from a multi-unit recording experiment. As shown by the examples the method is able to analyse interrelationships within a group of observed neurons without having to decompose the group into pairs as is usually done. The strength of interaction between neurons, reflected in the degree of firing synchrony, is manifested in the velocity of aggregation of the corresponding particles. The aggregation process can be studied by investigating the time dependence of interpair distances or, more visually directed, by projecting the trajectories onto an appropriately chosen plane in N-space.

A number of interesting issues concerning this representation could not be addressed here. These include, amongst others, the detection of inhibitory interactions and a modification of the gravitational clustering such that it allows "causal" inference: representation of the network in terms of a directed graph, rather than just identifying which neurons show associated firing behaviour. These and other issues are discussed at greater depth in the original papers (Gerstein et al. 1985, Gerstein and Aertsen 1985).

4.2 Sensitivity

The sensitivity of the method is extraordinary as becomes apparent from Fig. 7c, where a mere 50 spikes from each contributing neuron suffices to produce aggregation into a particle configuration which shows the essential characteristics of the circuit's Venn diagram. This sensitivity transcends at least by an order of magnitude that of usual measures like the Joint PST scatter diagram (Gerstein and Perkel 1972) or the three neuron "snowflake" (Perkel et al. 1975) which require hundreds to thousands of spikes from each neuron to produce clear pictures, moreover with the constraint that only two or three neurons are analyzed simultaneously.

The basic reason for this improvement is related to the different effects of time integration in these methods. In the gravitational representation the net distance traveled by a particle is proportional to the time integral of the total of "interactive force components" Q_iQ_j acting upon it. As a result of the definition of the charge function, especially the modification of the type as given in Eq. (9), the integrated effect of "noisy" spikes on the average cancels to a zero displacement; only the "near-coincident" firings result in a net effect on the particle positions. This "differential" behaviour of the gravitational representation is in contrast with usual correlation type methods as mentioned above, where time integration results in a steadily increasing noisy background in the scatter diagrams. The systematic effects of correlated firing usually are hardly discernible, unless many spikes have been processed to reduce the variance in the background to a comfortable enough level.

4.3 Type of Interactions

It should be observed that the force rule defined in Eqs. (3–6) only takes into account pair interactions, i.e., interactions of the type Q_iQ_j. This implies that the forces, and thus the trajectories of the particles, are determined by pair correlations of the corresponding spike trains, i.e., correlations of the type z_iz_j. In other words, specific higher order correlations in the neural activity, e.g., of the sort $z_iz_jz_k$, which go beyond combinations of participating doublets, are not considered in this representation.

The vectorial addition of force components from different sources (Eq. 5), although combining the influence of different doublets, by its linear nature does not address the possible presence of purely higher order interactions. Analogously to the correlation approach to nonlinear systems (.e.g., Marmarelis and Marmarelis 1978, Eggermont et al. 1983b) one might say that the gravity representation specifically aims at highlighting that component in the neural interaction which can be described by linear spatio-temporal integration (spatial = summation over different neurons; temporal = (leaky) integration over past spike activity from observed neurons (e.g., Johannesma and Van den Boogaard, 1985)). As a consequence specific nonlinear (e.g., Poggio and Torre, 1981) or composed (e.g., Shepherd 1974) synapses may go undetected.

The principal advantage of the gravitational representation as compared to normal cross correlation of pairs as applied extensively in the analysis of multi-unit data (e.g., Gerstein 1970), is that all possible pair interactions are analyzed at the same time. This is precisely what is accomplished by the multidimensionality of the present approach.

Another interesting observation to be made, is that the gravitational equations described above bear a close resemblance to the formalism describing plasticity of synaptic coupling between neurons, "Hebb's rule" (Hebb 1949, Palm 1982). In the latter context the principal drive for "learning", i.e., changing the synaptic weight, is supposed to reside in pair interactions of participating neurons (coherent pre- and postsynaptic firing). Like the gravitational representation, Hebb's rule does not take into account triple or higher order interactions. One might say that gravitational clustering considers neuronal interaction in a Hebbian universe.

This analogy might be pushed even further by imagining a somewhat different realization of the gravitational representation. Instead of working with particles in an N-space one might represent the N neurons under observation as a network of N nodes with some initially specified connectivity matrix (e.g., uniform or random). Applying the formalism described in the present paper, with only slight modifications, to gradually "update" the connectivity matrix, this network could be made to "learn" by changing its "synapses", such that the final state would more or less mimic the circuit of neurons that generated the measured spike sequences in the first place.

4.4 Stimulus-Dependent Neural Interaction

Until now it was implicitly assumed that the pattern of neural interactions remains fixed throughout the recording. An interesting question arises when this pattern, in fact, does change, for instance induced by changing stimulus conditions (see the example given in the Introduction). This should be reflected in a change in the aggregation process in the gravitational representation. Clusters may form and, later in the calculation, may "evaporate" to build new configurations, more or less mimicking the way the corresponding neurons are involved in a dynamic succession of different assemblies, "ignited" by changes in the external world (Braitenberg 1977). The dynamics of clustering thus represent aspects of structure (anatomy) as well as process (physiology). A third aspect, the function of neuronal interaction, can only be understood once the relation with the outside world, i.e., the stimulus ensemble giving rise to the observed phenomena, is explicitly taken into account. The analysis of multi-unit activity by gravitational-like methods emphasizes the "syntactical" aspects of neural activity. Stimulus-response correlation focuses on the "semantical" aspects. It is in the interrelation of both approaches that meaningful models of sensory information processing may be hoped to emerge.

Acknowledgements. This work was supported through grants from NIH (GLG NS 05606), the System Development Foundation (GLG SDF 0013), the Fulbright Program (AA travel grant) and ZWO (AA,PJ). The help of Volker Staiger in obtaining adequate copies of the Figures is gratefully acknowledged.

References

Aertsen AMHJ, Gerstein GL (1985) Evaluation of neuronal connectivity: sensitivity of cross-correlation. Brain Res 340:341–354

Bloom M, Gerstein G (1984) Multineuron analysis shows spatial tuning not found in single unit responses. Soc Neurosci Abstr 10:245

Boogaard HFP van den, Hesselmans GHFM, Johannesma PIM (1985a) Transformation of point processes, correlation functions and system identification. Submitted

Boogaard HFP van den (1985b) Neural interaction equations: Transformation of point processes. Ph.D. Thesis, University of Nijmegen, Nijmegen, The Netherlands

Braitenberg V (1977) Cell assemblies in the cerebral cortex. In: Heim R, Palm G (eds) Theoretical approaches to complex systems. Lecture notes in biomathematics, vol 21. Springer, Berlin Heidelberg New York, pp 171–188

Craik K (1943) The nature of explanation. Cambridge Univ Press, Cambridge

Eggermont JJ, Epping WJM, Aertsen AMHJ (1983a) Stimulus dependent neural correlations in the auditory midbrain of the grassfog (Rana temporaria L.). Biol Cybernet 47:103–117

Eggermont JJ, Johannesma PIM, Aertsen AMHJ (1983b) Reverse-correlation methods in auditory research. Q Rev Biophys 16:341–414

Espinosa I, Gerstein G (1984) Correlation among spike trains in cat's auditory cortex during presentation of three-tone sequences. Soc Neurosci Abstr 10:245

Evans EF (1974) Neural processes for the detection of acoustic patterns and for sound localization. In: Schmidt FO, Worden FG (eds) The neurosciences; 3rd Study Program, MIT Press, Cambridge, pp 131–145

Frostig RD, Gottlieb Y, Vaadia E, Abeles M (1983) The effects of stimuli on the activity and functional connectivity of local neuronal groups in the cat auditory cortex. Brain Res 272:211–221

Fukunaga K (1972) Introduction to statistical pattern recognition. Academic Press, London New York

Gerstein GL (1970) Functional association of neurons: detection and interpretation. In: Schmitt FO (ed) The neurosciences: 2nd Study Program. Rockefeller Univ Press, New York, pp 648–661

Gerstein GL, Perkel DH (1972) Mutual temporal relationships among neuronal spike trains. Biophys J 12:453–473

Gerstein GL, Bloom MJ, Espinosa IE, Evanczuk S, Turner MR (1983) Design of a laboratory for multineuron studies. IEEE Trans Syst Man Cybern SMC-13:668–676

Gerstein GL, Perkel DH, Dayhoff JE (1985) Cooperative firing activity in simultaneously recorded populations of neurons: detection and measurement. J Neurosci 5:881–889

Gerstein GL, Aertsen AMHJ (1985) Representation of cooperative firing activity among simultaneously recorded neurons. J Neurophysiol (in press)

Grinvald A (1984) Real-time optical imaging of neuronal activity. Trends Neurosci 7:143–150

Hebb DO (1949) The organization of behavior. A neuropsychological theory. Wiley, New York; Chapman and Hall, London

Johannesma PIM, Eggermont JJ (1983) Receptive fields of auditory neurons in the frog's midbrain as functional elements for acoustic communication. In: Ewert JP, Capranica RR, Ingle DJ (eds) Advances in vertebrate neuroethology. Plenum Press, New York London, pp 901–910

Johannesma PIM, Boogaard HFP van den (1985) Stochastic formulation of neural interaction. Acta Math Appl 4:201–224

Marmarelis PZ, Marmarelis VZ (1978) Analysis of physiological systems. The white-noise approach. Plenum Press, New York London

McCulloch WS (1965) Embodiments of mind. MIT Press, Cambridge

Palm G (1982) Neural assemblies, an alternative approach to artificial intelligence. Studies of brain function, vol VII. Springer, Berlin Heidelberg New York

Perkel DH, Gerstein GL, Moore GP (1967) Neuronal spike trains and stochastic point processes. II. Simultaneous spike trains. Biophys J 7:419–440

Perkel DH, Gerstein GL, Smith MS, Tatton WG (1975) Nerve-impulse patterns: a quantitative display technique for three neurons. Brain Res 100:271–296

Poggio T, Torre V (1981) A theory of synaptic interaction. In: Reichardt W, Poggio T (eds) Theoretical approaches in neurobiology. MIT Press, Cambridge, pp 28–38

Shepherd GM (1974) The synaptic organization of the brain. Oxford Univ Press, New York

Wright WE (1977) Gravitational clustering. Pattern Recognition 9:151–166