Biol. Cybern. 54, 29–40 (1986)

Biological Cybernetics

A Neural Cocktail-Party Processor*

Ch. von der Malsburg and W. Schneider

Abteilung Neurobiologie, Max-Planck-Institut für Biophysikalische Chemie, D-3400 Göttingen, Federal Republic of Germany

Abstract. Sensory segmentation is an outstanding unsolved problem of theoretical, practical and technical importance. The basic idea of a solution is described in the form of a model. The response of "neurons" within the sensory field is temporally unstable. Segmentation is expressed by synchronization within segments and desynchronization between segments. Correlations are generated by an autonomous pattern formation process. Neuronal coupling is the result both of peripheral evidence (similarity of local quality) and of central evidence (common membership in a stored pattern). The model is consistent with known anatomy and physiology. However, a new physiological function, synaptic modulation, has to be postulated. The present paper restricts explicit treatment to the peripheral evidence represented by amplitude modulations globally present in all components of a sound spectrum. Generalization to arbitrary sensory qualities will be the subject of a later paper. The model is an application and illustration of the Correlation Theory of brain function.

1 Introduction

The act of perception, in higher animals and in man, may be divided into three highly interdependent processes, segmentation, pattern recognition and integration of patterns into a scene. Segmentation separates the field of sensory information into pieces which form patterns. There are two sources of evidence for segmentation, peripheral and central. Peripheral evidence is based on similarity of local quality within a pattern. Central evidence is based on knowledge about patterns, i.e., about such constellations of local features which have proved of significance in the past. Whereas the sources of evidence are easily accessible to psychophysical experiment, both the type of process by which sensory segmentation is achieved, and the format into which the result of this process is cast are still unknown. (For the neural mechanism of figure-ground discrimination by motion in the fly see Reichardt et al. 1983.)

In technological contexts, the dominating segmentation device is the pattern-pass filter. In artificial intelligence, segmentation is often expressed by attaching labels to the elements of a pattern, or by copying them into specially designated lists. A format often proposed, also in the neural context, is the creation of a boundary enclosing a pattern, thus separating it from the ground. All of these ideas either are too special to serve as an explanation of segmentation in human (or animal) perception, or they cannot be implemented directly in neural architecture.

The explanation for sensory segmentation proposed in this paper can be broadly classified as a selective attention mechanism, although there are significant deviations from that idea as usually expressed (see, for instance, Treisman 1980 for psychophysical aspects, Crick 1984 for the discussion of a hypothesis regarding specific neural processes and anatomical structures to be involved in the mechanism). According to the idea of selective attention, neural activity is, in a given moment, limited to the elements of one segment. Usually, selective attention is imagined as a "spot-light" of neural excitation (or enhancement) projected from a central command structure into the set of neurons making up the sensory field. In contrast, the theory proposed here assumes temporal structure to be produced by instability and neural coupling within the field itself.

The "cocktail-party effect" (Cherry 1953; Cherry and Taylor 1954) refers to our remarkable ability to attend to and follow one speaker in the noisy environ-

* This work has been supported by Grant I/37-821 of the Stiftung Volkswagenwerk.

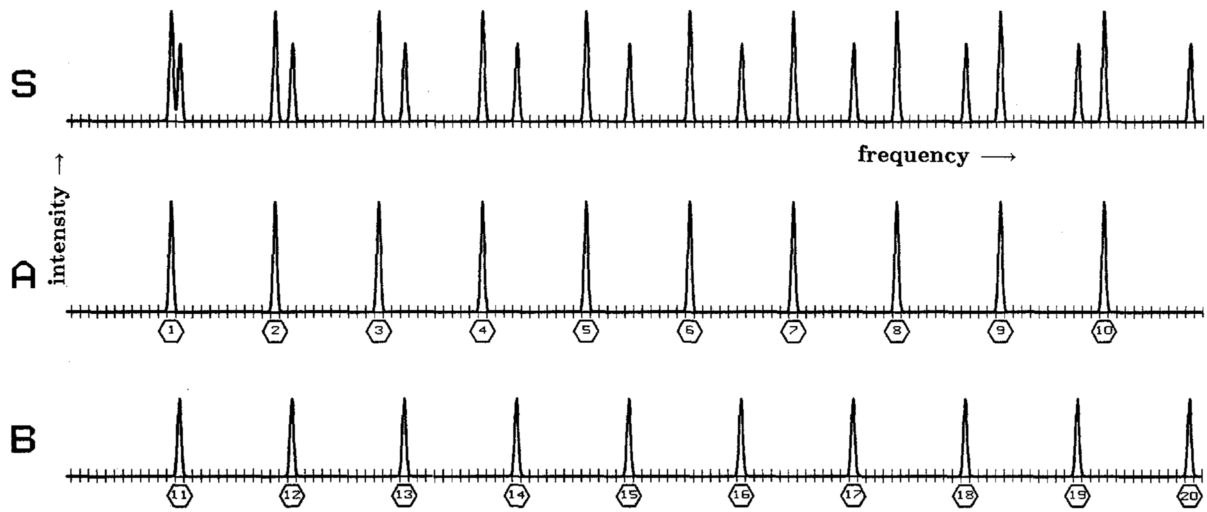

Fig. 1. *Superposed spectra. S*: Spectrum resulting from two simultaneous sound phenomena. It may be taken as a schematic rendering of the signal intensity distribution in a tonotopically organized area of the auditory system. *A* and *B*: Separated spectra. This separation has to be achieved by segmentation. Phonetic patterns can only be detected in the individual spectra *A* or *B*, not in the superposition *S*. The numbers underneath components refer to labels of *E*-cells in all subsequent figures. Tics are supposed to give an idea of the assumed cortical spectral resolution. Only those *E*-cells are modeled which are actually excited in a given stimulation

ment of a cocktail-party. The difficulty is schematically illustrated in Fig. 1. Our inner ear can be roughly described as performing a short-term frequency decomposition of the sound phenomenon hitting the eardrum. The cochlear nerve thus represents sounds by short-term frequency spectra. Each voice, taken by itself, consists of a series of discrete lines (forgetting, for a moment, the complication of insufficient spectral resolution at high frequencies). Phonetic information is encoded in the changing intensity distribution of the spectral components. Before this information can be evaluated, individual sound spectra must be recovered by segmentation, by a structure or device which may be called a "cocktail-party processor". The importance of auditory segmentation is not limited to cocktail-parties and human voices, to be shure. In any auditory environment there will be sound phenomena produced by independent sources, such that auditory recognition usually has to be preceded by segmentation. (In order to avoid misunderstandings it should be stressed at this point that the work described here does not attempt to deal with the problem of cutting the auditory stream into temporal segments, corresponding, for instance, to phonemes and words.)

The individual frequency components of an auditory phenomenon are distinguished from those of other sound phenomena by being similar in local quality (McAdams 1982). Among the relevant local qualities of a spectral component may be amplitude modulation (Helmholtz 1885), interaural delay (Cherry 1953; Cherry and Taylor 1954; Mitchell et al. 1971), frequency modulation, and local harmonic structure (McAdams 1982). Technical cocktail-party processors have been proposed on the basis of interaural delay (Strube 1981) and of harmonic structure (Parsons 1976). This paper is based on common amplitude modulation, or stimulus onset synchrony, as distinguishing mark of the components of one sound phenomenon (Helmholtz 1885, p 60; Bregman and Pinker 1978; Dannenbring and Bregman 1978; Rasch 1978).

This paper is meant to be a concise and simple introduction of the essential ideas of a general theory of sensory segmentation. This theory is independent of the peculiarities of the auditory modality, is based on the full range of local qualities, and provides for the integration of peripheral and of central evidence. The full theory will be the subject of a forthcoming paper (Schneider and von der Malsburg, in preparation). It is itself an application and illustration of a more comprehensive conceptual framework (von der Malsburg 1981). In order to be clear and simple, the specific model presented here leaves out complications such as overlap between the spectra to be separated, frequency modulation, and the restriction of temporal (and qualitative) correlations between partials to local neighborhoods.

Section 2 describes an abstract model and contains all essential ideas. Section 3 fills in all technical and quantitative detail necessary to repeat the work. It is not necessary to read Sect. 3 in order to qualitatively understand the model. Section 4 describes a number of experiments performed with the model as simulated on a digital computer.

2 The Abstract Model

The model, Fig. 2, consists of a set of excitatory cells ("E-cells") and one inhibitory cell ("H-cell"). Each spectral component is thought to be represented by one E-cell. (In reality each unit of spectral resolution is represented in vertebrates by a large number of neurons on all stages of the auditory system. The inhibitory system consists of many cells, which are here, however, treated as one pool, represented by the H-cell.) Each E-cell is connected to all other E-cells by an excitatory synaptic connection. The H-cell receives an excitatory connection from and sends an inhibitory connection to each E-cell.

Each E-cell receives a separate input from the auditory periphery. The inputs fluctuate between two levels, 0 and 1. If there are several sound phenomena, all inputs corresponding to one of them change their values at the same time. The inputs corresponding to different phenomena are uncorrelated with each other.

Groups of E-cells respond to afferent tonic excitation with an unstable activity level. After stimulus onset, the activity rises up to a certain level, and then drops sharply. Such an activity excursion is called a burst. The tendency to burst may be an intrinsic property of cells or it may result of external feed-back. Bursts in different E-cells are synchronized by the excitatory links between them. They are desynchronized by the inhibition mediated by the H-cell which tends to limit the total activity in the network. The balance is set such that an inter-burst delay above a certain threshold is quickly increased in magnitude from burst to burst, until activity is completely out of phase between the cells.

In consequence, those cells which are simultaneously activated, presumably by the components of one sound phenomenon, have a tendency to synchronize their activity bursts with each other, and to desynchronize them from the rest of the cells. Therefore, all cells hit by the same sound phenomenon are synchronized to form one block of activity, different such blocks being desynchronized. All subsequent stages of neural processing can get an unobstructed view of a single sound phenomenon by oscillating in phase with the group of cells representing it. Central evidence is integrated into the segmentation process with the help of the indirect couplings introduced into the set of E-cells by pattern evaluating circuitry, see Fig. 3.

Transients in the afferent activity appropriate to mark out membership of components in one spectrum may be relatively rare in the course of a sensory stimulation. Also, accidentally coinciding afferent transients may cause trouble. It therefore is important to have short-term memory. This is constituted by synaptic modulation (von der Malsburg 1981). According to this, synaptic coupling strengths between E-cells vary on a rapid timescale, i.e., during a single sensory stimulation: in the resting state, cells are coupled with medium strength. When there is synchronous activity in the two E-cells, the synaptic strength

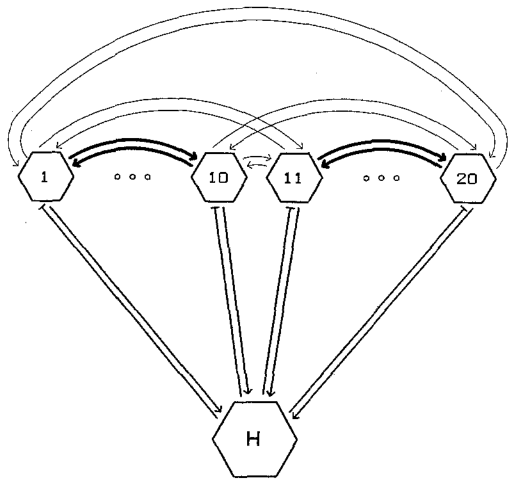

Fig. 2. *The Model.* Upper row of hexagons: E-cells. Lower hexagon: H-cell (pool of inhibitory cells). → excitatory connections, ⊣ inhibitory connections. The H-cell limits the total activity of the E-cells. In the resting state all links between E-cells have equal strength. Successful segmentation expresses itself by synchronous bursts of activity within the sets of cells excited by one spectrum, and desynchronization between the sets. With the stimulus in Fig. 1, cells 1 to 10 and cells 11 to 20 form blocks of synchronization. Synaptic modulation strengthens the connections (*thick arrows*) between cells with synchronous bursts, and weakens the connections (*thin arrows*) between cells with asynchronous bursts

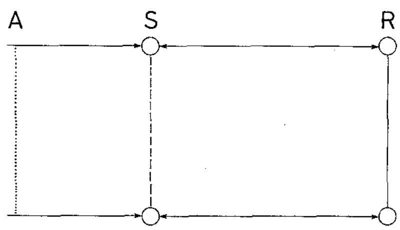

Fig. 3. *Coupling to pattern-evaluating circuitry.* Circuits are symbolically represented by circles (cells) and lines (connections and signal correlations). The segmentation network S, which is discussed in this paper, generates segmentation in the form of temporal correlations between cells belonging to one pattern and anticorrelations between cells belonging to different patterns, with the help of the positive feed-back loop between activation of synaptic connections and correlations (both represented by *dashed line*). Segmentation in S integrates influences from two sources: peripheral evidence, in the form of correlations (*dotted line*) between afferent signals A, and central evidence, in the form of correlations produced by connections in pattern circuits R (*vertical solid line*). R is supposed to represent patterns in a way similar to S, but with permanent connections within those sets of elements which form patterns. There is a positive feed-back loop between activation of connection patterns in S and in R

is increased, up to a maximum value which is characteristic for the synapse. When there is asynchronous activity in the two cells, the synaptic strength is decreased, ultimately down to zero. These changes can take place in fractions of a second. After activity ceases in the two cells, synaptic strength slowly relaxes back to the resting value, with a time-constant characteristic of short-term memory, about a minute.

In the resting state, the matrix of connections between E-cells is homogeneous, all synaptic strengths being equal. This coupling strength should be such that blocks of synchronous activity are marginally stable. Two sound phenomena with a slight onset asynchrony should lead to a decay of the corresponding global block of cells into local blocks. On the other hand, cells which are excited in synchrony should stay synchronized with high probability. A few bursts of activity change the matrix by synaptic modulation. Connections within a block grow stronger, connections between blocks become weaker. The now more tightly coupled blocks should be stable against decay into sub-blocks, whereas super-blocks formed by an accidental synchrony should be unstable against spontaneous decay into the segments formed earlier. (Segments can be caused to accidentally coincide in time by synchronous stimulus onsets or by a stimulus onset which happens to coincide with one of the spontaneously created bursts belonging to another segment.) In this way the original segmentation is stored in short-term memory. Even if synchronizing transients are absent from the afferent activity for a while, synchronicity relationships within and between the original blocks are stabilized by the differentiated synaptic couplings. Only a longer break (or some reset-command to the synapses) will allow the synapses to return to their resting state, so that the E-cells can be segmented in a new way.

How is synchrony evaluated in the brain? We know that neurons are coincidence detectors. The signals in several fibers converging on one neural dendrite can summate (and help each other to transcend the neural threshold) only if their postsynaptic potentials overlap in time. Thus, the signals from all the synchronized cells in a segment can cooperate with each other in firing the neurons of the pattern evaluation machinery of the brain, whereas signals from cells belonging to different segments are desynchronized and temporally miss each other (for a more thorough discussion see von der Malsburg 1985). The circuitry which is to evaluate the patterns in a given segment only has to synchronize its oscillating activity with that in the segment to get an unobstructed ("stroboscopic") view of it through a temporal mask. Synaptic modulation binds a segment temporarily to the relevant evaluation circuitry.

3 The Concrete Model

In the simulations, only those E-cells are actually represented which receive afferent excitation from a component in the spectra currently presented. There are two input stimuli, each having 10 spectral components. Thus there are 20 E-cells, whose activity is described by $E_i(t)$, $i=1, \ldots, 20$. There is one additional, inhibitory, cell, whose activity is designated $H(t)$. The signals $E_i(t)$ and $H(t)$ are meant to represent the momentary frequency of a statistical sequence of action potentials. They are modeled by smooth functions. Their underlying statistical nature is taken into account by a noise term $z_i(t)$ in the dynamical equations. Time t proceeds in discrete steps of size τ, one of which may be roughly interpreted as a millisecond, to fix ideas.

The dynamics of the model is described by the following difference equations (the numbers in square brackets being the parameter values used in the simulations discussed in the next section):

$$E_i(t+\tau)=N_i(t)\cdot \mathbf{1}\Big\{A_i(t)+\alpha E_i(t)+\sum_{j\neq i} s_{ij}(t)E_j(t) - s_{he}H(t)+z_i(t)\Big\}, \quad [\alpha=0.89;\ s_{he}=0.22], \tag{1}$$

$$H(t+\tau)=\mathbf{1}\Big\{\beta H(t)+s_{eh}\sum_j E_j(t)\Big\}, \quad [\beta=0.63;\ s_{eh}=0.036], \tag{2}$$

with afferent input

$$A_i(t)=\begin{cases}0.1, & \text{if } E\text{-cell } i \text{ receives input}\\ 0, & \text{else.}\end{cases} \tag{3}$$

The output-nonlinearity $\mathbf{1}(x)$ consists in simple clipping:

$$\mathbf{1}(x)=\begin{cases}0 & \text{for } x<0\\ 1 & \text{for } x>1\\ x & \text{else.}\end{cases} \tag{4}$$

(The upper threshold 1 is never reached during the simulations presented here.) The noise term $z_i(t)$ is modeled by independent sequences of pseudo-random numbers with identical flat distribution in the interval (0, 0.01). Noise is important for the function of the model, to break the symmetry between accidentally synchronized but weakly coupled E-cells.

The function $\mathbf{N}_i(t)$ in (1) takes the values of 1 or 0. It is 1 during bursts and is 0 during the inter-burst ("refractory") periods. When a gliding average $G_i(t)$ of the signal $E_i(t)$, defined by

$$G_i(t+\tau)=(1-\delta)G_i(t)+\delta E_i(t), \quad [\delta=0.35], \tag{5}$$

reaches the threshold $g_u=0.4$, $\mathbf{N}_i$, and correspondingly the cell's output $E_i(t)$, is put to zero so that $G_i(t)$ decays

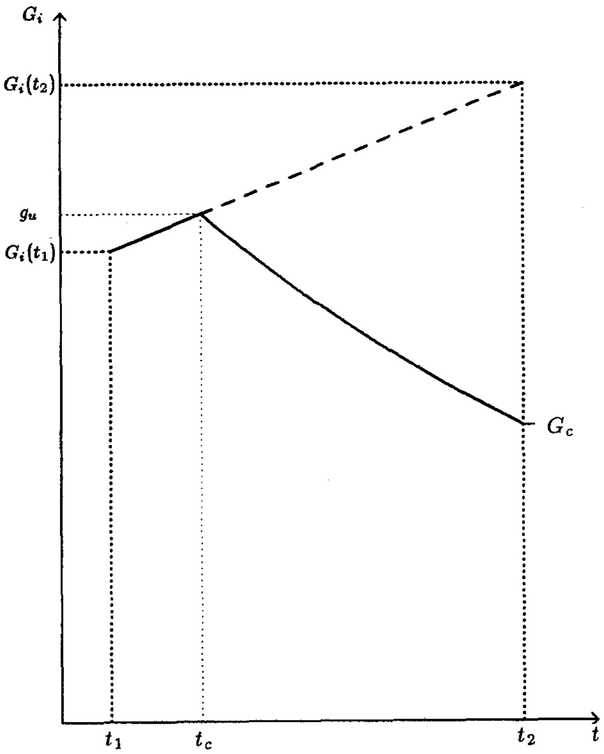

Fig. 4. *Interpolation of $G_i(t)$ between temporal steps.* Artificial synchronization due to discrete time is avoided by determination of t_c, a corrected point in time at which the gliding average $G_i(t)$ reaches the upper threshold g_u, by linearly interpolating between $G_i(t_1)$, the last value of $G_i(t)$ below threshold, and $G_i(t_2)$, the first value above threshold. E_i is then imagined to be switched off at t_c, and a new value G_c is calculated for $G_i(t)$ assuming an exponential decay with time-constant $(1-\delta)$, cf. (5). The interpolated break-off point t_c is used in the argument of the coactivity function which, together with the control function q(s), regulates synaptic modulation, cf. (7)

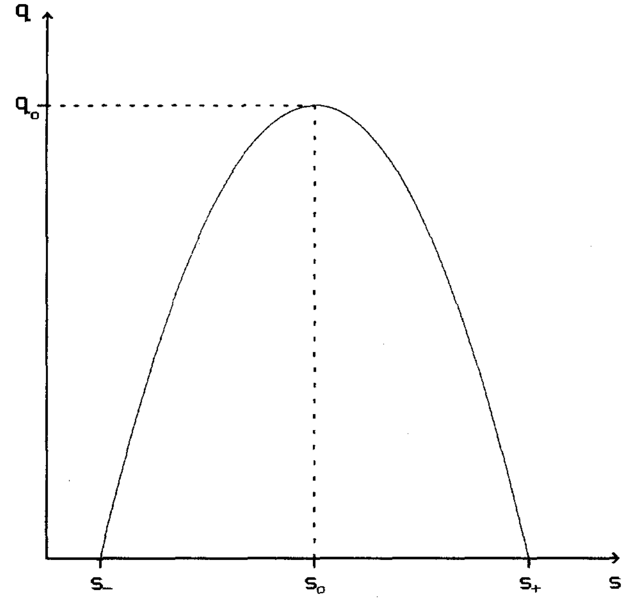

Fig. 5. *Control function of synaptic plasticity*, formula (8), used in (7). A synapse which has been driven into saturation near s_- or s_+ by repeated antisynchronous or synchronous events is thereby rendered insensitive to occasional episodes of erroneous synchronization or antisynchronization, respectively

with time constant $(1-\delta)$. When $G_i(t)$ reaches the lower threshold $g_l=0.01$, N_i is put back to 1, i.e.,

$$\mathrm{N}_i(t)=\begin{cases}0 & \text{if } G_i(t)>g_u \text{ or } (\mathrm{N}_i(t-\tau)=0 \\ & \text{and } G_i(t)>g_l) \\ 1 & \text{else}.\end{cases} \tag{6}$$

In order to keep temporal discretization from artificially synchronizing the cells, both start and end of the "refractory" period are interpolated between time-steps. This is done with the help of a linear estimation of $G_i(t)$ during the elemenatry interval, see Fig. 4. The values of $E_i(t)$ at the discrete times are then corrected accordingly by linear interpolation.

The strength $s_{ij}(t)$ of the excitatory synapse between "presynaptic" E-cell j and "postsynaptic" E-cell i evolves by synaptic modulation. Generally, the change of a given synapse is governed by the following rule: if one of the two cells is in a subliminal state (i.e., it has not produced bursts for the period $T+T_a/2$, where T is the repetition period, and T_a is the length of bursts), no change will occur. If both cells are bursting, $s_{ij}(t)$ is changed according to

$$\Delta s_{ij}(t)=q(s_{ij}(t))\cdot \mathrm{Co}(t_i-t_j;\ T;\ T_a)\,, \tag{7}$$

with the control function

$$q(s)=q_0(1-\{(s-s_0)/(s_0s_d)\}^2)\,,$$
$$[q_0=0.01;\ s_0=0.012;\ s_d=0.8]\,. \tag{8}$$

The latter keeps synaptic strengths within 80% of the resting value s_0, and is convex, i.e., small at high and low synaptic strength, see Fig. 5, to render short-term memory insensitive to short episodes of erroneous synchronization or desynchronization (such as for instance between about steps 660 to 700 in Fig. 8, or at the beginning of the periods displayed in Fig. 10). The "coactivity function" Co($\cdot$) estimates the correlation between cells i and j. Its evaluation was simplified under the assumption of regular bursts of uniform shape. This shape is parameterized by the average burst repetition period, T, and the average length of the bursts, T_a. Both T and T_a are computed as short-time ensemble averages, updated for each new burst. Co($\cdot$) then depends only on the delay $\Delta t=t_i-t_j$ between the break-off times of the bursts in cell i and cell j. For the special case of $T_a=T/2$ the shape of $\mathrm{Co}(t_i-t_j;\ T;\ T/2)$ was assumed to be a cosine with period T. In the general case of $T_a \neq T/2$, the domains of definition for the positive and negative half-wave of the cosine are

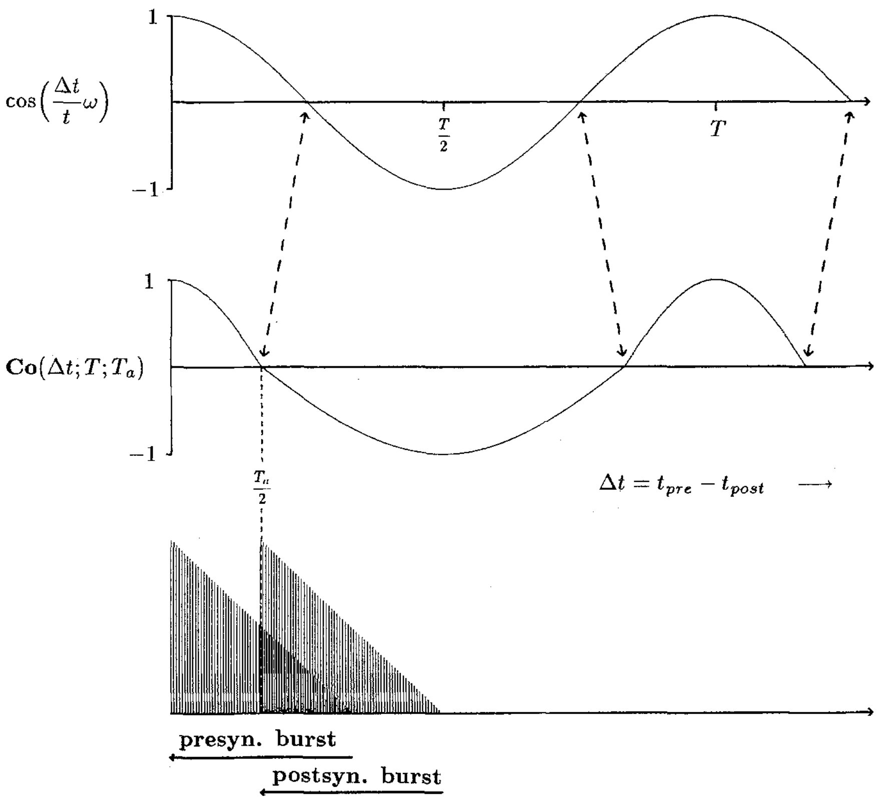

Fig. 6. *Evaluation of correlations.* A standard shape of the activity burst is assumed. It is parameterized by the period T and the duration of the burst, T_a, which are determined by a running ensemble average over all active cells. $\mathrm{Co}(\cdot)$ is evaluated for (presynaptic) cell i at the first time-step after each burst break-off. For each postsynaptic $j \neq i$ the balance between synchronous firing, i.e., burst overlap (double-hatched region in the schematic burst patterns in the lower graph), and asynchronous firing is estimated with the help of a cosine-shaped function, upper graph. In order to allow for variable burst length, the positive and negative parts of the cosine are compressed and stretched, such that crossing-over is when bursts overlap for half of their duration. Bursts appear inverted in shape since earlier times are to the right. The second maximum at T takes into account bursts with a delay of $T-\Delta t$, which are thus treated as advanced, i.e., corresponding to a "negative" delay $-\Delta t$). In order to be insensitive to small variations in the burst period of individual cells, the last burst in cell j is taken into account up to the delay $T+T_a/2$

linearly stretched and compressed, see Fig. 6. To economize computer-time, the $s_{ij}(t)$ are updated for a given cell i only when the cell just entered the inter-burst period. The slow relaxation of synaptic strength back to the resting state has been neglected for the relatively short periods of simulation.

The number of parameters in the model is too large for them to be determined by blind search. Some additional remarks may help to gain insight into the type of influence different parameters have. Some parameters simply fix the temporal scale. The inter-burst period is rigidly fixed by g_u, g_l and δ. During this period cells are silenced by $\mathbf{N}_i(t)=0$ and simply have to wait until G_i has decayed, according to (5), down to g_l. For the simplified case of a block of n precisely synchronized cells and homogeneous coupling strength, say s, the shape of the burst is determined by a two-dimensional (E and H) system of linear differential equations, which can be solved analytically (until a non-linearity occurs when a cell breaks off). The shape of the burst is then determined explicitely by the constants α, β, s, the product s_{eh}, s_{he} and n. The ratio s_{eh}/s_{he} can be used to scale $H(t)$ and keep it between its saturation levels 0 and 1, see (4). For the parameters chosen here and the synapses at resting level, the duration of bursts varies slowly with n, between $T=6.971$ for $n=1$ and $T=5.838$ for $n=20$.

If two blocks of cells are nearly synchronized with each other it is important to know how the relative phase between them will evolve in time. At the moment at which the cells of the block leading in phase switch off ($G_i(t)$ having reached the upper threshold, cf. (6)), the excitation reaching the cells of the trailing block suffers a sharp drop. $H(t)$ is, however, still at the high value corresponding to the total excitation from both blocks, and relaxes with time constant β, cf. (2). The rate of growth of $E_i(t)$, and correspondingly of $G_i(t)$, for the cells of the trailing block will be reduced and the phase-lag will be increased. This effect on the relative phase is counteracted by a decrease during the time both blocks are active. Let the blocks comprise n cells, let the strength of synapses within and between the blocks be s_1 and s_2, respectively. Then one can easily derive an equation for $\Delta S(t)$, the difference between the total excitation in the blocks:

$$\Delta S(t+\tau)=(\alpha-s_1+n(s_1-s_2))\Delta S(t)\,. \tag{9}$$

If $s_1=s_2=s_0$, $\Delta S(t)$ geometrically decreases with the rate $\alpha-s_0$. The parameter s_0 can be used to regulate the balance between this decrease in phase-lag between the two blocks and the increase in phase-lag produced after the leading block has switched off. This balance should be set such that on the one hand a small stimulus onset asynchrony between the two blocks suffices to let their phases drift apart completely after a few bursts, whereas, on the other hand, the relative phases within the blocks stay near to zero. Synaptic modulation will then increase s_1 and decrease s_2. The parameter q_0, which according to (7) and (8) regulates the sensitivity of synaptic modulation, should be made big enough so that after a few bursts the changes in s_1 and s_2 suffice, cf. (9), to stabilize the blocks against decay and to destabilize the phase between the blocks.

Many of the details described in this section are unimportant for the realization of the abstract model described in Sect. 2 and could be replaced by others. Also, the function of the model is fairly insensitive to changes in the parameters employed, except where the marginal stability of blocks is involved. There is ample space for tuning the model to experimental data, once they become available.

4 Simulations

Simulations were done under on-line control, with the help of a simple command language. All runs discussed in this paper are based on one set of parameters. Except when stated otherwise, stimuli consist of two "spectra" comprising ten components each, as exemplified in Fig. 1. The function of the system depends on a balance between synchronizing and desynchronizing effects. This balance is shifted by synaptic modulation (see end of Sect. 3). The system was tuned with the help of a few simplified noiseless test cases. In the first, only one block of 10 cells was activated. Synapses were all of the same strength. The stimulus to one of the cells led the others by 1 step. Synapses were made strong enough to keep the relative delay between the leading cell and the others low for at least a few bursts. A 60% increase in coupling strength between cells in the block (reached after 11 bursts) was then sufficient to "heal" even the worst case of a lead of 6 steps for the test cell.

The second test case involved the normal two stimuli to blocks of 10 cells each. The synaptic resting value just determined let a stimulus onset delay (and correspondingly a burst onset delay) of 1 step between the blocks increase to 2 steps at the end of the burst, see

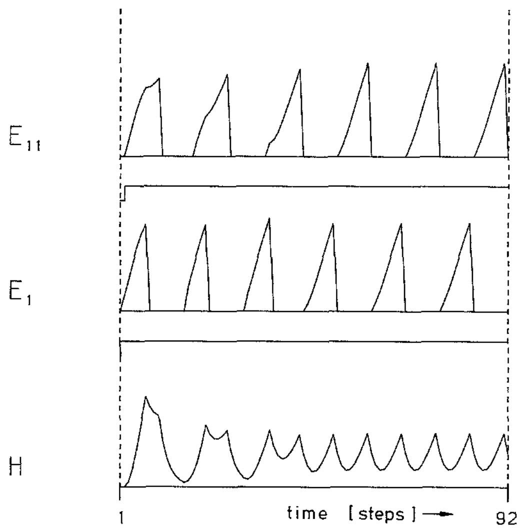

Fig. 7. *Desynchronization.* The spike rate is shown for three out of the 20 + 1 cells. The stimuli and the numbering of cells are as in Fig. 1. The stimulus activating E-cells 1 to 10 began at step 1; the other stimulus, activating E-cells 11 to 20, begins at step 2. Both stimuli last over the whole period of the diagram. The E-cells respond to this tonic input with bursts of spikes. The second group of cells (exemplified by E_{11}), having been stimulated 1 step later than the first group, reaches the switch-off point later and is delayed still further by the inhibition, cell H, stirred up together with the first group. The original delay between the two stimuli is thereby amplified gradually, until the two sequences of bursts are in complete antiphase. The last bit of temporal overlap can be observed near step 37

Fig. 7. The second burst starts with exactly this delay between the blocks, which is then further increased, until a stationary state is reached in which the bursts of the two blocks no longer overlap in time. The sensitivity to stimulus onset delays is further enhanced by synaptic modulation decreasing the coupling between the blocks.

Figure 8 shows the signals of all 21 cells in the model for a run 1000 steps long (which may correspond to 1 s). During this run synaptic strengths are modulated by the mechanism described in the last two sections. After a few bursts already, the coupling within the blocks is strengthened sufficiently to stabilize them against a decay which could be triggered by stimulus onset jitter within one of the stimuli. At the end of the run in Fig. 8 the matrix of synaptic connections between E-cells has the form shown in Fig. 9. Problems for the system are created when one stimulus is on and the onset of the other stimulus is synchronous with the bursts in the first group of cells. If this happens with sufficient precision and when the synapses are in the resting state, a single coherent stimulus is simulated and the system cannot segment it, to be shure. After segments have been formed, however, and the matrix of

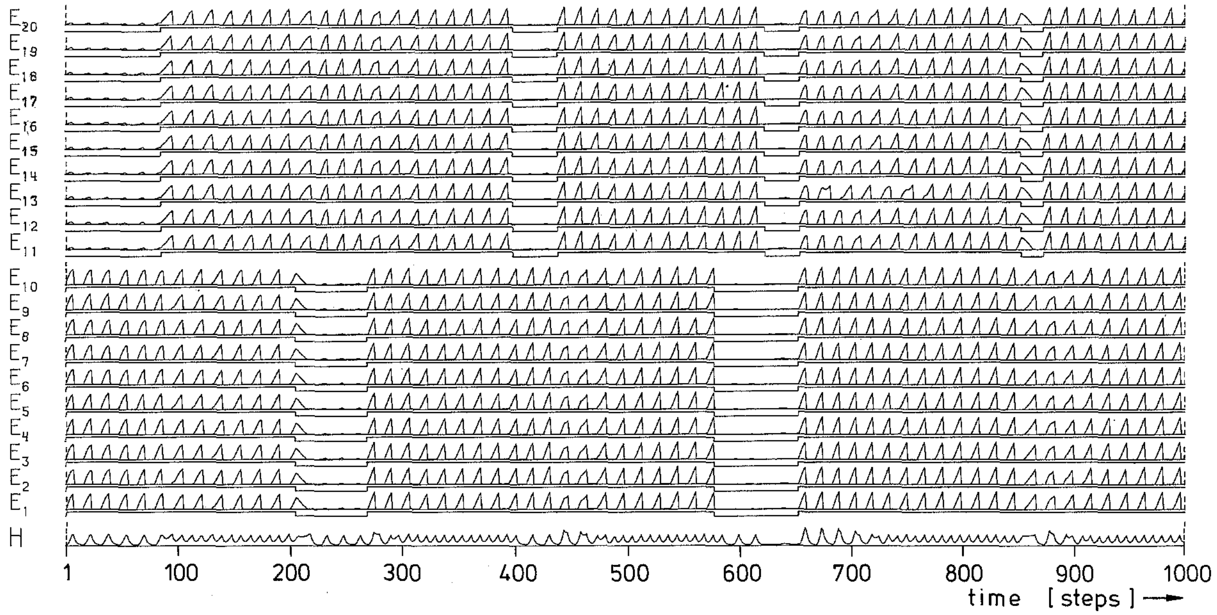

Fig. 8. *The complete set of responses for one second of stimulation.* The line underneath each trace indicates whether the stimulus is on or off. Cells are grouped according to stimulus, and within groups according to spectral frequency, as shown in Fig. 1. The two stimuli switch on and off independently of each other. At $t=0$ the matrix of connections between E-cells is in its resting state. During the run it is gradually modulated. At $t=1000$ it has the form shown in Fig. 9. In the beginning, group E_{11} to E_{20} receives no afferent stimulation but is weakly excited by the connections from cells E_1 to E_{10}. Later, around steps 250 and 420, this cross-coupling can already be seen to be weaker. At step 653 the two stimuli "happen" to be switched on at the same time, so that the first few bursts are synchronous between the groups. This state is, however, unstable since, already, the synapses between groups are weaker and synapses within groups are stronger than in the resting state. After a few bursts, the desynchronized state is reached again. A good indicator of desynchronization is frequency doubling and amplitude reduction in the H-cell

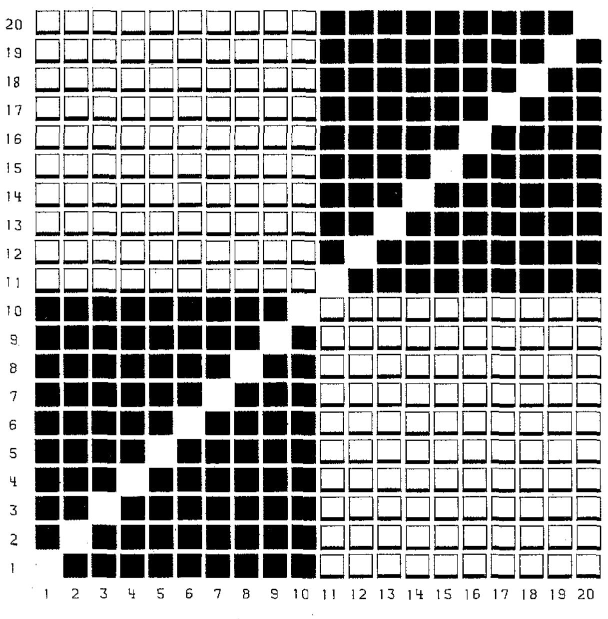

Fig. 9. *Synaptic matrix modified by modulation.* Strength of synapses between E-cells at $t=1000$ in the run of Fig. 8 is indicated by the hight of filling in the small squares. The synapse in row i and column j connects E-cell j to E-cell i. The strong synaptic strenghts within the groups E_1 to E_{10} and E_{11} to E_{20} reflect the synchronization within those groups in the recent past, the weak synaptic strengths between the groups reflect desynchronization

synaptic couplings has been decomposed into the corresponding blocks, even an accidental precise coincidence of stimulus onset or of one stimulus onset with a burst in the other group is successfully dealt with. This is illustrated by the runs in Fig. 10. The precise symmetry between bursts must in this case be broken by random fluctuations in the cellular signals. In most cases the period needed to break the symmetry is not longer than 80 to 100 steps. The influence of the noise level was tested by varying its amplitude from its usual value of 10^{-2} up to 10^{-1} – the maximum noise which led to a stably antisynchronised state – and down to 10^{-4}. In the latter case, the time to decay into blocks increased to 200 steps. Characteristically, cells stay synchronous within a block and phase-switching occurs for all cells simultaneously.

The following stability tests were performed with fixed $s_{ij}^{(r)}=s_0(1+r)$ within and $s_{ij}^{(r)}=s_0(1-r)$ between blocks. In one experiment, two groups of 10 cells each were activated with a relative delay corresponding to half a burst period. The ensuing pattern of alternate bursts in the two groups was stable over more than 3000 steps for fixed values of $r=0.4$ and $r=0$. In a second experiment, there was only one group of 10 cells. With $r=0$ the group stayed approximately synchronous for at least 6 bursts (which would be sufficient for synaptic modulation to reach $r=0.32$ if r

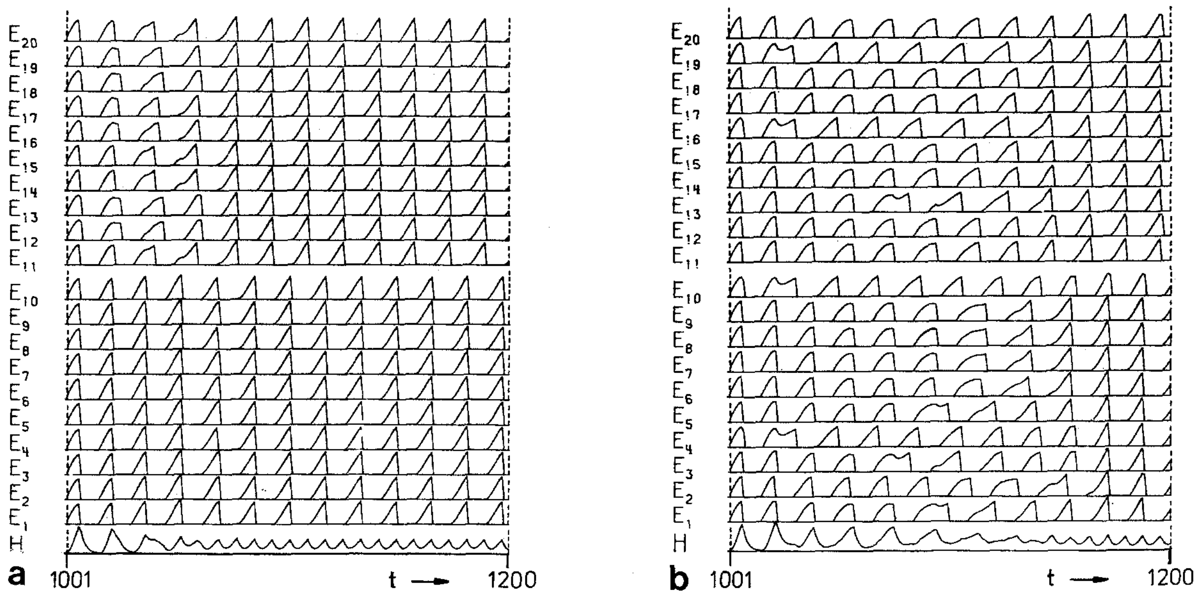

Fig. 10a, b. *Spontaneous symmetry breaks.* The two runs continue the one in Fig. 8 and are identical except for independent noise. Activities were reset to zero and both stimuli were switched on at $t = 1001$. Complete synchrony is unstable due to the blockstructure of the synaptic matrix, which is shown in Fig. 9. Both runs eventually reach desynchronization between the groups defined by the original stimulus onsets. Run **a** was chosen as the fastest, run **b** as the slowest to desynchronize among the more than 30 tests which were run. Transition in **a** is a collective wave, in **b** it starts with individual cells and is chaotic. Synaptic modulation continues during the runs

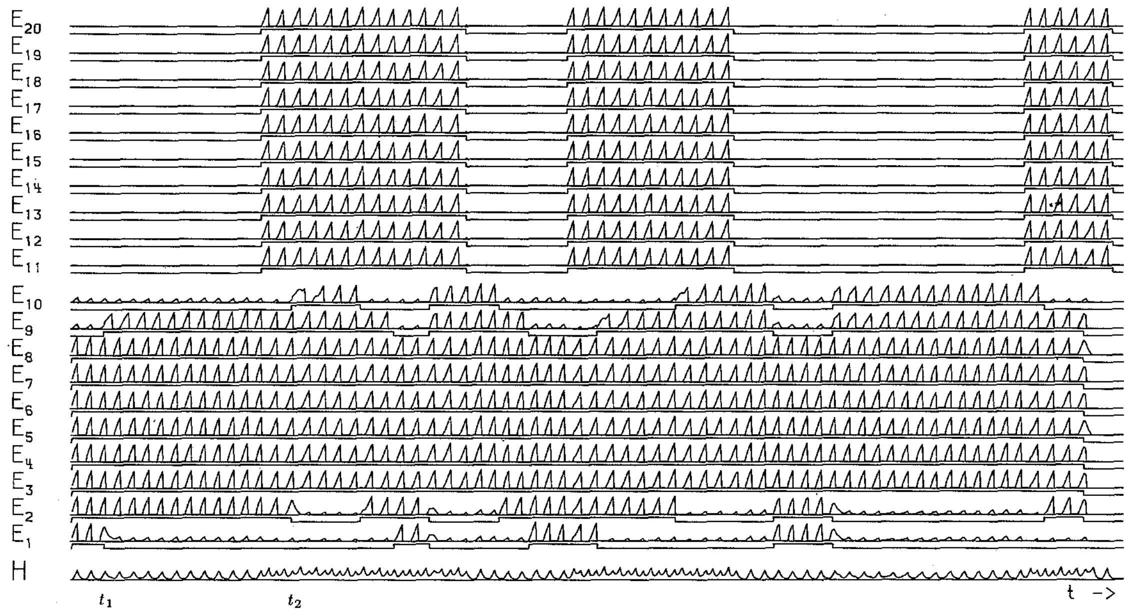

Fig. 11. *A simulation with a variable pattern.* Conventions are as in Fig. 8. All pairs of patterns in the range of cells 1 to 10 and in the range 11 to 20 have been presented to the system previously long enough to allow for full synaptic modulation (not shown). Asynchronously appearing cells are resynchronized with the correct group. Without the background signal ($E_{11} - E_{20}$) this is more difficult (e.g., for cell E_9 at time t_1) than with the background signal (e.g., for cell E_{10} at time t_2)

were not fixed). The block then decayed into two blocks of 5 cells each, the decay being complete after at least 700 steps. With $r = =0.2$ there are some excursions of individual cells, but the block of 10 stays together for several hundred steps. With $r = 0.4$ the block is stable for at least 10000 steps. All experiments were repeated 20 times with independent sequences of random numbers.

In real life, the size of stimuli in terms of numbers of activated cells will, of course, vary. In order to find out whether the function of the model critically depends on relative block size, a number of tests, comparable to the one shown in Fig. 8, were run with block sizes 11 to 9, 12 to 8, 13 to 7 and 14 to 6. At the latter test the series started to be problematic and a stimulus onset asynchrony of 2 steps was necessary to segment the two stimuli. Otherwise the function of the model was as smooth as in Fig. 8, except for an occasional escape of one or two cells from the bigger block to the smaller (depending on the random number sample).

The envelope of a spectrum may change. New spectral lines may appear and old may disappear. In order to simulate such effects, we superposed two signals. Signal 1 consisted, as usual, of 10 partials, which were switched on and off in synchrony at irregular intervals. Signal 2 activated 8 consecutive cells, belonging at different times to the series of partials E_1-E_8, E_2-E_9, or E_3-E_{10}, see Fig. 11; thus simulating a vowel formant moving around in frequency space. Each of the three modifications of signal 2 was switched on and maintained in the presence of signal 1 sufficiently long to allow for full modulation ($r=0.8$) of synapses. After this training (not shown) it was possible to shift signal 2 at arbitrary times by one or two cells without the newly activated cells being erroneously bound to the wrong signal (1), see Fig. 11.

5 Discussion

The model and the simulations described are only a caricature, intended to communicate an idea, not to represent reality. There are only two patterns, a figure and a ground. This is not a fundamental limitation, more patterns being possible. The model restricts itself to a single quality in the afferent field, which also is not a fundamental limitation. Stochastic nervous activity has been represented by smoothly varying curves and these curves are assumed to be quasi-periodic. Finally, a full, more complicated, model will have to represent arbitrary spaces of local quality.

The model implies a general mechanism of sensory segmentation. This mechanism and its properties are to be seen in contrast to those proposed in the literature. Some authors insist, with respect to the visual modality, that a segment should be connected in terms of the topology of the sensory field (Gurari and Wechsler 1982), or that a closed boundary is of fundamental importance for segmentation. The mechanism proposed here does not require any rigid property of patterns. It is not based on the existence of a closed boundary, and it requires only that a segment be connected in terms of nervous connections. This is in agreement with the fact that patterns to be separated from each other may overlap on the sensory surface and that one pattern may be divided by another pattern into disconnected regions. This is the normal case in the auditory modality, where the partials of two spectra interleave on the frequency axis, and it is also frequent in the visual modality, when, for instance, objects are seen through loose foliage.

It is also a wide-spread idea that recognition of patterns (possibly by a single cell) is necessary before a scene can be segmented. There are indeed images the segmentation of which is dominated by central evidence, no peripheral evidence – no regularity private to a figure – being present. In that case segmentation is only possible after recognition, which is extremely slow because it has to start with spontaneous hypotheses. This is a rare extreme case. There is also the opposite extreme, in which an as yet unknown pattern is defined exclusively on the basis of peripheral evidence. In all normal cases, central and peripheral evidence are integrated with each other, and segmentation is arrived at in a bootstrapping fashion. Integration of evidence from various sources into the segmentation process is natural to the model proposed here, simply by adding structured couplings between the cells involved, thereby increasing their tendency to synchronize, see Fig. 3. The way in which segmentation information is represented according to the model – differential tagging instead of suppression of the background – allows for iterative corrections with the help of information on recognized patterns. In contrast, the overwhelming majority of theories of pattern recognition presuppose segmentation machinery in the form of a selective filter: the pattern recognition machinery can only digest a figure on a blank ground.

It is instructive to discuss the present model as a mechanism of selective attention. Essential to that idea is the concentration of activity into one segment at a time. In one extreme formulation, selective attention is projected into the sensory field by a central command structure. The mechanism proposed here places the selective instability in the sensory field itself. This allows the definition of a segment to be based on all the information available in the sensory field. A hybrid theory, proposed by (Sejnowski and Hinton 1985), uses a relaxation mechanism within the sensory field to define the boundaries of a segment but lets a central command structure select the segment to be activated.

The auditory case has been selected for the pilot study presented here for several reasons. Auditory patterns in tonotopic representation are not topologically coherent, emphasizing the need for a more general mechanism than just enclosure within a boundary. A simplified version based on the presence of temporal markers already in the afferent signals is possible only in the auditory case. A sensory field of one-dimensional extension, convenient for first simulations, is natural in the auditory system but would be judged as too abstract and "theoretical" in the visual modality. Finally, a simulation study based exclusively on peripheral evidence might seem more acceptable for a modality in which central evidence is often of no avail (a fact which is often distressingly felt by elderly people in a cocktail-party). All of these considerations are superficial, and they are irrelevant for an application of the idea presented here to sensory segmentation in general. A more complete model based on a multi-dimensional space of local quality has already been simulated (Schneider 1986; Schneider and von der Malsburg, in preparation) and is able to segment in the absence of stimulus onset asynchrony. The integration of central evidence in the process has been demonstrated (von der Malsburg and Bienenstock, 1986). Extension of the simulations to two-dimensional sensory fields is limited only by computer-power.

We are aware of several simplifications in the present simulation study as applied to auditory segmentation. In particular, no collisions between components of different spectra (and no combination tones) have been allowed in the stimuli. This complication has been dealt with and will be described in the more complete study (Schneider 1986; Schneider and von der Malsburg, in preparation).

The auditory modality deals with temporal patterns. These must be represented centrally. Our mechanism of segmentation is based on temporal patterns created spontaneously within the sensory field. The collision between the two types of temporal patterns must be resolved somehow. This is possible if peripheral circuits within the auditory system (cochlear nuclei, inferior olive, colliculus inferior and possibly first stages of cortical processing) encode all temporal patterns which fall into the frequency range sequestered by the segmentation machinery, and represent them with the help of slowly varying pattern-specific signals. The frequency range needed for segmentation is thus freed. Such encoding is probably required anyway for the temporal storgae of auditory stimuli.

The model is consistent with the known anatomy and physiology of the brain. It has been known for a long time that central (thalamic, cortical) neurons respond to peripheral stimuli in an irregular way and that their signals must be averaged by post-stimulus-time-histograms before regular responses can be extracted experimentally. This has often been taken as a sign of unreliability of neural machinery. Our model, and more generally Correlation Theory, assign an important function to temporal fluctuations. The existence of such fluctuations and the fact that neurons are coincidence detectors prove that theories in terms of cellular mean frequencies are unrealistic and that the processing of fluctuations and of correlations (Sejnowski 1981) must be an important aspect of brain function. The requirement of direct connections between all E-cells is not critical. Sets of components which are to be classified as one segment only have to have sufficient connectivity to prevent block-instability. This, and the existence of an inhibitory system, are very natural requirements. The model is minimal in terms of anatomical structure.

It is not possible to predict with certainty where in the auditory system sensory segmentation is performed. However, it should be central to the stages of peripheral encoding of rapid temporal patterns alluded to above, and it should be in a structure with a sufficient density of recursive connections. Likely candidates are colliculus inferior and the first areae of the auditory thalamo-cortical complex. It should be stressed that a precise localization of the segmentation circuits is not possible or useful since all direct and indirect couplings between different parts of the sensory field contribute to segmentation decisions. Among these is all pattern evaluation circuitry, which certainly includes a large part of what is called auditory cortex.

How are the components of the model to be interpreted in terms of nervous hardware? The "cells" of the model are to be identified with the large sets of neurons which respond to resolvable frequency components. These comprise thousands of neurons. Let us call those sets "units". The signals $E_i(t)$ and $H(t)$ are to be interpreted as the combined rate of all spikes produced by the neurons in a unit and by the population of inhibitory cells, respectively. The variables $G_i(t)$ correspond to a system of delayed inhibition within units, different in anatomy and physiology from the one realizing $H(t)$. The synaptic couplings between cells are realized by the thousands of axons and synapses between each pair of units. The treatment of the activity of entire units by single variables is possible only under certain constraints regarding the structure of connectivity within and between units. These will not be discussed here. The envelope of the burst of activity on a block of units still leaves room for more finely structured correlations between subsets of neurons in different units.

Experimental verification of the model is straightforward in principle. According to the model, signals should be synchronized for those central neurons which are excited by spectral components belonging to one segment. The signals of neurons should be antisynchronized for components belonging to different segments. It may be necessary to work with waking animals trained to pay attention to the required segmentation. The experiment may necessitate extensive temporal averaging in order to detect significant correlations in small sets of cells. With luck it may be possible to detect ensemble-average signals, either in the form of mass-potentials, to be recorded by low-resistance electrodes (such potentials may have been detected by (Freeman 1977) in the olfactory modality), or in the form of signals of particular single cells which happen to poll whole populations (as the inhibitory cells of the model do). The direct detection of high order correlations, which alone are important for the function of the brain, and which alone can be significant in one-shot experiments, may remain difficult or impossible for some time to come.

There is no direct evidence in the literature for what is called here "synaptic modulation", i.e., for rapid control of synaptic efficacy by pre- and postsynaptic signals. However, there is a vast literature on fast synaptic modulation brought about by chemical signals (for citations see Kandel et al. 1983). It therefore seems more than worthwhile to look for activity-controlled synaptic modulation, e.g., in tissue-culture.

Very little psychophysical work has been done on auditory figure-ground separation. The main obstacle may have been technical difficulties in the past to produce appropriate auditory stimuli and to develop appropriate psychoacoustic criteria for successful segmentation by the subject. The situation may change radically, now that fast digital signal processors are readily available. A promising type of stimuli would be pairs of spectra which together have a flat envelope (M.R. Schroeder, personal communication), and which each have a recognizable timbre if separated. From the point of view of the present model it would be interesting to learn about the temporal characteristics of auditory segmentation: How long is the "incubation period" necessary to create reliable segmentation; what is the temporal resolution with which varying segmentation patterns can be perceived; what is the threshold on stimulus onset asynchrony (e.g., Dannenbring and Bregman 1978; Grey and Moorer 1977) beyond which segmentation occurs. These and related questions could be used to tune parameters of the model and, hopefully, to find critical tests for or against the applicability of the model. Also independently of the verification of the model presented here, more knowledge about the physical basis of auditory segmentation and especially about the factors responsible for the reduction of this ability with age and under pathological conditions, would be of enormous practical importance.

References

Bregman AS, Pinker S (1978) Auditory streaming and the building of timbre. Can J Psychol/Rev Can Psychol 31:151–159

Cherry EC (1953) Some experiments on the recognition of speech, with one and with two ears. J Acoust Soc Am 25:975–979

Cherry EC, Taylor WK (1954) Some further experiments upon the recognition of speech, with one and with two ears. J Acoust Soc Am 26:554–559

Crick F (1984) Function of the thalamic reticular complex: the searchlight hypothesis. Proc Natl Acad Sci USA 81:4586–4590

Dannenbring GL, Bregman AS (1978) Streaming vs. fusion of sinusoidal components of complex tones. Percept Psychophys 24:369–376

Freeman WJ (1977) Spatial properties of an EEG event in the olfactory bulb and cortex. Electroencephalogy Clin Neurophysiol 44:586–605

Grey JM, Moorer JA (1977) Perceptual evaluations of synthesized musical instrument tones. J Acoust Soc Am 62:454–462

Gurari EM, Wechsler H (1982) On the difficulties involved in the segmentation of pictures. IEEE Trans PAMI 4:304

Helmholtz H (1885) On the sensations of tone as a physiological basis for the theory of music. Trans. by A.J. Ellis from the 1877 German edition

Kandel ER, Abrams T, Bernier L, Carew TJ, Hawkins RD, Schwartz JH (1983) Classical conditioning and sensitization share aspects of the same molecular cascade in Aplysia. Cold Spring Harbor Symposium on Quant. Biology XLVIII

Malsburg C von der (1981) The correlation theory of brain function. Internal Report 81–2, Dept. of Neurobiology, Max-Planck-Institute for Biophysical Chemistry, Göttingen

Malsburg C von der (1985) Nervous structures with dynamical links. Ber Bunsenges Phys Chem 89:703–710

Malsburg C von der, Bienenstock E (1986) Statistical coding and short-term synaptic plasticity: a scheme for knowledge representation in the brain. In: Bienenstock E, Fogelman F, Weisbuch G (eds) Disordered systems and biological organization. Springer, Berlin Heidelberg New York Tokyo, pp. 247–272

McAdams S (1982) Spectral fusion and the creation of auditory images. In: Clynes M (ed) Music, mind and brain: the neuropsychology of music. Plenum, New York

Mitchell OMM, Ross CA, Yates GH (1971) Signal processing for a cocktail party effect. J Acoust Soc Am 50:656–660

Parsons TW (1976) Separation of speech from interfering speech by means of harmonic selection. J Acoust Soc Am 60:911–918

Rasch RA (1978) The perception of simultaneous notes such as in polyphonic music. Acoustica 40:21–33 (1978)

Reichardt W, Poggio T, Hausen K (1983) Figure-ground discrimination by relative movement in the visual system of the fly. Part II. Towards the neural circuitry. Biol Cybern 46:6 (Suppl.) 1–30

Schneider W (1986) Anwendung der Korrelationstheorie der Hirnfunktion auf das akustische Figur-Hintergrund-Problem (Cocktailparty Effekt). Doctoral thesis, Universität Göttingen

Schneider W, Malsburg C von der (in preparation). A neural cocktail-party processor based on the full spectrum of auditory qualities

Sejnowski TJ (1981) Skeleton filters in the brain. In: Hinton GE, Anderson JA (eds) Parallel models of associative memory. Lawrence Erlbaum, Hillsdale NJ, pp 189–211

Sejnowski TJ, Hinton GE (1985) Separating figure from ground with a Boltzmann machine. In: Arbib MA, Hanson AR (eds) Vision, brain and cooperative computation. MIT Press, Cambridge

Strube HW (1981) Separation of several speakers recorded by two microphones (cocktail-party processing). Signal Processing 3:355–364

Treisman AM (1980) A feature-integration theory of attention. Cogn Psychol 12:97–136

Received: November 22, 1985

Dr. Ch. von der Malsburg
Max-Planck-Institut für Biophysikalische Chemie
Abteilung Neurobiologie
Postfach 2841
D-3400 Göttingen
Federal Republic of Germany