Communicated by Ralph Linsker

Temporal Differentiation and Violation of Time-Reversal Invariance in Neurocomputation of Visual Information

D. S. Tang
Microelectronics and Computer Technology Corporation,
3500 West Balcones Center Drive, Austin, TX 78759-6509 USA

V. Menon
Department of Computer Sciences, University of Texas at Austin,
Austin, TX 78712 USA

Information-theoretic techniques have been employed to study the time-dependent connection strength of a three-layer feedforward neural network. The analysis shows (1) there is a natural emergence of time-dependent receptive field that performs temporal differentiation and (2) the result is shown to be a consequence of a mechanism based on violation of the time-reversal invariance in the visual information processing system. Both analytic and numerical studies are presented.

1 Introduction

A synthetic three-layer feedforward neural network is proposed and shown to have the ability to detect time-dependent changes in the input signal. Using a few physical assumptions about the properties of the transmission channels, we deduce purely within the information-theoretical framework (Shannon and Weaver 1949) that temporal differentiation is based on the violation of time-reversal invariance of the information rate. These results may be relevant to the study of early visual information processing in the retina and the construction of artificial neural network for motion detection. An earlier information-theoretic study (Linsker 1989) with a different model has indicated that the cell's output is approximately a linear combination of a smoothed input and a smoothed first time derivative of the input. Here, we show analytically how the temporal differentiation emerges. Section 2 discusses the information-theoretic formalism of time-dependent neural nets in general terms. A simple three-layer feedforward network is introduced. Analytic solutions to the time-dependent transfer function are derived in Section 3, which are followed by a study of the properties with further numerical calculations in Section 4. Section 5 describes the input–output relations between the layers. Section 6 summarizes the main results.

Neural Computation **2, 162–172** (1990)

2 Time-Dependent Information-Theoretic Formalism

Consider a three-layer feedforward network (Linsker 1986). Each layer of neurons is assumed to be two-dimensional in the X–Y plane spatially. Light signals with variable intensity are assumed to be incident on the first layer, layer A, in the Z-direction. Activities of the layer A neurons are relayed to the second layer, layer B. The output activities in layer B are then sent to layer C neurons.

Below we specify the assumptions of our model.

1. Each output neuron at location $\boldsymbol{r}$ is locally connected to its input neurons in the previous layer. The input neurons are spatially distributed according to the gaussian distribution density $\rho(\boldsymbol{R}) = C_\rho \exp(-R^2/2\alpha^2)$ with $C_\rho \equiv N/2\pi\alpha^2$. N is the total number of input neurons, α^2 is the variance, and $\boldsymbol{R}$ is measured relative to $\boldsymbol{r}$. All spatial vectors are two-dimensional in the X–Y plane.

2. The output signal, $Y_i(t)$, of the ith layer B neuron is assumed to be linear,

$$Y_i(t) = C_0 \sum_{j=1,\ldots,N}^{(i)} \sum_{\tau=t-\delta,\ldots,t} e^{-b(t-\tau)} X_j(\tau) \tag{2.1}$$

   Each connection has a constant weight C_0 modulated by a microscopic time delay factor $\exp[-b(t-\tau)]$ as in an RC circuit. This time delay models a physical transmission property of the channel, which possesses a simple form of memory, i.e., past signals persist in the channel for a time interval of the order $1/b$. Here, b is the reciprocal of the decay time constant. The temporal summation is from a finite past time $t-\delta$ to the present time t. From the causality principle, no future input signals $X_j(T), T > t$, contribute to the present output signal $Y_i(t)$. It is assumed that the time scale $\delta \gg 1/b$ is satisfied. The index (i) in the spatial summation means that the N input neurons are randomly generated according to the density ρ relative to the location $\boldsymbol{r}_i$ of the ith output neuron in accordance with assumption 1. The stochastic input signals $X_j(\tau)$ are assumed to satisfy the a priori probability distribution function

$$P(X) = \frac{e^{[-(X-\tilde{X})^T B^{-1}(X-\tilde{X})]/2}}{(2\pi)^{3\tilde{N}/2}\sqrt{\text{Det } B}} \tag{2.2}$$

   which is a gaussian and statistically independent, $B_{ij}^{t\tau} = v\delta_{ij}\delta_{t\tau}$. Here, $\tilde{N}$ is the total number of distinct space–time labels of $X_j(\tau)$. $\tilde{X}$ is the mean of $X_j(\tau)$.

3. The output signal, $Z_m(t)$, of the mth neuron in layer C is assumed to be

$$Z_m(t) = \sum_{i=1,\ldots,N}^{(m)} \sum_{\tau=t-\delta,\ldots,t} \mathcal{H}_i(\tau) Y_i(\tau) \tag{2.3}$$

where the transfer function $\mathcal{H}_i(\tau)$ satisfies the constraints (1) $\mathcal{H}^T\mathcal{H} = A_0$ in matrix notation, and (2) $[\sum_{i=1,\ldots,N;\tau=t-\delta,\ldots,t} \mathcal{H}_i(\tau)]^2 = A_1$. A_0 and A_1 are real constants. The first constraint is on the normalization of the transfer function. The second constraint restricts the value of the statistical mean of the transfer functions up to a sign. The resultant restriction on the transfer functions when these two constraints are considered together is a net constraint on the variance of the transfer functions since the variance is directly proportional to $A_0 - A_1$. In the far-past, $\tau < t - \delta$, and the future, $\tau > t$, $\mathcal{H}_i(\tau)$ is set to zero. These transfer functions will be determined by maximizing the information rate in the next section. Here, an additive noise $n_i(\tau)$ added to $Y_i(\tau)$ [i.e., $Y_i(\tau) \to Y_i(\tau) + n_i(\tau)$] in equation (2.3) is assumed for the information-theoretic study. The noises $n_i(\tau)$ indexed by the location label i and the time label τ are assumed to be statistically independent and satisfy a gaussian distribution of variance β^2 with zero mean.

Below we derive the information-theoretic equation that characterizes the behavior of the transfer function $\mathcal{H}_i(\tau)$. The information rate for the signal transferred from layer B to layer C can be shown to be

$$R(Z) = \frac{1}{2} ln[1 + \frac{\mathcal{H}^T W \mathcal{H}}{\beta^2 \mathcal{H}^T \mathcal{H}}] \tag{2.4}$$

with W being the spatiotemporal correlation matrix $\langle\langle Y_i(\tau) Y_j(\tau')\rangle\rangle$. From equations (2.1) and (2.2), the following expressions for the correlation matrix can be derived.

$$\langle\langle Y_i(\tau) Y_j(\tau')\rangle\rangle = h Q_{ij} G_{\tau\tau'} \tag{2.5}$$

with

$$Q_{ij} = e^{-\frac{|\boldsymbol{r}_i - \boldsymbol{r}_j|^2}{4\alpha^2}} \tag{2.6}$$

and

$$G_{\tau\tau'} = e^{-b|\tau-\tau'|} \theta(\delta - |\tau - \tau'|) \tag{2.7}$$

Here, θ denotes the Heaviside step function. To arrive at equation (2.7), terms of order $\exp(-b\delta)$ have been neglected as the time scale assumption $\delta b \gg 1$ is evoked. h is a constant independent of space and time. Its explicit form is irrelevant in subsequent discussions, as will be shown below.

Now, the method of Lagrange multiplier is employed to optimize the information rate. This is equivalent to performing the following variational calculation with respect to the transfer function

$$\delta \left\{ \mathcal{H}^T W \mathcal{H} + k_2 \left[\left[\sum_{i\tau} \mathcal{H}_i(\tau) \right]^2 - A_1 \right] \right\} = 0 \tag{2.8}$$

Here, k_2 is a Lagrange multiplier. It is a measure of the rate of change of the signal with respect to the variation of the value of the overall transfer function. The variational equation above produces the following eigenvalue problem governing the behavior of the transfer function

$$\lambda \mathcal{H}_i(\tau) = h \sum_{j\tau'} [Q_{ij} G_{\tau\tau'} + \frac{k_2}{h}] \mathcal{H}_j(\tau') \tag{2.9}$$

This equation defines the morphology of the receptive field of the neurons in layer C in space and time. For simplicity, one can absorb the multiplicative factor h in λ and k_2. We assume that this is done by setting $h = 1$.

3 Analytic Expressions of the Transfer Function

In continuous space and time variables, the eigenvalue problem is reduced with the use of the time-translational invariance of the temporal correlation function (equation 2.7) to the homogeneous Fredholm integral equation

$$\begin{aligned} \lambda \mathcal{H}(\boldsymbol{r}, \tau) &= \int_{-\infty}^{\infty} d\boldsymbol{R} \int_{-\delta/2}^{\delta/2} d\tau' \rho(\boldsymbol{R}) \left[Q(\boldsymbol{r} - \boldsymbol{R}) G(\tau - \tau') + k_2 \right] \\ & \quad \mathcal{H}(\boldsymbol{R}, \tau') \end{aligned} \tag{3.1}$$

Here, $Q(\boldsymbol{r} - \boldsymbol{R})$ and $G(\tau - \tau')$ are the continuous version of Q_{ij} and $G_{\tau\tau'}$ of equations (2.6) and (2.7), respectively. Note that the kernel respects both space- and time-reversal invariances. It means that the kernel itself does not have a preferential direction in time and space. Note also that the information rate also respects the same symmetry operation. However, if it turns out that the solutions $\mathcal{H}(\boldsymbol{r}, \tau)$ do not respect some or all of the symmetries of the information rate, a symmetry-breaking phenomenon is then said to appear. This can have unexpected effects on the input–output relationship in equation (2.3). In particular, the violation of time-reversal invariance underlies temporal differentiation, the ability to detect the temporal changes of the input signal. This will be discussed in the next section.

The solutions to this integral equation can be found by both the Green's function method and the eigenfunction expansion method (Morse and Feshbach 1953). They can be classified into different classes according to their time-reversal and space-reversal characteristics, $\mathcal{H}(\boldsymbol{r},\tau) = \mathcal{H}(-\boldsymbol{r},-\tau), \mathcal{H}(\boldsymbol{r},\tau) = -\mathcal{H}(-\boldsymbol{r},\tau), \mathcal{H}(\boldsymbol{r},\tau) = \mathcal{H}(\boldsymbol{r},-\tau)$, $\mathcal{H}(\boldsymbol{r},\tau) = \mathcal{H}(-\boldsymbol{r},-\tau)$, ..., etc. In this study, we are interested only in the solutions that correspond to the two largest eigenvalues (i.e., the larger one of these two has the maximum information rate). They are as follows.

3.1 Symmetric Solution. This solution obeys both space- and time-reversal invariance.

$$\begin{aligned} \mathcal{H}(\boldsymbol{r},\tau) &= \frac{h_0}{\eta}\Bigg\{1 + \frac{2D}{b\eta}[1 - e^{-b\Delta}\cosh(b\tau)]e^{-(r^2/2\sigma_0^2)} \\ &\quad + \left[-\Delta - \frac{2D}{b\eta}[1 - e^{-b\Delta}\cosh(b\tau)] + D\eta\mathcal{T}(\tau,\eta)\right] e^{-(r^2/2\sigma_\infty^2)}\Bigg\}, \\ &\quad -\delta/2 \le \tau \le \delta/2 \\ \mathcal{H}(\boldsymbol{r},\tau) &= 0, \quad \text{otherwise} \end{aligned} \tag{3.2}$$

with $\Delta = \delta/2$, $D = 1.244$, $\eta = \lambda/1.072C_\rho\pi\alpha^2$, and

$$\mathcal{T}(\tau,\eta) = \frac{b}{2 - b\eta}\left\{\frac{(b^2+\gamma^2)\cos\gamma\tau}{b[b\cos\Delta\gamma - \gamma\sin(\Delta\gamma)]} - 1\right\} \tag{3.3}$$

with γ being determined by

$$\eta = \frac{2b}{(b^2+\gamma^2)} \tag{3.4}$$

In equation (3.2), h_0 is the normalization constant. The eigenvalue is determined from η by

$$\begin{aligned} \frac{1}{k_2} &= \frac{2\alpha^2\pi C_\rho}{\lambda}\Bigg\{\left(1 - \frac{D}{[1+(\alpha^2/\sigma_\infty^2)]}\right)\delta \\ &\quad + \frac{2D}{b\eta}\left(\frac{1}{[1+(\alpha^2/\sigma_0^2)]} - \frac{1}{[1+(\alpha^2/\sigma_\infty^2)]}\right)\left(\delta + \frac{[\exp(-b\delta) - 1]}{b}\right) \\ &\quad + \frac{Db\eta}{[1+(\alpha^2/\sigma_\infty^2)](2-b\eta)}\left[\frac{2(b^2+\gamma^2)\sin\gamma\Delta}{b\gamma(b\cos\gamma\Delta - \gamma\sin\Delta\gamma)} - \delta\right]\Bigg\} \end{aligned} \tag{3.5}$$

Note that all the space- and time-dependent terms in equation (3.2) are even functions of both space and time. In obtaining these equations, we have used in the eigenfunction expansion the fact that the iteration equation $2\alpha^2/\sigma_{i+1}^2 = 1 - 1/(3 + 2\alpha^2/\sigma_i^2)$ with $\sigma_0^2 = 3\alpha^2$ converges rapidly to $2\alpha^2/\sigma_\infty^2$ so that the approximation $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_i^2 = \ldots = \sigma_\infty^2 = 2.7320\alpha^2$ is accurate with error less than one percent. Details may be found in Tang (1989).

3.2 Time-Antisymmetric–Space-Symmetric Solution. The transfer function is

$$\begin{aligned}\mathcal{H}(\boldsymbol{r},\tau) &= h_1 e^{-(r^2/2\sigma_\infty^2)} \sin \gamma\tau, \quad -\delta/2 \le \tau \le \delta/2 \\ \mathcal{H}(\boldsymbol{r},\tau) &= 0, \quad \text{otherwise}\end{aligned} \tag{3.6}$$

The eigenvalue is

$$\lambda = \frac{\lambda_T}{0.268(3 + 2\alpha^2/\sigma_\infty^2)} \tag{3.7}$$

with λ_T being determined by

$$\lambda_T = \frac{2b}{b^2 + \gamma^2}$$

in which γ satisfies the following self-consistent equation

$$\gamma = -b \tan \Delta\gamma \tag{3.8}$$

h_1 is the normalization constant.

4 Properties of the Solution to the Eigenvalue Problem

The behavior of the eigenvalue of the symmetric solution (equation 3.5) is shown as the curved line in Figure 1. In contrast, the eigenvalues of the antisymmetric solution (equation 3.7) are independent of the parameter k_2 since the integral $\int_{-\delta/2}^{\delta/2} \sin\ \gamma\tau\, d\tau$ is identically zero. Therefore, the eigenvalue for the antisymmetric case remains constant as k_2 varies, as depicted by the horizontal line in Figure 1. The statement that the eigenvalue does not depend on k_2 is true for any eigenfunctions that are antisymmetric with respect to either time- or space-reversal operation. The interesting result is that for positive and not too negative k_2 values, the eigenvalue of the symmetric solution is the largest. That is, the transfer function producing the maximum information rate respects the symmetry (time- and space-reversal invariances) of the information rate. However, as k_2 becomes sufficiently negative, the time-antisymmetric solution has the largest eigenvalue since the maximum symmetric eigenvalue decreases below that of the time-antisymmetric solution. In this regime, the symmetry breaking of the time-reversal invariance occurs. Within the range of arbitrarily large positive value of k_2 to the point of the transition of the symmetry breaking of the time-reversal invariance, parity conservation is respected. The transfer functions at the center of the receptive field plotted against time for different values of k_2 are shown in Figure 2. Transfer functions located farther from the center show similar behavior except for a decreasing magnitude, as a result of the spatial gaussian decay. It is found that no spatial center-surround morphology corresponding to maximum information rate appears for any given time in the k_2 regime we are considering.

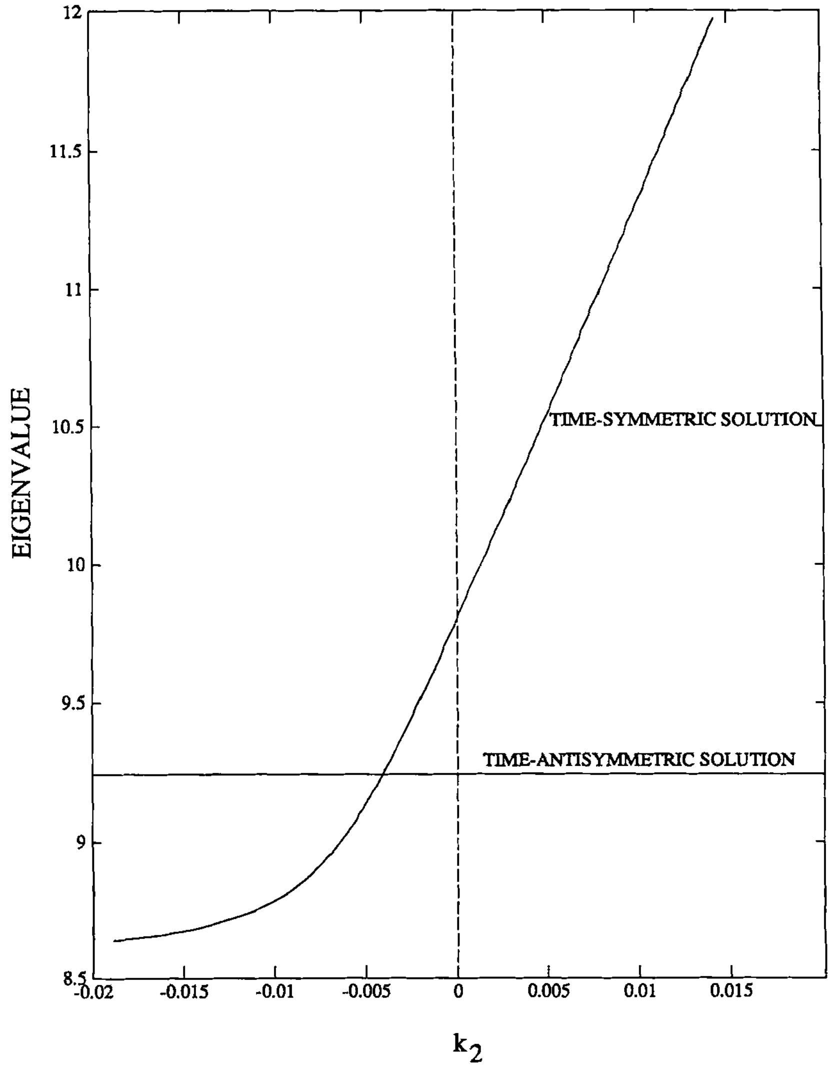

Figure 1: The maximum eigenvalues for the symmetric and time-antisymmetric solutions plotted against the Lagrange multiplier k_2. Here, the time constant $1/b$ is arbitrarily chosen as 5 msec.

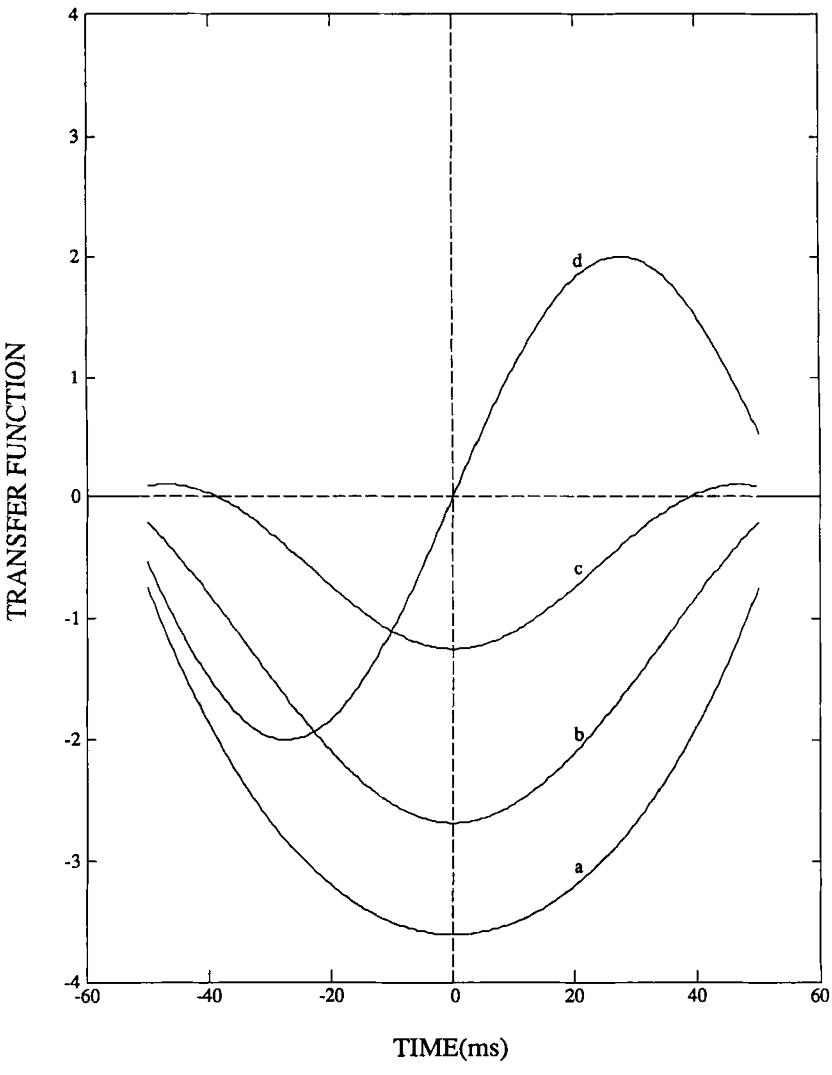

Figure 2: Time-dependent properties of the temporal transfer function. The present time is arbitrarily chosen to be at 50 msec. The time in the far-past is at −50 msec. Curves a, b, and c depict the symmetric transfer functions with eigenvalues 10.2, 9.5, and 9, respectively. Curve d is the time-antisymmetric transfer function with eigenvalue equal to 9.24. The time constant $1/b$ is 5 msec.

5 Input–Output Relations

From the results of last section, the output events can naturally be classified into time-symmetric or time-antisymmetric events. The time-symmetric/antisymmetric events are defined as the set of output activities obtained by the linear input–output relationship having a time-reversal invariant/noninvariant transfer function. They are statistically independent from each other, from equations (2.3) and (2.9)

$$\langle\langle Z_m^{\text{symmetric}}(t) Z_m^{\text{antisymmetric}}(t)\rangle\rangle = 0 \tag{5.1}$$

The symmetric transfer function (equation 3.2) does not discriminate the inputs in the near-past $[0, \delta/2]$ from those in the far-past $[-\delta/2, 0]$. The output event duplicates as much of the input signals as possibly allowed in this three-layer feedforward network operating in an optimal manner within the information-theoretic framework. This is illustrated by curves a, b, and c in Figure 3. In obtaining these figures, the light signal impinging on layer A is assumed to be a stationary light spot modeled by a step function $\theta(t)$, that is, it is off for time $t \leq 0$ and on for time $t \geq 0$. The curve labeled layer B output is a typical output signal of the response of an RC circuit with a step function input (equation 2.1). These results suggest that the input–output relation with the symmetric transfer function (equation 3.2) in this k_2 regime operates in the information-relaying mode. It acts as a passive relay. The speed of the response is determined by the width δ of the time window — the shorter the width the faster the speed. The behavior of the input-output relation with the time-antisymmetric transfer function (equation 3.6) is totally different. It is a simple form of temporal differentiation as illustrated by curve d in Figure 3. Temporally constant input signal produces zero output. Note that the peak of the output is time delayed by 0.5δ compared with the time defining the fastest change in the input function. This temporal differentiation is not identical to the time derivative in calculus, even though both can detect the temporal changes in the input signal and both are time-antisymmetric operations. These transfer functions process the input signals to form the output differently. In the regime of time-reversal invariance, the transfer functions (equation 3.2) simply relay the input to the output without actively processing the inputs, curves a, b, and c in Figure 3. However, in the regime of time-reversal noninvariance, the transfer functions (equation 3.6) have the capability to extract the temporal changes of the temporal input signals, curve d in Figure 3.

6 Conclusions

We have analyzed, from information-theoretic considerations, how a simple synthetic three-layer, feedforward neural network acquires the ability

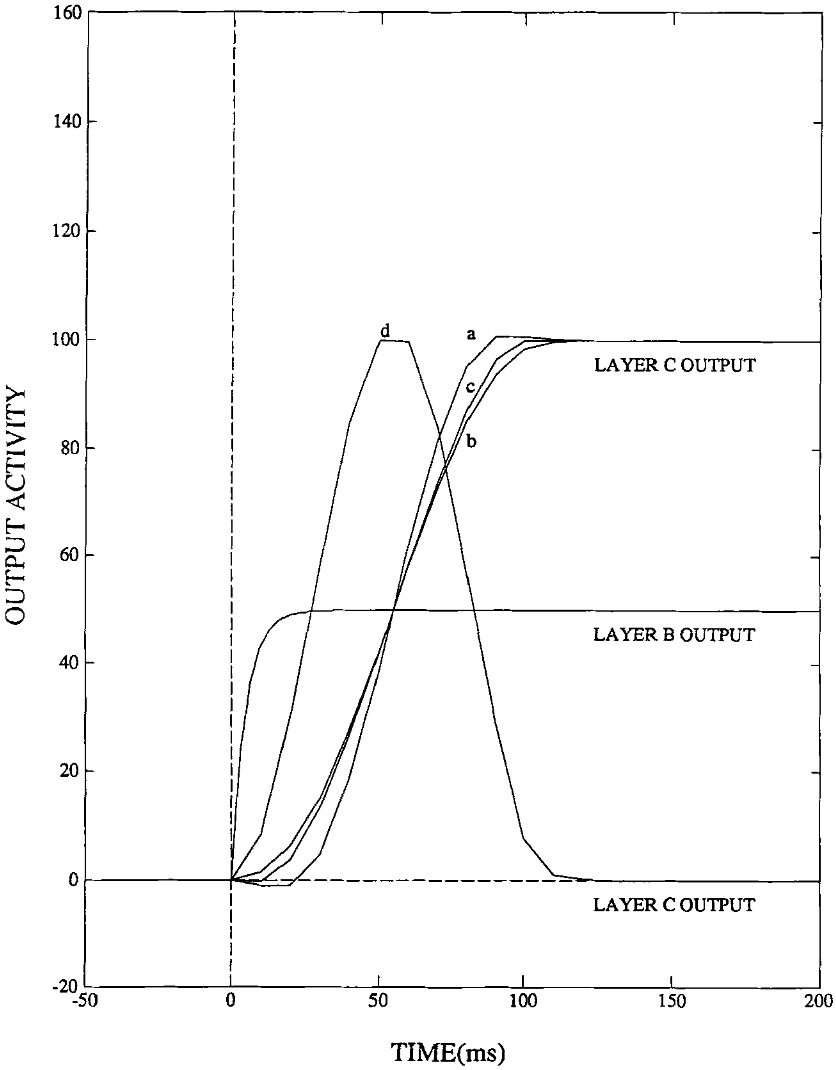

Figure 3: The input–output relations. The input to layer B is a stationary and local light spot defined by $\theta(t)$. Curves a, b, and c are layer C outputs that correspond to the symmetric transfer functions a, b, and c in Figure 2, respectively. Curve d is the layer C output with time-antisymmetric transfer function corresponding to the d curve in Figure 2. Note that only curve d signals the temporal change in the incoming signal.

to detect temporal changes. Symmetry breaking in time-reversal invariance has been identified as the source of such an ability in our model. Furthermore, the symmetry-classes to which the transfer function belongs define the distinct categories of the temporal events in the output sample space. We summarize below the main results of this paper.

1. The persistence of signals in the network is an important aspect of temporal signal processing. In the present study we have modeled this as a channel with memory (equation 2.1) and it underlies the results obtained.

2. Eigenvectors of the constrained spatial-temporal correlation function are the transfer functions of the three-layer, feedforward neural network model studied here.

3. The symmetries that the eigenvectors respect and define the classes of the output events.

4. There are two modes of the temporal information processing: one is the information-relaying mode defined by the time-symmetric transfer function (equation 3.2) and the other is the information-analyzing mode defined by the time-antisymmetric transfer function (equation 3.6).

5. Realization of the information-analyzing mode is done by a symmetry-breaking mechanism.

6. The breaking of the time-reversal invariance leads to temporal differentiation.

References

Linsker, R. 1986. From basic network principles to neural architecture: Emergence of spatial-opponent cells. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7508.

Linsker, R. 1988. Self-organization in a perpetual network. *Computer* **21**(3), 105.

Linsker, R. 1989. In *Advances in Neural Information Processing System* I, D. S. Touretzky, ed., p. 186. Morgan Kaufman, San Mateo, CA.

Morse, P. M., and Feshbach, H. 1953. *Methods of Theoretical Physics*, Vol. 1. McGraw-Hill, New York.

Shannon, C. E., and Weaver, W. 1949. *The Mathematical Theory of Communication*. Univ. of Illinois Press, Urbana.

Tang, D. S. 1989. Information-theoretic solutions to early visual information processing: Analytic results. *Phys. Rev. A* **40**, 6626.

Received 16 October 1989; accepted 23 February 1990.