

# 辨识阅读眼动注视点参数的级联神经网络<sup>1)</sup>

陈凌育 孙复川

(中国科学院上海生理研究所神经生物学开放实验室 上海 200031)  
及生物物理所视觉信息加工开放实验室

L. Stark

(美国加利福尼亚大学伯克利分校遥控机器人学和神经学部)

**摘 要** 提出了一种由三层感知机和单层线性网络构成的级联神经网络,能用于自动辨识眼动实验数据中的注视停顿并确定其时程. 计算机仿真结果表明,该网络经过学习训练后能快速、准确地实时处理眼动实验数据,相对误差小于 0.2%.

**关键词** 神经网络,图形辨识,阅读眼动,注视停顿,注视时程.

## 1 引言

文字阅读是一种视觉模式识别过程. 在阅读中,眼球并不是均匀地扫描每一行,它的运动模式是由一系列快速跳动(saccades)和注视停顿(fixation pauses)所组成,由快速跳动改变眼球位置,由注视停顿采集信息<sup>[1-3]</sup>. 在阅读眼动中,最重要的是注视停顿点的参数,因为只要求得注视停顿的次数、位置、时程,就可以从中计算出阅读过程中的其它参数如辨认跨距、眼球跳动幅度等<sup>[3]</sup>. 由于人在阅读过程中的眼部运动轨迹反映了人脑的信息处理过程,因此分析阅读眼动过程中各项参数不仅对研究阅读有重要意义,而且也利于探讨脑的信息处理机制<sup>[3-5]</sup>.

分析注视停顿参数的传统方法及存在问题:

1) 人工分析法:此法是直接对眼动位置的时间函数(图 1 所示)用人工进行分析,由于注视停顿是在二次眼球快速跳动之间眼球位置基本不变的一种状态,在时间函数上就表现为一个平台区(见图 1,虽然有的平台带有噪声,但仍属一个注视停顿),其平台宽度也就是注视停顿的时程. 人工处理时虽然可以反复斟酌判断,但很耗费时间,因此在实际工作中大量处理时容易出错.

2) 计算机程序法:随着计算机的迅速发展,开始引入计算机程序自动检测眼球跳动和注视停顿<sup>[6,7]</sup>,这也是对眼球位置的时间函数进行处理,根据统计结果从注视眼动速度、时程、位置等参数的临界范围上进行判别和测量,取得了一定效果. 但由于眼动变异性较大及噪声影响,此方法分析结果中注视停顿的点数有时会有遗漏,有时又会将眼球的其它运动或噪声误判成注视停顿,因而在可靠性和准确性上不能满足深入研究工作的要求.

1) 国家自然科学基金及中国科学院“神经科学前沿课题开拓”重大项目部分资助.

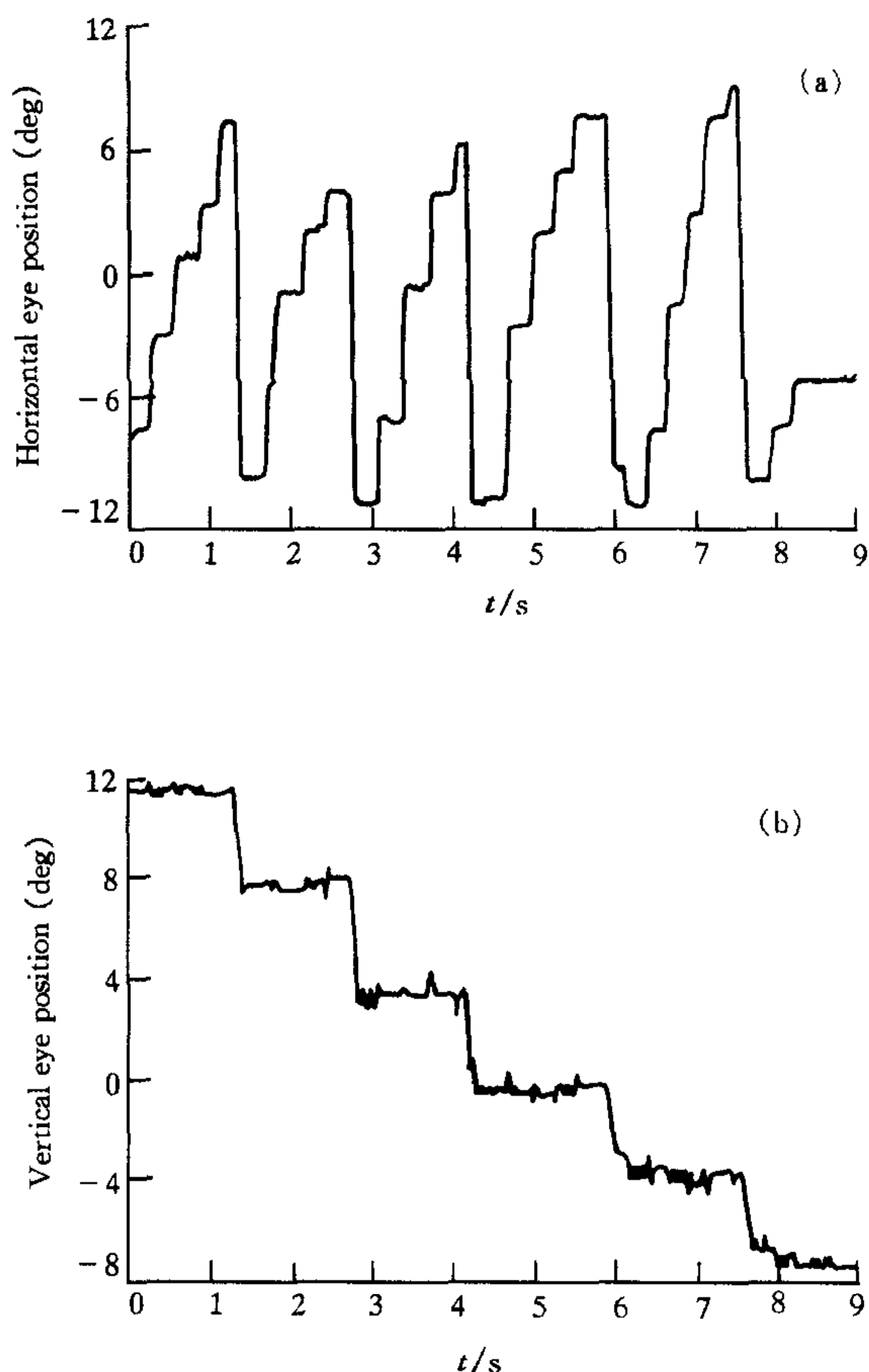


图 1 眼动位置的时间函数例图

本文提出神经网络法,由于注视停顿的含义是眼球注视在空间某一固定位置时停顿时间较长,而且仅在此固定位置附近作一些微小颤动,因此阅读眼动的研究者仔细观察眼动轨迹点的空间分布图象,就能直观地辨认出注视停顿的点数和位置.而神经网络在一定程度上能模拟人的视觉图形识别功能<sup>[8-10]</sup>,因此本工作目的就是试图利用它来实现辨认和确定眼动注视点的位置和时程.

## 2 神经网络结构及其权函数的学习规律

1) 神经网络的结构见图 2:本工作分析眼动信号参数分二步进行,即先求出一次阅读实验中眼动注视位置和注视点的总点数,然后再确定每一个注视点的注视时程.这二步任务分别由二级神经网络实现.该网络结构如图 2 所示.第一级是一个多层感知器(perceptron)网络,第二级是一个单

层线性网络<sup>[9]</sup>.前者由一个输入层、二个隐含层和一个输出层组成,主要功能是确定眼动注视点的总点数和注视点位置.后者仅包括一个输入层和一个输出层,它的输入层也就是第一级网络的第二个隐含层,此级网络主要根据每个注视停顿的时程与它所对应的眼动轨迹图中的注视点的面积大小的相关性来确定各个注视停顿的注视时程.

2) 第一级网络的具体结构及网络辨识阅读眼动注视点的学习训练:

(1) 将学习样本的阅读眼动空间分布矢量  $\{a_i\} i=1, 2, \dots, N$ , 加到网络输入端,并对它作模糊(blur)滤波及阈值处理<sup>[11]</sup>.设第一个隐含层第  $j$  个神经元的输出为  $b_j$ , 则

$$b_j = f(s_j) = f\left(\sum w_{ji} a_i - \theta_j\right). \quad (1)$$

其中  $w_{ji}$  为输入层第  $i$  个神经元与第一隐含层第  $j$  个神经元的权重,当 blur 半径为  $r$  时,

$$w_{ji} = \begin{cases} 1, & |i - j| \leq r, \\ 0, & |i - j| > r. \end{cases}$$

这里的激励函数(activation function)  $f(s_j)$  为 Heaviside 函数<sup>[8]</sup>,

$$f(s_j) = \begin{cases} 1, & s_j \geq 0, \\ 0, & s_j < 0. \end{cases}$$

$\theta_j$  为阈值,对该层各神经元相同,即  $\theta_j = \theta$ ,所以第一隐含层的输出相当于将眼动轨迹图象



转变为一种二进制数值的斑块(patch or local region)图象,它是  $r$  与  $\theta$  的函数,一个斑块粗略地对应于一个注视停顿。

(2) 由零阶及一阶矩(moments)计算,求斑块面积和 centroid<sup>[11]</sup>:

$$s_k = \sum_{j \in G_k} w_{kj} b_j \quad (2)$$

$s_k$  为第  $k$  个独立斑块区  $G_k$  的二值图象面积;并可得出各斑块图象的质心(centroid):

$$x_{ck} = \frac{\sum w_{kj} x_j b_j}{\sum w_{kj} b_j}, y_{ck} = \frac{\sum w_{kj} y_j b_j}{\sum w_{kj} b_j} \quad (3), (4)$$

其中  $w_{kj} = \begin{cases} 1, & \text{当 } j \text{ 在第 } k \text{ 个独立斑块内时,} \\ 0, & \text{其它 } j. \end{cases}$  再经阈

值激励函数的作用,则

$$c_k = f(s_k) = \begin{cases} 1, & s_k > 0, \\ 0, & s_k = 0. \end{cases}$$

即每一个  $c_k=1$  处对应于一个注视停顿。

(3) 注视停顿的总点数  $P_0$  可由输出层的神经元计算得到

$$p_0 = \sum_k w_k c_k \quad (5)$$

(4) 神经网络在训练期间的学习规则: 设网络实际输出为  $p_0$ , 网络目标输出为  $p_d$ , 由 LMS (Least Mean Square) 准则, 网络误差函数  $E_1$  为

$$E_1 = \frac{1/2(p_d - p_0)^2}{p_d} \quad (6)$$

则 blur 半径  $r$  和阈值  $\theta$  的改变应使误差函数减小。

$$\Delta r = -\beta' \frac{\partial E_1}{\partial r}, \quad \Delta \theta = -\eta' \frac{\partial E_1}{\partial \theta}.$$

在学习过程中,不断调整网络参数  $r, \theta$ , 当误差函数达到最小值时, 则有  $p_0$  趋近于  $p_d$ 。

3) 第二级神经网络的结构和确定注视时程的学习训练规则:

第二级神经网络是单层线性网络, 它的输入层也即第一级网络的第二个隐含层. 设第  $m$  个神经元的输出时程  $t_m$  和输入  $s_k$  的关系为

$$t_m = g(\sum w_{mk} s_k - \theta_m), \quad (7)$$

$w$  为对角矩阵,  $w_{mk} = \begin{cases} w, & m=k, \\ 0, & m \neq k. \end{cases}$  又因注视时程和所对应的斑块图象面积近似为线性关

系, 故可得

$$t_m = w s_m - \theta_m$$

设  $t_{dm}$  为第  $m$  个神经元的注视时程的目标期望值, 则网络误差函数为

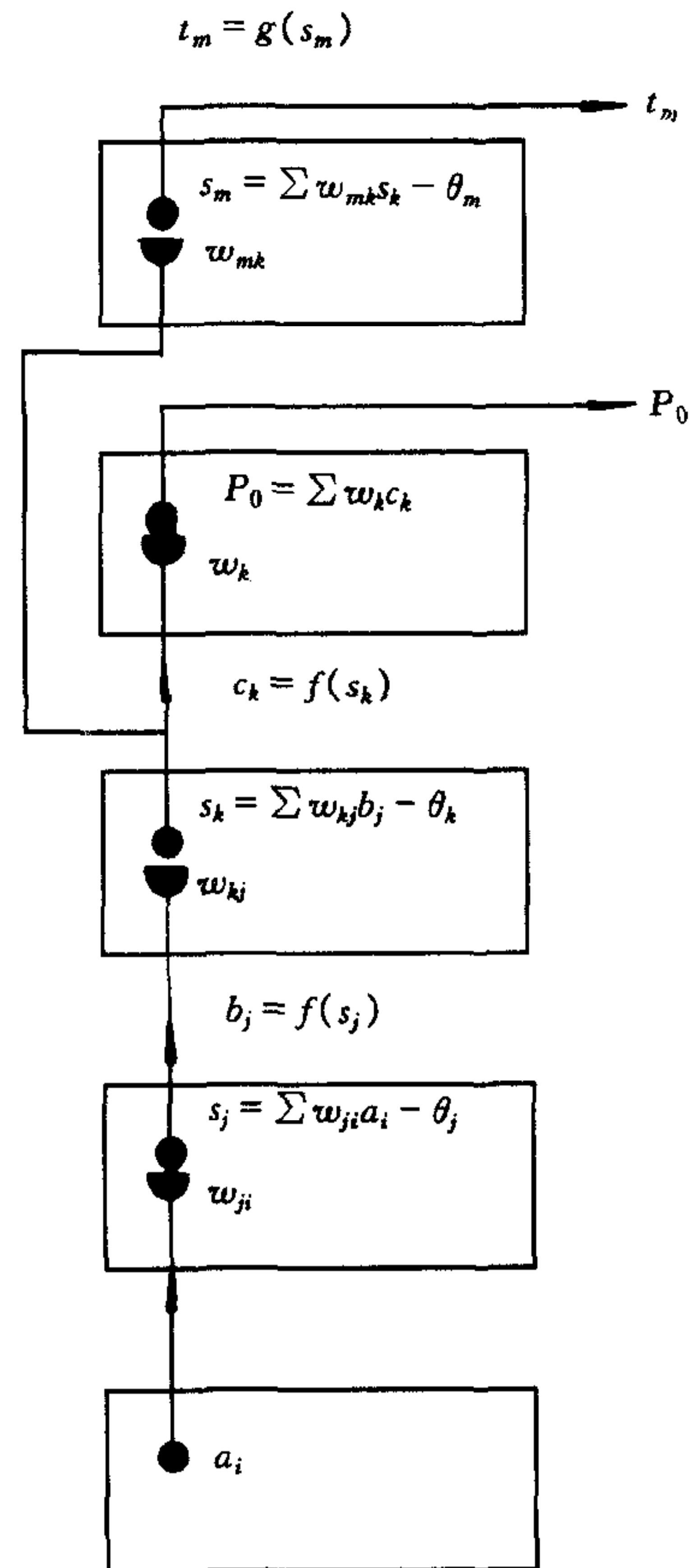


图 2 神经网络的结构图

$$E_2 = \frac{1}{2} \sum_m \frac{(t_{dm} - t_m)^2}{p_d} \quad (8)$$

权重改变为

$$\Delta w = -\eta' \frac{\partial E_2}{\partial w}$$

令  $\delta$  表示其中平均误差:  $\sum (t_{dm} - ws_m) / p_d$ , 则  $\Delta w \propto \delta$ , 可采用 sigmoid 函数以避免  $\Delta w$  在训练过程中的发散, 则

$$\Delta w = w_{old} \left( \frac{2}{1 + e^{-\lambda \delta}} - 1 \right),$$

式中  $\lambda$  是小于 1 的常数, 它决定训练步长. 经过二级网络的学习训练, 训练的结果参数  $r$ 、 $\theta$ 、 $w$  可作为网络的固定参数直接应用在其它的阅读眼动数据上.

### 3 神经网络的仿真实验与结果实例

#### 3.1 实验方法

神经网络由 C 语言编制程序进行仿真, 先后在 Silicon Graphics (IRIS-4D) 和 IBM486-66DX 计算机上进行仿真实验. 中文阅读材料选自小学课文, 眼球运动由 pup-1 红外电视眼动仪测量, 并由计算机程序控制 A/D 转换器进行数据采集. 本文记录了儿童和成人阅读中文材料的约 30 组眼动数据, 并在其中选择一组眼动模式数据作为学习训练样本, 而将其余数据作为学习训练样本成功后的结果参数的试验样本. 学习训练和试验样本的目标值(即该样本的期望值)——注视点总点数  $p_d$  和注视时程  $t_{dm}$  是由几个分析阅读眼动数据较有经验的研究工作者综合眼动位置的时间函数和空间分布图反复推敲得到的, 以用于网络的训练学习和验证比较.

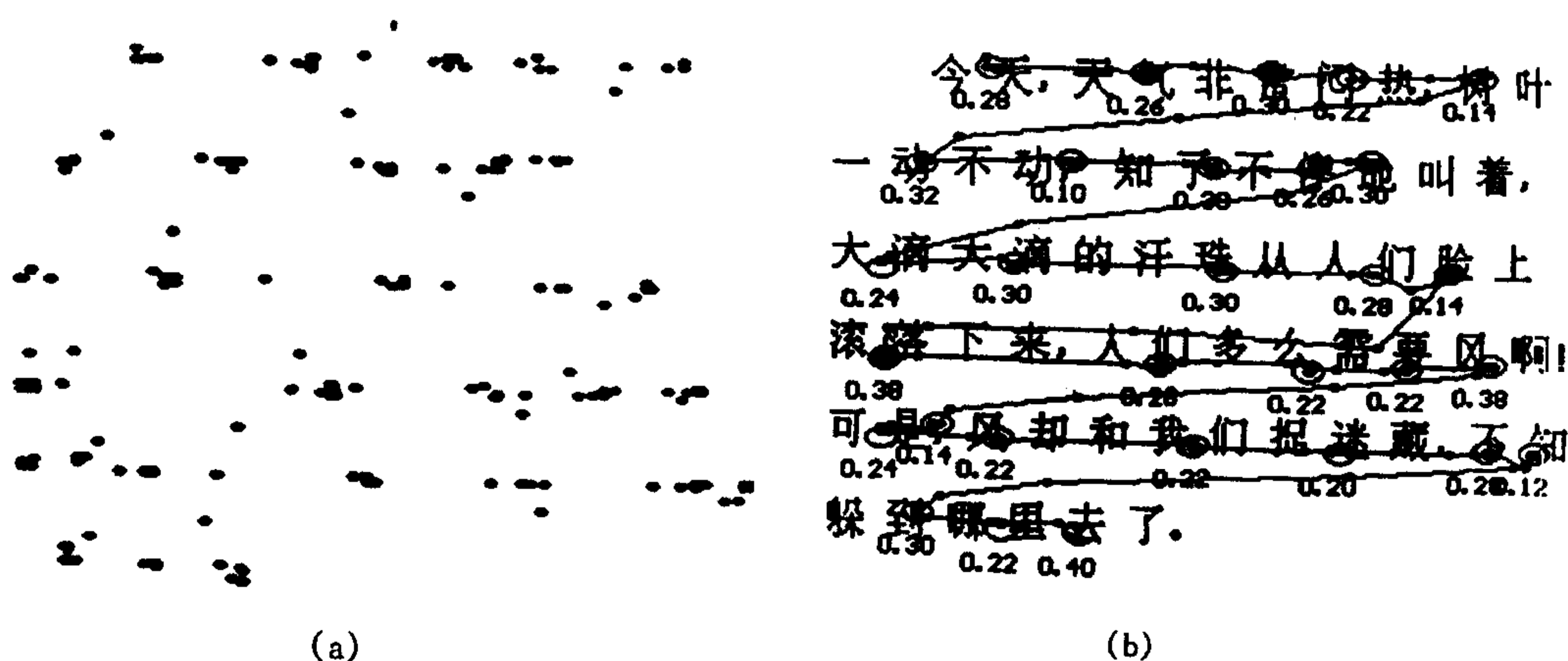


图 3 训练样本及学习完成后输出

(a) 阅读眼动空间分布图; (b) 网络学习完成后的输出与阅读课文重叠显示在一起的结果输出图. 图中圆圈为 fixation, 其下方的数字为时程, 单位为秒, 黑点为眼动轨迹.

#### 3.2 神经网络仿真过程与结果

1) 学习样本的训练过程及结果: 仿真实验中第一级 perceptron 网络的输入层及第一



隐含层神经元个数  $N$  取为  $420 \times 300$ , 第二隐含层取为 100, 决定点数的输出层为一个神经元; 第二级单层线性网络的输出层神经元个数为 100. 学习样本的阅读眼动数据的空间分布如图 3(a) 所示, 它相应的时间曲线如图 1 所示. 图 3(b) 是网络学习完成后的输出与课文重迭在一起显示. 图中黑点为眼动空间轨迹, 圆圈表示网络辨识的注视停顿, 下方的数字是它的时程, 以秒为单位. 阅读样本每个注视点的注视时程期望值和网络输出值二者拟合良好, 其均方误差仅为  $0.000147s^2$ , 相对误差为  $0.06\%$ . 第一级网络在学习训练中误差函数  $E_1$  的变化由图 4(a) 显示, 当它小于一定值时, 学习完成. 图 4(b) 是第二级网络决定注视时程的学习训练中误差函数  $E_2$  随迭代次数而收敛的情况.

2) 将学习成功的神经网络对该训练样本同一个受试者的其他阅读眼动数据进行辨识试验, 并将结果和这些数据用人工分析法测量所得期望值进行比较, 误差很小, 均方误差

小于  $0.0005s^2$ , 相对误差小于  $0.2\%$ . 图 5(a) 是其中一组数据的结果显示, 图 5(b) 是注视时程的网络输出值和期望值的对比曲线. 横轴代表 fixation 的注视点序号, 纵轴代表相应的注视点的注视时程.

3) 对其他受试者的阅读眼动数据, 网络输出值和期望值比较: 对一个样本学习训练成功后, 将经验参数  $r, \theta, w$  应用在其他受试者的共二十余组阅读数据时, 所得到的结果依然很好, 与人工测量值比较, 均方误差也均小于  $0.0005s^2$ , 相对误差小于  $0.2\%$ . 因篇幅有限图省略.

4) 该网络也曾试用于处理英文阅读眼动数据, 其结果的网络输出值和人工测量值也非常符合. 图省略.

## 4 讨论

1) 从仿真结果来看, 用神经网络确定阅读过程中眼动注视停顿的点数、位置、和注视时程, 是可行的有效途径. 无论是学习样本或是试验样本, 神经网络输出注视点总点数及

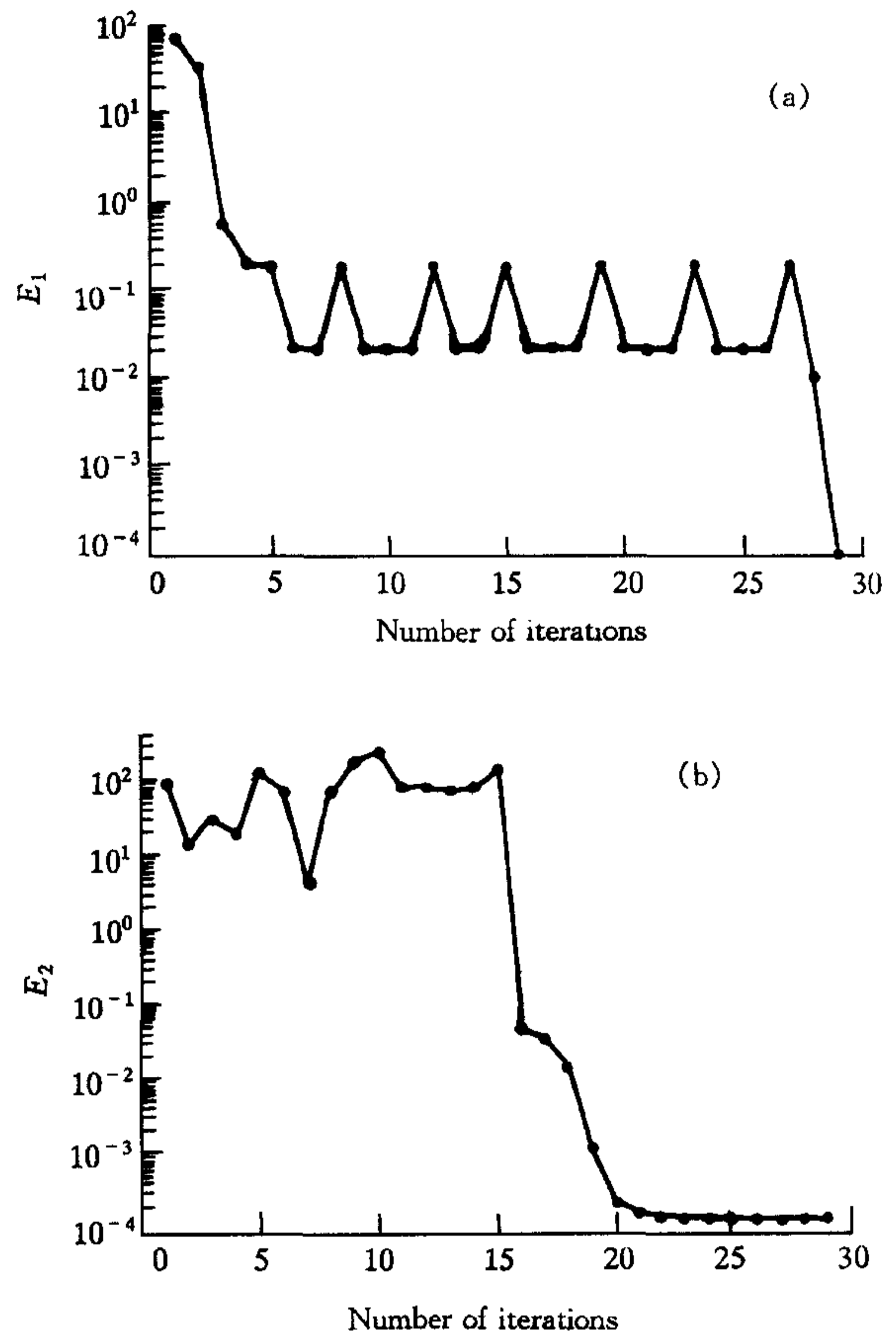


图 4 训练过程中误差函数的变化

(a) 第一级网络误差函数  $E_1$  随迭代次数而减少.

(b) 第二级网络误差函数  $E_2$  随迭代次数而减少.

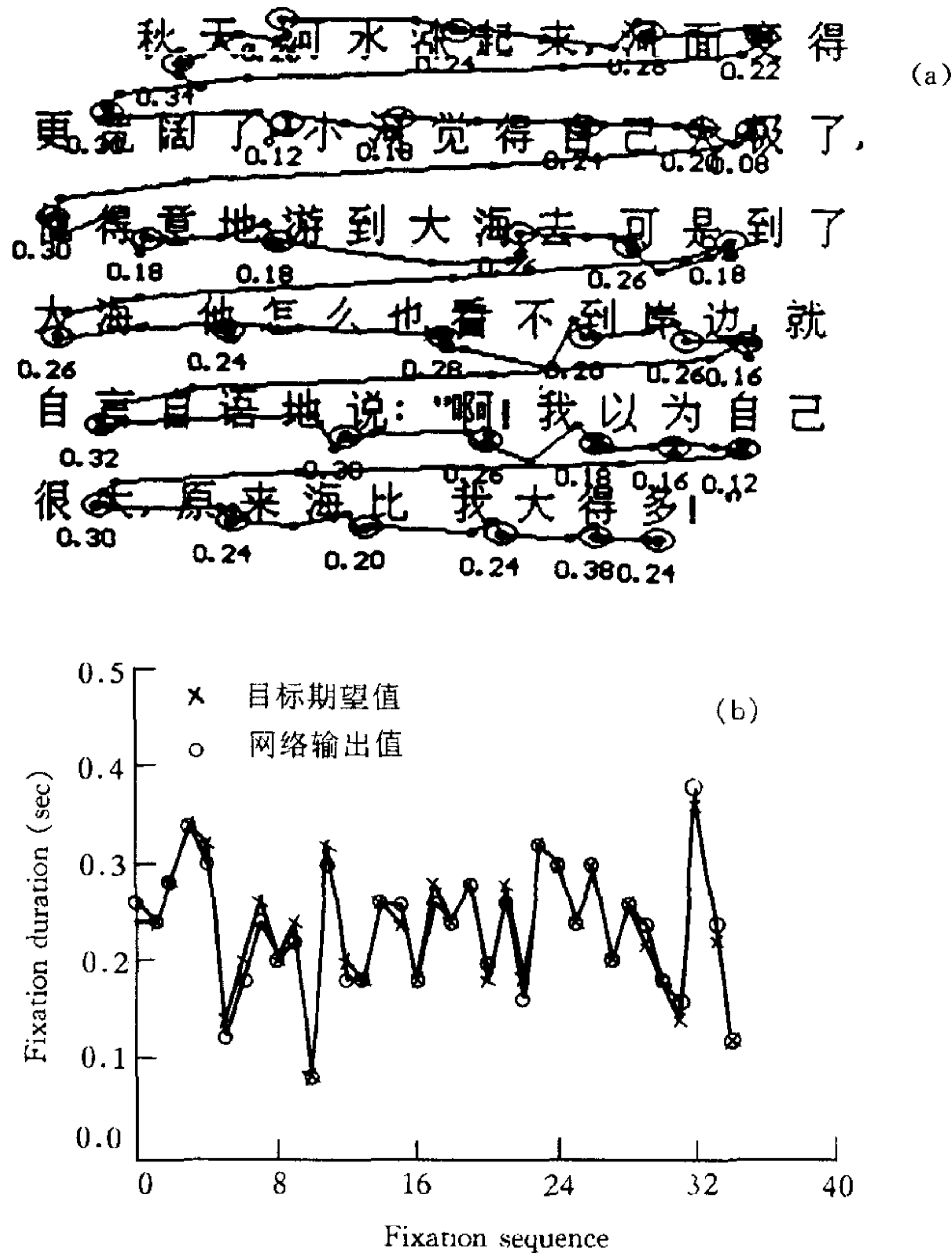


图5 神经网络处理同一受试者阅读其它课文的眼动数据

(a)经网络辨识后的结果输出例图;(b)注视时程的网络输出值和期望值的对比曲线。

时程和实测期望值非常一致,相对误差均小于0.2%,因此神经网络方法处理的正确性较高,速度快,可满足阅读研究工作需要。

2) 某个受试者的一个学习样本在训练成功后所得网络结果参数,可作为网络固定参数处理不同受试者阅读不同材料的眼动数据,因此该网络普遍适用性较强。

3) 本工作在IBM 486上模拟时,速度快,耗时短,有助于大规模阅读眼动研究的数据处理。

4) 本神经网络不仅可应用在阅读眼动研究中,也可应用到其他有关的眼动研究工作中,如图像辨认眼动和临床眼动检查等。另外,对一些类似的图像辨认问题,如血细胞自动计数及其截面积测量等,也可有实际应用价值。

5) 由于眼动的随机性和噪声存在,当二个不同时间的注视点的空间位置完全重合在一起时,(在 regression 眼动中可能发生)本方法中神经网络判断就会发生兼并性的误差,虽然由于概率不大而影响不严重,但从原理上讲应考虑到时间维度上的区分,这有待于今后进一步的工作来解决。

**致谢** 本文在选用部分中文阅读眼动数据作为输入样本时曾得到赵信珍同志的帮助,在此表示感谢。



## 参 考 文 献

- [1] Sun F, Morita M, Stark L. Comparative patterns of reading eye movement in Chinese and English. *Perception & Psychophysics*, 1985, **37**:502—506.
- [2] 孙复川. 眼球运动控制与视觉信息处理. 复旦神经生物学讲座, 1989, **8**:78—86.
- [3] O'Regan J K. Eye movements and their role in visual and cognitive processes. New York, E Kowler, Elsevier; 1990, 395—443.
- [4] 孙复川, Stark L. 视觉信息处理: 阅读中文和英文时眼动模式的对比研究. 生物物理学报, 1988, **4** (1):1—6.
- [5] Sun F. Eye movements in reading Chinese: paragraphs, single characters and pinyin. *Studies in Visual Information Processing*, 1993, **3**:245—255.
- [6] R Groner *et al.* Eye Movements and Psychological Functions: International Views. Lawrence Erlbaum Associates; Hillsdale 1983, 19—29.
- [7] O'Regan J Kand A Levy-Schoen. Eye movements from physiology & cognition. ESP, Amsterdam, 1987, 658—660.
- [8] Hertz J *et al.* Introduction to the theory of neural computation. Redwood city: Addison-wesley. c1991.
- [9] Carpenter G A, Grossberg S. Pattern recognition by self-organizing neural networks. Cambridge, Massachusetts, London: MIT Press, 1991. 243—257.
- [10] 张立明. 人工神经网络的模型及其应用. 上海: 复旦大学出版社, 1993.
- [11] Allen P K. Robotic object recognition using vision and touch. Boston: Kluwer Academic Pub., 1990.

## A CASCADE NEURAL NETWORK FOR IDENTIFYING FIXATION PARAMETERS OF READING EYE-MOVEMENTS

CHEN LINGYU SUN FUCHUAN

( *Shanghai Institute of Physiology, Chinese Academy of Sciences,*      *Shanghai 200031* )  
( *Lab. of Visual Information Processing, Institute of Biophysics* )

LAWRENCE STARK

( *Telerobotics and Neurology Units, University of California, Berkeley* )

**Abstract** A cascade neural network, which is constructed with a three-layer perceptron and a one-layer linear network, is provided for identifying fixation pauses and fixation durations in eye movements. Simulation results show that computation of fixation parameters of eye-movement data with this neural network is accurate and fast.

**Key words** Neural network, pattern identification, reading eye movement, fixation pause, fixation duration.

**陈凌育** 1944年生. 1965年毕业于中国科技大学无线电电子学系, 现为中国科学院上海生理所高级工程师, 从事专业为生物电子学, 生物信号及神经网络.

**孙复川** 1937年生. 北京大学无线电物理系毕业, 现为中国科学院上海生理所生物控制论及神经网络实验室研究员. 博士生导师, 电子学会会士, 自动化学会生物控制论与医学工程委员会委员, 加州大学伯克利分校 Telerobotics and Neurology wrists 和悉尼大学计算机系访问教授.