Hierarchical timescales in the neocortex: Mathematical mechanism and biological insights

Songting Li\textsuperscript{a,b,c,1} and Xiao-Jing Wang\textsuperscript{a,1}

\textsuperscript{a}School of Mathematical Sciences, Shanghai Jiao Tong University, Shanghai 200240, China; \textsuperscript{b}Institute of Natural Sciences, Shanghai Jiao Tong University, Shanghai 200240, China; \textsuperscript{c}Ministry of Education Key Laboratory of Scientific and Engineering Computing, Shanghai Jiao Tong University, Shanghai 200240, China; and \textsuperscript{d}Center for Neural Science, New York University, New York, NY 10003

Edited by Terrence Sejnowski, Computational Neurobiology Laboratory, Salk Institute for Biological Studies, La Jolla, CA; received June 3, 2021; accepted December 7, 2021

A cardinal feature of the neocortex is the progressive increase of the spatial receptive fields along the cortical hierarchy. Recently, theoretical and experimental findings have shown that the temporal response windows also gradually enlarge, so that early sensory neural circuits operate on short timescales whereas higher-association areas are capable of integrating information over a long period of time. While an increased receptive field is accounted for by spatial summation of inputs from neurons in an upstream area, the emergence of timescale hierarchy cannot be readily explained, especially given the dense interareal cortical connectivity known in the modern connectome. To uncover the required neurobiological properties, we carried out a rigorous analysis of an anatomically based large-scale cortex model of macaque monkeys. Using a perturbation method, we show that the segregation of disparate timescales is defined in terms of the localization of eigenvectors of the connectivity matrix, which depends on three circuit properties: 1) a macroscopic gradient of synaptic excitation, 2) distinct electrophysiological properties between excitatory and inhibitory neuronal populations, and 3) a detailed balance between long-range excitatory inputs and local inhibitory inputs for each area-to-area pathway. Our work thus provides a quantitative understanding of the mechanism underlying the emergence of timescale hierarchy in large-scale primate cortical networks.

---

Significance

In the neocortex, while early sensory areas encode and process external inputs rapidly, higher-association areas are endowed with slow dynamics suitable for accumulating information over time. Such a hierarchy of temporal response windows along the cortical hierarchy naturally emerges in a model of the macaque primate cortex. This finding raises the question of why diverse temporal modes are not mixed in roughly the same way across the whole cortex, despite high connection density and an abundance of feedback loops. We investigate this question by mathematically analyzing the anatomically based network model of macaque cortex and theoretically show that three sufficient conditions of synaptic excitation and inhibition give rise to timescale segregation in a hierarchy, a functionally important characteristic of the cortex.

Author contributions: S.L. and X.-J.W. designed research; S.L. performed research; and S.L. and X.-J.W. wrote the paper. The authors declare no competing interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

\textsuperscript{1}To whom correspondence may be addressed. Email: songting@sjtu.edu.cn or xjwang@nyu.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2110274119/-/DCSupplemental.

Published February 2, 2022.
contains a subnet of 29 areas widely distributed from sensory to association areas in the macaque cortex, and each area includes both excitatory and inhibitory neuronal populations. The neuronal population dynamics in the ith area are described as

\[
\tau_E \frac{d}{dt} \nu_E^i = -\nu_E^i + \beta_E \left[ I_{syn,E}^i + I_{ext,E}^i \right],
\]

\[
\tau_I \frac{d}{dt} \nu_I^i = -\nu_I^i + \beta_I \left[ I_{syn,I}^i + I_{ext,I}^i \right],
\]

where \( \nu_E^i \) and \( \nu_I^i \) are the firing rate of the excitatory and inhibitory populations in the ith area, respectively; \( \tau_E \) and \( \tau_I \) are their time constants, respectively; and \( \beta_E \) and \( \beta_I \) are the slope of the frequency–current (f-I) curve for the excitatory and inhibitory populations, respectively. The f-I curve takes the form of a rectified linear function with \( [I]_+ = \max(I, 0) \). In addition, \( I_{ext,E}^i \) and \( I_{ext,I}^i \) are the external currents, and \( I_{syn,E}^i \) and \( I_{syn,I}^i \) are the synaptic currents that follow

\[
I_{syn,E}^i = (1 + \eta h_i) \left( w_{EE} \nu_E^i + \mu_{EE} \sum_{j=1}^{N} FLN_{ij} \nu_E^j \right) - w_{EI} \nu_I^i,
\]

\[
I_{syn,I}^i = (1 + \eta h_i) \left( w_{IE} \nu_E^i + \mu_{IE} \sum_{j=1}^{N} FLN_{ij} \nu_E^j \right) - w_{II} \nu_I^i,
\]

where \( w_{pq} \), \( p, q \in \{ E, I \} \) is the local coupling strength from the \( q \) population to the \( p \) population within each area. \( FLN_{ij} \) is the fraction of labeled neurons (FLN) from area \( j \) to area \( i \) reflecting the strengths of long-range input (3), and \( \mu_{EE} \) and \( \mu_{IE} \) are scaling parameters that control the strengths of long-range input to the excitatory and inhibitory populations, respectively. Both local and long-range excitatory inputs to an area are scaled by its position in the hierarchy quantified by \( h_i \) (a value normalized between 0 and 1), based on the observation that the hierarchical position of an area highly correlates with the number of spines on pyramidal neurons in that area (2, 4). A constant \( \eta \) maps the hierarchy \( h_i \) into excitatory connection strengths. Note that both local and long-range projections are scaled by hierarchy, rather than just local projections, following the observation that the proportion of local to long-range connections is approximately conserved across areas (25). The values of all the model parameters are specified in Materials and Methods.

By simulating the model, it has been observed in ref. 2 that the decay time of neuronal response in each area increases progressively along the visual cortical hierarchy when a pulse input is given to area V1, as shown here in Fig. 1A. Early visual areas show fast and transient responses while prefrontal areas show slower responses and longer integration times with traces lasting for several seconds after the stimulation. In addition, white-noise input to V1 is also integrated with a hierarchy of timescales by computing the autocorrelation of neuronal activity at each area (2). As shown in Fig. 1B, the activity of early sensory areas shows rapid decay of autocorrelation with time lag while that of association areas shows slow decay. In Fig. 1C, by fitting single or double exponentials to the decay of the autocorrelation curves (2), the dominant timescale of each area tends to increase along the hierarchy approximately, and thus a hierarchy of widely disparate timescales emerges from this model. It is worth noting, however, that the timescale does not change monotonically with the anatomically defined hierarchy (\( x \) axis); the precise pattern is sculpted by the measured interareal wiring properties.

Note that, although the multiareal model (Eqs. 1 and 2) is nonlinear by taking into account a rectified linear f-I curve, the stimuli in our simulations drive all neuronal population activities above the firing threshold with positive input currents to all areas. Therefore, the stimuli essentially drive the network dynamics into the linear regime. Before we perform mathematical analysis to understand the mechanism underlying the emergence of hierarchical timescales in the simulations, to simplify the notation, we rewrite the network dynamics Eqs. 1 and 2 in the linear regime in the form

\[
\frac{d}{dt} \nu = W \nu + I_{ext},
\]

where

\[
\nu = [\nu_E^1, \ldots, \nu_E^n, \nu_I^1, \ldots, \nu_I^n]^T, \quad W = \begin{bmatrix} D_{EE} + F_{EE} & D_{EI} & D_{IE} \\ D_{IE} & D_{II} \end{bmatrix},
\]

with \( n = 29 \), and \( D_{EE}, D_{EI}, D_{IE}, D_{II} \) being four diagonal matrices whose ith element on their diagonal line is

\[
d_{EE} = \frac{\beta_E}{\tau_E} \left( 1 + \eta h_i \right) w_{EE} - \frac{1}{\beta_E}, \quad d_{EI} = -\frac{\beta_E}{\tau_E} \left[ w_{EI} \right],
\]

\[
d_{IE} = \frac{\beta_I}{\tau_I} \left( 1 + \eta h_i \right) w_{IE}, \quad d_{II} = -\frac{\beta_I}{\tau_I} \left[ w_{II} + 1 \right],
\]

respectively, and matrices \( F_{EE} \) and \( F_{IE} \) being two nondiagonal matrices whose ith-row-jth-column element is

\[
f_{EE}^{ij} = \frac{\beta_E}{\tau_E} \left( 1 + \eta h_i \right) \mu_{EE} FLN_{ij}, \quad f_{IE}^{ij} = \frac{\beta_I}{\tau_I} \left( 1 + \eta h_i \right) \mu_{IE} FLN_{ij},
\]

respectively. Note that matrices \( D_{EE}, D_{EI}, D_{IE}, D_{II} \) reflect local intraareal interactions, while matrices \( F_{EE} \) and \( F_{IE} \) reflect long-range interareal interactions. In addition, elements in \( D_{EE}, D_{IE}, F_{EE}, \) and \( F_{IE} \) depend on area hierarchy \( h_i \), while elements in \( D_{EI} \) and \( D_{II} \) are constant. Finally, the external input vector is

\[
I_{ext} = \begin{bmatrix} \frac{\beta_E}{\tau_E} I_{ext,E}^1, \ldots, \frac{\beta_E}{\tau_E} I_{ext,E}^n, \frac{\beta_I}{\tau_I} I_{ext,I}^1, \ldots, \frac{\beta_I}{\tau_I} I_{ext,I}^n \end{bmatrix}^T.
\]

Denoting the eigenvalues and eigenvectors of the connectivity matrix \( W \) as \( \lambda_i \) and \( \nu_i \) (\( i = 1, 2, \ldots, 2n \)), respectively, i.e., \( W \nu_i = \lambda_i \nu_i \), the analytical solution of Eq. 3 can be obtained as

\[
\nu'(t) = \sum_{j=1}^{2n} \left( \tilde{a}_j e^{\lambda_j t} + \int_0^t e^{\lambda_j (t-t')} \tilde{f}_j(t') dt' \right) \nu_i^j,
\]
where \( \nu^i \) and \( V_j \) are the \( i \)th element in \( \nu \) and \( V \), respectively, and \( \bar{a}_j \) and \( I_1 \) are the coefficients for the initial condition and the external input, respectively, represented in the coordinate system of the eigenvectors \( \{ V_j \} \). Note that, from Eq. 4, each area integrates input current with the same set of time constants \( \{ \tau_i \} \) determined by the real part of the eigenvalues, i.e., \( \tau_i = -1/Re{\lambda} \). Therefore, the characteristic timescale of each area across the network is expected to be similar in the general case. To obtain distinct timescales at each area, it requires 1) the localization of eigenvectors \( V_j \), i.e., most of the elements in \( V_j \) are close to zero, and 2) the orthogonality of all pairs of eigenvectors, i.e., the nonzero elements nearly nonoverlap for different \( V_j \).

By computing the eigenvalues and eigenvectors of matrix \( W \), as shown in Fig. 2A, the timescale pool \( \{ \tau_i \} \) derived from eigenvalues can be classified into two groups; one group shows a quite fast timescale of about 2 ms, and the other group includes relatively slow timescales ranging from tens to hundreds of milliseconds. In addition, we are particularly interested in the excitatory population because the majority of neurons in the cortex are excitatory neurons. We observe that the magnitude of the eigenvectors corresponding to the fast timescale is nearly zero for the excitatory population at each area, while that corresponding to the slow timescale is weakly localized and weakly orthogonal, i.e., each eigenvector has a few nonzero elements that almost do not overlap with other eigenvectors’ nonzero elements. According to Eq. 4, the pattern of eigenvectors gives rise to the disparate timescales for the excitatory neuronal population at each cortical area. We next perform mathematical analysis to investigate the sufficient conditions for 1) vanishing magnitude of the excitatory component of fast-eigenmode eigenvectors and 2) weak localization and orthogonality of the excitatory component of slow-eigenmode eigenvectors in this network system.

**Perturbation Analysis of the Model**

We note that the parameters of the model (specified in Materials and Methods) give

\[
\epsilon = \frac{\beta_E}{\tau_E} \approx 0.094, \quad \delta = \frac{\mu_{EE}}{\mu_{IE}} - \frac{\omega_{EI}}{\omega_{II} + 1/\tau_I} \approx 0.038,
\]

which can be viewed as two small parameters to allow us to perform perturbation analysis below.

We first study the network in the absence of the long-range interactions among areas. In this scenario, we study the \( 2 \times 2 \) block matrix \( D = \begin{bmatrix} D_{EE} & D_{EI} \\ D_{IE} & D_{II} \end{bmatrix} \) in which each block is a diagonal matrix defined above. By viewing \( \epsilon = \frac{\beta_E}{\tau_E} (\frac{\mu_{EE}}{\mu_{IE}}) \) as a small parameter, we have

\[
D_{II}, D_{II} \sim O(1); \quad D_{EI}, D_{EE} \sim O(\epsilon) \tag{5}
\]

from their definitions. Accordingly, we can prove the following proposition:

**Proposition 1.** If \( D_{II} \sim O(1), \ D_{II} \sim O(1), \ D_{EI} \sim O(\epsilon), \ D_{EE} \sim O(\epsilon) \), then \( D \) has \( n \) eigenvalues being \( O(\epsilon) \) and \( n \) eigenvalues being \( O(1) \).

**Proof.** It is straightforward to prove that matrix \( D \) can be diagonalized by matrix \( P \); i.e.,

\[
A = P^{-1}DP = \begin{bmatrix} A_U & O \\ O & A_L \end{bmatrix} = \begin{bmatrix} D_{EE} + AD_{IE} & O \\ O & -D_{IE}A + D_{II} \end{bmatrix},
\]

where

\[
P = \begin{bmatrix} I + AB & -A \\ -B & I \end{bmatrix},
\]

\( I \) is the identity matrix, and diagonal matrix \( A \) satisfies \( -D_{EE} + AD_{IE} + D_{II} = 0 \) and diagonal matrix \( B \) satisfies \( D_{II} + B(D_{EE} + AD_{IE}) + (D_{IE} - D_{II})B = 0 \).

We solve the equation of \( A \) and choose one of the two solutions of \( A \) as

\[
A = \frac{1}{2} D_{II}^{-1} [D_{II} - D_{EE} + \sqrt{(D_{II} - D_{EE})^2 + 4D_{IE}D_{II}D_{EE}}],
\]

where the square root of a diagonal matrix is defined as taking the square root of its elements. Due to the fact that \( D_{II} \sim O(1), \ D_{II} \sim O(1), \ D_{II} \sim O(\epsilon), \ D_{EE} \sim O(\epsilon) \), we have \( A = -D_{II}D_{II}^{-1} + O(\epsilon) \sim O(\epsilon) \), and \( B = D_{IE}D_{II}^{-1} + O(\epsilon) \sim O(1) \). Accordingly, we have
\[
\Lambda_U = D_{EE} - D_{EI} D_{EI}^{-1} D_{IE} + O(\epsilon^2) \sim O(\epsilon),
\]

\[
\Lambda_L = D_{II} + D_{EI} D_{EI}^{-1} D_{IE} + O(\epsilon^2) \sim O(1).
\]

From Proposition 1, the eigenvalues of \(D\) have two separated scales belonging to \(\Lambda_U\) and \(\Lambda_L\), respectively. As the timescales of the network system are given by \(\tau_i = -1/Re(\lambda_i)\) (\(\lambda_i\) is the \(i\)th diagonal element in matrix \(\Lambda\)), the separation of scales for eigenvalues in \(\Lambda_U\) and \(\Lambda_L\) explains that the intrinsic timescale pool can be classified into two groups with a separation of scales, which is mainly determined by the distinct electrophysiological properties between excitatory and inhibitory neuronal populations within each area described by \(\epsilon\). In addition, from the analysis, the eigenvalues in \(\Lambda_U\) with large magnitude (fast timescale) are less sensitive to the hierarchy level because \(\Lambda_U \approx D_{EE} - D_{EI} D_{EE}^{-1} D_{EI}\), and the elements in \(D_{II}\) do not depend on \(h_i\). Therefore, the gradient of \(h_i\) across areas barely affects the fast-timescale pool. In contrast, the eigenvalues in \(\Lambda_L\) with small magnitude (slow timescale) are more sensitive to the hierarchy level because \(\Lambda_L \approx D_{EE} - D_{EI} D_{EI}^{-1} D_{IE}\), and both the elements in \(D_{EE}\) and \(D_{IE}\) depend on \(h_i\). Therefore, the gradient of \(h_i\) across areas increases the range of the slow-timescale pool. Further, the slow timescales of each area in this disconnected network are segregated and follow the hierarchical order \(h_i\) as the corresponding eigenvectors are perfectly localized and orthogonal to each other.

Now consider the multiaireal network in the presence of long-range interactions. Adding long-range connectivity to local connectivity matrix \(D\) changes the eigenvalues and eigenvectors of matrix \(D\), which can be analyzed in the following.

By multiplying \(P\) and \(P^{-1}\) (given in the Proof of Proposition 1) on both sides of \(W\), we have

\[
\Gamma = P^{-1} WP = P^{-1} \left( \begin{bmatrix} D_{EE} & D_{EI} \\ D_{IE} & D_{II} \end{bmatrix} \right) + \left( \begin{bmatrix} F_{EE} \\ F_{IE} \end{bmatrix} \right) = \Lambda + \Sigma,
\]

where

\[
\Sigma = P^{-1} \begin{bmatrix} F_{EE} \\ F_{IE} \end{bmatrix} = \sum_{UL} + \sum_{UR},
\]

with \(\sum_{UL} = (F_{EE} + AF_{IE})(I + AB), \sum_{UR} = -(F_{EE} + AF_{IE})A, \sum_{LL} = (B_{EE} + (I + BA)F_{IE})(I + AB), \sum_{LR} = -(B_{EE} + (I + BA)F_{IE})A\), and with \(\Lambda, A,\) and \(B\) defined in the Proof of Proposition 1.

Denoting one of the eigenvectors of matrix \(\Gamma\) as \([u, v]^T\) and the corresponding eigenvalue as \(\lambda\), we have

\[
\begin{bmatrix} \Lambda_U & 0 \\ \sum_{UL} & \sum_{UR} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \lambda \begin{bmatrix} u \\ v \end{bmatrix}.
\]

According to the definition of \(\epsilon = \frac{\partial e_k}{\partial u} \frac{\partial u}{\partial v}\) and \(\delta = \frac{\partial \epsilon}{\partial u} = \frac{\partial \epsilon}{\partial v}\), \(\Lambda \sim O(\epsilon), B \sim O(1)\) and \(\sum_{UL} \sim O(\epsilon), F_{EE} \sim O(\epsilon), F_{IE} \sim O(1)\), we have \(F_{EE} \sim O(\delta^2), I + AB \sim O(1), B_{EE} + (I + BA)F_{IE} \sim O(1)\), and accordingly,

\[
\sum_{UL} \sim O(\epsilon), \quad \sum_{UR} \sim O(\epsilon^2 \delta), \quad \sum_{LL} \sim O(\epsilon), \quad \sum_{LR} \sim O(\epsilon).
\]

As \(\sum_{UL} \sim O(\epsilon^2 \delta)\) is a higher-order term compared with \(\Lambda_U, \Lambda_L, \sum_{UL}, \sum_{LL}, \) and \(\sum_{LR}\), it can be dropped out in Eq. 6 and the error of eigenvalue and eigenvector is at most \(O(\epsilon^2)\) (see SI Appendix, Proposition S1 for a detailed proof). Consequently, we can obtain two equations from Eq. 6 in the vector form,

\[
(\Lambda_U + \sum_{UL}) u = \lambda u,
\]

\[
\sum_{LL} u + (\Lambda_L + \sum_{LR}) v = \lambda v.
\]

To describe the eigenvector property of Eqs. 7 and 8, we first introduce the definitions of weak localization and weak orthogonality as follows:

**Definition 1.** A vector \(u(\delta)\) is weakly localized if it can be represented as \(u = a e_k + \delta b + O(\delta^2)\) for some \(k\), where \(a \sim O(1)\) is a constant number, \(b \sim O(1)\) is a constant vector, \(\delta\) is a small parameter, and \(e_k\) represents the natural basis with only the \(k\)th element being 1 and others being zero, i.e., \(e_k = [0, \ldots, 0, 1, 0, \ldots, 0]\).

**Definition 2.** Two vectors \(u(\delta)\) and \(v(\delta)\) are weakly orthogonal to each other if their inner product \(\langle u, v \rangle \sim O(\delta)\), where \(\delta\) is a small parameter.

With the concept of weak localization and weak orthogonality defined above, we introduce the following proposition that describes the property of \(u\) in the system of Eqs. 7 and 8.

**Proposition 2.** In the system described by Eqs. 7 and 8, if all matrices are analytic with respect to \(\epsilon\) and \(\delta\), and if \(\sum_{UL} \sim O(\epsilon^2), \sum_{LR} \sim O(1), \sum_{LL} \sim O(\epsilon), \Lambda_L \sim O(1), \Lambda_U \sim O(\epsilon), \Lambda_L \sim O(1);\) and if \(\Lambda_U\) has \(n\) simple eigenvalues; then

1) there exist \(n\) eigenvectors \([u, v]^T\) in which \(u = 0\), with \(\lambda \sim O(1)\) correspondingly, and

2) there exist \(n\) eigenvectors \([u, v]^T\) in which \(u\) is weakly localized and weakly orthogonal to each other, with \(\lambda \sim O(\epsilon)\) correspondingly.

**Proof.** 1) It is noted that \(u = 0\) is a trivial solution of Eq. 7. By defining \(\sum_{LR} = \sum_{UL}/\epsilon \sim O(1)\), Eq. 8 becomes

\[
(\Lambda_L + \epsilon \sum_{LR}) u = \lambda v,
\]

in which \(v\) is the eigenvector of matrix \(\Lambda_L + \epsilon \sum_{LR}\). By viewing \(\epsilon \sum_{LR}\) as a perturbation matrix to \(\Lambda_L\), then the leading order of \(\lambda\) shall be the same as that of \(n\) elements in the diagonal line of \(\Lambda_L\), which takes the order of \(O(1)\).

2) In Eq. 7, if \(u \neq 0\), by defining \(\tilde{\Lambda}_U = \Lambda_U/\epsilon, \sum_{UL} = \sum_{UL}/\epsilon^2\), and \(\tilde{\lambda} = \lambda/\epsilon\), we have

\[
(\tilde{\Lambda}_U + \delta \sum_{UL}) u = \lambda u.
\]

Therefore, \(u\) is also the eigenvector of matrix \(\tilde{\Lambda}_U + \delta \sum_{UL}\), and \(\tilde{\lambda}\) is the corresponding eigenvalue. As \(\Lambda_U\) has simple eigenvalues, so does \(\tilde{\Lambda}_U\), and then \(\tilde{\lambda}\) and \(\lambda\) are analytic with respect to the perturbation parameter \(\delta\) (26), i.e., \(\tau = \sum_{k=1}^n \delta^k u_k, \lambda = \sum_{k=1}^n \delta^k \tilde{\lambda}\), for \(\delta\) near zero. Therefore, to the leading order, we have

\[
\tilde{\Lambda}_U u_0 = \mu_0 u_0,
\]

in which \(\mu_0\) is the eigenvalue of the diagonal matrix \(\tilde{\Lambda}_U\), and \(u_0\) is the corresponding eigenvector. Accordingly, \(u_0 \in \{e_k\}\), and thereafter

\[
u = \epsilon k + \delta u_1 + O(\delta^2), \quad k = 1, 2, \ldots, n,
\]

where \(e_k\) represents the \(k\)th natural basis, and the leading order of \(\lambda = \epsilon \tau k\) is \(\epsilon \mu_0 \sim O(\epsilon)\). It follows straightforward to verify that \(u\) are weakly localized and weakly orthogonal to each other.

If we denote the unit-length eigenvector of the connectivity matrix \(W\) as \(r_{EE}, \tau E^T\), and denote the corresponding eigenvalue as \(\lambda\) (the same as that of matrix \(\Gamma\) after the similarity transform), then from Propositions 1 and 2 we have the following:

**Proposition 3.** Under the same conditions in Propositions 1 and 2, the unit-length eigenvector \(r_{EE}, \tau E^T\) of the connectivity matrix \(W\) has the following properties:

1) For eigenvalue \(\lambda \sim O(\epsilon)\), the corresponding \(r_{EE}\) is weakly localized and weakly orthogonal to each other, and

2) for eigenvalue \(\lambda \sim O(1)\), the corresponding \(r_{EE} \sim O(\epsilon)\).
Proof. We first consider $[r_E, r_i]^T$ with nonunit length. According to the similarity transform, we have the following linear relation between $[r_E, r_i]^T$ and $[u, v]^T$:

$$
[r_E, r_i] = P [u, v] = \begin{bmatrix} I + AB & -A \\ -B & I \end{bmatrix} [u, v],
$$

i.e., $r_E = (I + AB)u - Av$, and $r_i = -Bu + v$. From Proposition 1, we have $A \sim O(\epsilon)$, and $B \sim O(1)$.

1) From Proposition 2, we have $u = e_k + \delta u_i + O(\delta^2)$ for $\lambda \sim O(\epsilon)$ ($k = 1, 2, \ldots, n$). Accordingly, $v$ can be solved as $v = (\Lambda - \Lambda_{UL} - \Sigma_{LR})^{-1}\Sigma_{LU}u \sim O(1)$, which gives $r_E \sim O(1)$, and $r_i \sim O(1)$. Therefore, the length of $[r_E, r_i]^T$ denoted by $c$ is order $O(1)$. By normalizing the length of $[r_E, r_i]^T$ to be unity, we have $r_E = c^{-1}(I + AB)e_k + O(\epsilon) + O(\delta)$ being weakly localized and weakly orthogonal to each other.

2) From Proposition 2, we have $u = 0$ for $\lambda \sim O(1)$. Accordingly, $v$ can be solved as the eigenvector of matrix $(\Lambda_k + \Sigma_{UL})$ with unit length. Therefore, $[r_E, r_i]^T = [-Av, v]^T$, and the length of $[r_E, r_i]^T$ denoted by $c$ is order $O(1)$. By normalizing the length of $[r_E, r_i]^T$ to be unity, we have $r_E = -c^{-1}Av \sim O(\epsilon)$.

Note that Propositions 2 and 3 hold for sufficiently small $\epsilon$ and $\delta$ near zero. However, the convergence radius of the power series of $u$ in Proposition 2 is not specified yet. Although difficult to calculate the convergence radius, we can compute the analytical expression of $u_i$ in the power series $u = \sum_{i=0}^{\infty} \delta^i u_i$ in Proposition 2 to obtain the first-order perturbation solution of $u$ and thereby $r_E$, which could help us gain insight about when weak localization and orthogonality of $u$ and $r_E$ will break down approximately.

To the order of $\delta$ in Eq. 9, we have

$$
\bar{\Lambda}_U u_i + \bar{\Sigma}_{UL} u_0 = \mu_0 u_1 + \mu_2 u_0.
$$

Without loss of generality, we assume $u_0 = e_k$, and accordingly, $\mu_0 = \lambda_k$ is the $k$th element in the diagonal line of matrix $\bar{\Lambda}_U$. In addition, we normalize $u$ to make the $k$th element in $u$ denoted by $u^k$ to be unity, and correspondingly $u^i = 0$ for $i \geq 1$. By left multiplying $e_j^T$ ($j \neq k$) to Eq. 10, we have

$$
u_j^i = \frac{\Sigma_{UL}^k}{\lambda_k - \lambda_j}.
$$

where $\Sigma_{UL}^k$ is the element in the $j$th row and $k$th column in matrix $\Sigma_{UL}$. To make the first-order perturbation solution valid, we expect that $\nu_j^i$ is small compared with $\delta^{-1}$; otherwise the separation of orders will no longer hold in the power series (i.e., first-order term $\delta u_1$ becomes larger than the zeroth-order term $e_k$). In such a case, the spectral gap $\lambda_k - \lambda_1$ shall be large enough compared with elements in $\Sigma_{UL}$. The spectral gap of matrix $\bar{\Lambda}_U$ attributes to the gradient of excitation across areas or simply $h_i$. In Fig. 2 B and C, we show that the eigenvector $[r_E, r_i]^T$, which is solved using the perturbation theory in Propositions 2 and 3 to the first-order accuracy (Eq. 11), agrees well with the original eigenvector in most cases. However, some eigenvectors show less similarity to the original eigenvectors when the first-order perturbation theory breaks down for the reason discussed above.

Biological Interpretations of the Three Requirements

From the above analysis, three conditions are required to obtain weakly localized and orthogonal eigenvectors to maintain the hierarchy of timescales: 1) small $\epsilon$, 2) small $\delta$, and 3) the gradient of $h_i$ across areas. We briefly summarize the important roles of the three conditions in proving eigenvector localization and orthogonality in the perturbation analysis illustrated in Fig. 3. As shown in Fig. 3 A and B, to remove intraareal interactions between the excitatory and inhibitory populations within each area, we first change the coordinate system from $(r_E, r_i)$ to $(u, v)$ with a transform matrix $P$ given in Proposition 1. In the new coordinate system, there is no local interaction between the dynamical variables $u$ and $v$. Furthermore, considering a directed long-range projection from area $j$ to area $i$ in Fig. 3C, it has been shown that small $\delta$ leads to weak interaction from $u^j$ to $u^i$, and small $\epsilon$ additionally leads to even weaker interaction from $v^j$ to $v^i$ that can be removed with an ignorable error (SI Appendix, Proposition S1). And a gradient of $h_i$ leads to a nonzero spectral gap between area $i$ and area $j$. All three conditions result in the weak localization and orthogonality of the $u$ component in the $(u, v)$ coordinate system, as proved by Proposition 2 and Eq. 11. Finally, as shown in Fig. 3D, small $\epsilon$ gives rise to the fact that the leading orders of $u$ and $r_E$ are the same, and thus $r_E$ is also weakly localized and orthogonalized similar to $u$, as proved by Proposition 3.

We next discuss the biological interpretation of the three conditions. First, according to the definition of $\epsilon = \frac{2\epsilon}{\pi^2} / h_i$, small $\epsilon$ indicates that the electrophysiological properties of excitatory and inhibitory neurons are different, in particular, their
membrane time constant and the slope of the gain function. The substantial difference of electrophysiological properties between the excitatory and inhibitory neurons has been supported by experimental evidence, i.e., inhibitory neurons have larger slope of the gain function and smaller membrane time constant (27–30).

Second, small $\delta = \frac{\Delta r_{E}}{r_{E}} = \frac{\Delta r_{I}}{r_{I}}$ indicates the balanced condition between the interareal excitatory and intraareal inhibitory inputs. When the presynaptic excitatory input from the jth area is increased by $\Delta r_{Ej}$, its influence on the excitatory population activity in the ith area in the steady state can be calculated in a straightforward way as $\Delta r_{Ej} = C^j \Delta r_{Ej}$, in which $C^j \sim O(\delta)$. This indicates that the signal from the jth area has a small influence on the activity of the excitatory population in the ith area, because the global long-range excitatory input is balanced with and canceled by the local inhibitory synaptic input, leading to small net inputs in each signal pathway, as shown in Fig. 4. This condition corresponds to a detailed balance of excitation and inhibition that may benefit signal control and gating, as proposed in previous studies (31). The importance of excitation–inhibition balance on timescale hierarchy is supported by a recent study showing that the imbalance of excitation and inhibition could have a substantial effect on the change of intrinsic timescales across brain areas, which is a manifestation of psychosis such as hallucination and delusion (32).

Third, the gradient of $h$ parameterizes the gradient of synaptic excitation across areas in the model, supported by the fact that $h_i$ is proportional to the spine count per pyramidal neuron across areas (2, 4) in the form of a macroscopic gradient (33). The gradient of synaptic excitation leads to two consequences: 1) It gives rise to the hierarchy of intrinsic timescale for each area while being disconnected to other parts of the cortex and 2) it stabilizes the localization of intrinsic timescale for each area in the presence of long-range connections. From the perturbation analysis and Eq. 11, the degree of eigenvector localization is determined by the competition between the strength of long-range connections encoded in matrix $\Sigma_{UL}$ and the spectral gap of matrix $\Lambda_U$. Therefore, the long-range connections tend to delocalize eigenvectors and thus break the timescale hierarchy, but the heterogeneity of local recurrent excitation level weakens its effect on eigenvector delocalization in a divisive fashion. In fact, the heterogeneity or randomness in local node properties has been shown to give localized eigenvectors in models of a physical system, for instance, a phenomenon known as Anderson localization (34) that describes the transition from a conducting medium (corresponding to delocalized eigenvectors) to an insulating medium (corresponding to localized eigenvectors). A similar mechanism has been identified in studying the eigenvector localization of an idealized neural network with simple nodes in each cortical area (35).

**Discussion**

In this work, we investigated the requirements for the emergence of a hierarchy of temporal response windows in a multiareal model of the macaque cortex (2). The original model is a nonlinear dynamical system by including a rectified linear f-I curve, and it becomes essentially linear when neural population activities are all above the firing threshold, as happened in our simulations of a hierarchical timescale phenomenon. This fact enabled us to define the time constants precisely from the eigenvectors of the connectivity matrix and carry out a detailed mathematical analysis to identify biologically interpretable conditions. (Rectified) linear models have been broadly used in theoretical and experimental neuroscience studies (36–38). Although microscopic neural activity is nonlinear in general, it has been shown in a recent study (39) that linear models can capture macroscopic cortical dynamics in the resting state more accurately than nonlinear model families, including neural field models for describing the spatiotemporal average of individual neuronal activities. Nonlinear models are more general for capturing neural circuits. However, for a nonlinear model, the time constants of the system are not uniquely defined. A linear model can be understood as a linearization of a nonlinear dynamical system around an internal state such as the resting state of the brain.

In contrast to previous computational models studying the emergence of timescales (35), the model we studied is anatomically more realistic as it incorporates 1) experimental measurements of directed and weighted anatomical connectivity, 2) a gradient of synaptic excitation reflected by spine counts in pyramidal neurons across areas, and 3) both excitatory and inhibitory neuronal populations. By performing rigorous perturbation analysis, we show that the segregation of timescales is attributable to the localization of eigenvectors of the connectivity matrix, and the parameter regime that makes this happen has three crucial properties: 1) a macroscopic gradient of synaptic excitation, 2) distinct electrophysiological properties between excitatory and inhibitory neuronal populations, and 3) a detailed balance between long-range excitatory inputs and local inhibitory inputs for each area-to-area pathway.

The theoretically identified biological conditions for the segregation of timescales enable us to make experimentally testable predictions. First, the condition of the macroscopic gradient of synaptic excitation suggests that a shallower gradient of synaptic excitation shall lead to less localized eigenvectors. Consequently, the difference of time constants for a pair of areas shall be larger if their synaptic excitation or hierarchical levels are less similar, which can be directly tested in experiments. Second, the condition of distinct electrophysiological properties between excitatory and inhibitory neuronal populations suggests that the change of neuronal physiology will affect the segregation of timescales. This condition can be tested experimentally by using genetic tools to knock down or knock out specific genes to change the firing properties of neurons (40). Third, the condition of detailed balance of excitation and inhibition suggests that areas with unbalanced excitation and inhibition could also alter hierarchical time constants. With growing evidence for excitation/inhibition imbalance in schizophrenia (41, 42), this condition is supported by recent experiments showing that the intrinsic time constants of schizophrenia patients have been substantially changed (32). And this condition can be further tested in animal models with genetic tools to disturb the excitation–inhibition balance (43).
It is worth mentioning that, although the specific pattern of interareal connectivity does not affect the eigenvector localization substantially based on the perturbation analysis, it shapes the timescale hierarchy qualitatively. In particular, the timescale hierarchy does not exactly follow monotonically the areal anatomical hierarchy in the presence of long-range connections, as shown in Fig. 1C. Furthermore, within a brain region time constants are heterogeneous across individual neurons (44, 45). To better relate the model with experimentally observed timescales in various specific cortical areas, the roles of long-range connections, cell types, and other circuit properties require further elucidation.

It has been noted that the neuronal activity propagates along the hierarchy with significant attenuation in the model (2). The attenuation can be alleviated by tuning the model parameters to the regime of strong global balanced amplification (GBA) (46) (parameters in Materials and Methods). Balanced amplification was originally introduced for a local neural network, associated with strong nonnormality of the system where amplification was originally introduced for a local neural network (GBA) (46) (parameters in Materials and Methods). In the strong GBA regime, the enhancement of signal propagation in the model correlates with the increase of the nonorthogonality of the eigenvectors or the nonnormality of the connectivity matrix. In the strong GBA regime, δ ≈ 0.38, which is 10 times larger than its original value, suggesting that the detailed balance condition is less well satisfied. In such a case, the localization of timescales may no longer exist in this linear model. The situation is different in nonlinear models (2, 49, 50), where inputs may be amplified by strongly recurrent circuit dynamics to enhance signal propagation or routing of information is selectively gated (for a subset of connection pathways in a goal-directed manner) (31, 51), while the conditions for a timescale hierarchy are satisfied. For a nonlinear system, however, eigenmodes can be defined only with respect to a particular network state. Consequently, the time constants observed in single neurons are no longer unique and may differ, for instance, when the brain is at rest or during a cognitive process. It remains to be seen to what extent the connectivity pattern identified here in the brain’s various internal states, while the precise pattern of timescales can be flexibly varied to meet behavioral demands.

Materials and Methods

Model Parameters. In the macaque cortical network model, we set τ = 20 ms, n = 10 ms, 1/β = 0.066 Hz/pA, κ = 0.351 Hz/pA, μ = 24.4 pA/Hz, μ/τ = 12.2 pA/Hz, τE = 19.7 pA/Hz, τI = 12.5 pA/Hz, μE = 33.7 pA/Hz, μI = 25.5 pA/Hz, and δ = 0.68. We set wE = 25.2 pA/Hz and μE = 51.5 pA/Hz for the strong balanced amplification regime (46) introduced in Discussion. Some of the parameters are derived from experimental measurements of primary visual cortex (37). The FLN values are obtained from the experimental measurements of macaque cortical connectivity (3). The hierarchy values h are set to a cortical area are obtained by fitting a generalized linear model that assigns hierarchical values to areas (2) such that the differences in hierarchical values predict the supragranular layer neurons (SLNs) measured in experiment (52).

Data Availability. The python code for model simulation and perturbation analysis is publicly available on Github (https://github.com/sonting858/mechanism-of-hierarchical-time-constants) (53). Previously published data were used for this work (4).