

RESEARCH ARTICLE

# Mixed Signal Learning by Spike Correlation Propagation in Feedback Inhibitory Circuits

Naoki Hiratani<sup>1,2\*</sup>, Tomoki Fukai<sup>2\*</sup>

**1** Department of Complexity Science and Engineering, The University of Tokyo, Kashiwa, Chiba, Japan, **2** Laboratory for Neural Circuit Theory, RIKEN Brain Science Institute, Wako, Saitama, Japan

\* [N.Hiratani@gmail.com](mailto:N.Hiratani@gmail.com) (NH); [tfukai@riken.jp](mailto:tfukai@riken.jp) (TF)



**OPEN ACCESS**

**Citation:** Hiratani N, Fukai T (2015) Mixed Signal Learning by Spike Correlation Propagation in Feedback Inhibitory Circuits. *PLoS Comput Biol* 11(4): e1004227. doi:10.1371/journal.pcbi.1004227

**Editor:** Abigail Morrison, Research Center Jülich, GERMANY

**Received:** December 2, 2014

**Accepted:** March 6, 2015

**Published:** April 24, 2015

**Copyright:** © 2015 Hiratani, Fukai. This is an open access article distributed under the terms of the [Creative Commons Attribution License](http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This work was partly supported by Grants-in-Aid for Scientific Research on Innovative Areas (no. 22115013, <http://www.jsps.go.jp/english/e-grants/index.html>), and JST CREST (<http://www.jst.go.jp/kisoken/crest/en/>). NH was supported by JSPS Research Fellowship for Young Scientists (<https://www.jsps.go.jp/english/e-pdf/index.html>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

The brain can learn and detect mixed input signals masked by various types of noise, and spike-timing-dependent plasticity (STDP) is the candidate synaptic level mechanism. Because sensory inputs typically have spike correlation, and local circuits have dense feedback connections, input spikes cause the propagation of spike correlation in lateral circuits; however, it is largely unknown how this secondary correlation generated by lateral circuits influences learning processes through STDP, or whether it is beneficial to achieve efficient spike-based learning from uncertain stimuli. To explore the answers to these questions, we construct models of feedforward networks with lateral inhibitory circuits and study how propagated correlation influences STDP learning, and what kind of learning algorithm such circuits achieve. We derive analytical conditions at which neurons detect minor signals with STDP, and show that depending on the origin of the noise, different correlation timescales are useful for learning. In particular, we show that non-precise spike correlation is beneficial for learning in the presence of cross-talk noise. We also show that by considering excitatory and inhibitory STDP at lateral connections, the circuit can acquire a lateral structure optimal for signal detection. In addition, we demonstrate that the model performs blind source separation in a manner similar to the sequential sampling approximation of the Bayesian independent component analysis algorithm. Our results provide a basic understanding of STDP learning in feedback circuits by integrating analyses from both dynamical systems and information theory.

## Author Summary

In natural environments, although sensory inputs are often highly mixed with one another and obscured by noise, animals can detect and learn discrete signals from this mixture. For example, humans easily detect the mention of their names from across a noisy room, a phenomenon known as the cocktail party effect. Spike-timing-dependent plasticity (STDP) is a learning mechanism ubiquitously observed in the brain across various species and is considered to be the neural basis of such learning; however, it is still unclear how STDP enables efficient learning from uncertain stimuli and whether spike-based learning offers benefits beyond those provided by standard machine learning

**Competing Interests:** The authors have declared that no competing interests exist.

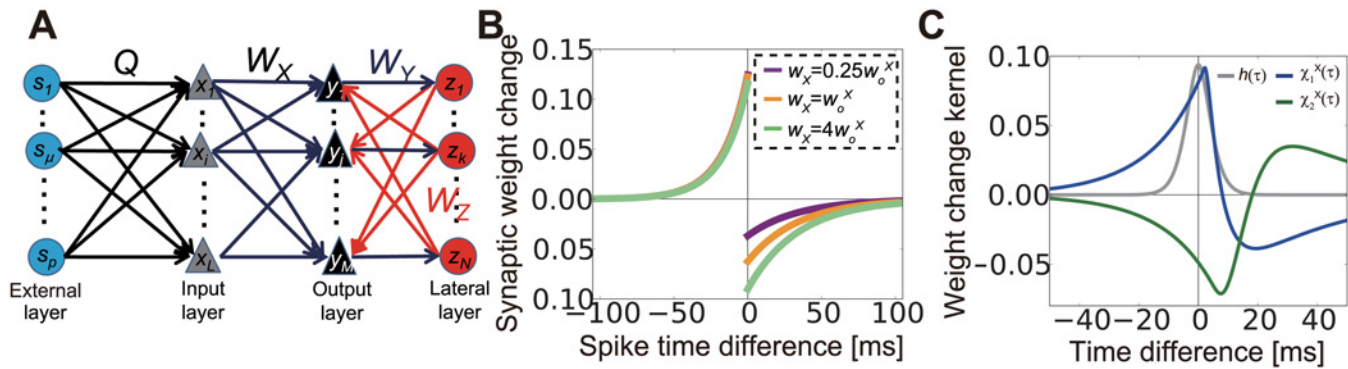
methods for signal decomposition. To begin to answer these questions, we conducted analytical and simulation studies examining the propagation of spike correlation in feedback neural circuits. We show that non-precise spike correlation is useful for handling noise during the learning process. Our results also suggest that neural circuits make use of stochastic membrane dynamics to approximate computationally complex Bayesian learning algorithms, progressing our understanding of the principles of stochastic computation by the brain.

## Introduction

Neurons receive inputs from a large number of other neurons encoding a variety of information about various signals. Despite the diversity and variability of input spike trains, neurons can learn and represent specific information during developmental processes and according to specific task requirements. Spike-timing-dependent plasticity (STDP) [1,2] is a candidate mechanism of neural learning. Extensive studies have revealed the type of information that a single neuron can learn through STDP [3–7]; however, the type of information that a population of neurons interacting with each other learns through STDP has not yet been determined. Understanding this extension from a single neuron to a population of neurons is crucial because a single neuron learns and represents only a limited amount of information that may be transmitted to it from thousands of inputs.

Among neural interactions, lateral inhibition is a basic interaction widely observed in various regions, such as the olfactory bulb [8], visual cortex [9], somatosensory cortex [10], and entorhinal cortex [11]. Previous theoretical results showed that neural circuits with lateral inhibition enhance signal detection [12,13] and improve learning performance [14–16]. Several simulation studies further revealed that neurons acquire receptive field [17–19] or spike patterns [20] through STDP by introducing lateral inhibition; yet, those studies were limited to simplified cases for which a large population of independent neurons was suggested to be sufficient [5,21,22]. Therefore, it remains unclear whether lateral inhibition plays a crucial role in STDP learning; in particular, the spike level effects of lateral inhibition remain elusive. Moreover, recent experimental results suggest that animals learn and discriminate mixed olfactory signals [23–25] or auditory signals masked by noise [26,27], but it is still unknown how feedback interactions contribute to such learning.

Here, by considering a simple feedback network model of spiking neurons, we investigated the algorithm inherent to STDP in neural circuits containing feedback. We analyzed the propagation of spike correlations through inhibitory circuits, and revealed how such secondary correlations influence STDP learning at both feedforward and feedback connections. We discovered that the timescale of spike correlation preferable for learning depends on whether the noise is independent from any signal (random noise) or generated from the mixing of signals (cross-talk noise). We also found that excitatory and inhibitory STDP cooperatively shapes lateral circuit structure, making it suitable for signal detection. We further found a possible link between stochastic membrane dynamics and sampling process, which is necessary for neural approximation of learning algorithm of Bayesian independent component analysis (ICA). We applied our findings by demonstrating that STDP implements a spike-based solution in neural circuits for the cocktail party problem [26,28,29].



**Fig 1. Description of the model.** (A) Schematic figure of the model. (B) Spike-time dependent synaptic weight change in log-spike-timing-dependent plasticity (STDP). (C) Normalized temporal cross-correlogram of input neurons receiving common sources (gray line), and kernel functions of plasticity propagated by feedforward correlation (blue line) and feedback correlation (green line).

doi:10.1371/journal.pcbi.1004227.g001

## Results

### Model

We constructed a network model with three feedforward layers as shown in Fig 1A (see Neural dynamics in Methods for details). The external source layer represents the external environment or neural activity at sensory systems. The external layer also provides common inputs to the input layer and induces correlations in the neurons in the input layer. The input layer shows rate-modulated Poisson firing based on events at the external layer and external noise, which is approximated with the constant firing rate  $\{r_i^o\}$ . Subsequently, spike activity at the input layer projects to the output layer, which also receives inhibitory feedback from the lateral layer. Neurons in the lateral layers are excited by inputs from the output layer. We assumed that all neurons in the input layer and the output layer are excitatory, whereas lateral-layer neurons are assumed to be inhibitory. Although excitatory lateral interactions also exist in the sensory cortex, they are typically sparse [30] and weak [10] compared with inhibitory interactions; thus we concentrated on the latter. For the analytical treatment, the neurons in the output and lateral layers were modeled with a linear Poisson model. We first studied synaptic plasticity at the feedforward connections (connections from the input layer to the output layer), while fixing lateral connections (i.e., connections from the output layer to the lateral layer and connections from the lateral layer to the output layer). For STDP, we used pairwise log-STDP (Fig 1B) [31], which replicates the experimentally observed long-tailed synaptic weight distribution [32,33].

We considered the case for information encoded in the correlated activity of input neurons [34,35], and fixed the average firing rate of all input neurons at the constant value  $v_o^x$  (See Table 1 and 2 for the list of variables and parameters). If the firing rate of input neuron  $i$  is

given as  $r_i^o + \sum_{\mu=1}^p q_{i\mu} \int_0^\infty \phi(t') s_\mu(t-t') dt'$ , for external event  $s_\mu(t)$  and the response probability

of the neuron  $q_{i\mu}$ , then common inputs from the external layer induce a temporal correlation proportional to

$$h(\tau; \theta_i) \equiv \int_{\max(\tau, 0)}^\infty dt' \phi(t') \phi(t' - \tau). \quad (1)$$

where  $\phi(t)$  is a response kernel (see Eqs (14) and (24) in Methods for details). If we use  $\phi(t) = t^2 e^{-t/\theta_t} / 2\theta_t^3$ , where  $\theta_t$  is the parameter that controls the timescale of spike correlations, then  $h(\tau; \theta_t) = \frac{1}{16\theta_t^3} (\tau^2 + 3\theta_t|\tau| + 3\theta_t^2) e^{-|\tau|/\theta_t}$  (gray line in Fig 1C). For the kernel function, we

**Table 1. Definition of variables.**

$s_\mu(t)$	The activity of external source $\mu$	$s_\mu(t) = \sigma[v_\mu^S]$
$x_i(t)$	The spiking activity of input neuron $i$	Eq (14)
$u_j^E(t)$	Membrane potential of output neuron $j$	Eq (15)
$y_j(t)$	The spiking activity of output neuron $j$	$y_j(t) = \sigma[u_j^E(t)]$
$u_k^I(t)$	Membrane potential of inhibitory neuron $k$	Eq (16)
$z_k(t)$	The spiking activity of inhibitory neuron $k$	$z_k(t) = \sigma[u_k^I(t)]$
$w_{ji}^X$	The synaptic weight of a feed-forward excitatory connection from $i$ to $j$	Eq (17)
$q_{i\mu}$	Response probability of input neuron $i$ to external source $\mu$	Eq (14)
$C_{ii}$	Non-normalized correlation between input neuron $i$ and $i$	$C_{ii} = \sum_\mu q_{i\mu} q_{i\mu}$
$C_{ij}(s)$	Cross correlation between input neuron $i$ and $j$	Eq (24)
$G_1^X(w), G_2^X(w)$	Coefficients of correlation-based synaptic weight change	Eq (30)
$\chi_1^X, \chi_2^X$	The correlation kernel functions	Eqs (3) and (4)

doi:10.1371/journal.pcbi.1004227.t001

used the gamma distribution with shape parameter  $k_g = 3$  in order to reproduce broad spike correlations typically observed in cortical neurons [36,37]. Synaptic weight dynamics by STDP is written as

$$\frac{dw_{ji}^X}{dt} = x_i(t - d_{ji}^{Xa}) \int_0^\infty F_d(w_{ji}^X, s) y_j(t - s - d_{ji}^{Xd}) ds + y_j(t - d_{ji}^{Xd}) \int_0^\infty F_p(w_{ji}^X, s) x_i(t - s - d_{ji}^{Xa}) ds$$

for  $F_d(w_{ij}^X, s) = f_d(w_{ij}^X) e^{-s/\tau_d}$ ,  $F_p(w_{ij}^X, s) = f_p(w_{ij}^X) e^{-s/\tau_p}$ , where  $f_d(w)$  and  $f_p(w)$  are synaptic weight dependence of LTD/LTP (long-term depression/potential), respectively. By taking the average of above equation over time and ensemble (see Average synaptic weight velocity in [Methods](#) for details), the weight change of the feedforward connection  $W_X$  can be approximated as

$$\dot{W}_X \approx W_X (g_1^X E - g_2^X W_Z W_Y) C^t, \tag{2}$$

**Table 2. Parameter settings.**

T	Simulation time	3000 s (for Figs 5C, 5D, 5E, 6 and 7: T = 4000 s)
L, M, N	Neural population	400, 20, 20 (for Figs 7 and 8: M = N = 40)
$L_a, M_a, N_a$	Neural subpopulation	100, 10, 10
$\tau_A^X, \tau_B^X, \tau_A^Y, \tau_B^Y, \tau_A^Z, \tau_B^Z$	EPSP/IPSP time constants	5.0, 1.0, 4.0, 0.8, 2.5, 0.5 ms
$w_o^X, w_o^Y, w_o^Z$	Synaptic weights	2.5, 100.0, 50.0 (for Figs 7 and 8: $w_o^Z = 80.0$ )
$d_{\min}^{Xa}, d_{\max}^{Xa}$	Axonal delays	2.0, 4.0 ms
$d_{\min}^{Xd}, d_{\max}^{Xd}$	Dendritic delays	0.5, 1.5 ms
$d_{\min}^Y, d_{\max}^Y, d_{\min}^Z, d_{\max}^Z$	Synaptic (axonal) delays	0.2, 1.2, 0.2, 1.2 ms
$\theta_t$	Correlation timescale	2.0 ms
$v_o^S, v_o^X$	Firing rates	10, 10 Hz
$\eta^X$	Learning rate	0.05 $w_o^X$
$\sigma_{sig}$	Noise amplitude of plasticity	0.3
$\tau_p, \tau_d$	STDP time windows	17, 34 ms
$\alpha, \beta$	Parameters for log-STDP	20.0, 50.0
$\sigma_W^{init}$	Initial variance of synaptic weights	0.1
$\gamma^Y, \gamma^Z$	LTD/LTP balance	1.4, 0.7

doi:10.1371/journal.pcbi.1004227.t002

where  $g_1^X$  and  $g_2^X$  are scalar coefficients,  $C$  is the correlation matrix, and  $E$  is the identity matrix (see Eqs (25)–(30) for derivation). The first term describes the synaptic weight change directly caused by an input spike correlation and can be rewritten into the convolution of the temporal correlation and correlation kernel function  $\chi_1^X$  as

$$g_1^X \equiv G_1^X(w_o^X), G_1^X(w) \equiv \int_{-\infty}^{\infty} \chi_1^X(\tau; w)h(\tau)d\tau, \quad (3)$$

$$\chi_1^X(\tau; w) = \int_{-\tau+2d_{Xd}}^{\infty} dsF(w, s)\varepsilon_X(\tau + s - 2d_{Xd}),$$

where  $F(w, s) = F_d(w, s)$  if  $s < 0$ , else  $F(w, s) = F_p(w, s)$ , and  $\varepsilon_X$  is the EPSP curve of input neurons (see Eqs (15) and (31) in the Methods). By the deconvolution of  $G_1^X(w)$ , we can separate the effect of the intrinsic network property  $\chi_1^X$  and that of the input correlation  $h(\tau)$  for STDP-based learning. Due to causality, LTP/LTD balance, and dendritic delay,  $\chi_1^X(\tau; w)$  typically becomes LTP-dominant around  $\tau = 0$  (blue line in Fig 1C; we set  $w = w_o^X$ ), so that  $g_1^X$  takes positive values, which enables coincidence-based learning [4,5,38]. The second term of Eq (2), which is of particular interest in this model, describes how the input correlation influences STDP learning at feedforward connections through lateral inhibition:

$$g_2^X \equiv G_2^X(w_o^X), G_2^X(w) \equiv \int_{-\infty}^{\infty} \chi_2^X(\tau; w)h(\tau)d\tau, \quad (4)$$

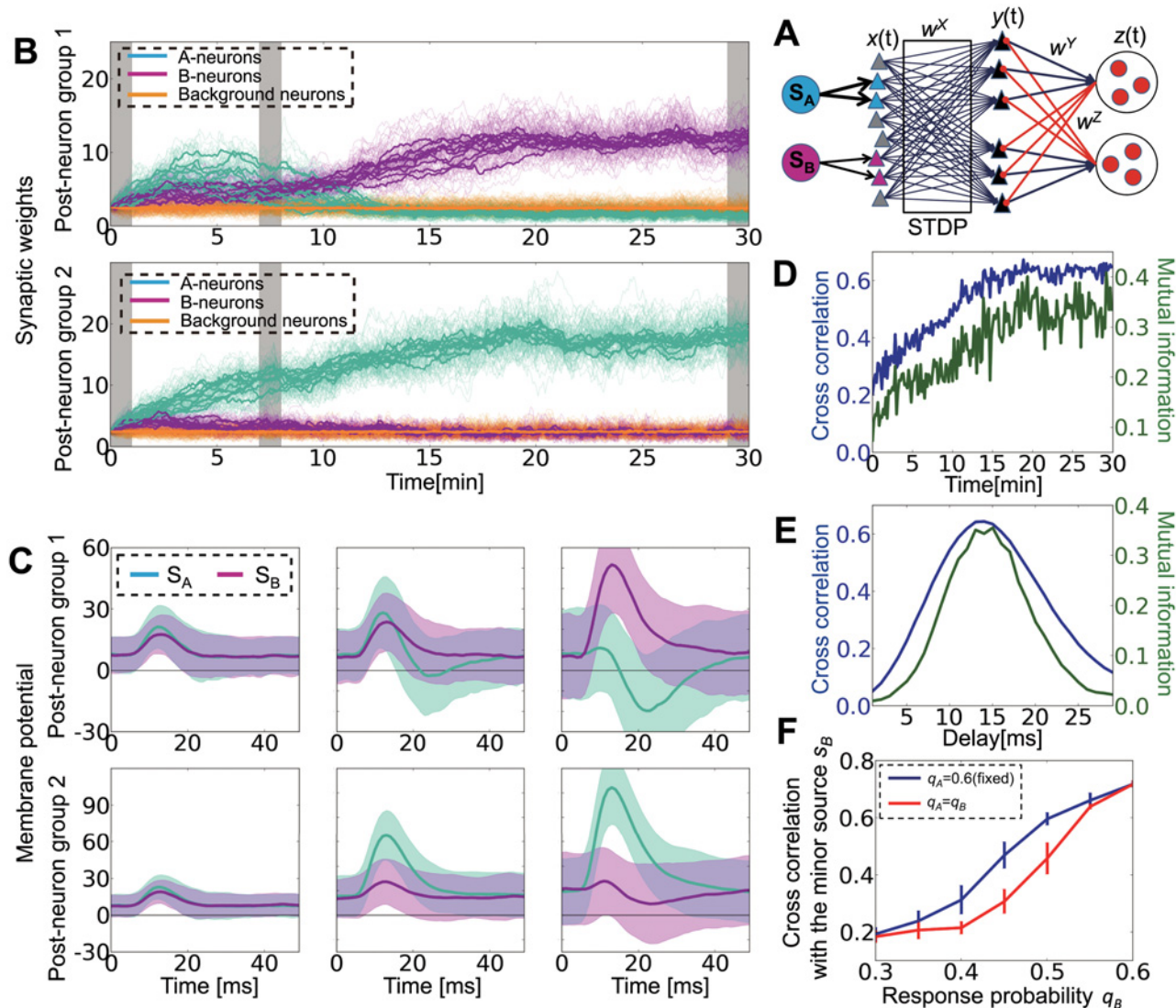
$$\chi_2^X(\tau; w) = \int_{-\tau+D}^{\infty} dsF(w, s) \int_0^{\tau+s-D} dr\varepsilon_Z(r) \int_0^{\tau+s-r-D} dq\varepsilon_Y(q)\varepsilon_X(\tau + s - r - q - D),$$

where  $D = 2d_{Xd} + d_Y + d_Z$ , and  $\varepsilon_Y$  and  $\varepsilon_Z$  are EPSP/IPSP curves of output/inhibitory neurons, respectively. This term primarily causes LTD as the sign flips through lateral inhibition ( $-\chi_2^X(\tau; w)$ ; shown as the green line in Fig 1C). Previous simulation studies showed lateral inhibition has critical effects on excitatory STDP learning [17–19]; however, it has not yet been well studied how a secondary correlation generated through the lateral circuits influences STDP at feedforward connections, and it is still largely unknown how lateral inhibition functions with various stimuli in different neural circuits. For example, the correlation kernel of the feedback term exhibits a delay as the signal propagates through the inhibitory circuit; yet, we do not know how much delay is permitted for effective learning or if realistic synaptic delays satisfy such a condition. Furthermore, it is also unknown what information a circuit can learn if there are several mixed signals with different amplitudes for which symmetry-breaking learning [5,39] is not valid. Therefore, using theoretical analysis and simulation, we first investigated the properties of the inhibitory kernel  $-\chi_2^X(\tau; w)$  in STDP learning.

### Lateral inhibition enhances minor source detection by STDP

In Eq (2), if lateral inhibition is negligible (i.e.,  $g_2^X/g_1^X = 0$ ), all output neurons acquire the principal component of the response probability matrix  $Q$ , and the other information is neglected [7,40,41]. On the other hand, if lateral inhibition is effective, different output neurons may acquire various components of the external structure. We first examined that point in a simple network model with two independent external sources (Fig 2A). In the model, each external source drives an independent subgroup of input neurons (we defined those input neurons as A-neurons and B-neurons), which project excitatory inputs to all of the output neurons. Here, we assume that source A drives input neurons with a higher probability than source B ( $q_A = 0.6, q_B = 0.5$ ), so that input neurons projected by source A show higher correlations





**Fig 2. Lateral inhibition enables minor source detection by spike-timing-dependent plasticity (STDP) through membrane hyperpolarization.** (A) Schematic figure of the simplified model.  $S_A$  and  $S_B$  (on the left side) are the sources that project to subsets of input neurons (colored triangles). Gray triangles are background neurons, black triangles (on the right) are output neurons, and red circles are inhibitory neurons. (B) Development of synaptic weights. Thick lines are mean synaptic weights from A-neurons (blue), B-neurons (red), and Background-neurons (orange) to each output neuron. Thin lines are traces of individual synaptic weights. Gray bar shows the timing at which figure C is calculated. (C) Peristimulus time histograms (PSTHs) of membrane potentials averaged within output neuron groups.  $T = 0$  indicates the timing of events at external layers. The three figures are calculated from the data at  $t = 0-1$  min,  $7-8$  min, and  $29-30$  min. (D) Development of mean cross-correlation and mutual information between external sources and population activity of output neurons for the simulation depicted in panels B and C. (E) Delay dependence of mean cross-correlation and mutual information. Both values were calculated from five simulations. (F) Cross-correlation between the output group that detected the minor source and the minor source activity for various response probabilities  $q_B$  with a fixed  $q_A (= 0.6)$ . When none of output groups detected the minor source, the larger value calculated for the two output groups was used. Throughout the study, error bars represent standard deviation calculated from five simulations, unless otherwise indicated.

doi:10.1371/journal.pcbi.1004227.g002

( $c_A = 0.36$ ) than those receiving the output of source B ( $c_B = 0.25$ ). In the matrix form,

$$Q = \begin{pmatrix} q_A & 0 \\ 0 & q_B \\ 0 & 0 \end{pmatrix}, C = \begin{pmatrix} c_A & 0 & 0 \\ 0 & c_B & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

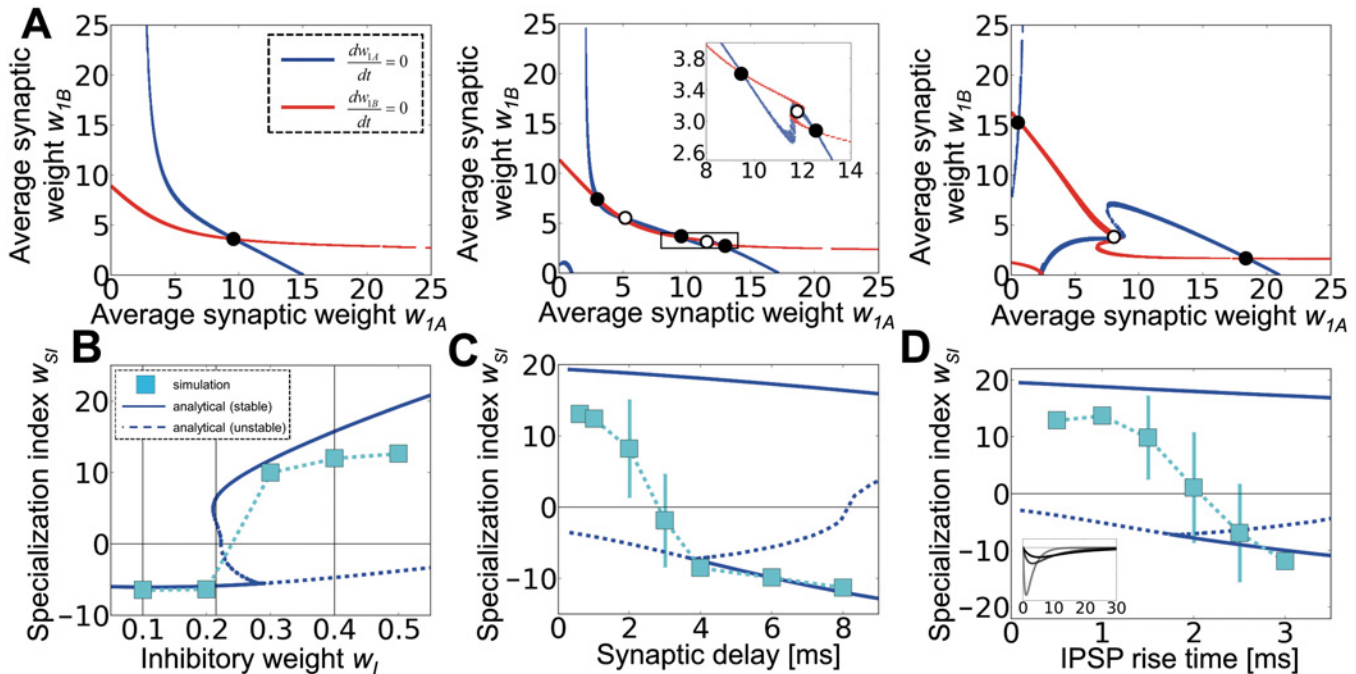
The third row in  $Q$  represents response probabilities of background neurons in the input layer (gray triangles in Fig 2A; note that  $C = QQ^t$ ). We refer to this as the minor source detection task below. Here, for lateral connections, we assumed that both excitatory-to-inhibitory (E-to-I) and inhibitory-to-excitatory (I-to-E) connections are well organized such that inhibition only works mutually between two output neuron groups (Fig 2A; blue lines are E-to-I and red lines are I-to-E connections. See also Eq (30) in Methods). The origin of these structured lateral connections is discussed later. When the network is excited by inputs from external sources, excitatory postsynaptic potential (EPSP) sizes of feedforward connections  $W_x$  change according to STDP rules. Initially, in all output neurons, synaptic weights from A-neurons (blue triangles in Fig 2A) become larger because A-neurons are more strongly correlated with one another than B-neurons are. However, as learning proceeds, one of the output neuron groups becomes selective for the minor source B (Fig 2B). After 30 min, the network successfully learns both sources. If we focus on the peristimulus time histogram (PSTH) for the average membrane potential of output neurons aligned to external events, both neuron groups initially show weak responses to both correlation events, and yet the depolarization is relatively higher for source A than for source B (Fig 2C left). After 10 min of learning, both neuron groups show relatively stronger initial responses for source A, but group 1 shows a hyperpolarization soon after the initial response (Fig 2C middle). As a result, synaptic weights from A-neurons to group 1 become weaker, and group 1 neurons eventually become selective for the minor source B (Fig 2C right). The mean cross-correlation (see cross-correlation in Methods for details) between the external sources and the population activity of output neurons is maximized when the delay is approximately 10–15 ms (Fig 2E). If we fix the delay at 14 ms, then the cross-correlation gradually increases as the network learns both sources (Fig 2D). The same argument holds if mutual information is used for performance evaluation (green lines in Fig 2D and 2E). Interestingly, the network better detects the minor source when it is learned with a highly correlated source compared with when it is learned with another minor source (Fig 2F), because a highly correlated opponent source causes strong lateral inhibition on the output neurons, which enhances minor source learning. Similar results are also obtained for conductance-based leaky integrate-and-fire (LIF) neurons (S1 Fig).

### Lateral inhibition should be strong, fast, and sharp

To investigate how and when the network can acquire multiple sources represented by correlated inputs, we further analyzed the model above (see Mean-field approximation of a two-source model in Methods for details). Because both output excitatory neurons and lateral inhibitory neurons are bundled into groups, in the mean-field approximation, we can approximate  $M$  excitatory populations and  $N$  inhibitory populations into two representative output neurons and two inhibitory neurons. Similarly, input neurons can be bundled into three groups (A-neurons, B-neurons, and Background-neurons). In addition, we assumed that the synaptic connections from Background-neurons to output neurons are fixed because they showed little weight change in the simulation (orange lines in Fig 2B). In this approximation, by inserting Eq (32) into Eq (29), the mean synaptic weight changes of feedforward connections follow

$$\begin{aligned} \frac{dw_{\mu\nu}^x}{dt} \cong & \sum_{\nu'=1}^{L/L_a} L_a w_{\mu\nu'}^x v_o^S G_1^X(w_{\mu\nu}^x) \sum_{\rho} q_{\nu\rho} q_{\nu'\rho} - N_a w_z M_a w_Y \sum_{\nu'=1}^{L/L_a} L_a w_{\mu\nu'}^x v_o^S G_2^X(w_{\mu\nu}^x) \sum_{\rho} q_{\nu\rho} q_{\nu'\rho} \\ & + \bar{F}(w_{\mu\nu}^x) \left[ (v_o^X)^2 \sum_{\nu'=1}^{L/L_a} L_a w_{\mu\nu'}^x - (v_o^X)^2 N_a w_z M_a w_Y \sum_{\nu'=1}^{L/L_a} L_a w_{\mu\nu'}^x + (N_a w_z)^2 M_a w_Y v_o^X v_{\mu}^Z \right], \end{aligned} \tag{5}$$

where  $\mu = 1, 2$  and  $\bar{\mu} = 2, 1 (\mu \neq \bar{\mu})$ , and  $\nu = A, B$ . The first two terms are correlation-based



**Fig 3. Lateral inhibition is strong, fast, and sharp.** (A) Nullclines of the average synaptic weight changes at different inhibitory amplitudes  $w_Z = 0.1, 0.215, 0.4$ . The inset in the middle graph is a magnified view of boxed area. (B) Specialization indices  $w_{Si}$  for various inhibitory weights. Positive  $w_{Si}$  indicates the winner-share-all state, whereas negative  $w_{Si}$  indicates the winner-take-all state. Blue lines are analytical estimations and cyan squares are the results of simulations. Vertical lines correspond to the values at which the nullclines in Figure A are calculated. (C) The same graphs for various synaptic delays. The average synaptic delay of both lateral excitatory  $(d_{min}^Y + d_{max}^Y)/2$  and inhibitory  $(d_{min}^Z + d_{max}^Z)/2$  connections was changed, while the variability was kept at  $d_{max}^Y - d_{min}^Y = d_{max}^Z - d_{min}^Z = 1.0$  ms. (D) IPSP rise time dependence. The inset shows IPSP curves at  $\{\tau_A^Z, \tau_B^Z\} = \{0.5, 2.5\}$  (gray line),  $\{1.5, 7.5\}$  (dark gray line), and  $\{2.5, 12.5\}$  (black line).

doi:10.1371/journal.pcbi.1004227.g003

learning, and the last term is the homeostatic effect intrinsic to STDP [5].  $G_1^X$  and  $G_2^X$  are coefficients determined by synaptic delays, EPSP/IPSP (Inhibitory postsynaptic potential) shapes, and correlation structure, as shown in Eqs (3) and (4). By solving the self-consistency condition (Eq (34) in Methods), the firing rates of inhibitory neurons are approximated as

$$v_1^Z = \frac{M_a w_Y v_o^X}{1 - (M_a w_Y N_a w_Z)^2} [(L_a w_{1A} + L_a w_{1B} + 2L_a w_o^X) - (M_a w_Y N_a w_Z)(L_a w_{2A} + L_a w_{2B} + 2L_a w_o^X)] \quad (6)$$

$$v_2^Z = \frac{M_a w_Y v_o^X}{1 - (M_a w_Y N_a w_Z)^2} [(L_a w_{2A} + L_a w_{2B} + 2L_a w_o^X) - (M_a w_Y N_a w_Z)(L_a w_{1A} + L_a w_{1B} + 2L_a w_o^X)].$$

We estimated the nullclines by calculating the lines that satisfy  $\dot{w}_{1\mu}(w_{1A}, w_{1B}, w_{2A}^*(w_{1A}, w_{1B}), w_{2B}^*(w_{1A}, w_{1B})) = 0$  for  $\mu = A$  or  $B$ . As a result, we found that when the mutual inhibition is weak ( $w_I = 10$ ), the system has only one stable point at which  $w_{1A}$  is larger than  $w_{1B}$  (Fig 3A left). At this point,  $w_{2A}$  is also larger than  $w_{2B}$  ( $w_{2A} = 9.64, w_{2B} = 3.60$ ; not shown in the figure), which means that both output neuron groups are specialized for the major source A (we call this state a winner-take-all state or T-state); however, if the inhibition is moderately strong ( $w_I = 21.5$ ), two new stable fixed points and two unstable fixed points appear in the system (Fig 3A middle). In the stable point on the left, neuron group 1 picks up source B while neuron group 2 picks up source A ( $w_{2A} = 12.52, w_{2B} = 2.87$ ). On the right-hand side, neuron group 1 selects source A while neuron group 2 selects source B (we denote those two states as winners-share-all states or S-states below). At the stable point in the middle, both groups detect source A ( $w_{1A} = w_{2A} = 9.47, w_{1B} = w_{2B} = 3.61$ ). Note that because of the mutual



inhibition, the synaptic weight from A-neuron is smaller when both groups learn A than it is when only group 1 learns A. For strong inhibition ( $w_I = 40.0$ ), the stable point in the middle disappears, and the system is stable only when two neuron groups detect different sources (Fig 3A right). Simulation results confirm this analysis because strong inhibition indeed causes a winner-share-all state in which multiple neuron groups survive in competition [15], whereas the network tends to show a winner-take-all learning when the inhibition is weak (Fig 3B). We measured the degree of winner-share-all/winner-take-all states by defining the specialization index  $w_{SI}$  as

$$w'_{SI} = (w_{1A} - w_{1B})(w_{2B} - w_{2A}), w_{SI} = w'_{SI} / \sqrt{|w'_{SI}|}. \tag{7}$$

If  $w'_{SI} = 0$ , we set  $w_{SI} = 0$ . If two output groups are specialized for different sources,  $w_{SI}$  becomes positive, whereas if two groups are specialized for the same source,  $w_{SI}$  becomes negative. When the synaptic delay in the lateral connections is small, only S-states are stable, whereas at longer delays, both S-states and T-states are stable. In the simulation, the network typically grows toward the latter state in the bistable strategy (Fig 3C). Moreover, if we change the shape of the IPSP curve while keeping  $\tau^Z_B = 5 \tau^Z_A$ , for steep IPSP curves (i.e., both  $\tau^Z_A$  and  $\tau^Z_B$  are small), only the S-states are stable, whereas T-states also become stable for slower IPSPs (Fig 3D). Therefore, both analytical and simulation studies indicate that lateral inhibition should be strong, fast and sharp to detect higher correlation structure. Moreover, lateral inhibition does not need to be pathologically strong because the I/E balance of  $N_a w_z / L w_o^x \cong 20\%$  is sufficient to cause multistability.

### Optimal correlation timescale changes depend on the noise source

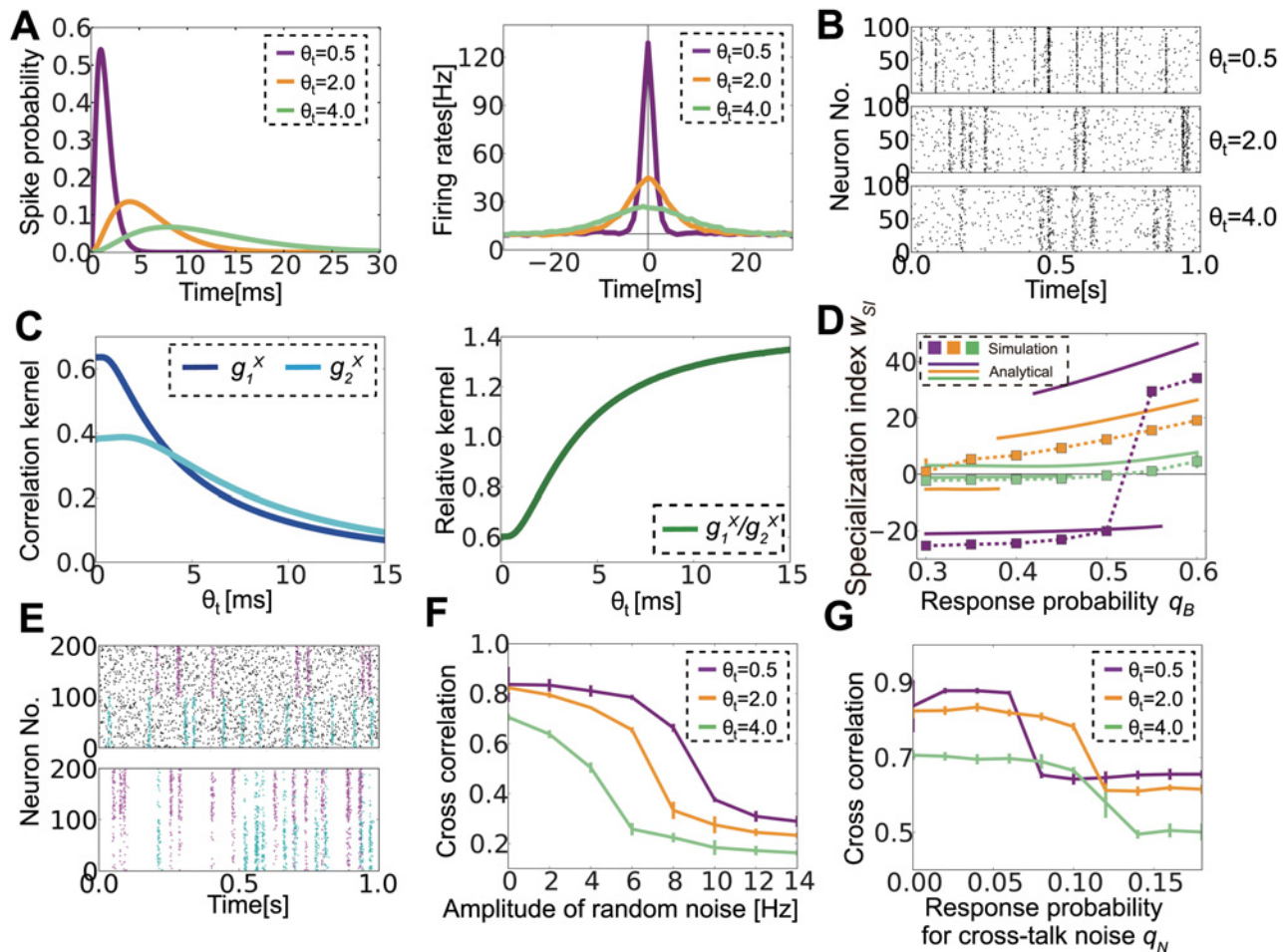
In the previous section, we revealed the effects of network properties for a fixed input correlation structure; however, actual neurons show various timescales for correlations depending on the brain region [37,42] and characteristics of the stimuli [43,44], and it is largely unknown how different timescales influence correlation-driven learning. Therefore, we next considered the effect of correlation timescales, especially on noise tolerance. In our current model, input neurons respond to external sources with input kernel  $\phi(t) = t^2 e^{-t/\theta_t} / 2\theta_t^3$  (Fig 4A left), and so the correlation between input neuron  $i$  and  $l$  is given as

$$C_{il}(s) = v_o^s \sum_{\mu=1}^p q_{i\mu} q_{l\mu} h(s).$$

By changing the parameter  $\theta_t$ , we studied the effect of the correlation timescale on learning. The correlation is precise when  $\theta_t$  is small, whereas it becomes broad at large values of  $\theta_t$  (Fig 4A right, Fig 4B). Because STDP causes homeostatic plasticity that does not depend on a correlation, as shown in the third term of Eq (5), in a more precise approximation, Eq (2) should be written as

$$\dot{W}_X \approx W_X (g_1^X E - g_2^X W_Z W_Y) C^t + \langle \text{homeostatic term} \rangle. \tag{8}$$

We first calculated  $g_1^X$  and  $g_2^X$  at various  $\theta_t$ . Both  $g_1^X$  and  $g_2^X$  become smaller for a larger  $\theta_t$ , but decreases in  $g_2^X$  are slower than those in  $g_1^X$ , and, as a result,  $\kappa = g_2^X / g_1^X$  becomes larger for a longer correlation timescale (Fig 4C). This means that a longer temporal correlation is more suitable for the detection of multi-components. This is indeed confirmed in the simulation (Fig 4D). When  $\theta_t = 0.5$  and the minor component is slightly weaker than the major one ( $c_A = 0.36$ ,  $c_B = 0.25$ ), the minor component is no longer detectable. On the other hand, at  $\theta_t = 2.0$ , the minor component is detectable even if the strength of the induced correlation is less than half



**Fig 4. Optimal correlation timescale changes depending on noise characteristics.** (A) Response kernels of input neurons to external events (left) and cross-correlation among input neurons responding to the same source calculated from simulated data (right) for three different correlation timescale parameters  $\theta_t$ . (B) Raster plots of input neurons for various  $\theta_t$ . Only 100 correlated neurons are plotted although there are 400 input neurons in total. (C) Analytically calculated correlation kernels  $g_1^x, g_2^x$  (left), and their ratio  $g_1^x/g_2^x$ . (D) Specialization index  $w_{SI}$  for various response probabilities  $q_B$  while fixing  $q_A = 0.6$ . Lines represent  $w_R$  at analytically estimated stable points, and dotted squares represent simulation results. (E) Raster plots of two types of noise. The upper panel shows random noise, whereas the lower panel depicts crosstalk noise. In both panels, the first 100 neurons respond primarily to the cyan source, and the next 100 neurons respond to the purple source. For random noise, the noise (black dots) is independent from the signals, whereas the crosstalk noise (purple dots in the lower half, cyan dots in the upper half) is correlated with the signal for the other population. (F, G) The effects of random noise (F) and crosstalk noise (G) at various correlation timescales.

doi:10.1371/journal.pcbi.1004227.g004

( $c_A = 0.36, c_B = 0.16$ ). At  $\theta_t = 4.0$ ,  $g_1^x$  becomes smaller so that even the major signal is not fully detectable.

Similar results hold for crosstalk noise. In the model above, the noise is provided through the spontaneous Poisson firing of input neurons as random noise (Fig 4E top, black dots are spikes caused by random noise). In reality, however, there would be crosstalk noise among input spike trains caused by the interference of external sources. We implemented this crosstalk noise by introducing non-diagonal components in the response probability matrix as

$$Q = \begin{pmatrix} q_s & q_N \\ q_N & q_s \\ 0 & 0 \end{pmatrix},$$

where  $q_S$  is the response probability to the preferred signal and  $q_N$  is that to the non-preferred signal (Fig 4E bottom). We refer to this as the noisy source detection task below. To make a clear comparison, in the simulation of random noise, we kept  $q_N = 0$  and changed the spontaneous firing rate of the input neurons ( $r_i^o$ ) to modify the noise intensity, whereas in simulation of crosstalk noise we removed random noise (i.e.,  $r_i^o = 0$ ) and changed  $q_N$ . For random noise, a smaller  $\theta_t$  enables better learning because a large  $g_I^X$  competes with the homeostatic force (Fig 4F). By contrast, for crosstalk noise, the performance is better at  $\theta_t = 2.0$  than at  $\theta_t = 0.5$  because strong lateral inhibition suppresses crosstalk noise (Fig 4G). Although for small noise regimes, the network performs better at  $\theta_t = 0.5$  than at  $\theta_t = 2.0$ , but the difference is almost negligible. Therefore, to cope with crosstalk noise, the spike correlation needs to be broad, whereas a narrow spike correlation is better for random noise. We note that qualitatively the same arguments as above also hold for the exponential kernel  $\phi_e(t) = e^{-t/\theta_t}/\theta_t$  (S3D and S3E Fig). However, the ratio of two coefficients (i.e.,  $\kappa_e = g_{e2}^X/g_{e1}^X$ ) is typically smaller for this kernel than for the kernel we used throughout this study (S3B and S3C Fig vs. Fig 4D) because lateral inhibition is less effective due to highly peaked spike correlation (S3A Fig).

### Excitatory and inhibitory STDP cooperatively shape structured lateral connections

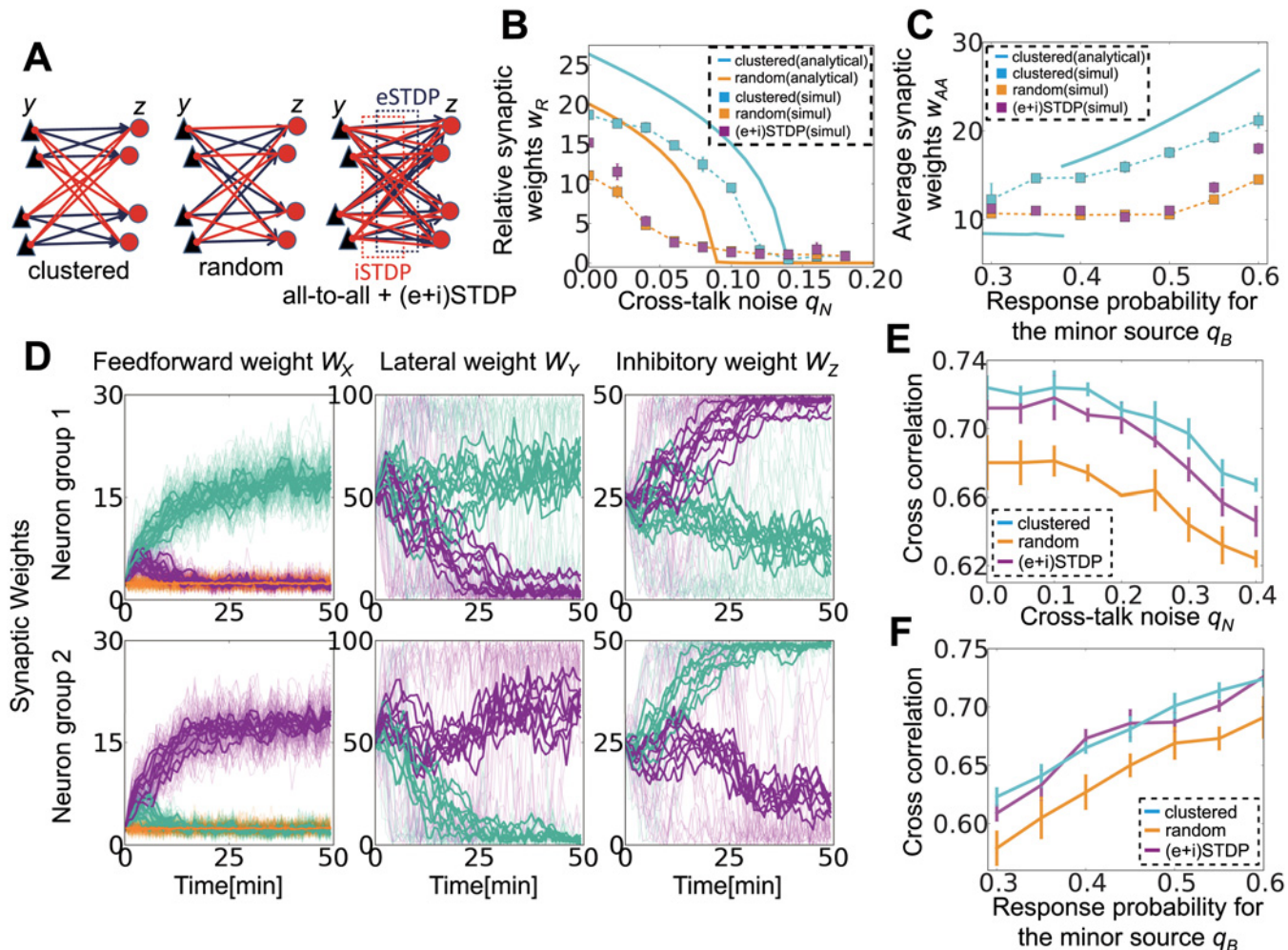
To this point, we have considered a network already clustered into two assemblies that inhibit one another (as in Fig 5A left). This means that the network somehow knows a priori that the number of external sources is two; however, in reality, a randomly connected network should also learn such information. To test this idea, we introduced STDP-type synaptic plasticity in lateral excitatory connections and feedback inhibitory connections and investigated how different STDP rules cause different structures in the circuit.

We first checked whether structured lateral connections were helpful for learning. For comparison, we also considered a model with random lateral connections in which all output neurons and inhibitory neurons are randomly connected with probability 0.5 (Fig 5A middle). When lateral connections are random, mean-field equations are modified as

$$\begin{aligned} \frac{dw_{\mu\nu}^X}{dt} &\cong \sum_{\nu'=1}^{L/L_a} L_a w_{\mu\nu'}^X v_o^S G^X(w_{\mu\nu}^X) \sum_{\rho} q_{\nu\rho} q_{\nu'\rho} - N_a w_Z M_a w_Y \sum_{\mu'=1}^p \sum_{\nu'=1}^{L/L_a} L_a w_{\mu'\nu'}^X v_o^S G^Y(w_{\mu\nu}^X) \sum_{\rho} q_{\nu\rho} q_{\nu'\rho} \\ &+ \bar{F}(w_{\mu\nu}^X) \left[ (v_o^X)^2 \sum_{\nu'=1}^{L/L_a} L_a w_{\mu\nu'}^X - (v_o^X)^2 N_a w_Z M_a w_Y \sum_{\mu'=1}^p \sum_{\nu'=1}^{L/L_a} L_a w_{\mu'\nu'}^X + (N_a w_Z)^2 M_a w_Y v_o^X v_{tot}^Z \right], \quad (9) \\ v_{tot}^Z &\equiv \frac{2M_a w_Y v_o^X (L_a w_{1A} + L_a w_{1B} + 2L_a w_o^X)}{1 + 2M_a w_Y N_a w_Z}. \end{aligned}$$

We separated lateral connections into two groups as in the previous case, but this approximation is legitimate only when two input sources are symmetrical (i.e.,  $q_A = q_B$ ). In other cases, neurons are often organized into two groups with different population sizes. In such cases, for evaluating performance, we measured average weights from source A on the output neurons receiving stronger inputs from A-neurons than from B-neurons or Background-neurons. For randomly connected lateral inhibition, learning performance dropped significantly in noisy source detection (Fig 5B) and in minor source detection (Fig 5C); thus clustered connectivity is indeed advantageous for learning.

We next investigated whether such structure can be learned using STDP rules. We first introduced Hebbian STDP for both E-to-I and I-to-E connections. With these learning rules, the lateral connections successfully learn a mutual inhibition structure (Fig 5D); however, this



**Fig 5. Lateral connection structuring by excitatory and inhibitory spike-timing-dependent plasticity (STDP).** (A) Schematic figures of connections between the output layer and the lateral layer. In the simulation, each layer consists of 20 neurons. (B) The effect of crosstalk noise on different lateral structures. Analytical results are shown as bold lines, and the results from simulations are shown as dotted lines. (C) Minor source detection with different lateral structures, the average synaptic weights from source A to those output neurons that prefer source A were measured instead. (D) Synaptic weight development at three connections. In the left and right columns, panels show synaptic weights of excitatory/inhibitory synapses projected to the neuron group 1 (top) and group 2 (bottom). In the middle graph, panels correspond to excitatory synapses projected from the neuron group 1 (top) and group 2 (bottom). In all panels, thin lines indicate the development of individual synapses, thick lines represent average weights onto output neurons, and colors indicate A-neurons (blue), B-neurons (red), and Background-neurons (orange). (E, F) Performance of the network with different lateral structures in noisy signal detection (E) and minor signal detection (F). Here (and only here), a pre-learned network is used to investigate responses for various inputs.

doi:10.1371/journal.pcbi.1004227.g005

learning is achievable only when the learning of a hidden external structure is possible from the random lateral connections (magenta lines in Fig 5B and 5C; note that orange points are hidden by magenta points because they show similar behaviors in noisy cases), which means either when crosstalk noise is low or two sources have similar amplitudes. Nevertheless, once a structure is obtained in easy settings ( $q_N = 0$  or  $q_A = q_B$ ), that network outperforms the network with random lateral connections in both noisy source detection (Fig 5E) and minor source detection (Fig 5F). In Fig 5E, we evaluated the performance of noisy source detection by first conducting STDP learning at  $q_N = 0$ , and then we terminated STDP and performed simulations at the various noise levels  $q_N$ . Similarly, in the minor source detection task depicted in Fig 5F, we first performed STDP learning with  $q_A = q_B = 0.6$ , and then evaluated the performance for a



smaller  $q_B$ . STDP can also generate similar lateral connection structures when the total number of input sources is larger than two (S2A and S2B Fig). Therefore, STDP at lateral connections helps signal detection by efficiently organizing the connection structure.

We next studied the analytical conditions for learning of the clustered structure (see Analytical approach for STDP in lateral and inhibitory connections in Methods for details). The synaptic weight dynamics of lateral excitatory and inhibitory connections are approximately given as

$$\begin{aligned} \dot{W}_Y &\approx g_1^Y W_Y W_X C^t W_X^t g_1^Y \equiv \int_{-\infty}^{\infty} ds F^Y(s) \int D_r^X \int D_u^Y \int D_r^X h(u+r'-s-r) \\ \dot{W}_Z &\approx g_1^Z W_X C W_X^t W_Y^t g_1^Z \equiv \int_{-\infty}^{\infty} ds F^Z(s) \int D_r^X \int D_u^Y \int D_r^X h(r-s-u-r'-d_z-d_y). \end{aligned} \tag{10}$$

Both equations represent indirect effects of the input correlation propagated into the lateral circuit. From a linear analysis, we can expect that when  $g_1^Y$  is positive, E-to-I connections tend to be feature selective (see Eq (35) in Methods). Each inhibitory neuron receives stronger inputs from one of the output neuron groups and, as a result, shows a higher firing rate for the corresponding external signal. On the other hand, if  $g_1^Z$  is positive, I-to-E connections are organized in reciprocal form, where one of the reciprocal connections is enhanced and the other is suppressed (see Eq (36) in Methods). We can evaluate feature selectivity of inhibitory neurons by

$$\phi^Y = \frac{1}{N} \sum_{k=1}^N \left( \frac{1}{|\Omega_A^Y|} \sum_{j \in \Omega_A^Y} w_{kj}^Y - \frac{1}{|\Omega_B^Y|} \sum_{j \in \Omega_B^Y} w_{kj}^Y \right) / \left( \frac{1}{M} \sum_{j=1}^M w_{kj}^Y \right), \tag{11}$$

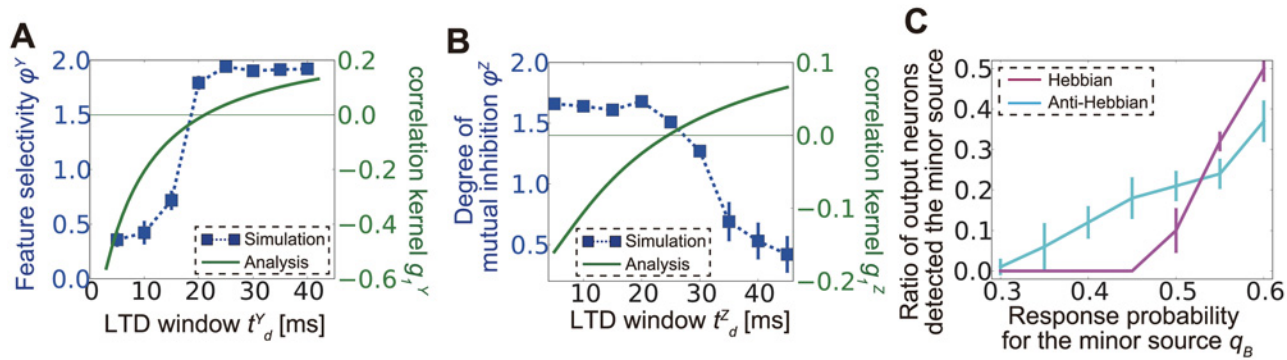
where  $\Omega_A^Y$  and  $\Omega_B^Y$  are the sets of excitatory neurons responding preferentially to sources A and B, respectively. Indeed, when the LTD time window is narrow, analytically calculated  $g_1^Y$  tends to take negative values (the green line in Fig 6A), and E-to-I connections organized in the simulation are not feature selective (the blue points in Fig 6A). By contrast, for a long LTD time window (i.e., when LTD is weakly spike-timing dependent),  $g_1^Y$  tends to take positive values, and E-to-I connections become clustered. In the simulation,  $W_Z$  is also plastic, but as shown in Eq (10), the effect of  $W_Z$  on the plasticity of  $W_Y$  is negligible in first-order approximations.

Similarly, for I-to-E connections, we measure the degree of mutual inhibition (non-reciprocity) with

$$\phi^Z = \frac{1}{N} \sum_{k=1}^N \left| \frac{w_{kj}^Y}{\sum_{j=1}^M w_{kj}^Y} - \frac{w_{jk}^Z}{\sum_{j=1}^M w_{jk}^Z} \right|. \tag{12}$$

When LTD is strongly spike-timing dependent,  $g_1^Z$  is negative and  $\phi^Z$  calculated from the simulation data tends to be large (Fig 6B), which means that inhibitory connections are organized such that the inhibition functions as mutual inhibition between excitatory neuron groups. Note that the organized neuronal wiring patterns are not a pure product of the pre-post causality of STDP but the effect of spike correlations propagating through lateral inhibitory circuits. If the structural plasticity is merely caused by the pre-post causality, both  $\phi^Y$  and  $\phi^Z$  can decrease with increases in the inhibitory population while maintaining the total synaptic weights because the causal effect becomes weaker as each synaptic weight becomes smaller [45]; however, in our simulations, the values of both quantities generally increased for larger inhibitory populations (S2C Fig).





**Fig 6. Correlation propagation shapes lateral connection structure.** (A) Comparison between feature selectivity  $\phi^Y$  (blue dots) calculated from simulation results and analytically calculated correlation kernel function  $g_1^Y$  (green line) for lateral excitatory connections. Thin green horizontal line represents  $g_1^Y = 0$ . (B) Comparison between the degree of mutual inhibition  $\phi^Z$  (blue dots) calculated from the simulation and analytically calculated correlation kernel  $g_1^Z$  (green line) for lateral inhibitory connections. Negative  $g_1^Z$  is correlated with a high degree of mutual inhibition, as expected (see [Methods](#)). (C) Ratio of output neurons tuned for the minor source in a minor source detection task under Hebbian and anti-Hebbian inhibitory spike-timing-dependent plasticity.

doi:10.1371/journal.pcbi.1004227.g006

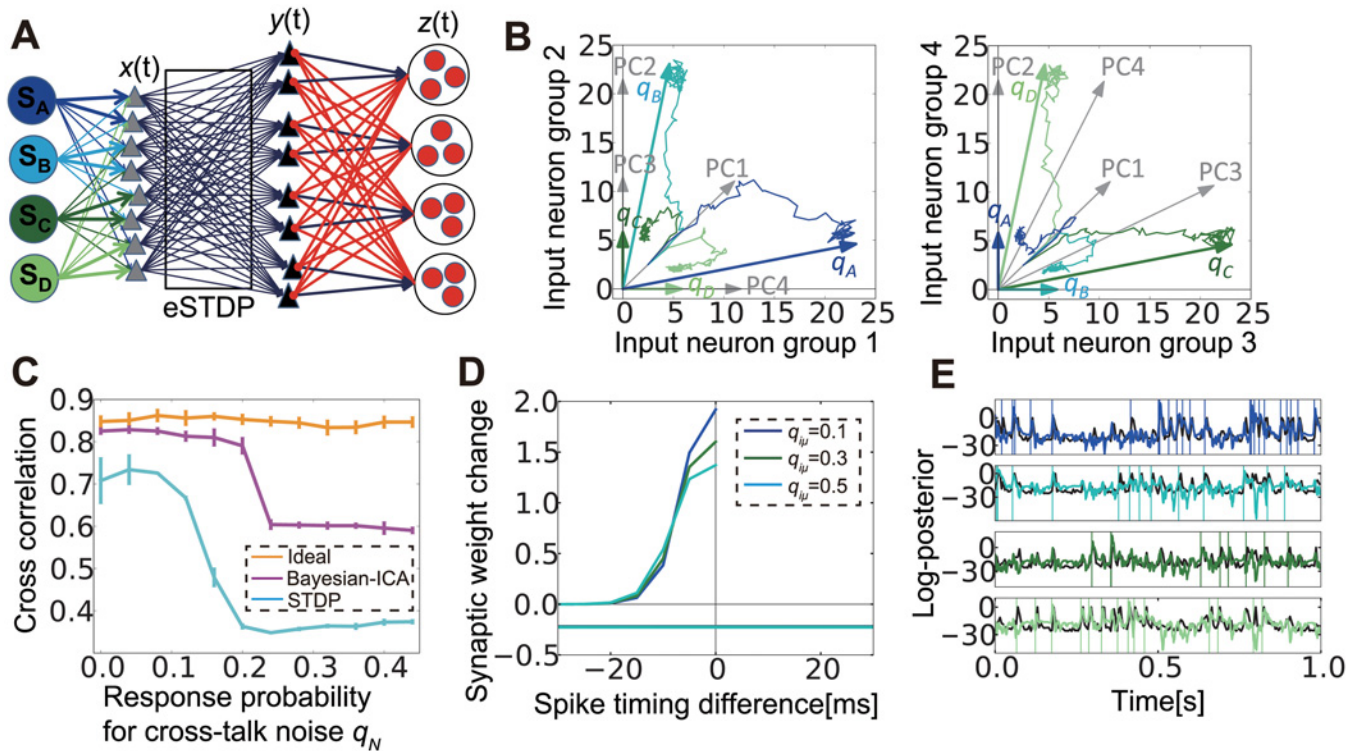
Hebbian inhibitory STDP at lateral connections is not always beneficial for learning. For example, in minor source detection, if we use Hebbian inhibitory STDP, a slightly minor source is not detectable, whereas for anti-Hebbian STDP, a small number of neurons still detect the minor source because reciprocal connections from strong-source responsive inhibitory neurons to strong-source responsive output neurons inhibit synaptic weight development for the stronger source ([Fig 6C](#)).

### Neural Bayesian ICA and blind source separation

Our results to this point have revealed that correlation-based STDP learning combined with lateral inhibition can successfully detect signals from mixed inputs masked by noises. To confirm this mechanism is indeed effective in realistic tasks, we applied the above method to blind source separation. We first examined the condition in which the network could capture external sources. We extended the previous network to include four independent sources mixed at the input layer ([Fig 7A](#)). In the present application, we used structured lateral connections because learning for clustered structures is difficult with noisy stimuli, as shown in the preceding section. The response probability matrix  $Q$  and correlation matrix  $C$  are given as

$$Q = \begin{pmatrix} q_S & q_N & 0 & q_N \\ q_N & q_S & q_N & 0 \\ 0 & q_N & q_S & q_N \\ q_N & 0 & q_N & q_S \end{pmatrix}, C = \begin{pmatrix} q_S^2 + 2q_N^2 & 2q_Sq_N & 2q_N^2 & 2q_Sq_N \\ 2q_Sq_N & q_S^2 + 2q_N^2 & 2q_Sq_N & 2q_N^2 \\ 2q_N^2 & 2q_Sq_N & q_S^2 + 2q_N^2 & 2q_Sq_N \\ 2q_Sq_N & 2q_N^2 & 2q_Sq_N & q_S^2 + 2q_N^2 \end{pmatrix}.$$

Therefore, the principal components of matrix  $Q$  (i.e., eigenvectors of  $C$ ) are  $\{1, 1, 1, 1\}$ ,  $\{-1, 0, 1, 0\}$ ,  $\{0, -1, 0, 1\}$ ,  $\{-1, 1, -1, 1\}$ . Because the first-order approximation of synaptic weight dynamics follows  $\dot{W}_X \approx g_1^X W_X C^t$ , we may expect that synaptic weight vectors converge to the eigenvectors of the principal components; however, this was not the case in our simulations, even if we took into account the non-negativity of synaptic weights (see [Fig 7B](#), where we renormalized the principal vectors to the region between 0 and 1). Instead, each weight vector converged to a column of the response probability matrix  $Q$  ([Fig 7B](#), the left panel is the projection to the first two dimensions, and the right panel is the projection to the other two



**Fig 7. With lateral inhibition, spike-timing-dependent plasticity (STDP) mimics Bayesian independent component analysis (ICA).** (A) Schematic figure of the model with four sources. (B) Synaptic weight development in input neuron space. Arrows  $q_A$  to  $q_D$  are response probability vectors of the four sources, and PC1 to PC4 are normalized principal components of the correlation matrix C. Lines represent traces of average synaptic weight from each input group to the output groups that learned corresponding sources during the learning process. (C) Comparison of performance among the ideal observer, Bayesian ICA learning, and STDP learning. (D) LTP/LTD time window of Bayesian ICA learning. (E) Behaviors of log-membrane potential (color lines) in the STDP model, and estimated log-posterior (black lines) in the Bayesian ICA algorithm for the same stimuli. Vertical lines represent timings of external events. Log-membrane potentials are normalized to align the mean and the variance to the corresponding log-posteriors.

doi:10.1371/journal.pcbi.1004227.g007

dimensions). This result implies that the network can extract independent sources, rather than principal components, from multiple intermixed inputs.

We next evaluated the performance of hidden external source detection, especially its tolerance against crosstalk noise. To this end, we compared the performance of the model with that of the Bayesian ICA algorithm, in which independence of external sources is treated as a prior [46,47]. In the algorithm, the learned mixing matrix may converge to its Bayesian optimal value estimated from a stream of inputs. Although we cannot directly argue the optimality of cross-correlations, if the mixing matrix is accurately estimated, external activity is also well inferred, and thus we can use the mean cross-correlation as a measure for the optimality of learning. In terms of discretized input activity  $X$ , the external source activity  $S$  and prior information  $I$ , we can express the conditional probability of the estimated response probability

matrix  $\tilde{Q}$  as  $P[\tilde{Q}|X, I] = \frac{P[\tilde{Q}|I]}{P[X|I]} \int P[X|S, \tilde{Q}, I] P[S|I] dS$  (see Bayesian ICA in [Methods](#) for details).

This means that even if no prior information is given for  $\tilde{Q}$  itself (i.e.  $P[\tilde{Q}|I] = const.$ ), posterior  $P[\tilde{Q}|X, I]$  still depends on a prior given for  $S$ . If we introduce a prior that each external source follows an independent Bernoulli Process

(i.e.  $P[S|I] = \prod_{k=1}^{T/\Delta t} \prod_{i=1}^L (r_s \Delta t)^{s_{ki}} (1 - r_s \Delta t)^{1-s_{ki}}$ ), then the stochastic gradient descent of posterior

function is given as,

$$\frac{\partial}{\partial \tilde{q}_{i\mu}} \log P[\tilde{Q}|X, I] = \frac{1}{Z_p} \sum_{k=1}^{T/\Delta t} \int P[S, X|\tilde{Q}, I] \frac{2x_i^k - 1}{x_i^k p_i^k / (1 - p_i^k) + (1 - x_i^k)} \frac{\sum_{k'=0}^{\infty} \phi_{k'} s_{\mu}^{k-k'}}{1 - \tilde{q}_{i\mu} \sum_{k'=0}^{\infty} \phi_{k'} s_{\mu}^{k-k'}} dS,$$

where

$$p_i^k = 1 - (1 - r_i^o \Delta t) \prod_{\mu=1}^p \left[ 1 - \tilde{q}_{i\mu} \sum_{k'=0}^{\infty} \phi_{k'} s_{\mu}^{k-k'} \right], \phi_k = \frac{1}{2\theta_t^3} [(k + 1/2)\Delta t]^2 \exp[-(k + 1/2)\Delta t/\theta_t].$$

We approximated this Bayesian ICA algorithm by a sequential sampling source activity instead of calculating the integral over all possible combinations in the estimation of the log-posterior of the response probability matrix Q. In this approximation, the learning rule of the estimated response probability matrix  $\tilde{Q}$  obeys

$$\Delta \tilde{q}_{i\mu}^k \propto \frac{2x_i^k - 1}{x_i^k p_i^k(Y^{1:k-1}) / (1 - p_i^k(Y^{1:k-1})) + (1 - x_i^k)} \times \frac{\sum_{k'=0}^{\infty} \phi_{k'} y_{\mu}^{k-k'}}{1 - \tilde{q}_{i\mu}^k \sum_{k'=0}^{\infty} \phi_{k'} y_{\mu}^{k-k'}} \quad (13)$$

$$p_i^k(Y^{1:k-1}) = 1 - (1 - r_i^o \Delta t) \prod_{\mu=1}^p \left[ 1 - \tilde{q}_{i\mu}^k \sum_{k'=0}^{\infty} \phi_{k'} y_{\mu}^{k-k'} \right],$$

where Y is the sampled sequence, and  $p_i^k(Y^{1:k-1})$  is the sample based approximation of  $p_i^k$  in the previous equation. This rule has spike-timing and weight dependence similar to those seen in STDP (Fig 7D). Although the performance of STDP is much worse than the ideal case (when the true Q is given), this performance is similar to that for the sample-based learning algorithm discussed above (Fig 7C). Therefore, the network detects independent sources if cross-talk noise is not large. We further studied the response of the models for the same inputs and

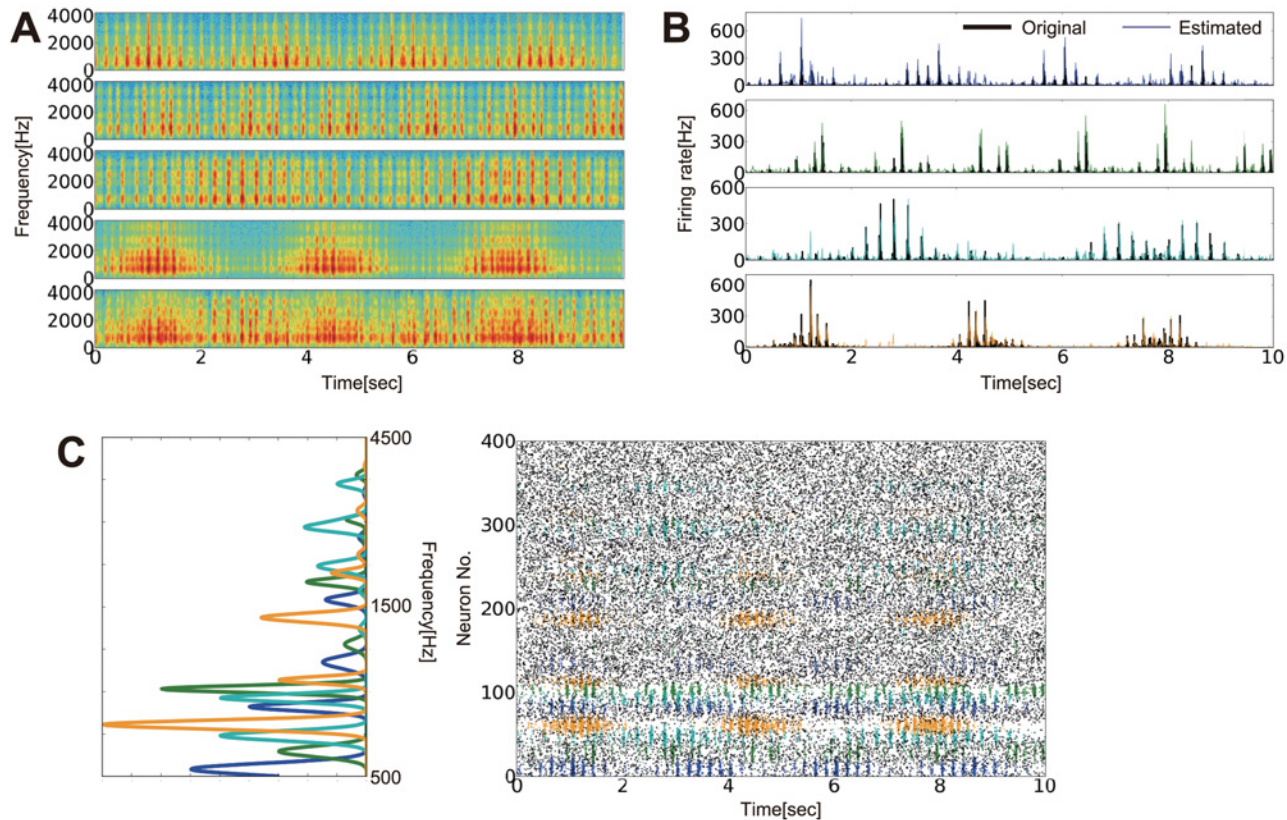
found that the logarithm of the average membrane potential  $u_{\mu}^E = \frac{1}{|\Omega_{\mu}|} \sum_{j \in \Omega_{\mu}} u_j^E$  well approxi-

mates the log-posterior estimated in Bayesian ICA, even in the absence of a stimulus (Fig 7E). This result suggests that in the STDP model, expected external states are naturally sampled through membrane dynamics that are generated through the interplay of feedforward and feedback inputs.

We finally performed the blind separation task using the same network as shown in Fig 7A. We created “sensory” inputs by mixing four artificially created auditory sequences (Fig 8A and S1 Auditory File). In the auditory cortex, various frequency components of a sound, particularly high-frequency components, are represented by specific neurons typically organized in a tonotopic map structure [48], whereas low-frequency components are expected to be perceived as a change in sound pressure. Furthermore, populations of neurons in the primary auditory cortex are known to synchronize the relative timing of their spikes during auditory stimuli and provide correlated spike inputs for higher cortical areas in which the auditory scene is fully analyzed and perceived [49,50]. We modeled these features by assuming that input neurons have a preferred frequency  $\{f_i\}$  defined as

$$f_i = \exp \left[ \frac{i}{L} (\log f_{\max} - \log f_{\min}) + \log f_{\min} \right],$$

and auditory inputs are provided as time-dependent response probabilities, which follow  $q_i(t) = q_o \sum_q a_i^q(t) a_h^q(f_i)$ , where  $a_h^q(f)$  is the spectrum of auditory source q (left panel of



**Fig 8. Blind source separation by spike-timing-dependent plasticity (STDP).** (A) Four original auditory signals (from the top to the fourth set of signals) and one mixed signal (bottom). (B) Amplitudes of original signals (black lines) and those estimated from output firing rates (colored lines). (C) Spectra of auditory sources  $a^q_h(f)$  (left). Raster plots of input neuron activity. Colors were probabilistically assigned based on expected sources. All figures were calculated from the 30'00"–30'10" portion of the auditory signals and simulation.

doi:10.1371/journal.pcbi.1004227.g008

Fig 8C), and  $a^q(t)$  is the temporal change of the sound pressure (black lines in Fig 8B). In this representation, each sound source is represented by correlated spikes of neural populations (right panel of Fig 8C). Even if signals have overlapping frequency components  $\{a^q_h(f)\}_q$ , blind separation is possible as long as  $\{a^q_h(t)\}_q$  are independent and have sharp rising profiles sufficient to cause spike correlations. After learning, four output neuron groups successfully detected changes in the sound pressure of the four original auditory signals (colored lines in Fig 8B) by correctly identifying the input neurons that encoded the signals. Therefore, STDP rules implemented in a feedforward neural network with lateral inhibition serve as a spike-based solution to the blind source separation or cocktail party effect problem.

## Discussion

By analytically investigating the propagation of input correlations through feedback circuits, we revealed how lateral inhibition influenced plasticity at feedforward connections. We showed that a population of neurons could learn multiple signals with different strengths or mixed levels. In addition, we found that to perform learning from signals corrupted with random noise, the timescale of the input correlations needed to be in the range of milliseconds, whereas the timescale was broader for crosstalk noise, which may explain why the spike correlation of cortical neurons often exhibits a large jitter (approximately 10 ms) [36,37]. We also investigated the functional roles of STDP at lateral excitatory and inhibitory connections to demonstrate that



Hebbian STDP shaped the lateral structure to improve signal detection performance. Our results also suggested that anti-Hebbian plasticity was helpful for learning from minor sources and implied that different STDP rules at lateral connections induced different algorithms at feedforward connections. Furthermore, we derived an STDP-like online learning rule by considering an approximation of Bayesian ICA with sequence sampling. This result suggested that lateral inhibition adjusted the membrane potentials of postsynaptic neurons so that their spiking processes accurately performed sequence sampling. We also demonstrated that this mechanism was applicable to blind source separation of auditory signals.

## Noise characteristics and correlation timescales

Simultaneously recorded neurons in close proximity often show correlated spiking, yet the precision of these correlations varies across brain regions. Neurons in the lateral geniculate nucleus show strong spike correlations [42,51], while correlations in V1 [36,52] or higher visual areas [37] are less precise. Our results indicate the interesting possibility that these differences may reflect the different characteristics of the noise with which the various cortical areas need to contend. At an early stage of sensory processing, the major noise component may be environmentally produced background noise from various sources; thus precise spike correlation is beneficial at this stage for noise reduction during signal detection and learning (Fig 4G). By contrast, in higher sensory cortices, crosstalk noise accumulated through signal propagation in circuits may form the primary noise source, so less precise spike correlation is preferable (Fig 4H). It would be intriguing to examine whether lower and higher cortical areas similarly change the strength of spike correlations for other sensory modalities.

## STDP in E-to-I and I-to-E connections

It is known that both glutamergic synapses on inhibitory neurons [53,54] and GABAergic synapses on excitatory neurons [55,56] show STDP, and it is also known that STDP at E-to-I connections plays an important role in developmental plasticity [57]; however, detailed properties of these plasticities are still largely disputable [58,59] and, reportedly, highly dependent on inhibitory cell type [60], neuromodulator [61], and region [58]. We showed that in a feedback circuit, Hebbian inhibitory STDP preferred winner-take-all while anti-Hebbian inhibitory STDP tended to cause winner-share-all (see Fukai and Tanaka 1997 for winner-share-all) at excitatory neurons (Fig 6D). This result indicates that different inhibitory STDP imposes different functions for excitatory STDP, which suggests that a neural circuit may select optimal inhibitory STDP for a specific purpose or strategy of learning, and this may differ across regions and be modified by neuromodulators. A recent study showed that inhibitory plasticity even directly influences the plasticity at excitatory synapses of the postsynaptic neuron [62]. In such cases, algorithm selection would play a more important role than it did for the standard STDP implemented in our model.

Recently, inhibitory neurons in the rodent hippocampus CA1 were shown to display context-dependent activity rate changes during a spatial learning task, in association with the activity rate changes in excitatory cells [63]. In addition, the authors suggested the candidate mechanism for this change in activity is STDP at E-to-I synapses. Our results examining E-to-I STDP confirmed this configuration of inhibitory cells modulated by plasticity at feedforward excitatory connections (Fig 5D, S2A and S2B Fig). In our model, although inhibitory neurons are not directly projected from input sources, as excitatory neurons learn a specific input source (Fig 5D, left panel), inhibitory neurons acquire feature selectivity through Hebbian STDP at synaptic connections from those excitatory neurons (Fig 5D, middle panel). Furthermore, our results indicate an important function of these feature-selective inhibitory neurons.



Once an adequate circuit structure is learned and inhibitory connections are organized into a feature-selective pattern, even if the input to the network becomes noisy or faint, the network can still robustly detect signals (Fig 5E and 5F). This robustness would be useful for spatial learning, as contextual information is often uncertain.

## STDP and Bayesian ICA

Our results indicated that STDP in a lateral inhibition circuit mimicked Bayesian ICA [46,47]. First, output neurons were able to detect hidden external sources, without capturing principal components (Fig 7B). Previous results suggest that for a single output neuron, an additional homeostatic competition mechanism is necessary to detect an independent component [7,22]. In addition, when information is coded by firing rate, homeostatic plasticity is critically important, because STDP itself does not mimic Bienenstock-Cooper-Munro learning [18]. However in our model, information was encoded by correlation, and mutual inhibition naturally induced intercellular competition so that intracellular competition through homeostatic plasticity was unnecessary. Moreover, our analytical results suggested the reason that independent sources are detected. To perform a principal components analysis using neural units, the synaptic weight change needs to follow

$$\dot{W}_x = W_x C - LT[W_x C W_x^t] W_x,$$

where  $LT[]$  means lower triangle matrix [64,65]. This LT transformation protects principal components caused by the lateral modification from higher order components; however in our model, because all output neurons receive the same number of inhibitory inputs Eq (2), all neurons are decorrelated with one another and develop into independent components.

Recently, it was shown that STDP can perform Bayesian optimal learning [66,67]. In the model used by those authors, the synaptic weight matrix is treated as a hyper parameter and estimated by considering the maximum likelihood estimation of input spike trains. By contrast, in the Bayesian ICA framework, the mixing matrix (corresponding to synaptic weight matrix) is treated as a probabilistic variable. Using this framework, we needed to calculate an integral over all possible source activities in the past to derive stochastic gradient descent; however, as shown in Fig 7C, the stochastic learning was well performed by employing an approximation with sequential sampling. Moreover, we naturally derived an adequate LTP time window from the response kernel of input neurons to external events (Fig 7D). We also found that STDP self-organized a lateral circuit structure that performed better than a random global inhibition (Fig 5E and 5F). Mathematically, to perform sampling from a probabilistic distribution, we first needed to calculate the occurrence probability of each state; however, in a neural model, membrane potentials of output neurons approximately represent the occurrence probability through membrane dynamics. In machine learning methods, integration over possible source activities is often approximated using Markov chain Monte Carlo (MCMC) sampling [68]. Interestingly, a recent study showed that a recurrent network performed MCMC sampling [69,70], suggesting that our network may perform a more accurate sampling in the presence of recurrent excitatory connections.

## Suboptimality of STDP

Previous theoretical results suggest that STDP can modulate synaptic weights in a way that optimizes information transmission between pre- and postsynaptic neurons [71,72]. In the Bayesian ICA framework, blind source separation can be formulized as an optimization problem, but, in this case, the problem itself is ill-defined because optimality does not guarantee the true

solution. In addition, local minima are often unavoidable for online learning rules. Nevertheless, the problems faced by the brain are often ill-defined, and suboptimality is inevitable [73]. Because we performed both nonlinear dynamics-based and machine learning-based analyses, we can offer some insights regarding the origins of local minima in stochastic gradient descent learning. In the initial state, synaptic weights are typically homogeneously distributed, and this state is often locally stable. As a result, the homogeneous stable point is more likely to be selected in learning (Fig 2C and 2D) than the non-homogenous, more desirable, points; however, introducing additional noise may change this situation. Indeed, in Fig 4B and Fig 7C, the performance of the model was improved by adding a small amount of noise to input activities, although the improvement was not significant; however, because a large amount of noise is harmful for computations and stable learning, the benefit of noise addition is highly limited, and the brain may recruit other mechanisms for near optimal learning.

## Neural mechanism of blind source separation

Humans and nonhuman animals can detect a specific auditory sequence from a mixed, noisy auditory stimulus, a phenomenon often called the cocktail party effect. The mechanism underlying the cocktail party effect remains elusive [26,28,29], although several solutions have been proposed [74,75]. An effective solution for this problem is ICA [76–78], and the neural implementation of the algorithm has been studied by several authors [14,18,79,80]. Our study extended these results through a rigorous analytical treatment on biologically plausible STDP learning of spiking neurons, and our analyses enabled us to discover interesting functions of correlation coding. Moreover, by explicitly modeling inhibitory neurons, we found that STDP at E-to-I and I-to-E connections cooperatively organized a lateral structure suitable for blind source separation. In addition, we successfully extended a previous model for the formation of static visual receptive fields [18,19] to a more dynamic model in an auditory blind source separation task. In realistic auditory scene analysis, the frequency spectrum of acoustic signals is first analyzed in the cochlea, where each frequency component is the mixture of sound components from independent sources. Components belonging to the same source may be separated and integrated by downstream auditory neurons for the perception of the original signal. These frequency components can be considered a mixed signal in the ICA problem [81]; thus even if signals are mixed in frequency space, if the amplitudes of the signals are temporally independent, blind separation is still achievable. In the neural implementation of the problem, if two frequencies are commonly activated in the same signal, neurons representing those frequencies show spike correlation under the presence of the signal; thus the learning process is naturally achieved by STDP learning. These results indicate an active role of spike correlation and STDP in efficient biological learning.

## Methods

### Model

**Neural dynamics.** Based on the previous study [7], we constructed a network model with one external layer and three layers of neurons (Fig 1A). The first layer is the external layer that corresponds to external stimuli or the sensory system's response to these stimuli. For simplicity, we approximated the activity of external sources using a Poisson process with the constant rate  $v_o^S$ . If we define the Poisson process with rate  $r$  as  $\hat{\sigma}(r)$ , the activity of the external source  $\mu$  at time  $t$  is written as  $s_\mu(t) = \hat{\sigma}(v_o^S)$  (see Table 1 for the list of variables). Neurons in the input layer fire spikes in response to activity in the external layer. By assuming a rate-modulated

Poisson process, the spiking activity of the input neuron  $i$  follows

$$x_i(t) = \hat{\sigma} \left[ r_i^o + \sum_{\mu=1}^P q_{i\mu} \int_0^\infty \phi(t') s_\mu(t-t') dt' \right], \quad (14)$$

where  $r_i^o$  is the instantaneous firing rate defined with  $r_i^o = v_o^x - \sum_{\mu=1}^P q_{i\mu} v_o^s$ ,  $q_{i\mu}$  is the response probability for the hidden external source  $\mu$ , and  $\phi(t) = t^2 e^{-t/\theta_t} / 2\theta_t^3$  is the response kernel for each external event. In most theoretical studies, cross-correlations give an exponential decay or a delta function [5,38], but here we used a response kernel that produces broader correlations (Fig 4A right), because the actual correlations observed in the cortex are usually not sharply peaked [36,37]. For instance, for the exponential kernel  $\phi_e(t) = e^{-t/\theta_t} / \theta_t$ , correlations show a peaked distribution even if the timescale parameter  $\theta_t$  is several milliseconds (S3A Fig). Because of the common inputs from the external layer, input neurons show highly correlated activity, which enables population coding of the hidden structure. Although here we explicitly assumed the presence of the external layer, these analytical results can also be applied for arbitrary realization of a spatiotemporal correlation.

Output neurons are modeled with the Poisson neuron model [5,38,45] in which the membrane potential of neuron  $j$  at time  $t$  is described as

$$u_j^E(t) = \sum_{i=1}^M w_{ji}^X \int_0^\infty \varepsilon_X(r) x_i(t-r-d_{ji}^X) dr - \sum_{k=1}^N w_{jk}^Z \int_0^\infty \varepsilon_Z(r) z_k(t-r-d_{jk}^Z) dr, \quad (15)$$

where  $w_{ji}^X$  and  $w_{jk}^Z$  are the EPSPs/IPSPs of input currents from input neuron  $x_i$  and lateral neuron  $z_k$ , respectively, convolution functions are defined as  $\varepsilon_X(r) = \frac{e^{-r/\tau_A^X} - e^{-r/\tau_B^X}}{\tau_A^X - \tau_B^X}$  and  $\varepsilon_Z(r) = \frac{e^{-r/\tau_A^Z} - e^{-r/\tau_B^Z}}{\tau_A^Z - \tau_B^Z}$ , and synaptic delays in the feedforward excitatory and feedback inhibitory connections are  $d_{ij}^X$  and  $d_{jk}^Z$ . For feedforward excitatory connections, the synaptic delay  $d_{ij}^X$  is given by the sum of the axonal delay  $d_{ij}^a$  and dendritic delay  $d_{ij}^d$ , whereas for inhibitory connections, we assume for simplicity that the delay is purely axonal. The response of the output neuron follows  $y_j(t) = \hat{\sigma}[g_E(u_j^E)]$ . Similarly, inhibitory neurons in the lateral layer show Poisson firing based on the membrane potential  $\{u_k^I\}_{k=1, \dots, N}$  which is defined as

$$u_k^I(t) = \sum_{j=1}^M w_{kj}^Y \int_0^\infty \varepsilon_Y(r) y_j(t-r-d_{kj}^Y) dr, \quad (16)$$

for EPSPs of a lateral connection  $w_{kj}^Y$ , convolution function  $\varepsilon_Y(r) = \frac{e^{-r/\tau_A^Y} - e^{-r/\tau_B^Y}}{\tau_A^Y - \tau_B^Y}$ , and synaptic delay of the lateral connection  $d_{kj}^Y$ . The synaptic delay of the excitatory lateral connection is also approximated as the axonal delay. The spiking activity of the inhibitory neurons is given with  $z_k(t) = \hat{\sigma}[g_I(u_k^I)]$ . For analytical tractability, we use a linear response curve  $g_E(u) = u$  and  $g_I(u) = u$ .

**Synaptic plasticity.** For most of this study, we focused on synaptic plasticity in the feedforward connection  $W_X$ , with fixed lateral synaptic weights  $W_Y$  and  $W_Z$ . When the timing of the spikes at the cell bodies of pre- and postsynaptic neurons is  $t_{pre}$  and  $t_{post}$ , spike timings at the synaptic sites are  $t_{pre}^s = t_{pre} + d_{ji}^a$  and  $t_{post}^s = t_{post} + d_{ji}^d$  with axonal and dendritic delays of  $d_{ji}^a$

and  $d_{ji}^d$ . For every pair of  $t_{pre}^s$  and  $t_{post}^s$ , synaptic weight change is given with

$$\Delta w_{ji}^X = \begin{cases} \eta^X f_p(w_{ji}^X) \exp[-(t_{post}^s - t_{pre}^s)/\tau_p] & (\text{if } t_{post}^s > t_{pre}^s) \\ \eta^X f_d(w_{ji}^X) \exp[-(t_{pre}^s - t_{post}^s)/\tau_d] & (\text{if } t_{post}^s < t_{pre}^s) \end{cases} \quad (17)$$

For the synaptic weight dependence of STDP, we considered a pairwise log-STDP [31] in which LTP/LTD follows

$$f_p(w) = C_p(1 + \sigma_{stdp} \xi) e^{-w/(\beta w_o)}, f_d(w) = -C_d(1 + \sigma_{stdp} \xi) \frac{\log(1 + \alpha w/w_o)}{\log(1 + \alpha)}, \quad (18)$$

where  $\xi$  is a Gaussian random variable. The log-weight dependence well replicates experimentally observed synaptic weight distributions [32,33] and is suggested to have an important function in memory modulation [82]. Analytical treatment below is applicable to other types of synaptic weight dependence, yet in the additive STDP (i.e.  $f_p(w) = C_p$  and  $f_d(w) = C_d$ ), the mean-field equation typically does not have any stable fixed point. In addition, under the multiplicative STDP in which LTD has a linear rather than a logarithmic dependence on synaptic weight, strong correlation is often necessary to induce salient LTP [31]. The coefficients  $C_p = 1$  and  $C_d = C_p \tau_p^X / \tau_d^X$  are chosen so that total LTP and LTD are balanced around the referential synaptic weight.

The STDP at E-to-I connections and I-to-E connections is similarly defined. For simplicity, we assume that synaptic delays are solely axonal (i.e.,  $d_{kj}^Y = d_{kj}^{Y,a}, d_{kj}^Z = d_{kj}^{Z,a}$ ), and the change in synaptic weight does not depend on the synaptic weight. To maintain the balance between LTP and LTD, coefficients are chosen as  $C_p^Y = 1, C_d^Y = \gamma^Y C_p^Y \tau_p^Y / \tau_d^Y, \eta^Y = 0.3 \eta w_o^Y / w_o^X$ . Similarly, for I-to-E connections,  $C_p^Z = 1, C_d^Z = \gamma^Z C_p^Z \tau_p^Z / \tau_d^Z, \eta^Z = 0.3 \eta w_o^Z / w_o^X$ . We also modify constant (initial) synaptic weights to  $w_o^Y = 50.0$  and  $w_o^Z = 25.0$ , and bounded synaptic weights with  $w_{max}^Y = 100.0$  and  $w_{max}^Z = 50.0$ . In this normalization, the total lateral inhibition takes the same value as that in the non-plastic model at the initial state. Time windows are defined as  $\tau_p^Y = \tau_d^Y = \tau_p^Z = \tau_d^Z = 20.0$  ms.

In Fig 6C, anti-Hebbian STDP was calculated by

$$\Delta w^Q = \begin{cases} -\eta^Q \exp[-(t_{post}^s - t_{pre}^s)/\tau_d^Q] & (\text{if } t_{post}^s > t_{pre}^s) \\ \eta^Q \gamma^Q (\tau_d^Q / \tau_p^Q) \exp[-(t_{pre}^s - t_{post}^s)/\tau_p^Q] & (\text{if } t_{post}^s < t_{pre}^s) \end{cases}$$

for  $Q = Y$  or  $Z$ . Similarly, the correlation detector type of STDP in S2 Fig was defined as

$$\Delta w^Q = \begin{cases} \eta^Q (\exp[-(t_{post}^s - t_{pre}^s)/\tau_p^Q] - (\tau_p^Q / \tau_d^Q) \exp[-(t_{post}^s - t_{pre}^s)/\tau_d^Q]) & (\text{if } t_{post}^s > t_{pre}^s) \\ \eta^Q \gamma^Q (\exp[-(t_{post}^s - t_{pre}^s)/\tau_p^Q] - (\tau_p^Q / \tau_d^Q) \exp[-(t_{post}^s - t_{pre}^s)/\tau_d^Q]) & (\text{if } t_{post}^s < t_{pre}^s) \end{cases}$$

The anti-correlation detector was calculated by changing the sign of above equations.

**Leaky Integrate-and-Fire (LIF) model.** In the main text, we performed all simulations with a linear Poisson model for analytical purposes, although we also confirmed those results with a conductance-based LIF model (S1 Fig). In the LIF model, the membrane potentials of excitatory neurons follow

$$\frac{dv_j^E}{dt} = -\frac{1}{\tau_m^E} (v_j^E - V_L) - g_j^{EE} (v_j^E - V_E) - g_j^{EI} (v_j^E - V_I),$$

$$\frac{dg_j^{EE}}{dt} = -\frac{g_j^{EE}}{\tau_s^{EE}} + \sum_{i=1}^L w_{ji}^X \sum_{t_i^s} \delta(t - t_i^s), \text{ and } \frac{dg_j^{EI}}{dt} = -\frac{g_j^{EI}}{\tau_s^{EI}} + \sum_{k=1}^N w_{jk}^Z \sum_{t_k^s} \delta(t - t_k^s),$$

where  $g_j^{EE}$  and  $g_j^{EI}$  are excitatory and inhibitory conductances, respectively, and  $t_i^s$  and  $t_k^s$  are the spike timings of input neuron  $i$  and lateral neuron  $k$ . Similarly, for inhibitory neurons in the lateral layer,

$$\frac{dv_k^I}{dt} = -\frac{1}{\tau_m^I} (v_k^I - V_L) - g_k^{IE} (v_k^I - V_E) - g_k^{II} (v_k^I - V_I),$$

$$\frac{dg_k^{IE}}{dt} = -\frac{g_k^{IE}}{\tau_s^{IE}} + \sum_{j=1}^M w_{kj}^Y \sum_{t_j^s} \delta(t - t_j^s), \quad \text{and} \quad \frac{dg_k^{II}}{dt} = -\frac{g_k^{II}}{\tau_s^{II}} + w_{II} \sum_{t_r^s} \delta(t - t_r^s).$$

In addition to the excitatory inputs from the output layer, we added random inhibitory inputs as Poisson processes with a fixed firing rate  $r_o^{II}$  for inhibitory neurons. A neuron fires if the membrane potential exceeds the threshold  $V_{th}$ , and immediately goes into a refractory period in which the membrane potential stays at  $V_{ref}$  for 1 ms after spiking. Plasticity was implemented for  $w_{ji}^X$  in the same manner as that for the Poisson model. Parameters were chosen as  $V_L = -70.0$ ,  $V_E = 0.0$ ,  $V_I = -80.0$ ,  $V_{ref} = -60.0$ ,  $V_{th} = -50.0$  mV,  $t_m^E = 20.0$ ,  $t_m^I = 10.0$ ,  $ts^{EE} = 5.0$ ,  $t_s^{EI} = 2.5$ ,  $t_s^{IE} = 4.0$ ,  $t_s^{II} = 5.0$  ms,  $w_o^X = 0.001$ ,  $w_o^I = 0.008$ ,  $w_o^L = 1.0$ ,  $r_o^{II} = 1000.0$  Hz,  $w_o^{II} = 0.005$ ,  $C_d = 1.8C_p\tau_p^X/\tau_d^X$ , and  $\alpha = 50.0$ . All other parameters were the same as those used in the Poisson model (Table 2).

In the LIF model, synaptic weights develop in a manner similar to that for the linear Poisson model, although change occurs more rapidly (Fig 1B, S1A Fig). Both cross-correlation and mutual information behave as they do in the Poisson model, but the performance is slightly better, possibly because the dynamics are deterministic (Fig 1D and 1E, S1B and S1C Fig); however, membrane potentials show different responses for correlation events (S1D Fig) because output neurons are constantly in high-conductance states, so that correlation events immediately cause spikes. As a result, membrane potentials drop to the  $V_{ref}$ , and the average potential goes down. Interestingly, after neuron groups detect different signals, a preferred signal initially causes hyperpolarization due to firing, but, subsequently, a non-preferred signal causes hyperpolarization due to lateral inhibition (Fig 1D right). The PSTH of firing shows that the behavior of the membrane potential in the Poisson model is similar (Fig 1C and S1E Fig). This is natural, because in the linear Poisson model, the firing rate has linear relationship with the membrane potential, whereas in LIF model relationship between the average membrane potential and firing rate is highly non-linear.

**Bayesian ICA.** If discretized with  $\Delta t$ , the time series of the external source activity is written as  $S = \{s_{\mu k}\}_{\mu=1, \dots, p}^{k=1, \dots, T/\Delta t}$ , and input activity becomes  $X = \{x_{ik}\}_{i=1, \dots, L}^{k=1, \dots, T/\Delta t}$ . Therefore, for prior information  $I$ , the joint probability of sources  $S$  and the estimated response probability matrix  $Q$  is

$$P[S, \tilde{Q}|X, I] = P[X|S, \tilde{Q}, I]P[S, \tilde{Q}|I]/P[X|I].$$

Therefore, by considering marginal probability,

$$P[\tilde{Q}|X, I] = \frac{P[\tilde{Q}|I]}{P[X|I]} \int P[X|S, \tilde{Q}, I]P[S|I]dS. \tag{19}$$

By considering maximum likelihood estimation for a given prior  $P[S|I]$ ,  $Q$  can be optimally estimated [46,47]. In our problem setting, by assuming that external signals are independent,



and input neurons respond to signals with a Bernoulli process,

$$P[X|S, \tilde{Q}, I] = \prod_{k=1}^{T/\Delta t} \prod_{i=1}^L [x_i^k p_i^k + (1 - x_i^k)(1 - p_i^k)], P[S|I] = \prod_{k=1}^{T/\Delta t} \prod_{i=1}^L (r_s \Delta t)^{s_i^k} (1 - r_s \Delta t)^{1-s_i^k},$$

where

$$p_i^k = 1 - (1 - r_i^o \Delta t) \prod_{\mu=1}^p \left[ 1 - \tilde{q}_{i\mu} \sum_{k'=0}^{\infty} \phi_{k'} s_{\mu}^{k-k'} \right], \phi_k = \frac{1}{2\theta_t^3} [(k + 1/2)\Delta t]^2 \exp[-(k + 1/2)\Delta t/\theta_t].$$

Therefore, log-likelihood becomes

$$\log P[\tilde{Q}|X, I] = \log \left( \int dS \prod_{k=1}^{T/\Delta t} \left[ \prod_{i=1}^L (x_i^k p_i^k + (1 - x_i^k)(1 - p_i^k)) \times \prod_{\mu=1}^p (r_s \Delta t)^{s_{\mu}^k} (1 - r_s \Delta t)^{1-s_{\mu}^k} \right] \right). \quad (20)$$

By taking gradient descent,

$$\frac{\partial}{\partial \tilde{q}_{i\mu}} \log P[\tilde{Q}|X, I] = \frac{1}{Z_p} \sum_{k=1}^{T/\Delta t} \int P[S, X|\tilde{Q}, I] \frac{2x_i^k - 1}{x_i^k p_i^k / (1 - p_i^k) + (1 - x_i^k)} \frac{\sum_{k'=0}^{\infty} \phi_{k'} s_{\mu}^{k-k'}}{1 - \tilde{q}_{i\mu} \sum_{k'=0}^{\infty} \phi_{k'} s_{\mu}^{k-k'}} dS.$$

Therefore, we need to calculate the integral over all possible combinations of sources in the past to obtain stochastic gradient descent; however, such a calculation is computationally difficult and incompatible with neural computation. Instead, we used sequential sampling of

$Y = \{y_{\mu k}\}_{\mu=1, \dots, p}^{k=1, \dots, T/\Delta t}$ , which is randomly sampled from

$$\begin{aligned} P[y^k = s^k] &\propto P[s^k, x^k | Y^{1:k-1}, \tilde{Q}, I] \\ &= \prod_{i=1}^L (x_i^k p_i^k(s^k, Y^{1:k-1}) + (1 - x_i^k)(1 - p_i^k(s^k, Y^{1:k-1}))) \times \prod_{\mu=1}^p (r_s \Delta t)^{y_{\mu}^k} (1 - r_s \Delta t)^{1-y_{\mu}^k}, \end{aligned} \quad (21)$$

where

$$p_i^k(y^k, Y^{1:k-1}) = 1 - (1 - r_i^o \Delta t) \prod_{\mu=1}^p \left[ 1 - \tilde{q}_{i\mu} \sum_{k'=0}^{\infty} \phi_{k'} y_{\mu}^{k-k'} \right].$$

Note in the above equations,  $x^k$  is given as a fixed value and not a random variable. Under this sample-based approximation, the stochastic gradient descent follows

$$\Delta \tilde{q}_{i\mu}^k \propto \frac{2x_i^k - 1}{x_i^k p_i^k(y^k, Y^{1:k-1}) / (1 - p_i^k(y^k, Y^{1:k-1})) + (1 - x_i^k)} \times \frac{\sum_{k'=0}^{\infty} \phi_{k'} y_{\mu}^{k-k'}}{1 - \tilde{q}_{i\mu} \sum_{k'=0}^{\infty} \phi_{k'} y_{\mu}^{k-k'}}. \quad (22)$$

For Fig 7C, we discretized the activity of hidden sources and input neurons with 5 ms bins, and performed learning with a learning rate  $\eta^{SGD} = 0.001$ . Cross-correlation was evaluated using the sample sequence Y. For the ideal case, we performed sequential sampling from the true response probability Q.

If  $y_{\mu}^{k-k'} = 1$  and  $y_{\mu}^{k-k''} = 0$  for all other nearby  $k'' (\neq k')$ , and if  $q_{i\nu} = 0$  for all  $\nu (\neq \mu)$ , then LTP at the connection  $q_{i\mu}$  caused by an output spike  $y_{\mu}^{k-k'} = 1$  for  $x_i^k = 1$  is written as

$$\Delta q_{i\mu}^{k,k',LTP} = \frac{(1 - [r_o^x - r_o^s \tilde{q}_{i\mu}]) (1 - \tilde{q}_{i\mu} \phi_{k'})}{1 - (1 - [r_o^x - r_o^s \tilde{q}_{i\mu}]) (1 - \tilde{q}_{i\mu} \phi_{k'})} \times \frac{\phi_{k'}}{1 - \tilde{q}_{i\mu} \phi_{k'}}. \quad (23)$$

In the absence of the input spike ( $x_i^k = 0$ ), an output spike  $y_{\mu}^{k-k'} = 1$  causes LTD in total

$\Delta q_{iu}^{LTD} = -\sum_{k'=0}^{\infty} \frac{\phi_{k'}}{1 - \tilde{q}_{iu} \phi_{k'}}$ . Therefore, this learning rule has weight dependence and temporal dependence similar to those in STDP. In Fig 7D, we plotted  $\Delta q_{iu}^{k,k',LTP}$  and  $\Delta q_{iu}^{LTD}$  for different  $\tilde{q}_{iu}$  ( $\tilde{q}_{iu} = 0.1, 0.3, 0.5$ ).

**Blind source separation.** In the blind source separation task, we created the original source by calculating high-frequency and low-frequency components separately. First, the spectrum of the signal  $q$  at a high frequency was defined as

$$a_h^q(f) = \sum_i \sum_k \frac{a_{h,i}^q b_{h,k}^q}{\sqrt{2\pi\sigma_{hf}^2}} \exp\left[-(f - kf_{h,i}^q)^2 / (2\sigma_{hf}^2)\right],$$

where  $f_{h,i}^q$  is a characteristic frequency of signal  $q$ , and  $kf_{h,i}^q$  are the harmonics of that frequency. The standard deviation was defined as  $\sigma_{hf}^o = k\sigma_{h,f}^o$  for  $\sigma_{h,f}^o = 20\text{Hz}$ . Low-frequency components were directly given as an exponential oscillation as below.

$$a_i^q(t) = \frac{1}{Z_i} \exp\left[\beta_i \sum_i a_{i,i}^q \cos(2\pi f_{i,i}^q (t - \delta_{i,i}^q))\right],$$

$f_{i,i}^q$  is a characteristic frequency, and  $\delta_{i,i}^q$  is the delay. By combining these two components, the amplitude of a mixed sound is given as

$$a(t) = \sum_q a_i^q(t) \sum_i a_h^q(f_i) \cos(2\pi f_i (t - \delta_i^q)).$$

Summation over frequency  $f$  is performed using 400 representative values that correspond to the tuned frequency of each input neuron:

$$f_i = \exp\left[\frac{i}{L} (\log f_{\max} - \log f_{\min}) + \log f_{\min}\right].$$

In neural implementation, input neurons were stimulated with the response probability  $q_i(t) = q_o \sum_q a_i^q(t) a_h^q(f_i)$  where  $q_o = 0.05$ .

In the simulated example, for high-frequency components, we defined  $f_{h,I}^q = \{\{523.3, 784.0\}, \{587.4, 880.0\}, \{650.0, 830.6\}, \{698.5, 932.4\}\}$ ,  $a_{h,I}^q = \{\{0.6, 0.4\}, \{0.3, 0.7\}, \{0.5, 0.5\}, \{0.9, 0.3\}\}$ ,  $b_{h,k}^q = \{\{1.0, 0.5, 0.2, 0.1\}, \{1.0, 0.5, 0.3, 0.2\}, \{1.0, 0.1, 1.0, 0.8\}, \{1.0, 0.8, 0.1, 0.1\}\}$ , and  $\sigma_{h,f}^o = 20\text{ Hz}$ . Each column represents four different sources. Similarly for low-frequency components, we used  $f_{l,I}^q = \{\{0.4, 5.0, 10.0, 40.0, 88.0\}, \{0.6, 6.0, 8.0, 42.0, 86.0\}, \{0.2, 4.0, 7.5, 44.0, 84.0\}, \{0.3, 6.0, 7.0, 46.0, 82.0\}\}$ ,  $a_{l,I}^q = \{\{0.3, 0.4, 0.2, 0.5, 0.5\}, \{0.25, 0.5, 0.2, 0.5, 0.5\}, \{0.24, 0.3, 0.4, 0.5, 0.5\}, \{0.61, 0.2, 0.2, 0.5, 0.5\}\}$ ,  $\delta_{l,I}^q = \{\{1.0, 0.25, 0.65, 0.17, 0.01\}, \{3.0, 0.12, 0.32, 0.13, 0.02\}, \{7.8, 0.55, 0.40, 0.11, 0.03\}, \{4.5, 0.22, 0.71, 0.07, 0.05\}\}$ ,  $\beta_l = 5.0$ , and  $Z_l = 27.24$ . We chose  $f_{\min} = 500\text{ Hz}$ ,  $f_{\max} = 4,500\text{ Hz}$ , and  $\delta_{l,I}^q$  was randomly selected from 0 to  $1/f_{\min}$ . Fig 8A was generated by performing Fourier transformations with 25 ms sliding bins at every 2.5 ms.

**Details of the simulation.** Simulations were calculated using the Runge-Kutta method, with a 0.05 ms time step. Initial synaptic weights were randomly chosen with  $w_{ij}^Q = w_o^Q (1 + \sigma_W^{init} \xi)$  for  $Q = X, Y, Z$  and a random Gaussian variable  $\xi$ . Similarly, synaptic delays were decided as  $d_{ij}^Q = d_{\min}^Q + (d_{\min}^Q - d_{\max}^Q) \xi$  for a random variable  $\xi$  uniformly chosen from  $[0, 1]$ .

### Analytical consideration of synaptic weight dynamics

**Correlation among input neurons.** Because input neurons receive common inputs from external sources, we define cross-correlation among input neurons as  $C_{il}(s) \equiv \langle x_i(t)x_l(t-s) - \langle x_i(t) \rangle \langle x_l(t) \rangle \rangle$ , and cross-correlation among input neurons satisfies

$$C_{il}(s) = \left\langle \hat{\sigma} \left[ r_i^o + \sum_{\mu=1}^p q_{i\mu} \int_0^\infty \phi(t') s_\mu(t-t') dt' \right] \times \hat{\sigma} \left[ r_l^o + \sum_{\mu=1}^p q_{l\mu} \int_0^\infty \phi(t'') s_\mu(t-s-t'') dt'' \right] \right\rangle - (v_o^x)^2 \quad (24)$$

$$\cong v_o^s \sum_{\mu=1}^p q_{i\mu} q_{l\mu} \int_0^\infty dt' \int_0^\infty dt'' \phi(t') \phi(t'') \delta(t' - t'' - s) = v_o^s \sum_{\mu=1}^p q_{i\mu} q_{l\mu} \int_{\max(0,s)}^\infty dt' \phi(t') \phi(t' - s).$$

When  $\phi(t) = t^2 e^{-t/\theta_t} / 2\theta_t^3$ ,  $C_{il}(s)$  becomes

$$C_{il}(s) = v_o^s \sum_{\mu=1}^p q_{i\mu} q_{l\mu} \frac{1}{16\theta_t^3} (s^2 + 3\theta_t |s| + 3\theta_t^2) e^{-|s|/\theta_t} = v_o^s \sum_{\mu=1}^p q_{i\mu} q_{l\mu} h(s),$$

where  $h(s) \equiv \frac{1}{16\theta_t^3} (s^2 + 3\theta_t |s| + 3\theta_t^2) e^{-|s|/\theta_t}$ .

**Average synaptic weight velocity.** The synaptic weight dynamics defined above can be re-written as

$$\frac{dw_{ji}^x}{dt} = x_i(t - d_{ji}^{xa}) \int_0^\infty F_d(w_{ji}^x, s) y_j(t - s - d_{ji}^{xd}) ds + y_j(t - d_{ji}^{xd}) \int_0^\infty F_p(w_{ji}^x, s) x_i(t - s - d_{ji}^{xa}) ds \quad (25)$$

for  $F_d(w_{ij}^x, s) = f_d(w_{ij}^x) e^{-s/\tau_d}$ ,  $F_p(w_{ij}^x, s) = f_p(w_{ij}^x) e^{-s/\tau_p}$ . By taking an average over a short period of time and also using a stochastic Poisson process, synaptic weight change follows

$$\begin{aligned} \left\langle \frac{dw_{ji}^x}{dt} \right\rangle &= \left\langle x_i(t - d_{ji}^{xa}) \int_0^\infty F_d(w_{ji}^x, s) y_j(t - s - d_{ji}^{xd}) ds \right\rangle + \left\langle y_j(t - d_{ji}^{xd}) \int_0^\infty F_p(w_{ji}^x, s) x_i(t - s - d_{ji}^{xa}) ds \right\rangle \\ &= \left\langle x_i(t - d_{ji}^{xa}) \int_{-\infty}^0 F_d(w_{ji}^x, -s') y_j(t + s' - d_{ji}^{xd}) ds' \right\rangle + \left\langle y_j(t - d_{ji}^{xd}) \int_0^\infty F_p(w_{ji}^x, s) x_i(t - s - d_{ji}^{xa}) ds \right\rangle \\ &= \left\langle \int_{-\infty}^0 F_d(w_{ji}^x, -s') x_i(t' - s' - d_{ji}^{xa}) y_j(t' - d_{ji}^{xd}) ds' \right\rangle + \left\langle \int_0^\infty F_p(w_{ji}^x, s) x_i(t - s - d_{ji}^{xa}) y_j(t - d_{ji}^{xd}) ds \right\rangle \\ &\cong \int_{-\infty}^\infty F(w_{ji}^x, s) \langle x_i(t - s - d_{ji}^{xa}) y_j(t - d_{ji}^{xd}) \rangle ds \end{aligned}$$

$$\text{where } F(w, s) \equiv \begin{cases} F_p(w, s) & \text{if } (s \geq 0) \\ F_d(w, -s) & \text{if } (s < 0) \end{cases}.$$

Therefore, by calculating the cross-correlation between pre-spikes  $x_i$  and post-spikes  $y_j$ , synaptic weight dynamics can be analytically estimated. Because the spike probability linearly depends on the membrane potential in our model, cross-correlation follows

$$\begin{aligned} \langle x_i(t - s - d_{ji}^{xa}) y_j(t - d_{ji}^{xd}) \rangle &\cong \langle x_i(t - s - d_{ji}^{xa}) u_j^E(t - d_{ji}^{xd}) \rangle \\ &\cong \sum_{l=1}^L w_{jl}^x \int_0^\infty dr \varepsilon_x(r) \langle x_i(t - s - d_{ji}^{xa}) x_l(t - d_{ji}^{xd} - r - d_{jl}^x) \rangle - \sum_{k=1}^N w_{jk}^z \int_0^\infty dr \varepsilon_z(r) \langle x_i(t - s - d_{ji}^{xa}) z_k(t - d_{ji}^{xd} - r - d_{jk}^z) \rangle. \end{aligned}$$

Since we define cross-correlation among input neurons as

$$C_{il}(s) \equiv \langle x_i(t)x_l(t-s) \rangle - \langle x_i(t) \rangle \langle x_l(t) \rangle,$$

the first term is written as

$$\begin{aligned} & \sum_{l=1}^L w_{jl}^X \int_0^\infty dr \varepsilon_X(r) \langle x_i(t-s-d_{ji}^{Xa}) x_l(t-d_{ji}^{Xd}-r-d_{ji}^{Xl}) \rangle \\ & \cong \sum_{l=1}^L w_{jl}^X [(v_o^X)^2 + \int_0^\infty dr \varepsilon_X(r) C_{il}(r-s+2d_{Xd})]. \end{aligned} \quad (26)$$

This result is consistent with that in previous studies [5,38,45]. The analysis can be extended to the cross-correlation between an input neuron and a lateral inhibitory neuron as

$$\begin{aligned} & \langle x_i(t-s-d_{ji}^{Xa}) z_k(t-d_{ji}^{Xd}-r-d_{jk}^{Zl}) \rangle \cong \langle x_i(t-s-d_{ji}^{Xa}) u_k^l(t-d_{ji}^{Xd}-r-d_{jk}^{Zl}) \rangle \\ & \cong \sum_{m=1}^M w_{km}^Y \int_0^\infty dq \varepsilon_Y(q) \langle x_i(t-s-d_{ji}^{Xa}) y_m(t-d_{ji}^{Xd}-r-d_{jk}^{Zl}-q-d_{km}^{Yl}) \rangle \\ & \cong \sum_{m=1}^M w_{km}^Y \sum_{l=1}^L w_{ml}^X \int_0^\infty dq \varepsilon_Y(q) \int_0^\infty dr' \varepsilon_X(r') \\ & \quad \langle x_i(t-s-d_{ji}^{Xa}) x_l(t-d_{ji}^{Xd}-r-d_{jk}^{Zl}-q-d_{km}^{Yl}-r'-d_{ml}^{Xl}) \rangle \\ & - \sum_{m=1}^M w_{km}^Y \sum_{n=1}^N w_{mn}^Z \int_0^\infty dq \varepsilon_Y(q) \int_0^\infty dr' \varepsilon_Z(r') \\ & \quad \langle x_i(t-s-d_{ji}^{Xa}) z_n(t-d_{ji}^{Xd}-r-d_{jk}^{Zl}-q-d_{km}^{Yl}-r'-d_{mn}^{Zl}) \rangle \\ & \cong \sum_{m=1}^M w_{km}^Y \sum_{l=1}^L w_{ml}^X [(v_o^X)^2 + \int_0^\infty dq \varepsilon_Y(q) \int_0^\infty dr' \varepsilon_X(r') C_{il}(r+q+r'-s+2d_{Xd}+d_Z+d_Y)] \\ & - \sum_{m=1}^M w_{km}^Y \sum_{n=1}^N w_{mn}^Z v_o^X v_n^Z. \end{aligned} \quad (27)$$

Theoretically, expansion over a lateral connection should be performed infinite times to obtain the exact solution, but at each expansion, the delay caused by synaptic delay  $d_Z+d_Y$  and EPSP/IPSP rise times is accumulated so that the effect on correlation rapidly becomes small, especially when the original input cross-correlation  $C(t)$  is narrow; however, even if  $C(t)$  is broad, the effect for learning is bounded by the STDP time window. Therefore, higher order terms practically influence weight dynamics only through firing rates, so that by applying the approximation

$$\int_0^\infty dq \varepsilon_Y(q) \int_0^\infty dr' \varepsilon_Z(r') \langle x_i(t-s-d_{ji}^{Xa}) z_n(t-d_{ji}^{Xd}-r-d_{jk}^{Zl}-q-d_{km}^{Yl}-r'-d_{mn}^{Zl}) \rangle \cong v_o^X v_n^Z,$$

the last term can be obtained. In general,  $v_n^Z$  is not analytically calculable, but by considering

the balanced condition, it can be estimated. Therefore, the second term is given as

$$\begin{aligned} & \sum_{k=1}^N w_{jk}^Z \int_0^\infty dr \varepsilon_Z(r) \langle x_i(t-s-d_{ji}^{Xa}) z_k(t-d_{ji}^{Xd}-r-d_{jk}^Z) \rangle \\ & \cong \sum_{k=1}^N w_{jk}^Z \sum_{m=1}^M w_{km}^Y \sum_{l=1}^L w_{ml}^X \\ & \quad \left[ (v_o^X)^2 + \int_0^\infty dr \varepsilon_Z(r) \int_0^\infty dq \varepsilon_Y(q) \int_0^\infty dr' \varepsilon_X(r') C_{il}(r+q+r'-s+2d_{Xd}+d_Z+d_Y) \right] \\ & \quad - \sum_{k=1}^N w_{jk}^Z \sum_{m=1}^M w_{km}^Y \sum_{n=1}^N w_{mn}^Z v_o^X v_n^Z \end{aligned}$$

Therefore, if we denote

$$\begin{aligned} \Gamma_{il}^{X1}(w_{ji}^X) & \equiv \int_{-\infty}^\infty ds F(w_{ji}^X, s) \int_0^\infty dr \varepsilon_X(r) C_{il}(r-s+2d_{Xd}) \\ \Gamma_{il}^{X2}(w_{ji}^X) & \equiv \int_{-\infty}^\infty ds F(w_{ji}^X, s) \int_0^\infty dr \varepsilon_Z(r) \int_0^\infty dq \varepsilon_Y(q) \int_0^\infty dr' \varepsilon_X(r') \\ & \quad C_{il}(r+q+r'-s+2d_{Xd}+d_Z+d_Y) \\ \bar{F}(w_{ji}^X) & \equiv \int_{-\infty}^\infty F(w_{ji}^X, s) ds, \end{aligned} \tag{28}$$

average synaptic weight dynamics satisfy

$$\begin{aligned} \left\langle \frac{dw_{ji}^X}{dt} \right\rangle & \cong \sum_{l=1}^L w_{jl}^X \Gamma_{il}^{X1}(w_{ji}^X) - \sum_{k=1}^N w_{jk}^Z \sum_{m=1}^M w_{km}^Y \sum_{l=1}^L w_{ml}^X \Gamma_{il}^{X2}(w_{ji}^X) \\ & \quad + \bar{F}(w_{ji}^X) \left[ \sum_{l=1}^L w_{jl}^X (v_o^X)^2 - \sum_{k=1}^N w_{jk}^Z \sum_{m=1}^M w_{km}^Y \sum_{l=1}^L w_{ml}^X (v_o^X)^2 + \sum_{k=1}^N w_{jk}^Z \sum_{m=1}^M w_{km}^Y \sum_{n=1}^N w_{mn}^X v_o^X v_n^Z \right]. \end{aligned} \tag{29}$$

The first two terms are Hebbian terms that depend on correlation by  $\Gamma^{X1}$  and  $\Gamma^{X2}$ , whereas the remainders are homeostatic terms. In all terms, synaptic weight dependence is primarily caused by  $w_{ji}^X$  and not by other synapses.

By inserting the explicit representation of correlation into the equation above,  $\Gamma^{X1}$  and  $\Gamma^{X2}$  can be rewritten as

$$\begin{aligned} \Gamma_{il}^{X1}(w_{ji}^X) & = v_o^S G_1^X(w_{ji}^X) \sum_{\mu=1}^p q_{i\mu} q_{l\mu}, \quad \Gamma_{il}^{X2}(w_{ji}^X) = v_o^S G_2^X(w_{ji}^X) \sum_{\mu=1}^p q_{i\mu} q_{l\mu}, \\ G_1^X(w_{ji}^X) & \equiv \int_{-\infty}^\infty ds F(w_{ji}^X, s) \int_0^\infty dr \varepsilon_X(r) \int_{\max(0, r-s+2d_{Xd})}^\infty dt' \phi(t') \phi(t' - (r-s+2d_{Xd})), \\ G_2^X(w_{ji}^X) & \equiv \int_{-\infty}^\infty ds F(w_{ji}^X, s) \int_0^\infty dr \varepsilon_Z(r) \int_0^\infty dq \varepsilon_Y(q) \int_0^\infty dr' \varepsilon_X(r') \\ & \quad \int_{\max(0, r')}^\infty dt' \phi(t') \phi(t' - (r+q+r'-s+2d_{Xd}+d_Z+d_Y)) \end{aligned} \tag{30}$$

where  $t'' = r+q+r'-s+2d_{Xd}+d_Z+d_Y$ . Note that  $G_1^X$  and  $G_2^X$  do not depend on any indexes of the neurons, except for synaptic weight dependency, and so the two values are considered basic constants that decide how correlation shapes learning.



If we ignore the homeostatic term, then the synaptic weight dynamic is written in the matrix form as  $\dot{W}_X \approx (W_X C^t) \cdot G_1^X - (W_X W_Z W_Y C^t) \cdot G_2^X$ , where the dot product is defined as  $(A \cdot B)_{ij} = A_{ij} B_{ij}$ . Especially if we approximate  $G_1^X$  and  $G_2^X$  with  $g_1^X \equiv G_1^X(w_o^X)$  and  $g_2^X \equiv G_2^X(w_o^X)$  (or if weight dependence is negligible as in additive-STDP),  $\dot{W}_X \approx W_X (g_1^X E - g_2^X W_Z W_Y) C^t$ .

The correlation kernel  $\chi_1^X$  was derived from

$$\begin{aligned} G_1^X(w_{ji}^X) &= \int_{-\infty}^{\infty} ds F(w_{ji}^X, s) \int_0^{\infty} dr \varepsilon_X(r) \int_{\max(0, r-s+2d_{Xd})}^{\infty} dt' \int_{-\infty}^{\infty} d\tau \phi(t') \phi(t' - (r - s + 2d_{Xd})) \\ &\quad \delta(\tau - (r - s + 2d_{Xd})) \\ &= \int_{-\infty}^{\infty} d\tau \int_{-\tau+2d_{Xd}}^{\infty} ds F(w_{ji}^X, s) \varepsilon_X(s - 2d_{Xd}) \int_{\max(0, \tau)}^{\infty} dt' \phi(t') \phi(t' - \tau) \\ &= \int_{-\infty}^{\infty} \chi_1^X(\tau; w_{ji}^X) h(\tau) d\tau \end{aligned} \tag{31}$$

where  $\chi_1^X(\tau; w) = \int_{-\tau+2d_{Xd}}^{\infty} ds F(w, s) \varepsilon_X(\tau + s - 2d_{Xd})$ , and  $h(\tau; \theta_i) \equiv \int_{\max(\tau, 0)}^{\infty} dt' \phi(t') \phi(t' - \tau)$ .

The second correlation kernel  $\chi_2^X$  was calculated in a similar way.

**Mean-field approximation of a two-source model.** If the correlation structure  $C(s)$  is simply organized, further analytical consideration is possible. In the two-source model shown in Fig 2A, lateral connections are structured non-reciprocally, and EPSP/IPSP sizes are constants. The synaptic weight matrices are written as

$$W_{km}^Y = \begin{cases} w_Y & (\text{if } \lfloor k/N_a \rfloor = \lfloor m/M_a \rfloor) \\ 0 & (\text{otherwise}) \end{cases}, W_{jk}^Z = \begin{cases} w_Z & (\text{if } \lfloor j/M_a \rfloor \neq \lfloor k/N_a \rfloor) \\ 0 & (\text{otherwise}) \end{cases}. \tag{32}$$

Therefore, the original  $L \times M$  differential equations can be reduced into  $2 \times 2$  equations of representative neurons as

$$\begin{aligned} \frac{dw_{\mu\nu}^X}{dt} &\cong \sum_{\nu'=1}^{L/L_a} L_a w_{\mu\nu'}^X v_o^S G^X(w_{\nu\nu'}) \sum_{\rho} q_{\nu\rho} q_{\nu'\rho} - N_a w_Z M_a w_Y \sum_{\nu'=1}^{L/L_a} L_a w_{\mu\nu'}^X v_o^S G^Y(w_{\nu\nu'}) \sum_{\rho} q_{\nu\rho} q_{\nu'\rho} \\ &\quad + \bar{F}(w_{\mu\nu}^X) \left[ (v_o^X)^2 \sum_{\nu'=1}^{L/L_a} L_a w_{\mu\nu'}^X - (v_o^X)^2 N_a w_Z M_a w_Y \sum_{\nu'=1}^{L/L_a} L_a w_{\mu\nu'}^X + (N_a w_Z)^2 M_a w_Y v_o^X v_{\mu}^Z \right]. \end{aligned} \tag{33}$$

The firing rates of inhibitory neurons can be approximated as

$$v_{\mu}^Z \cong \frac{1}{N_a} \sum_{k \in \Omega_{\mu}^Z} u_k^I \cong M_a w_Y v_{\mu}^Y \cong M_a w_Y ((L_a w_{\mu A} + L_a w_{\mu B} + 2L_a w_o^X) v_o^X - N_a w_Z v_{\mu}^Z). \tag{34}$$

Therefore, by solving the simultaneous equations for  $v_1^Z$  and  $v_2^Z$ ,

$$\begin{aligned} v_1^Z &= \frac{M_a w_Y v_o^X}{1 - (M_a w_Y N_a w_Z)^2} [(L_a w_{1A} + L_a w_{1B} + 2L_a w_o^X) - (M_a w_Y N_a w_Z)(L_a w_{2A} + L_a w_{2B} + 2L_a w_o^X)] \\ v_2^Z &= \frac{M_a w_Y v_o^X}{1 - (M_a w_Y N_a w_Z)^2} [(L_a w_{2A} + L_a w_{2B} + 2L_a w_o^X) - (M_a w_Y N_a w_Z)(L_a w_{1A} + L_a w_{1B} + 2L_a w_o^X)] \end{aligned}$$

This analytical approach is applicable only when the synaptic weight change is sufficiently slow relative to the neural dynamics. Also, because we ignored the variance in the synaptic weights, numerically the accuracy is limited.

**Analytic approach for STDP in lateral and inhibitory connections.** Using a similar calculation as above, synaptic weight development of the lateral connections is given as

$$\dot{W}_Y \approx g_1^Y W_Y W_X C^t W_X^t - g_2^Y W_Y W_X C^t W_X^t W_Y^t W_z^t - g_3^Y W_Y W_Z W_Y W_X C^t W_X^t, \quad (35)$$

where

$$\begin{aligned} g_1^Y &\equiv \int_{-\infty}^{\infty} ds F^Y(s) \int D_r^X \int D_u^Y \int D_r^X h(u + r' - s - r) \\ g_2^Y &\equiv \int_{-\infty}^{\infty} ds F^Y(s) \int D_r^X \int D_u^Y \int D_q^Z \int D_u^Y \int D_r^X h(u' + r' - s - q - u - r - d_Y - d_Z) \\ g_3^Y &\equiv \int_{-\infty}^{\infty} ds F^Y(s) \int D_r^X \int D_u^Y \int D_q^Z \int D_u^Y \int D_r^X h(u + q + u' + r' + d_Y + d_Z - s - r), \end{aligned}$$

where  $\int D_r^X \equiv \int_0^{\infty} dr \varepsilon_X(r)$ . The meaning of these equations is made clear by summarizing the correlation propagation in the diagrams (S2D i–iii Fig). In the diagram, blue wavy lines represent intrinsic correlation, and arrows are synaptic connections. To estimate how a blue correlation influences STDP at a red arrow, we need to determine all the major trajectories in which the correlation reaches pre- and postsynaptic neurons. In the linear Poisson framework, for a given trajectory, the propagation of a correlation is calculated by simply using integrals as above. From this diagram, we can safely assume that  $g_2^Y$  and  $g_3^Y$  are negligibly smaller than  $g_1^Y$ , because trajectories (ii) and (iii) are secondary correlations and also contain synaptic delays. In this approximation, we additionally assume that

$$C = \begin{pmatrix} c_s & 0 \\ 0 & c_s \end{pmatrix}, W_X = \begin{pmatrix} w_s^X & w_w^X \\ w_w^X & w_s^X \end{pmatrix}.$$

Then,

$$\begin{aligned} \frac{d}{dt} \begin{pmatrix} w_{11}^Y \\ w_{12}^Y \\ w_{21}^Y \\ w_{22}^Y \end{pmatrix} &\approx \begin{pmatrix} A^L & B^L & 0 & 0 \\ B^L & A^L & 0 & 0 \\ 0 & 0 & A^L & B^L \\ 0 & 0 & B^L & A^L \end{pmatrix} \begin{pmatrix} w_{11}^Y \\ w_{12}^Y \\ w_{21}^Y \\ w_{22}^Y \end{pmatrix} \\ A^L &\equiv c_s g_1^Y ((w_s^X)^2 + (w_w^X)^2), B^L \equiv 2c_s g_1^Y w_s^X w_w^X \end{aligned}$$

Therefore,  $(w_{11}^Y, w_{12}^Y, w_{21}^Y, w_{22}^Y) \propto (+1, -1, -1, +1)$  is an eigenvector of the transition matrix, and the eigenvalue is  $c_s g_1^Y (w_s^X - w_w^X)^2$ . Because the eigenvector develops by  $\exp[c_s g_1^Y (w_s^X - w_w^X)^2 t]$ , when  $g_1^Y$  is positive, the E-to-I connections are more likely to be structured in a way that the inhibitory neurons become feature selective. On the other hand, if that value is negative, such structure may not be obtained. Note that  $(1, -1, -1, 1)$  is not the principal eigenvector in this simple analysis, because the eigensystem of the matrix is  $\{A^L + B^L, A^L - B^L, A^L - B^L, A^L + B^L\}; \{1, 1, 0, 0\}, \{0, 0, 1, 1\}, \{1, -1, 0, 0\}, \{0, 0, 1, -1\}$ .

Similarly, for inhibitory connections

$$\begin{aligned} \dot{W}_Z &\approx g_1^Z W_X C W_X^t W_Y^t - g_2^Z W_Z W_Y W_X C W_X^t W_Y^t \\ g_1^Z &\equiv \int_{-\infty}^{\infty} F^Z(s) \int D_r^X \int D_u^Y \int D_r^X h(r-s-u-r'-d_Z-d_Y) \\ g_2^Z &\equiv \int_{-\infty}^{\infty} F^Z(s) \int D_q^Z \int D_u^Y \int D_r^X \int D_{u'}^Y \int D_r^X h(r+u+q-s-u'-r'). \end{aligned} \tag{36}$$

We approximated with only two terms because the third term is negligible (S2D iv-vi Fig).

If we assume  $W_Y = \begin{pmatrix} w_d^Y & w_r^Y \\ w_r^Y & w_d^Y \end{pmatrix}$ , and  $g_2^Z = 0$ , then the synaptic weight change follows

$\Delta w_{11}^Z - \Delta w_{12}^Z = \Delta w_{22}^Z - \Delta w_{21}^Z = c_s g_1^Z (w_s^X - w_s^X)^2 (w_d^Y - w_r^Y)$ . This means that if  $g_1^Z$  is positive, reciprocal connections are enhanced (or inhibitory connections to the neurons coding a similar feature are enhanced), whereas for negative  $g_1^Z$ , inhibitory connections develop non-reciprocally (i.e., lateral connections function as mutual inhibition between output excitatory neuron groups).

We have restricted our consideration to Hebbian STDP, but the properties of STDP on E-to-I and I-to-E connections are still debatable [58,59]. Although it is difficult to study all combinations of STDPs, we still provide analytical insights by investigating the behaviors of  $g_1^Y$  and  $g_1^Z$ . S2E Fig shows the behaviors of four different types of STDP. This indicates that the anti-correlation detector type of E-to-I STDP [53] tends to cause non-feature-selective lateral connections. In addition, under the anti-coincidence detector type of I-to-E STDP [55], mutual inhibition structures would be preferred; however, the implication of our analytical method is limited, and further study will be necessary to fully understand the functions of the various types of STDP.

## Evaluation of the performance

**Cross-correlation.** We evaluated the performance by measuring the mean cross-correlation between the external sources and population activity of the output neurons. For time bin  $\Delta t = 10$  ms, the activity of source  $\mu$  is defined as  $s_\mu^k = \frac{1}{\Delta t} \int_{k\Delta t}^{(k+1)\Delta t} s_\mu(t) dt$ , and, similarly, the population activity of the output neuron group  $v$  is  $y_v^k(\tau_D) = \sum_{j \in \Omega_v^Y} \frac{1}{\Delta t} \int_{k\Delta t}^{(k+1)\Delta t} y_j(t + \tau_D) dt$ , where  $\Omega_v^Y$  is a set of output neurons coding a source  $v$ . For these, cross-correlation is defined as

$$c_{\mu v}(\tau_D) \equiv \frac{1}{\sigma_\mu^s \sigma_v^y} \sum_{k=1}^{T_c/\Delta t} (s_\mu^k - \bar{s}_\mu)(y_v^k - \bar{y}_v),$$

where  $\bar{s}_\mu \equiv \frac{1}{T_c} \int_{T_o}^{T_o+T_c} s_\mu(t) dt$ ,  $\bar{y}_v \equiv \frac{1}{T_c} \int_{T_o}^{T_o+T_c} y_v(t) dt$ ,  $\sigma_\mu^s \equiv \sqrt{\sum_{k=1}^{T_c/\Delta t} (s_\mu^k - \bar{s}_\mu)^2}$ , and

$\sigma_v^y \equiv \sqrt{\sum_{k=1}^{T_c/\Delta t} (y_v^k - \bar{y}_v)^2}$ . We used  $T_c = 10$  ms for the analysis. Correspondence between

sources and output groups are arbitrary, and so the learned correlation should be given

as  $c(\tau_D) \equiv \max_{\psi} \frac{1}{p} \sum_{\mu=1}^p c_{\mu\psi(\mu)}(\tau_D)$  for all the  $p!$  number of combinations with function  $\Psi$

between sources and output groups. For example, when

$p = 2, c(\tau_D) = \max\{\frac{1}{2}[c_{A1}(\tau_D) + c_{B2}(\tau_D)], \frac{1}{2}[c_{A2}(\tau_D) + c_{B1}(\tau_D)]\}$ . Although, in reality, supervised or reinforcement learning is necessary to perform this readout, for simplicity we did not implement readout neurons explicitly. In Fig 2F, we plotted  $\max_v c_{Bv}(\tau_D)$  for the minor source B.

For the models with randomly connected lateral inhibition and (e+i) STDP, we defined output neuron  $j$  as belonging to  $\Omega_\mu^Y$  if

$$\frac{1}{|\Omega_\mu^X|} \sum_{i \in \Omega_\mu^X} w_{j\mu}^X > \alpha_{th} \max_{v \neq \mu} \left\{ \frac{1}{|\Omega_v^X|} \sum_{i \in \Omega_v^X} w_{jv}^X \right\}$$

for  $\alpha_{th} = 1.5$ , and the cross-correlation was calculated based on  $\Omega_\mu^Y$ .

**Mutual information.** Based on the discretized hidden external source/output neuron activity  $s_\mu^k, y_v^k$ , we defined the binary variables

$$\hat{s}_\mu^k \equiv \begin{cases} 1 & (\text{if } s_\mu^k > \bar{s}_\mu^k + \sigma_\mu^s) \\ 0 & (\text{otherwise}) \end{cases}, \hat{y}_v^k \equiv \begin{cases} 1 & (\text{if } y_v^k > \bar{y}_v^k + \sigma_v^y) \\ 0 & (\text{otherwise}) \end{cases}$$

Based on these variables, the states at time  $k$  can be defined as  $\hat{s}^k \equiv (\hat{s}_1^k, \dots, \hat{s}_p^k), \hat{y}^k \equiv (\hat{y}_1^k, \dots, \hat{y}_p^k)$ . Therefore, the probability that the external state takes one particular state is

$$p_s(\hat{s} = \hat{s}') \equiv \frac{1}{T_c/\Delta t} \sum_{k=1}^{T_c/\Delta t} [\hat{s}^k = \hat{s}']_{tof}, \text{ where } [X]_{tof} \text{ takes 1 if X is true, otherwise it takes 0, for the}$$

statement X. Therefore, mutual information can be defined as

$$MI \equiv \sum_{\hat{s}'} \sum_{\hat{y}'} p_{sy}(\hat{s} = \hat{s}', \hat{y} = \hat{y}') \log_2 \left( \frac{p_{sy}(\hat{s} = \hat{s}', \hat{y} = \hat{y}')}{p_s(\hat{s} = \hat{s}') p_y(\hat{y} = \hat{y}')} \right).$$

## Supporting Information

**S1 Fig. Simulations with the leaky integrate-and-fire model.** (A) Synaptic weight developments at the feedforward connection. (B) Cross-correlation and mutual information calculated for various delays. Both values were calculated by averaging five independent simulation results. (C) Development of two values for the simulation shown in (A). (D) PSTH of the membrane potential calculated for gray areas in (A). (E) Peristimulus time histogram (PSTH) of the firing probability for the same simulation. (EPS)

**S2 Fig. Spike-timing-dependent plasticity (STDP) at lateral connections shapes network structure.** (A, B) Synaptic weight development when the number of external inputs is three (A) and four (B). Thick lines represent averages over all synapses, and thin lines represent individual synaptic weights. Colors represent detected sources for output neurons (left) and inhibitory neurons (middle right). (C) Relationship between the number of inhibitory neurons and the lateral structure. (D) Propagation of structure. i to iii correspond to lateral excitatory connections, and iv to vi correspond to feedback inhibitory connections. (E) Analytic results for various types of STDP. (EPS)

**S3 Fig. The effects of noise in the model with exponential correlation kernel.** (A) Cross-correlations among input neurons responding to the same source calculated from simulated

data for three different correlation timescale parameters  $\theta_t$ . Note that in Fig 3, we used  $\theta_t = 0.5, 2.0, 4.0$  ms, while here we used  $\theta_t = 1.0, 3.0, 5.0$ ms. (B, C) The correlation kernels  $g_{e1}^X, g_{e2}^X$  (B) and their ratio  $g_{e1}^X/g_{e2}^X$  (C) are shown for the kernels  $g_{e1}^X$  and  $g_{e2}^X$  that were calculated from Eq (30) with  $\phi_e(t) = e^{-t/\theta_t}/\theta_t$ . (D,E) The effects of random noise (D) and crosstalk noise (E) at various correlation timescales.

(EPS)

**S1 Auditory File. Demonstration of blind source separation.** 0'00"–0'43": Independent auditory signals (10 s each); 0'44"–0'54": Mixed auditory signal; 0'55"–1'38": Decoded independent signals.

(MP3)

## Acknowledgments

We thank Matthieu Gilson and Florence Kleberg for helpful discussions and comments on the manuscript.

## Author Contributions

Conceived and designed the experiments: NH TF. Performed the experiments: NH. Analyzed the data: NH. Wrote the paper: NH TF.

## References

1. Markram H, Lübke J, Frotscher M, Sakmann B. Regulation of synaptic efficacy by coincidence of post-synaptic APs and EPSPs. *Science*. 1997; 275: 213–215. PMID: [8985014](#)
2. Bi GQ, Poo MM. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J Neurosci Off J Soc Neurosci*. 1998; 18: 10464–10472. PMID: [9852584](#)
3. Gerstner W, Kempter R, van Hemmen JL, Wagner H. A neuronal learning rule for sub-millisecond temporal coding. *Nature*. 1996; 383: 76–81. doi: [10.1038/383076a0](#) PMID: [8779718](#)
4. Song S, Miller KD, Abbott LF. Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nat Neurosci*. 2000; 3: 919–926. doi: [10.1038/78829](#) PMID: [10966623](#)
5. Gütig R, Aharonov R, Rotter S, Sompolinsky H. Learning input correlations through nonlinear temporally asymmetric Hebbian plasticity. *J Neurosci Off J Soc Neurosci*. 2003; 23: 3697–3714.
6. Legenstein R, Naeger C, Maass W. What can a neuron learn with spike-timing-dependent plasticity? *Neural Comput*. 2005; 17: 2337–2382. doi: [10.1162/0899766054796888](#) PMID: [16156932](#)
7. Gilson M, Fukai T, Burkitt AN. Spectral analysis of input spike trains by spike-timing-dependent plasticity. *PLoS Comput Biol*. 2012; 8: e1002584. doi: [10.1371/journal.pcbi.1002584](#) PMID: [22792056](#)
8. Arevian AC, Kapoor V, Urban NN. Activity-dependent gating of lateral inhibition in the mouse olfactory bulb. *Nat Neurosci*. 2008; 11: 80–87. doi: [10.1038/nn2030](#) PMID: [18084286](#)
9. Lee S-H, Kwan AC, Zhang S, Phoumthippavong V, Flannery JG, Masmanidis SC, et al. Activation of specific interneurons improves V1 feature selectivity and visual perception. *Nature*. 2012; 488: 379–383. doi: [10.1038/nature11312](#) PMID: [22878719](#)
10. Adesnik H, Scanziani M. Lateral competition for cortical space by layer-specific horizontal circuits. *Nature*. 2010; 464: 1155–1160. doi: [10.1038/nature08935](#) PMID: [20414303](#)
11. Couey JJ, Witoelar A, Zhang S-J, Zheng K, Ye J, Dunn B, et al. Recurrent inhibitory circuitry as a mechanism for grid formation. *Nat Neurosci*. 2013; 16: 318–324. doi: [10.1038/nn.3310](#) PMID: [23334580](#)
12. Amari S. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biol Cybern*. 1977; 27: 77–87. PMID: [911931](#)
13. Wiechert MT, Judkewitz B, Riecke H, Friedrich RW. Mechanisms of pattern decorrelation by recurrent neuronal circuits. *Nat Neurosci*. 2010; 13: 1003–1010. doi: [10.1038/nn.2591](#) PMID: [20581841](#)
14. Molgedey L, Schuster HG. Separation of a mixture of independent signals using time delayed correlations. *Phys Rev Lett*. 1994; 72: 3634–3637. PMID: [10056251](#)



15. Fukai T, Tanaka S. A simple neural network exhibiting selective activation of neuronal ensembles: from winner-take-all to winners-share-all. *Neural Comput.* 1997; 9: 77–97. PMID: [9117902](#)
16. Bartsch AP, van Hemmen JL. Combined Hebbian development of geniculocortical and lateral connectivity in a model of primary visual cortex. *Biol Cybern.* 2001; 84: 41–55. PMID: [11204398](#)
17. Wensch OG, Noll J, van Hemmen JL. Spontaneously emerging direction selectivity maps in visual cortex through STDP. *Biol Cybern.* 2005; 93: 239–247. doi: [10.1007/s00422-005-0006-z](#) PMID: [16195915](#)
18. Savin C, Joshi P, Triesch J. Independent component analysis in spiking neurons. *PLoS Comput Biol.* 2010; 6: e1000757. doi: [10.1371/journal.pcbi.1000757](#) PMID: [20421937](#)
19. King PD, Zylberberg J, DeWeese MR. Inhibitory interneurons decorrelate excitatory cells to drive sparse code formation in a spiking model of V1. *J Neurosci Off J Soc Neurosci.* 2013; 33: 5475–5485. doi: [10.1523/JNEUROSCI.4188-12.2013](#)
20. Masquelier T, Guyonneau R, Thorpe SJ. Competitive STDP-based spike pattern learning. *Neural Comput.* 2009; 21: 1259–1276. doi: [10.1162/neco.2008.06-08-804](#) PMID: [19718815](#)
21. Masquelier T, Guyonneau R, Thorpe SJ. Spike timing dependent plasticity finds the start of repeating patterns in continuous spike trains. *PLoS One.* 2008; 3: e1377. doi: [10.1371/journal.pone.0001377](#) PMID: [18167538](#)
22. Clopath C, Büsing L, Vasilaki E, Gerstner W. Connectivity reflects coding: a model of voltage-based STDP with homeostasis. *Nat Neurosci.* 2010; 13: 344–352. doi: [10.1038/nn.2479](#) PMID: [20098420](#)
23. Wilson RI, Mainen ZF. Early events in olfactory processing. *Annu Rev Neurosci.* 2006; 29: 163–201. doi: [10.1146/annurev.neuro.29.051605.112950](#) PMID: [16776583](#)
24. Niessing J, Friedrich RW. Olfactory pattern classification by discrete neuronal network states. *Nature.* 2010; 465: 47–52. doi: [10.1038/nature08961](#) PMID: [20393466](#)
25. Rokni D, Hemmelder V, Kapoor V, Murthy VN. An olfactory cocktail party: figure-ground segregation of odorants in rodents. *Nat Neurosci.* 2014; 17: 1225–1232. doi: [10.1038/nn.3775](#) PMID: [25086608](#)
26. McDermott JH. The cocktail party problem. *Curr Biol CB.* 2009; 19: R1024–1027. doi: [10.1016/j.cub.2009.09.005](#) PMID: [19948136](#)
27. Rodgers CC, DeWeese MR. Neural correlates of task switching in prefrontal cortex and primary auditory cortex in a novel stimulus selection task for rodents. *Neuron.* 2014; 82: 1157–1170. doi: [10.1016/j.neuron.2014.04.031](#) PMID: [24908492](#)
28. Cherry EC. Some Experiments on the Recognition of Speech, with One and with Two Ears. *J Acoust Soc Am.* 1953; 25: 975–979. doi: [10.1121/1.1907229](#)
29. Haykin S, Chen Z. The Cocktail Party Problem. *Neural Comput.* 2005; 17: 1875–1902. doi: [10.1162/0899766054322964](#) PMID: [15992485](#)
30. Hofer SB, Ko H, Pichler B, Vogelstein J, Ros H, Zeng H, et al. Differential connectivity and response dynamics of excitatory and inhibitory neurons in visual cortex. *Nat Neurosci.* 2011; 14: 1045–1052. doi: [10.1038/nn.2876](#) PMID: [21765421](#)
31. Gilson M, Fukai T. Stability versus neuronal specialization for STDP: long-tail weight distributions solve the dilemma. *PLoS One.* 2011; 6: e25339. doi: [10.1371/journal.pone.0025339](#) PMID: [22003389](#)
32. Song S, Sjöström PJ, Reigl M, Nelson S, Chklovskii DB. Highly Nonrandom Features of Synaptic Connectivity in Local Cortical Circuits. *PLoS Biol.* 2005; 3: e68. doi: [10.1371/journal.pbio.0030068](#) PMID: [15737062](#)
33. Buzsáki G, Mizuseki K. The log-dynamic brain: how skewed distributions affect network operations. *Nat Rev Neurosci.* 2014; 15: 264–278. doi: [10.1038/nrn3687](#) PMID: [24569488](#)
34. von der Malsburg C (1994) The Correlation Theory of Brain Function. In: Domany PE, Hemmen PDJL van, Schulten PK, editors. *Models of Neural Networks. Physics of Neural Networks.* Springer New York. pp. 95–119.
35. Kumar A, Rotter S, Aertsen A. Spiking activity propagation in neuronal networks: reconciling different perspectives on neural coding. *Nat Rev Neurosci.* 2010; 11: 615–627. doi: [10.1038/nrn2886](#) PMID: [20725095](#)
36. Lampl I, Reichova I, Ferster D. Synchronous Membrane Potential Fluctuations in Neurons of the Cat Visual Cortex. *Neuron.* 1999; 22: 361–374. doi: [10.1016/S0896-6273\(00\)81096-X](#) PMID: [10069341](#)
37. Bair W, Zohary E, Newsome WT. Correlated Firing in Macaque Visual Area MT: Time Scales and Relationship to Behavior. *J Neurosci.* 2001; 21: 1676–1697. PMID: [11222658](#)
38. Gilson M, Burkitt AN, Grayden DB, Thomas DA, van Hemmen JL. Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks. I. Input selectivity—strengthening correlated input pathways. *Biol Cybern.* 2009; 101: 81–102. doi: [10.1007/s00422-009-0319-4](#) PMID: [19536560](#)

39. Gilson M, Burkitt AN, Grayden DB, Thomas DA, van Hemmen JL. Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks. II. Input selectivity—symmetry breaking. *Biol Cybern.* 2009; 101: 103–114. doi: [10.1007/s00422-009-0320-y](https://doi.org/10.1007/s00422-009-0320-y) PMID: [19536559](https://pubmed.ncbi.nlm.nih.gov/19536559/)
40. Oja E. Simplified neuron model as a principal component analyzer. *J Math Biol.* 1982; 15: 267–273. doi: [10.1007/BF00275687](https://doi.org/10.1007/BF00275687) PMID: [7153672](https://pubmed.ncbi.nlm.nih.gov/7153672/)
41. Van Rossum MCW, Turrigiano GG. Correlation based learning from spike timing dependent plasticity. *Neurocomputing.* 2001;38–40: 409–415. doi: [10.1016/S0925-2312\(01\)00360-5](https://doi.org/10.1016/S0925-2312(01)00360-5)
42. Alonso J-M, Usrey WM, Reid RC. Precisely correlated firing in cells of the lateral geniculate nucleus. *Nature.* 1996; 383: 815–819. doi: [10.1038/383815a0](https://doi.org/10.1038/383815a0) PMID: [8893005](https://pubmed.ncbi.nlm.nih.gov/8893005/)
43. Mainen ZF, Sejnowski TJ. Reliability of spike timing in neocortical neurons. *Science.* 1995; 268: 1503–1506. doi: [10.1126/science.7770778](https://doi.org/10.1126/science.7770778) PMID: [7770778](https://pubmed.ncbi.nlm.nih.gov/7770778/)
44. Kohn A, Smith MA. Stimulus Dependence of Neuronal Correlation in Primary Visual Cortex of the Macaque. *J Neurosci.* 2005; 25: 3661–3673. doi: [10.1523/JNEUROSCI.5106-04.2005](https://doi.org/10.1523/JNEUROSCI.5106-04.2005) PMID: [15814797](https://pubmed.ncbi.nlm.nih.gov/15814797/)
45. Kempter R, Gerstner W, van Hemmen JL. Hebbian learning and spiking neurons. *Phys Rev E.* 1999; 59: 4498–4514. doi: [10.1103/PhysRevE.59.4498](https://doi.org/10.1103/PhysRevE.59.4498)
46. Roberts SJ. Independent component analysis: source assessment and separation, a Bayesian approach. *Vis Image Signal Process IEE Proc-.* 1998; 145: 149–154. doi: [10.1049/ip-vis:19981928](https://doi.org/10.1049/ip-vis:19981928)
47. Knuth KH (2002) A Bayesian approach to source separation. arXiv:physics/0205032.
48. Schreiner CE, Read HL, Sutter ML. Modular Organization of Frequency Integration in Primary Auditory Cortex. *Annu Rev Neurosci.* 2000; 23: 501–529. doi: [10.1146/annurev.neuro.23.1.501](https://doi.org/10.1146/annurev.neuro.23.1.501) PMID: [10845073](https://pubmed.ncbi.nlm.nih.gov/10845073/)
49. deCharms RC, Merzenich MM. Primary cortical representation of sounds by the coordination of action-potential timing. *Nature.* 1996; 381: 610–613. doi: [10.1038/381610a0](https://doi.org/10.1038/381610a0) PMID: [8637597](https://pubmed.ncbi.nlm.nih.gov/8637597/)
50. Atencio CA, Schreiner CE. Spectrotemporal Processing Differences between Auditory Cortical Fast-Spiking and Regular-Spiking Neurons. *J Neurosci.* 2008; 28: 3897–3910. doi: [10.1523/JNEUROSCI.5366-07.2008](https://doi.org/10.1523/JNEUROSCI.5366-07.2008) PMID: [18400888](https://pubmed.ncbi.nlm.nih.gov/18400888/)
51. Dan Y, Alonso J-M, Usrey WM, Reid RC. Coding of visual information by precisely correlated spikes in the lateral geniculate nucleus. *Nat Neurosci.* 1998; 1: 501–507. doi: [10.1038/2217](https://doi.org/10.1038/2217) PMID: [10196548](https://pubmed.ncbi.nlm.nih.gov/10196548/)
52. Toyama K, Kimura M, Tanaka K. Cross-Correlation Analysis of Interneuronal Connectivity in cat visual cortex. *J Neurophysiol.* 1981; 46: 191–201. PMID: [6267211](https://pubmed.ncbi.nlm.nih.gov/6267211/)
53. Lu J, Li C, Zhao J-P, Poo M, Zhang X. Spike-timing-dependent plasticity of neocortical excitatory synapses on inhibitory interneurons depends on target cell type. *J Neurosci Off J Soc Neurosci.* 2007; 27: 9711–9720. doi: [10.1523/JNEUROSCI.2513-07.2007](https://doi.org/10.1523/JNEUROSCI.2513-07.2007)
54. Fino E, Paille V, Deniau J-M, Venance L. Asymmetric spike-timing dependent plasticity of striatal nitric oxide-synthase interneurons. *Neuroscience.* 2009; 160: 744–754. doi: [10.1016/j.neuroscience.2009.03.015](https://doi.org/10.1016/j.neuroscience.2009.03.015) PMID: [19303912](https://pubmed.ncbi.nlm.nih.gov/19303912/)
55. Woodin MA, Ganguly K, Poo M. Coincident pre- and postsynaptic activity modifies GABAergic synapses by postsynaptic changes in Cl<sup>-</sup> transporter activity. *Neuron.* 2003; 39: 807–820. PMID: [12948447](https://pubmed.ncbi.nlm.nih.gov/12948447/)
56. Haas JS, Nowotny T, Abarbanel HDI. Spike-timing-dependent plasticity of inhibitory synapses in the entorhinal cortex. *J Neurophysiol.* 2006; 96: 3305–3313. doi: [10.1152/jn.00551.2006](https://doi.org/10.1152/jn.00551.2006) PMID: [16928795](https://pubmed.ncbi.nlm.nih.gov/16928795/)
57. Yazaki-Sugiyama Y, Kang S, Câteau H, Fukai T, Hensch TK. Bidirectional plasticity in fast-spiking GABA circuits by visual experience. *Nature.* 2009; 462: 218–221. doi: [10.1038/nature08485](https://doi.org/10.1038/nature08485) PMID: [19907494](https://pubmed.ncbi.nlm.nih.gov/19907494/)
58. Lamsa KP, Kullmann DM, Woodin MA. Spike-timing dependent plasticity in inhibitory circuits. *Front Synaptic Neurosci.* 2010; 2: 8. doi: [10.3389/fnsyn.2010.00008](https://doi.org/10.3389/fnsyn.2010.00008) PMID: [21423494](https://pubmed.ncbi.nlm.nih.gov/21423494/)
59. Vogels TP, Froemke RC, Doyon N, Gilson M, Haas JS, Liu R, et al. Inhibitory synaptic plasticity: spike timing-dependence and putative network function. *Front Neural Circuits.* 2013; 7: 119. doi: [10.3389/fncir.2013.00119](https://doi.org/10.3389/fncir.2013.00119) PMID: [23882186](https://pubmed.ncbi.nlm.nih.gov/23882186/)
60. Nissen W, Szabo A, Somogyi J, Somogyi P, Lamsa KP. Cell Type-Specific Long-Term Plasticity at Glutamatergic Synapses onto Hippocampal Interneurons Expressing either Parvalbumin or CB1 Cannabinoid Receptor. *J Neurosci.* 2010; 30: 1337–1347. doi: [10.1523/JNEUROSCI.3481-09.2010](https://doi.org/10.1523/JNEUROSCI.3481-09.2010) PMID: [20107060](https://pubmed.ncbi.nlm.nih.gov/20107060/)
61. Huang S, Hugarir RL, Kirkwood A. Adrenergic gating of Hebbian spike-timing-dependent plasticity in cortical interneurons. *J Neurosci Off J Soc Neurosci.* 2013; 33: 13171–13178. doi: [10.1523/JNEUROSCI.5741-12.2013](https://doi.org/10.1523/JNEUROSCI.5741-12.2013) PMID: [23926270](https://pubmed.ncbi.nlm.nih.gov/23926270/)
62. Wang L, Maffei A. Inhibitory plasticity dictates the sign of plasticity at excitatory synapses. *J Neurosci Off J Soc Neurosci.* 2014; 34: 1083–1093. doi: [10.1523/JNEUROSCI.4711-13.2014](https://doi.org/10.1523/JNEUROSCI.4711-13.2014)

63. Dupret D, O'Neill J, Csicsvari J. Dynamic reconfiguration of hippocampal interneuron circuits during spatial learning. *Neuron*. 2013; 78: 166–180. doi: [10.1016/j.neuron.2013.01.033](https://doi.org/10.1016/j.neuron.2013.01.033) PMID: [23523593](https://pubmed.ncbi.nlm.nih.gov/23523593/)
64. Sanger TD. Optimal unsupervised learning in a single-layer linear feedforward neural network. *Neural Netw.* 1989; 2: 459–473. doi: [10.1016/0893-6080\(89\)90044-0](https://doi.org/10.1016/0893-6080(89)90044-0)
65. Oja E. Neural networks, principal components, and subspaces. *Int J Neural Syst.* 1989; 01: 61–68. doi: [10.1142/S0129065789000475](https://doi.org/10.1142/S0129065789000475)
66. Nessler B, Pfeiffer M, Buesing L, Maass W. Bayesian Computation Emerges in Generic Cortical Microcircuits through Spike-Timing-Dependent Plasticity. *PLoS Comput Biol.* 2013; 9: e1003037. doi: [10.1371/journal.pcbi.1003037](https://doi.org/10.1371/journal.pcbi.1003037) PMID: [23633941](https://pubmed.ncbi.nlm.nih.gov/23633941/)
67. Habenschuss S, Puhf H, Maass W. Emergence of Optimal Decoding of Population Codes Through STDP. *Neural Comput.* 2013; 25: 1371–1407. doi: [10.1162/NECO\\_a\\_00446](https://doi.org/10.1162/NECO_a_00446) PMID: [23517096](https://pubmed.ncbi.nlm.nih.gov/23517096/)
68. Moussaoui S, Brie D, Mohammad-Djafari A, Carteret C. Separation of Non-Negative Mixture of Non-Negative Sources Using a Bayesian Approach and MCMC Sampling. *IEEE Trans Signal Process.* 2006; 54: 4133–4145. doi: [10.1109/TSP.2006.880310](https://doi.org/10.1109/TSP.2006.880310)
69. Buesing L, Bill J, Nessler B, Maass W. Neural Dynamics as Sampling: A Model for Stochastic Computation in Recurrent Networks of Spiking Neurons. *PLoS Comput Biol.* 2011; 7: e1002211. doi: [10.1371/journal.pcbi.1002211](https://doi.org/10.1371/journal.pcbi.1002211) PMID: [22096452](https://pubmed.ncbi.nlm.nih.gov/22096452/)
70. Petrovici MA, Bill J, Bytschok I, Schemmel J, Meier K (2013) Stochastic inference with deterministic spiking neurons. *ArXiv13113211 Cond-Mat Physicsphysics Q-Bio*.
71. Toyozumi T, Pfister J-P, Aihara K, Gerstner W. Optimality Model of Unsupervised Spike-Timing-Dependent Plasticity: Synaptic Memory and Weight Distribution. *Neural Comput.* 2007; 19: 639–671. doi: [10.1162/neco.2007.19.3.639](https://doi.org/10.1162/neco.2007.19.3.639) PMID: [17298228](https://pubmed.ncbi.nlm.nih.gov/17298228/)
72. Hennequin G, Gerstner W, Pfister J-P (2010) STDP in Adaptive Neurons Gives Close-To-Optimal Information Transmission. *Front Comput Neurosci* 4:143. doi: [10.3389/fncom.2010.00143](https://doi.org/10.3389/fncom.2010.00143) PMID: [21160559](https://pubmed.ncbi.nlm.nih.gov/21160559/)
73. Beck JM, Ma WJ, Pitkow X, Latham PE, Pouget A. Not Noisy, Just Wrong: The Role of Suboptimal Inference in Behavioral Variability. *Neuron*. 2012; 74: 30–39. doi: [10.1016/j.neuron.2012.03.016](https://doi.org/10.1016/j.neuron.2012.03.016) PMID: [22500627](https://pubmed.ncbi.nlm.nih.gov/22500627/)
74. von der Malsburg C, Schneider W. A neural cocktail-party processor. *Biol Cybern.* 1986; 54: 29–40. doi: [10.1007/BF00337113](https://doi.org/10.1007/BF00337113) PMID: [3719028](https://pubmed.ncbi.nlm.nih.gov/3719028/)
75. Asari H, Pearlmutter BA, Zador AM. Sparse Representations for the Cocktail Party Problem. *J Neurosci.* 2006; 26: 7477–7490. doi: [10.1523/JNEUROSCI.1563-06.2006](https://doi.org/10.1523/JNEUROSCI.1563-06.2006) PMID: [16837596](https://pubmed.ncbi.nlm.nih.gov/16837596/)
76. Comon P. Independent component analysis, A new concept? *Signal Process.* 1994; 36: 287–314. doi: [10.1016/0165-1684\(94\)90029-9](https://doi.org/10.1016/0165-1684(94)90029-9)
77. Bell AJ, Sejnowski TJ. An Information-Maximization Approach to Blind Separation and Blind Deconvolution. *Neural Comput.* 1995; 7: 1129–1159. doi: [10.1162/neco.1995.7.6.1129](https://doi.org/10.1162/neco.1995.7.6.1129) PMID: [7584893](https://pubmed.ncbi.nlm.nih.gov/7584893/)
78. Amari S. Natural Gradient Works Efficiently in Learning. *Neural Comput.* 1998; 10: 251–276. doi: [10.1162/089976698300017746](https://doi.org/10.1162/089976698300017746)
79. Jutten C, Herault J. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Process.* 1991; 24: 1–10. doi: [10.1016/0165-1684\(91\)90079-X](https://doi.org/10.1016/0165-1684(91)90079-X)
80. Girolami M, Gyfe C. Extraction of independent signal sources using a deflationary exploratory projection pursuit network with lateral inhibition. *Vis Image Signal Process IEE Proc-.* 1997; 144: 299–306. doi: [10.1049/ip-vis:19971418](https://doi.org/10.1049/ip-vis:19971418)
81. Smaragdīs P. Blind separation of convolved mixtures in the frequency domain. *Neurocomputing.* 1998; 22: 21–34. doi: [10.1016/S0925-2312\(98\)00047-2](https://doi.org/10.1016/S0925-2312(98)00047-2)
82. Hiratani N, Fukai T. Interplay between short- and long-term plasticity in cell-assembly formation. *PLoS One.* 2014; 9: e101535. doi: [10.1371/journal.pone.0101535](https://doi.org/10.1371/journal.pone.0101535) PMID: [25007209](https://pubmed.ncbi.nlm.nih.gov/25007209/)