



# A Spiking Neural Network for Visual Causal Inference with the Hidden Markov Model

Ruihuan Ren<sup>1</sup>, Pengcheng Cui<sup>2</sup>, Jinping Yuan<sup>2</sup>, Jiaheng Song<sup>2</sup>, Yiran Li<sup>2,\*</sup>, Yutong Lu<sup>3</sup>, Zixuan Huang<sup>3</sup>, Jianyu Wang<sup>3</sup>, Weisi Liu<sup>2,\*</sup>

<sup>1</sup>School of Artificial Intelligence, Yantai Institute of Technology, Yantai 264005, Shandong, China.

<sup>2</sup>School of Mathematics and Information Sciences, Yantai University, Yantai 264005, Shandong, China.

<sup>3</sup>School of Life Sciences, Yantai University, Yantai 264005, Shandong, China.

**How to cite this paper:** Ruihuan Ren, Pengcheng Cui, Jinping Yuan, Jiaheng Song, Yiran Li, Yutong Lu, Zixuan Huang, Jianyu Wang, Weisi Liu. (2025) A Spiking Neural Network for Visual Causal Inference with the Hidden Markov Model. *Advances in Computer and Communication*, 6(5), 310-316.  
DOI: 10.26855/acc.2025.12.009

**Received:** October 31, 2025

**Accepted:** November 30, 2025

**Published:** December 29, 2025

\***Corresponding author:** Yiran Li, School of Mathematics and Information Sciences, Yantai University, Yantai 264005, Shandong, China; Weisi Liu, School of Mathematics and Information Sciences, Yantai University, Yantai 264005, Shandong, China.

## Abstract

Causal inference serves as a fundamental cortical function that inherently involves neural communication across various cortical areas. However, the computational principles through which neurons communicate with each other to implement causal inference remain unclear. To address this question, this paper presents a two-layer spiking neural network under the hidden Markov model (HMM) framework to investigate basic visual causal inference. A hidden cause determines whether directions of paired visual stimuli share a common source, randomly. Receiving stimuli, the network incorporates neural communication to generate random responses and employs a distance-based method to infer the hidden cause. With rewards from inference, the network updates its parameters during training. The trained network achieves acceptable performance in designed visual causal inference while reproducing key neurodynamic phenomena of sparse coding and neural variability quenching. The spiking neural network integrates neural communication into an HMM-based unified framework for sparse coding, neural variability quenching, and visual causal inference.

## Keywords

Causal inference; Clustering; Spiking neural network; Hidden Markov model

## 1. Introduction

Causal inference serves as an important cortical function [1-3]. Previous experiments have established the implementation of causal inference involves neural communication across multiple cortical regions and areas [2, 4, 5]. Experimental observations indicate that it is feasible to consider neural communication in investigating the principles of causal inference.

Bayesian causal inference performs as an important probabilistic method to explore causal inference [6, 3, 7]. Our previous work designed a feedforward spiking neural network to implement visual causal inference with the Bayesian model [8]. Previous studies indicate that the probabilistic framework is suitable and feasible for causal inference.

Experimental studies have demonstrated that neurons in different cortical areas communicate with each other during causal inference [2, 4]. Particularly, the processing of form, color, and texture in visual experiments involves multiple regions in the ventral stream [9, 10]. The experimental evidence indicates that neural communication and cooperation should be considered in exploring causal inference. The winner-take-all (WTA) spiking neural networks consider neural communication under a probabilistic framework while investigating cortical functions [11-15]. Our previous studies have explored sparse coding, neural variability quenching, and concept learning within the hidden

Markov model (HMM), which incorporates both feedforward and lateral neural communication [11-13]. However, the principles of causal inference have not been explored within an HMM-based framework.

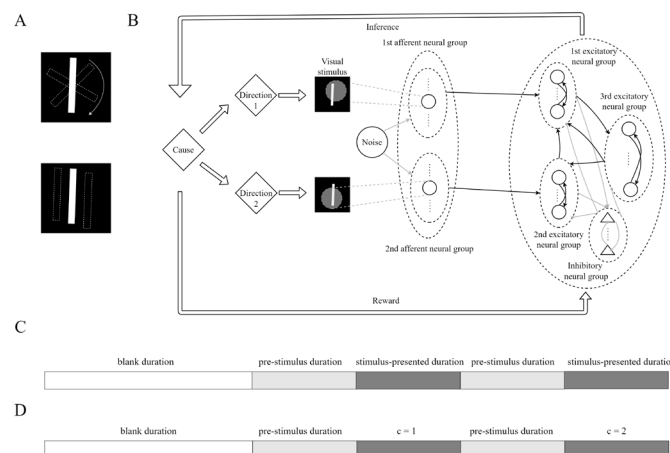
This paper designs a two-layer spiking neural network following the basic structure of the primary visual cortical neural circuit. A basic procedure of visual causal inference is designed as an example. For paired visual stimuli, a hidden cause determines whether their directions share a common source. Receiving stimuli, the network attempts to infer the state of the hidden cause. With random responses, the network employs a distance-based clustering method fusing information from both neural spikes and rates for inference. Based on rewards from inference, the network modifies both its plastic weights and clustering sets during the training phase. The trained network achieves acceptable performances in the visual causal inference and simulates basic phenomena, including sparse neural responses and neural variability quenching. Overall, our network provides a possible unified HMM-based framework for sparse coding, neural variability quenching, and visual causal inference.

## 2. Research Methodology

### 2.1 Designs of Simulations and the Spiking Neural Network

The basic design of visual causal inference follows that introduced in previous studies [3, 7]. The original gray-scale image consists of a  $50 \times 50$ -pixel background and a  $32 \times 8$ -pixel rectangular graphic. To obtain paired visual stimuli, the rectangle is randomly oriented and positioned within the square background. Besides images, a variable following  $U(0,1)$  performs as the temporal noisy stimulus.

The simulating procedure consists of  $L$  training simulations and  $L$  testing simulations (Table 1). Each simulation includes various phases with different stimuli [14]. A training simulation consists of one  $2T$ -time steps blank duration and two  $2T$ -time steps stimulus-associated durations (Table 1, Fig. 1C). The network receives no inputs during the blank duration. For each pair of visual images, its stimulus-associated duration is separated into a pre-stimulus duration and a stimulus-presented duration. The network only receives the white noisy stimulus in each pre-stimulus duration, simulating neural spontaneous activities [16]. Visual stimuli are presented to the network with noisy inputs during the stimulus-presented duration, simultaneously. Sources generating directions of stimuli are determined by a variable  $c \in \{1,2\}$  with equal probabilities (Fig. 1B). If  $c = 1$ , directions of paired visual stimuli derive from a common source and a variable  $D$  is sampled from  $U(0, \pi)$ . Directions of rectangular stimuli are generated from  $N(D, 3)$ , independently. If  $c = 2$ , two variables  $\{D_1, D_2\}$  are sampled from  $U(0, \pi)$  with directions generated from  $N(D_1, 3)$  and  $N(D_2, 3)$ . The rectangle is rotated with  $\{D_1, D_2\}$  (Fig. 1A). Besides, the rectangle in each image is randomly displaced from the center with a displacement vector sampled from  $N(0, \Sigma)$ ,  $\Sigma = \text{diag}(3,3)$  (Fig. 1A). In the testing phase, directions and locations of visual stimuli are determined as those in training simulations. Particularly,  $c = 1$  in the 1st stimulus-presented duration while  $c = 2$  in the other stimulus-presented duration.



**Figure 1. Designs of simulations and the spiking neural network. (A) Visual images. (B) The spiking neural network. Afferent neural groups (circles) receive paired visual stimuli. Three excitatory neural groups (circles) and one inhibitory neural group (triangles) compose the downstream network. Synaptic connections include excitatory connections (arrows) and inhibitory ones (circular dots). Connections among excitatory neurons are plastic (thick) while others keep fixed (thin). (C) The basic procedure of a training simulation. (D) The basic procedure of a testing simulation.**

Table 1. Parameters

$T$	$L$	$\tau_r$	$\tau_f$	$N_1$	$N_2$	$K_1$	$K_2$	$K_3$	$J$	$p_{EA}$
50	500	1	10	100	100	200	200	200	150	100%
$p_{EE}$	$p_{EI}$	$p_{EI}$	$p_{II}$	$v^{EI}$	$v^{IE}$	$v^{II}$	$\eta$	$R$	$N_{clu}$	$\sigma_{rate}^2$
50%	60%	57.5%	55%	0.5	0.5	0.5	0.005	50	20	20

With stimuli designed above, a two-layer spiking neural network is constructed to implement visual causal inference (Fig. 1B). The first-layer network consists of two afferent neural groups with their sizes as  $N_1$  and  $N_2$ . To process visual images as retinal ganglion cells, the  $n_1$ th afferent neuron has its receptive field simulated by a basic Difference-of-Gaussian (DoG) function [17]. To cover 50×50-pixel images (Fig. 1A), afferent neural receptive fields are located on a grid with the spacing distance as 5 pixels [19]. Processing the image with its receptive field, the  $n_1$ th afferent neuron obtains a strength of Poisson spikes and generates exponential synaptic trace  $\tilde{x}_{tn_1}$ :

$$\tilde{x}_{tn_1} = \sum_{t_{n_1}^{(l)} < t} \varepsilon(t - t_{n_1}^{(l)}), \varepsilon(s) = \begin{cases} \exp(-s/\tau_f) - \exp(-s/\tau_r), & s \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where  $\tau_r$  is the rise-time constant and  $\tau_f$  is the fall-time constant (Table 1) [15]. The downstream network includes three excitatory neural populations and one inhibitory population, distinguished by their corresponding inputs. Sizes of excitatory neural populations are  $K_1, K_2, K_3$  while that of the inhibitory population is  $J$ . Downstream neural inter-connections are determined with connective probabilities [14, 15]. For the  $k_1$ th neuron in the 1st excitatory neural population, its temporal membrane potential  $u_{tk_1}^z$ , as well as the soft-max distribution of its active state  $z_{tk_1}$ , can be expressed as:

$$u_{tk_1}^z = \sum_n w_{k_1 n_1} \tilde{x}_{tn_1} + \sum_{k_1'} v_{k_1 k_1'} \tilde{z}_{tk_1'} + \sum_{k_2} v_{k_1 k_2} \tilde{z}_{tk_2} + \sum_{k_3} v_{k_1 k_3} \tilde{z}_{tk_3} - \sum_j v^{EI} \tilde{y}_{tj},$$

$$p(z_{tk_1}) = \frac{\exp(u_{tk_1}^z z_{tk_1})}{1 + \exp(u_{tk_1}^z)}, \quad z_{tk_1} \in \{0,1\} \quad (2)$$

where  $w_{k_1 n_1} \tilde{x}_{tn_1}, v_{k_1 k_1'} \tilde{z}_{tk_1'}, v_{k_1 k_2} \tilde{z}_{tk_2}$  and  $v_{k_1 k_3} \tilde{z}_{tk_3}$  are excitatory postsynaptic potentials (EPSPs) evoked by different neurons.  $v^{EI} \tilde{y}_{tj}$  is the lateral inhibitory postsynaptic potential (IPSP). Feedforward connections to the  $k_1$ th neuron are set with  $p_{EA}$  (Table 1) [15]. Lateral connections from downstream excitatory and inhibitory neurons are also determined with  $p_{EE}$  and  $p_{EI}$ , respectively (Table 1) [14, 15]. For the existing connections,  $w_{k_1 n_1}, v_{k_1 k_1'}, v_{k_1 k_2}$  and  $v_{k_1 k_3}$  are plastic while  $v^{EI}$  is fixed.  $\tilde{x}_{tn_1}, \tilde{z}_{tk_1'}, \tilde{z}_{tk_2}, \tilde{z}_{tk_3}$  and  $\tilde{y}_{tj}$  are temporal synaptic traces (Eq. (1)). The  $k_1$ th excitatory neuron generates a spike at  $t$  if  $z_{tk_1} = 1$ . For neurons in the other two excitatory populations, neural membrane potentials, as well as distributions of active states, depend on their received inputs and can be expressed in the similar way.

The  $j$ th inhibitory neuron has its temporal membrane potential  $u_{tj}^y$  and temporal strength  $\rho_{tj}^y$  expressed as:

$$u_{tj}^y = \sum_{k_1} v^{IE} z_{tk_1} + \sum_{k_2} v^{IE} z_{tk_2} + \sum_{k_3} v^{IE} z_{tk_3} - \sum_{j'} v^{II} \tilde{y}_{tj'},$$

$$\rho_{tj}^y = \sigma_{rect}(u_{tj}^y), \quad (3)$$

where  $\tilde{z}_{tk_1}, \tilde{z}_{tk_2}, \tilde{z}_{tk_3}$  and  $\tilde{y}_{tj'}$  are temporal synaptic traces from downstream neurons. With  $p_{IE}$  and  $p_{II}$  (Table 1) [15], lateral excitatory and inhibitory connections are determined and marked as fixed weights  $v^{IE}, v^{II}$  (Table 1). With the strength  $\rho_{tj}^y$ , the inhibitory neuron generates Poisson spikes with its active state as  $y_{tj} \in \{0,1\}$ .

## 2.2 The Probabilistic Model

In the probabilistic model, the hidden cause is indicated by  $c$ . Visual directions are marked as  $D_1$  and  $D_2$ . Afferent

neural responses are observed variables while downstream network responses as hidden variables. Afferent inputs up to  $t$  are collected as  $O_t = \{o_1, \dots, o_t\}$ , where  $o_{t'} = \{\tilde{x}_{t'N_1}, \dots, \tilde{x}_{t'N_1}, \tilde{x}_{t'N_2}, \dots, \tilde{x}_{t'N_2}\}$  are temporal afferent synaptic inputs. Downstream network responses at  $t'$  are marked as  $h_{t'}$ , including  $\{z_{t'1}, \dots, z_{t'K_1}, z_{t'1}, \dots, z_{t'K_2}, z_{t'1}, \dots, z_{t'K_3}, y_{t'1}, \dots, y_{t'J}\}$  and  $\{\tilde{z}_{t'K_1}, \tilde{z}_{t'1}, \dots, \tilde{z}_{t'K_2}, \tilde{z}_{t'1}, \dots, \tilde{z}_{t'K_3}, \tilde{y}_{t'1}, \dots, \tilde{y}_{t'J}\}$ .  $\{h_1, \dots, h_t\}$  are collected as  $H_t$ . In the visual causal inference, the network infers  $c$  through a distance-based clustering method and gets a temporal reward  $r_t$  (Fig. 1B). Once the network makes the correct inference, it can get a positive reward  $r_t = 1$ . Otherwise, it receives  $r_t = 0$ .

Neural networks with reward-modulated learning rules have been proposed [11, 12, 20]. In this network,  $\{W, V\}$  are sets of plastic feedforward and lateral weights.  $\{C_1, C_2\}$  are clustering sets in the distance-based method. Parameters in the model are collected as  $\theta = \{W, V, C_1, C_2\}$ . The joint distribution of  $\{O_t, H_t, D_1, D_2, c\}$  depending on parameters is marked as  $p(O_t, H_t, D_1, D_2, c; \theta)$ . While  $p(O_t, H_t, D_1, D_2, c | r_t = 1; \theta)$  reflects the distribution of network responses which can make correct inferences. The network modifies its weights to minimize a Kullback-Leibler (KL) divergence  $D_{KL}\{p(O_t, H_t, D_1, D_2, c | r_t = 1; \theta) \| p(O_t, H_t, D_1, D_2, c; \theta)\}$ . Minimization of KL divergence is equal to maximization of a likelihood function  $L(\theta)$ :

$$L(\theta) = \mathbb{E}_{p(O_t, H_t, D_1, D_2, c, r_t; \theta)} [r_t \log p(O_t, H_t, D_1, D_2, c; \theta)]. \quad (4)$$

The network implements a stochastic online variant of the Expectation-Maximization (EM) algorithm upon  $\{W, V\}$  [21]. In the E-step, the network generates its responses as a single sample from  $p(O_t, H_t, D_1, D_2, c, r_t; \theta)$  which can be factorized with the conditional independence [21]:

$$\begin{aligned} p(O_t, H_t, D_1, D_2, c, r_t; \theta) &\propto p(O_t, H_t | D_1, D_2; \theta) \cdot p(r_t | O_t, H_t; \theta) \\ &\propto p(O_t | D_1, D_2) \cdot \frac{p(O_t, H_t; \theta)}{\int p(O_t, H_t; \theta) dH_t} \cdot p(r_t | O_t, H_t; \theta), \end{aligned} \quad (5)$$

where  $p(O_t, H_t; \theta)$  can be factorized with the Markov property as:

$$\begin{aligned} p(O_t, H_t; \theta) &= \prod_{t'=1}^t p(o_{t'} | h_{t'}; \theta) p(h_{t'} | h_{t'-1}; \theta) \\ &\propto \prod_{t'=1}^t \left[ \prod_{k_1} \exp(u_{t'k_1}^z z_{t'k_1}) \prod_{k_2} \exp(u_{t'k_2}^z z_{t'k_2}) \prod_{k_3} \exp(u_{t'k_3}^z z_{t'k_3}) \right]. \end{aligned} \quad (6)$$

The network generates excitatory neural spikes from  $p(O_t, H_t | D_1, D_2; \theta)$  at  $t$  based on values of plastic weights.  $L(\theta)$  can be approximated during the E-step as:

$$L(\theta) \simeq r_t \cdot [\log p(c) + \log p(D_1, D_2 | c) + \log p(O_t, H_t | D_1, D_2, c; \theta)]. \quad (7)$$

### 2.3 Estimation Upon the Cause and Modifications of Parameters

To modify parameters, the network infers  $c$  and gets  $r_t$ . Instead of the MAP (maximum-a-posteriori) estimation in the Bayesian inference [3, 7], the network obtains a point estimation upon  $c$  through an unsupervised online distance-based clustering method [22]. The main idea of this clustering method is introduced here while its detail illustration is in our previous study [11].

The temporal synaptic trace of the  $k_1$ th excitatory neuron  $\tilde{z}_{tk_1}$  performs as a kind of exponential spiking rate reflecting information of spikes before  $t$  (Eq. (1)). For all the downstream excitatory neurons, a vector  $\vec{Z}_t = (z_{t1}, \dots, z_{tK_1}, z_{t1}, \dots, z_{tK_2}, z_{t1}, \dots, z_{tK_3}, \tilde{z}_{t1}, \dots, \tilde{z}_{tK_1}, \tilde{z}_{t1}, \dots, \tilde{z}_{tK_2}, \tilde{z}_{t1}, \dots, \tilde{z}_{tK_3})$  performs as evoked network responses for inference. In the distance-based clustering method, the initial sets  $\{C_1, C_2\}$  are blank. To each pair of images, the first non-empty  $\vec{Z}_t$  is added to the clustering set. Distances between the later  $\vec{Z}_t$  and  $\{C_1, C_2\}$  are quantified through a re-sampling procedure [22]. These distances are based on samples are feasible for situations with small sample sizes [22]. With  $R$  re-samples (Table 1), a common threshold  $e_{threshold}(\vec{Z}_t)$  is found to quantify the divergence between  $\vec{Z}_t$  and each set. Likelihoods of  $\vec{Z}_t$  and  $\{C_1, C_2\}$  are measured through another re-sampling procedure during which resampling-based distances below  $e_{threshold}(\vec{Z}_t)$  are compared. With  $R$  novel re-samples, a clustering set is deemed to have a greater likelihood of matching  $\vec{Z}_t$  if a larger proportion of its resampling-based distances falls

below  $e_{threshold}(\vec{Z}_t)$ . In this way, the network achieves a point estimate upon the hidden cause with evoked neural responses.

In simulations, estimations are assumed to begin from the appearance of paired visual stimuli. When the network receives  $r_t = 0$ , it is assumed to know the temporal estimation is incorrect and continue to infer the hidden cause. Once the network receives  $r_t = 1$ , it is assumed to hold its correct estimation and, as a result, realizes visual casual inference. With rewards from inference, the network modifies its plastic weights and clustering sets with a reward-modulated rule. Modifications of weights derive from partial derivatives of  $L(\theta)$  (Eq. (7)). For instance, modifications of  $w_{k_1 n_1}$  and  $v_{k_1 k'_1}$  can be approximated as:

$$\begin{aligned} \Delta w_{k_1 n_1} &\approx \eta \cdot r_t \cdot z_{tk_1} \cdot \left\{ \tilde{x}_{tn_1} - \left[ 1 - \frac{1}{w_{k_1 n_1}} + \frac{1}{\exp(w_{k_1 n_1}) - 1} \right] \right\} \\ \Delta v_{k_1 k'_1} &\approx \eta \cdot r_t \cdot z_{tk_1} \cdot \left\{ \tilde{z}_{tk'_1} - \left[ 1 - \frac{1}{v_{k_1 k'_1}} + \frac{1}{\exp(v_{k_1 k'_1}) - 1} \right] \right\} \end{aligned} \quad (8)$$

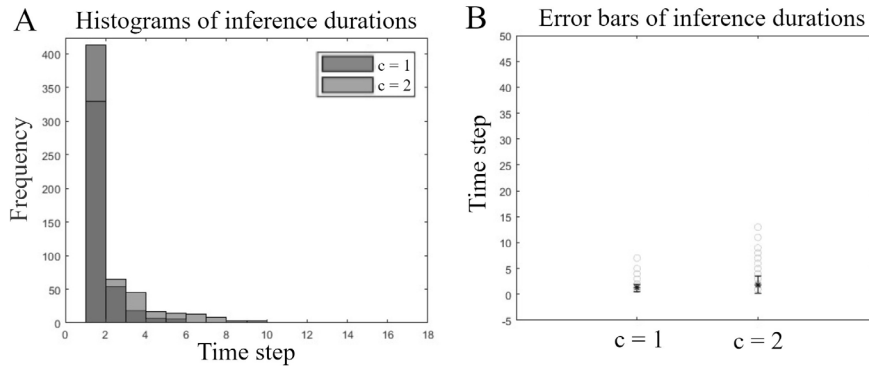
where  $\eta$  is the learning rate (Table 1). Modifications of other plastic weights can be expressed, similarly.

Adaptations of clustering sets are cautious. During training, only  $\vec{Z}_t$  making the first correct inference will be added to the corresponding clustering set. Each set removes redundant components with a First-In-First-Out (FIFO) policy once its size exceeds the limitation  $N_{clu}$  (Table 1).

### 3. Results

#### 3.1 Network Performance

In the previous experiment, periods between target onset and button press are regarded as reaction times to quantify performance of subjects [23]. To evaluate network performance in visual causal inference, the interval from stimulus onset to the first correct inference is marked as the inference duration. Upon each value of  $c$ , inference durations are collected over testing simulations. The empirical distributions of inference durations are measured by histograms and error bars (Fig. 2). The trained network completes inference within 15 time steps across all the causal states, demonstrating acceptable performance in the designed task.



**Figure 2. Inference durations upon the hidden cause. (A) Histograms of inference durations. (B) Error bars of inference durations. Circles are historical data of inference durations over testing simulations.**

#### 3.2 Network responses

Across two values of  $c$ , excitatory neural responses are compared to measure the neural responding sparseness [24]. In the  $l$ th testing simulation, if the  $k_1$ th excitatory neuron generates its spikes at  $\{t_{k_1 l}^{(1)}, t_{k_1 l}^{(2)}, \dots\}$ , its temporal spiking rate can be calculated as:

$$rate_{k_1}(l, t) = \sum_{t_{k_1 l}^{(g)} < t} \frac{1}{\sqrt{2\pi\sigma_{rate}^2}} \exp \left[ -\frac{(t - t_{k_1 l}^{(g)})^2}{2\sigma_{rate}^2} \right] \quad (9)$$

where  $\sigma_{\text{rate}}^2$  is the variance in the Gaussian function (Table 1). With averaged values of  $\text{rate}_{k_1}(l, t)$  over time and simulations for two causal states marked as  $\text{rate}_{k_11}$  and  $\text{rate}_{k_12}$ , responding sparseness of the  $k_1$ th neuron is quantified by  $s_{k_1}$  [24]:

$$s_{k_1} = 1 - \frac{\left(\frac{\text{rate}_{k_11} + \text{rate}_{k_12}}{2}\right)^2}{\frac{(\text{rate}_{k_11})^2 + (\text{rate}_{k_12})^2}{2}}, \quad (10)$$

where a statistic  $s_{k_1}$  close to 1 indicates the neuron generates sparse responses during the visual causal inference. For each downstream excitatory neuron, its responding sparseness can be measured by Eq. (10). Through the reward-modulated learning, neurons with sparse responses emerge in different downstream excitatory groups (Fig. 3A).

Temporal spiking rates of excitatory neural groups are averaged values of neural temporal spiking rates over included neurons, marked as variables  $\{\text{rate}^{(i)}(l, t), i = 1, 2, 3\}$ . For  $\{\text{rate}^{(i)}(l, t), i = 1, 2, 3\}$ , their standard variances over simulations are measured to explore neural variability [16]. The trained network can simulate basic phenomena of neural variability quenching in all the excitatory neural groups during the visual causal inference (Fig. 3B).

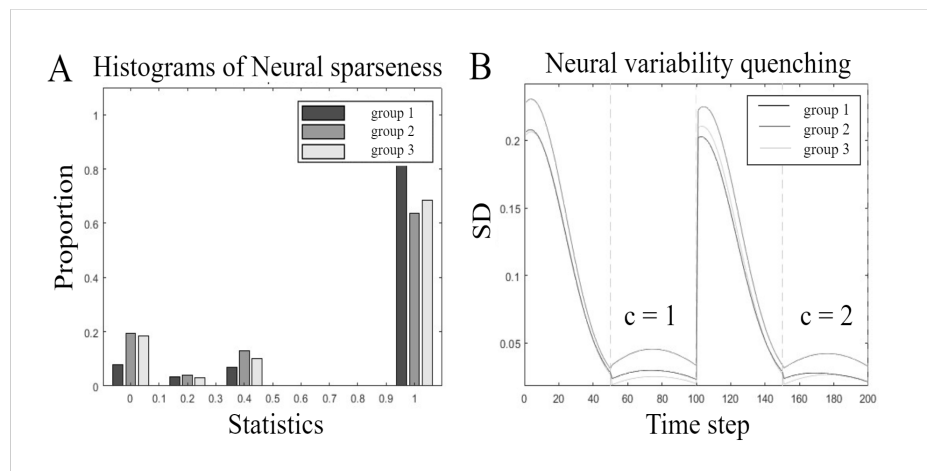


Figure 3. Network responses in visual causal inference. (A) Neural sparseness. (B) Neural variability quenching.

## 4. Conclusion

Experimental observations have demonstrated that causal inference involves neural communication [2, 4, 5]. Probabilistic inference is suitable for exploring causal inference [3, 6, 7]. To incorporate neural communication into the probabilistic framework, this paper designs an HMM-based spiking neural network for basic visual causal inference. Compared to the MAP estimate in Bayesian causal inference [3, 7], the network employs a distance-based clustering method for the point estimation of the hidden cause. Compared with previous WTA networks [14, 15, 21], the network incorporates both neural spikes and rates into the distance-based framework for causal inference. Based on inference, the network modifies its parameters through a reward-modulated rule during training. The trained network achieves acceptable performance through the distance-based method, simulating sparse coding and neural variability quenching. The HMM-based spiking neural network provides a unified probabilistic framework for sparse coding, neural variability quenching, and visual causal inference.

## Funding

This work is supported by Natural Science Foundation of Shandong Province, China (Grant Number: ZR2025QC09).

## Conflict of Interest

All authors reviewed the manuscript and declared that there were no competing financial interests in this work.

## Data Availability Statement

Source codes are available at <https://github.com/WSiLiu/Ren-ACC-2025.git>.

## References

- [1] Shams L, Beierholm U. Bayesian causal inference: A unifying neuroscience theory. *Neurosci Biobehav Rev.* 2022;137:104619.
- [2] Mohl JT, Pearson JM, Groh JM. Monkeys and humans implement causal inference to simultaneously localize auditory and visual stimuli. *J Neurophysiol.* 2020;124:715-27.
- [3] Fang Y, Yu Z, Liu JK, Chen F. A unified neural circuit of causal inference and multisensory integration. *Neurocomputing.* 2019;358:355-68.
- [4] French RL, DeAngelis GC. Multisensory neural processing: from cue integration to causal inference. *Curr Opin Physiol.* 2020;16:8-13.
- [5] Cao Y, Summerfield C, Park H, Giordano BL, Kayser C. Causal inference in the multisensory brain. *Neuron.* 2019;102:1076-87.
- [6] Huo H, Liu X, Tang Z, Dong Y, Zhao D, Chen D, et al. Interhemispheric multisensory perception and Bayesian causal inference. *iScience.* 2023;5:106706.
- [7] Ma WJ, Rahmati M. Towards a neural implementation of causal inference in cue combination. *Multisens Res.* 2013;26:159-76.
- [8] Liu W, Pan X. Effects of neural assembles in causal inference based on an entropy-maximization Bayesian neural network. *IEEE Access.* 2024;12:184442-55.
- [9] Cant JS, Arnott SR, Goodale MA. fmr-adaptation reveals separate processing regions for the perception of form and texture in the human ventral stream. *Exp Brain Res.* 2009;192:391-405.
- [10] Cant JS, Large ME, McCall L, Goodale MA. Independent processing of form, colour, and texture in object perception. *Perception.* 2008;37:57-78.
- [11] Liu W, Ren R. Disinhibition maintains network performances in concept learning through regulating neural responses. *Neurocomputing.* 2025;649:130760.
- [12] Liu W, Liu X. Pre-stimulus network responses affect information coding in neural variability quenching. *Neurocomputing.* 2023;531:1-20.
- [13] Liu W, Liu X. The effects of eye movements on the visual cortical responding variability based on a spiking network. *Neurocomputing.* 2021;436:58-73.
- [14] Pokorny C, Ison MJ, Rao A, Legenstein R, Papadimitriou C, Maass W. STDP forms associations between memory traces in networks of spiking neurons. *Cereb Cortex.* 2020;30:952-68.
- [15] Jonke Z, Legenstein R, Habenschuss S, Maass W. Feedback inhibition shapes emergent computational properties of cortical microcircuit motifs. *J Neurosci.* 2017;37:8511-23.
- [16] Wolff A, Chen L, Tumati S, Golesorkhi M, Gomez-Pilar J, Hu J, et al. Prestimulus dynamics blend with the stimulus in neural variability quenching. *Neuroimage.* 2021;238:118160.
- [17] Field DJ. What is the goal of sensory coding? *Neural Comput.* 1994;6:559-601.
- [18] Enroth-Cugell C, Robson JG. The contrast sensitivity of retinal ganglion cells of the cat. *J Physiol.* 1966;187:517-52.
- [19] Segal IY, Giladi C, Gedalin M, Rucci M, Ben-Tov M, Kushinsky Y, et al. Decorrelation of retinal response to natural scenes by fixational eye movements. *Proc Natl Acad Sci U S A.* 2015;112:3110-5.
- [20] Rueckert E, Kappel D, Tanneberg D, Pecevski D, Peters J. Recurrent spiking networks solve planning tasks. *Sci Rep.* 2016;6:21142.
- [21] Kappel D, Nessler B, Maass W. STDP installs in winner-take-all circuits an online approximation to hidden markov model learning. *PLoS Comput Biol.* 2014;10:e1003511.
- [22] Heinerman J, Haasdijk E, Eiben A. Unsupervised identification and recognition of situations for high-dimensional sensori-motor streams. *Neurocomputing.* 2017;262:90-107.
- [23] Kloosterman NA, de Gee JW, Werkle-Bergner M, Lindenberger U, Garrett DD, Fahrenfort JJ. Humans strategically shift decision bias by flexibly adjusting sensory evidence accumulation. *Elife.* 2019;8:e37321.
- [24] Olshausen BA, Field DJ. Sparse coding of sensory inputs. *Curr Opin Neurobiol.* 2004;14:481-7.