



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

Doctoral Thesis

# Spatio-temporally efficient coding: A computational principle of biological neural networks

Duho Sihm

Department of Biomedical Engineering  
(Human Factors Engineering)

Ulsan National Institute of Science and Technology

2022

# Spatio-temporally efficient coding: A computational principle of biological neural networks

Duho Sihn

Department of Biomedical Engineering  
(Human Factors Engineering)

Ulsan National Institute of Science and Technology

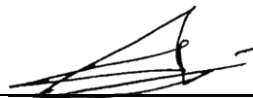
# Spatio-temporally efficient coding: A computational principle of biological neural networks

A thesis/dissertation submitted to  
Ulsan National Institute of Science and Technology  
in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy

Duho Sihn

06.15.2022 of submission

Approved by



---

Advisor

Sung-Phil Kim

# Spatio-temporally efficient coding: A computational principle of biological neural networks

Duho Sihm

This certifies that the thesis/dissertation of Duho Sihm is approved.

06.15.2022 of submission

Signature



Advisor: Sung-Phil Kim

Signature



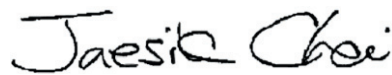
Oh-Sang Kwon

Signature



Kwang In Kim

Signature



Jaesik Choi

Signature



Se-Bum Paik

## Abstract

One of the major goals of neuroscience is to understand how the external world is represented in the brain. This is a neural coding problem: the coding from the external world to its neural representations. There are two different kinds of problems with neural coding. One is to study the types of neuronal activity that represent the external world. Representative examples here are rate coding and temporal coding. In this study, we will present the spike distance method that reads temporal coding-related information from neural data. Another is to study what principles make such neural representations possible. This is an approach to the computational principle and the main topic of the present study. The brain sensory system has hierarchical structures. It is important to find the principles assigning functions to the hierarchical structures. On the one hand, the hierarchical structures of the brain sensory system contain both bottom-up and top-down pathways. In this bidirectional hierarchical structure, two types of neuronal noise are generated. One of them is noise generated as neural information fluctuates across the hierarchy according to the initial condition of the neural response, even if the external sensory input is static. Another is noise, precisely error, caused by coding different information in each hierarchy because of the transmission delay of information when external sensory input is dynamic. Despite these noise problems, it seems that sensory information processing is performed without any major problems in the sensory system of the real brain. Therefore, a neural coding principle that can overcome these noise problems is needed; How can the brain overcome these noise problems? Efficient coding is one of representative neural coding principles, however, existing efficient coding does not take into account these noise problems. To treat these noise problems, as one of efficient coding principles, we devised spatio-temporal efficient coding, which was inspired by the efficient use of given space and time resources, to optimize bidirectional information transmission on the hierarchical structures. This optimization is to learn smooth neural responses on time domain. In simulations, we showed spatio-temporal efficient coding was able to solve above two noise problems. We expect that spatio-temporal efficient coding helps us to understand how the brain computes.

**Keywords** – Efficient coding, Hierarchical structure, Neural coding, Neural representation, Neuronal noise, Sensory system, Spike distance, Temporal coding.



## Contents

Abstract .....	i
Contents .....	ii
List of Figures .....	iii
1. Introduction .....	1
2. Scheme of neural coding .....	5
2.1. Rate coding and temporal coding .....	5
2.2. Spike distances: measures for rate coding and temporal coding .....	8
2.2.1. Spike distance focused on temporal coding .....	9
3. Principle of neural coding .....	33
3.1. Efficient coding and predictive coding .....	34
3.2. Spatio-temporally efficient coding .....	35
3.3. Implementation of spatio-temporally efficient coding .....	39
4. Spatio-temporally efficient coding: a case of static sensory input .....	45
5. Spatio-temporally efficient coding: a case of dynamic sensory input .....	57
6. Discussion .....	63
References .....	68
Acknowledgment .....	82



## List of Figures

Figure 1. Spike trains for grating stimuli. ....	6
Figure 2. Example of both rate coding and temporal coding. ....	7
Figure 3. Calculation of the spike distance based on the earth mover's distance. ....	12
Figure 4. Spike distance results for the measurement of spike timing differences. ....	15
Figure 5. Spike distance results for the measurement of temporal similarity. ....	18
Figure 6. Spike distance results for the measurement of spike time synchrony. ....	21
Figure 7. Comparison with the Victor-Purpura distance in terms of suitability for temporal coding with different firing rates. ....	24
Figure 8. Application of the spike distance to real neuronal data in the primary motor cortex in a non-human primate. ....	29
Figure 9. Application of the spike distance to resampled neuronal data. ....	31
Figure 10. Spatio-temporally efficient coding. ....	39
Figure 11. Implementation of spatio-temporally efficient coding. ....	44
Figure 12. A simulation method for static sensory input. ....	47
Figure 13. Decodable and stable neural representations. ....	50
Figure 14. Neural response distributions. ....	52
Figure 15. Neural response distributions for learned and unlearned inputs. ....	53
Figure 16. Orientation preference. ....	55
Figure 17. A simulation method for dynamic sensory input. ....	59
Figure 18. Distance between neural responses denoting consistent neural representations. ....	61
Figure 19. Decoding of bar stimuli denoting consistent neural representations. ....	62

# 1. Introduction

## 1.A. Backgrounds and motivations

One of the most important goals of neuroscience is to understand how the external world is represented in the brain, and how the brain adapts to the external world based on those neural representations (deCharms and Zador, 2000; Kriegeskorte and Diedrichsen, 2019). This is a neural coding problem: the coding from the external world to its neural representations. There are two different kinds of problems with neural coding. One is to study the types of neuronal activity that represent the external world. Roughly speaking, this is a data-driven approach. Based on neural data, it has been studied that the firing rate and firing timing of neurons code such information. This is called rate coding and temporal coding, respectively. In the section 2, we will first review this rate coding and temporal coding, and also review the spike distance, a method for measuring information of rate coding and temporal coding from data. Another is to study what principles make such neural representations possible. This is an approach to the computational principle and the main topic of the present study.

It has long been accepted that the brain, especially sensory system, has a hierarchical structure (Felleman and Van Essen, 1991; Mesulam, 1998; Harris et al., 2019; Hilgetag and Goulas, 2020). This hierarchical structure is related to gene expression (Burt et al., 2018; Hansen et al., 2021), suggesting that the hierarchical structure is genetically determined and a priori. Then, how can a prior hierarchical brain structure be given the function to represent the external world?

The hierarchical structure of the sensory system is bidirectional; The hierarchical structure has not only bottom-up pathway but also top-down pathway, even in early sensory (visual) system such as the lateral geniculate nucleus (Murphy and Sillito, 1987; Wang et al., 2006) and the primary visual cortex (Zhang et al., 2014; Muckli et al., 2015; Huh et al., 2018). In this bidirectional hierarchical structure, two types of neuronal noise can be generated, where neuronal noise is defined as the uncertainty of neural responses for given sensory input (Borst and Theunissen, 1999). One of them is noise generated as neural information fluctuates across the hierarchy according to the initial condition of the neural response, even if the external sensory input is static. It is known that this can occur in interconnected structures as a chaotic dynamics sensitive to the initial condition (Rubinov et al., 2009; Tomov et al., 2014). This is neuronal noise when static sensory inputs are given. Another is noise, precisely error, caused by coding different information in each hierarchy because of the transmission delay of information (Berry et al., 1999) when external sensory input is dynamic. Because of the information transmission delay, (lower) hierarchies close to the sensory organ represent relatively recent information,

and (upper) hierarchies distant from the sensory organ represent information relatively old, if external inputs are dynamic (changed). By the top-down pathway, old information in the upper hierarchy affects recent information in the lower hierarchy, and this becomes neuronal noise when dynamic sensory inputs are given. Despite these noise (error) problems, it seems that sensory information processing is performed without any major problems in the sensory system of the real brain. For the first noise problem (static sensory input), neural responses in sensory systems are decodable in both neuronal spikes (Berens et al., 2012; Zavitz et al., 2016) and blood-oxygen-level-dependent responses (Kamitani and Tong, 2005; Brouwer and Heeger 2009), indicating that the real brain is robust to noise of this type because this means the uncertainty of neural responses are not very large. For the second noise (error) problem (dynamic sensory input), the sensory system has information of future sensory input (Palmer et al., 2015; Chen et al., 2017; Sederberg et al., 2018; Liu et al., 2021), minimizing information discrepancy across the hierarchy. This suggests that the real brain is robust to noise (error) of this type because the miss-informed noise (erroneous information) across hierarchy decreases as decreasing of information discrepancy across the hierarchy. Therefore, a neural coding principle that can overcome these noise problems is needed; How can the brain overcome these noise problems?

### 1.B. Related works

The traditional view of information processing on the hierarchical structure is that bottom-up information processing, where simple features are processed in lower hierarchy and more complex features created from simple features are processed in higher hierarchy (Hubel and Wiesel, 1962; Hubel and Wiesel, 1968; Riesenhuber and Poggio, 1999; Riesenhuber and Poggio, 2000; Serre et al., 2007; DiCarlo et al., 2012; Yamins et al., 2014). However this view cannot take account of the role of top-down pathways which are abundant in the brain sensory system.

The role of top-down visual processing is emphasized in predictive coding (Rao and Ballard, 1999; Spratling, 2017) which is one of neural coding principles. According to predictive coding higher hierarchy performs top-down predictions on the response of lower hierarchy. Both inference and learning of predictive coding are to minimize the bottom-up prediction errors. Predictive coding has been shown to be able to explain neural responses corresponding to prediction errors (Friston, 2005) and has been extended from the explanation of perceptions to actions (Friston, 2010; Clark, 2013).

Prediction error minimization process in predictive coding may reduce the aforementioned neuronal noise where neuronal noise when static and dynamic sensory inputs are given. Nonetheless predictive coding has several theoretical disadvantages. Since inference in predictive coding is to minimize the

prediction errors, it seems that the brain should have an additional information processing subsystem to perform the inference. Also, since bottom-up transmitted information is only prediction error, predictive coding requires the error units that are hypothetical entities while in some cases it is even difficult to observe prediction error responses (Solomon et al., 2021). Moreover, existing predictive coding is the problem on real-time information processing (Hogendoorn and Burkitt, 2019).

Efficient coding, is one of such principles, aims to minimize informational redundancy of neural representations for the external world (Attneave, 1954; Barlow, 1961). The informational redundancy reduction is to use efficiently the space of neural responses subject to the maximal range of neural responses. Its biological plausibility has been verified in sensory systems (Laughlin, 1981). This has worked well for several sensory (visual) processing problems (Simoncelli and Olshausen, 2001).

Efficient coding does not have the theoretical disadvantages of predictive coding; neither it needs additional information processing subsystems, hypothetical error units, nor the problem on real-time information processing. Efficient coding has been applied to several studies on the hierarchical structure: the studies for complex cell property (Karklin and Lewicki, 2009), visual recognition (Hu et al., 2014), and acoustic feature encoding (Zhang et al., 2019). However, these studies did not take into account the passage of time or studied in a unidirectional hierarchical structure where only bottom-up processing exists. In this environment, the aforementioned noise problems do not occur. These noise problems arise in bidirectional hierarchical structures on the time domain. The real brain environment is a bidirectional hierarchical structure on the time domain.

### 1.C. Objectives

The goal of the present study is to present a neural coding principle to overcome the aforementioned neuronal noise problem; neuronal noise when static and dynamic sensory inputs are given. We, therefore, propose a novel neural coding principle, spatio-temporally efficient coding, for bidirectional hierarchical structures on time domain. Spatio-temporally efficient coding, as one of efficient coding, underscored by the efficient use of given resources in both neural activity space and processing time. Spatio-temporally efficient coding minimizes temporal differences of neural responses (temporally efficient coding), and maximizes activity space of neural responses for different external sensory input (spatially efficient coding). We call the combination of temporally efficient coding and spatially efficient coding as spatiotemporal efficient coding. By spatio-temporally efficient coding, neural responses change smoothly but dynamically.

In the section 3, we will review efficient coding and predictive coding, then will introduce spatio-

temporally efficient coding. In the section 4 and 5, we will show that spatio-temporally efficient coding solves the aforementioned neuronal noise problems: static sensory input (section 4) and dynamic sensory input (section 5). In the section 6, we will discuss all materials from the present study.

Contents of the present study were based on the author's previous studies (Sihn and Kim, 2019; Sihn and Kim, 2021; Sihn and Kim, 2022) and descriptions in the studies.

## 2. Scheme of neural coding

In this section, we will cover types of neural activity that represent the external world, which is a kind of neural coding problem. These studies mainly deal with neural coding models of how neural data as a result of experiments represent experimental stimuli. Typical such neural coding models are rate coding and temporal coding. We will introduce these two coding principles first, and then we will cover the spike distance, which measures information according to this coding in real data.

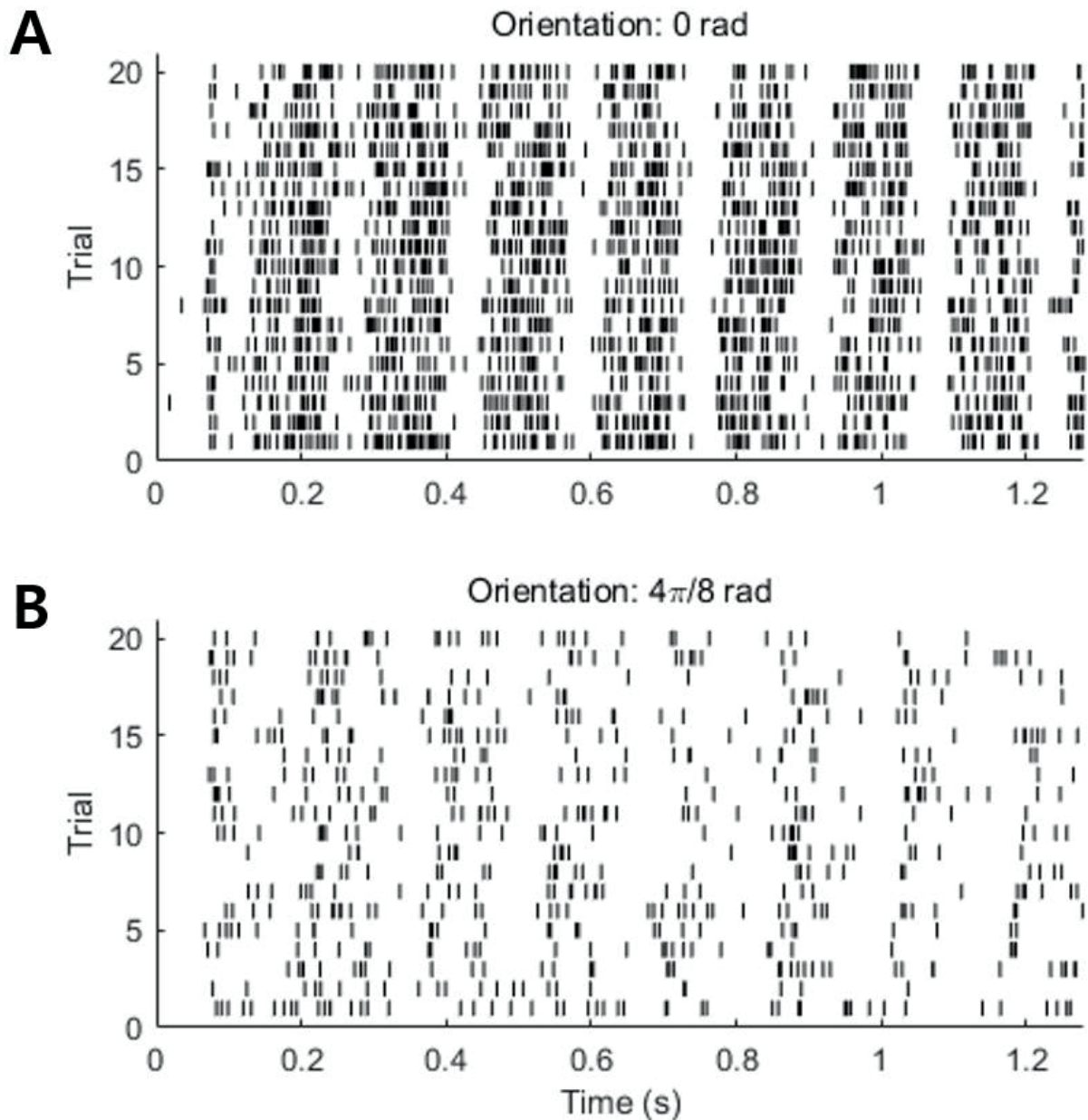
### 2.1. Rate coding and temporal coding

A neuronal action potential is also called a spike. Rate coding claims that information for the external world is coded by the spike rate (firing rate) in the brain (deCharms and Zador, 2000). This is perhaps the most commonly accepted view. An extreme version of rate coding is that information is coded only at the spike rate (Adrian, 1928). Whereas temporal coding asserts that the spike timing also codes information (deCharms and Zador, 2000). For example, the spike timing pattern can encode information (Tabuchi et al., 2018). Currently, not only rate coding but also temporal coding is widely accepted (Butts et al., 2007; Gollisch and Meister, 2008; Johansson and Flanagan, 2009).

Examples of information on rate coding and temporal coding can be found in figure 1 and 2. For figure 1 and 2, the extracellular unit recording dataset (Zandvakili and Kohn, 2015; Semedo et al., 2019; Zandvakili and Kohn, 2019) was used. The dataset was publicly available at [https://crcns.org/datasets/vc/v1v2-1/about\\_v1v2-1](https://crcns.org/datasets/vc/v1v2-1/about_v1v2-1). The dataset contains the extracellular unit recording data which recorded in the primary visual cortex (V1) of the anesthetized monkey when the oriented grating stimuli were presented. Grating stimuli had eight orientations:  $0, \pi/8, 2\pi/8, 3\pi/8, 4\pi/8, 5\pi/8, 6\pi/8$ , and  $7\pi/8$  (rad). For more experimental information, see Zandvakili and Kohn (2015). We used the 1071003p143 data, which contains 111 V1 units.

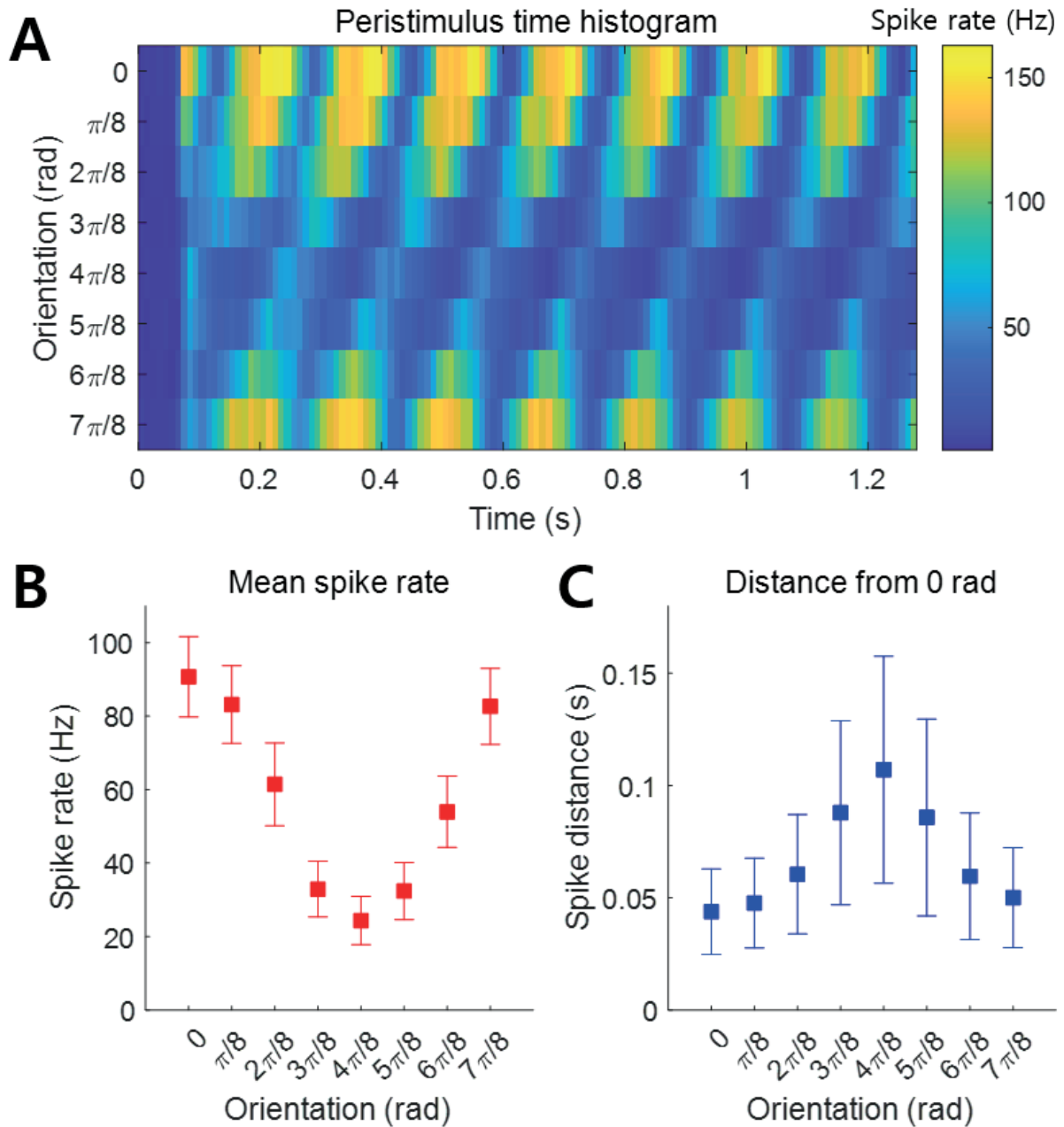
A spike train is the sequence of spike timings. Figure 1 shows the spike trains of unit #20 while the monkey is exposed to two grating stimuli in the  $0$  and  $4\pi/8$  rad orientations, respectively. It can be seen that the spike trains for the two orientation stimuli are different in both spike rate and spike timing. This can be more easily seen in the peristimulus time histogram (trial-averaged spike trains) of the same unit (Figure 2A). As the orientation gradually changes, the spike rate also gradually changes, and the

spike timing shifts. To quantify information on rate coding, the spike count was used (Figure 2B). To quantify information on temporal coding, the spike distance which is invariant to the spike rate changes was used (Figure 2C). This spike distance, is based on the earth mover's distance, will be covered in the section 2.2.1. The unit which was used in figure 1 and 2 is an example of both rate coding and temporal coding.



**Figure 1. Spike trains for grating stimuli.** (A) First 20 spike trains for the 0 rad orientation stimuli. (B) First 20 spike trains for the  $4\pi/8$  rad orientation stimuli.





**Figure 2. Example of both rate coding and temporal coding.** (A) The peristimulus time histogram for the eight orientation stimuli. (B) The spike counts of spike trains for each stimulus, which indicates rate coding. Filled square indicates the mean. Scale bar indicates the standard deviation. (C) The spike distances between spike trains from 0 rad stimulus to other stimuli, which indicates temporal coding. Filled square indicates the mean. Scale bar indicates the standard deviation.



## 2.2. Spike distances: measures for rate coding and temporal coding

A spike distance is a dissimilarity measure between two or more spike trains. The lower value of the spike distance indicates that the two spike trains are similar. The higher value of the spike distance indicates that the two spike trains are dissimilar. The measurement of spike distances can be designed to represent rate coding, temporal coding, or both (Satuvuori and Kreuz, 2018).

Several methods to measure a spike distance have been proposed. Victor and Purpura introduced a cost-based distance that assigns a cost to shifting, adding, or deleting a spike (Victor and Purpura 1996). In this method, the spike distance is defined as the minimum of all possible sums of costs. The Victor and Purpura distance depends on a time-scale parameter where the smaller value of the time-scale parameter emphasizes temporal coding while the larger value does rate coding. van Rossum (2001) also developed a spike distance that measures a difference between spike trains convolved with exponential functions. Most spike distances are rate-sensitive, reflecting an overall rate difference between spike trains to a certain extent even with an extreme time-scale parameter (Satuvuori and Kreuz, 2018). Here, the overall rate denotes the total number of spikes in a spike train divided by the time length of the train. If one aims to measure a distance between a pair of spike trains independent of the overall rate difference, which we call as purely timing-sensitive, the distance should reflect only a difference of spike timing distributions, no matter how different the overall spike rate is between trains.

A purely timing-sensitive spike distance is important to neuroscience studies on temporal coding, which assumes that neurons code information in spike timing patterns (Tabuchi et al., 2018). If a spike distance is rate-sensitive, it would be difficult to clarify whether a given result from a neural spike train analysis is based only on the temporal information. It has been suggested that precise spike timing plays a crucial role in neural information processing (Butts et al., 2007; Gollisch and Meister, 2008; Johansson and Flanagan, 2009).

Kreuz et al. developed the rate-sensitive ISI-distance, a spike distance based on a ratio between the inter-spike intervals of two spike trains (Kreuz et al. 2007). This was followed by the SPIKE-distance, a complementary distance which is still sensitive to rates but with a heightened sensitivity to spike timing (Kreuz et al. 2013). Finally, by removing rate dependence from the SPIKE-distance, Satuvuori et al. (2017) proposed the RI-SPIKE-distance as a distance purely sensitive to timing. The spike distances developed so far have been used in a number of studies for the analysis of neural firing patterns (MacLeod et al., 1998; Mechler et al., 1998; Victor and Purpura, 1998; Machens et al., 2001; Di Lorenzo and Victor, 2003; Narayan et al., 2006; Wang et al., 2007; Reich et al., 2001; Harvey et al., 2013;

Fukushima et al., 2015; Logiaco et al., 2015; Vargas-Irwin et al., 2015; Jamali et al., 2016; Krause et al., 2017).

### **2.2.1. Spike distance focused on temporal coding**

#### 2.2.1.A. Introduction: a spike distance based on the earth mover's distance

If one wants to analyze data by focusing only on rate coding, one can simply use the spike count instead of the spike distance. If rate coding and temporal coding are to be considered together, an existing spike distance such as the Victor and Purpura distance can be used with the time-scale parameter which regulates rate-temporal coding (Victor and Purpura 1996).

If one wants to focus only on temporal coding, it is necessary to use a rate-invariant spike distance. Unfortunately most spike distances are rate-sensitive (Satuvuori and Kreuz, 2018). Although Satuvuori et al. (2017) proposed the RI-SPIKE-distance as a distance purely sensitive to timing, it focuses on the local similarity between spike trains. Victor and Purpura distance measures the global similarity between spike trains; however it is rate-sensitive even in the extreme time-scale parameter (Satuvuori and Kreuz, 2018). In the present study, therefore, we adopt the earth mover's distance (EMD) to measure spike distance with a unique shortest length.

The EMD is also called the Wasserstein metric, which defines the distance between a pair of probability distributions. Here, a metric refers to a distance satisfying nonnegativity, symmetry and the triangle inequality. It measures the minimal cost based on an underlying distance taken to transfer from a probability distribution to another. It initially dealt with transportation problems (Kantorovich, 1940) and later modified toward today's form (Vaserstein, 1969). The EMD also has been implemented as an algorithm in the field of computer science for the comparison between two images (Rubner et al., 2000). The main idea underlying the EMD is that the shortest distance between two objects is equal to the length of the shortest delivery path from one object to the other. For neural spike data, delivery in a spike train operates by moving a part of the spike train from one location to another, with a goal to match one spike train with the other. A delivery path length is then calculated by summing the delivery distance between two locations multiplied by the amount of a delivered part. If we deal with a spike train as a distribution with a sum of 1, then the EMD measures a unique shortest distance between a pair of spike trains in a non-parametric way. A notable difference of the EMD from that of Victor and Purpura (Victor and Purpura 1996) is that delivery in the method of Victor and Purpura moves an entire

spike at once while delivery in the EMD can move a part of a spike.

Two different spike trains may contain a different number of spikes. However, the total number of spikes of each spike train should be equalized to measure the distance between them based only on shifting spikes in time. Victor and Purpura (1996) solved this problem by assigning a cost to adding/deleting a spike and to shifting a spike in time. However, this solution cannot produce a unique distance because it varies with the ratio of two different costs. To address this shortcoming, in the proposed method, we first define a spike train in which each spike is assigned a fixed quantity of 1. Then, we normalized individual spikes by the total number of spikes,  $N$ , so that each spike's quantity becomes  $1/N$  after normalization. For the normalization, we consider a spike train as a function  $f$  of time  $t$  such that

$$f(x) = \begin{cases} \frac{1}{N}, & \text{if a spike occurs at time } t \\ 0, & \text{otherwise} \end{cases} \quad (2.2.1.1)$$

where  $N$  is the number of spikes in the spike train. The overall summation of  $f$  must be one except the case of  $N = 0$ . Hereafter, a spike train will be expressed as functions  $f$  or  $g$ .

In our method, the EMD between  $f$  and  $g$  proposed in Rubner et al. (2000) was adjusted for one-dimensional data (i.e. a spike train) with a constraint that the sum of  $f$  or  $g$  should be equal to 1 (Rubner et al., 2000). The EMD is described as follows. We first rewrite the spike trains,  $f = \{(x_1, 1/N), (x_2, 1/N), \dots, (x_N, 1/N)\}$  and  $g = \{(y_1, 1/M), (y_2, 1/M), \dots, (y_M, 1/M)\}$  from Equation (1) where  $x_i$  and  $y_j$  are a sequence of spike timings. Let  $d(x_i, y_j)$  be an absolute difference between two spike timings  $x_i$  and  $y_j$ . Let  $\xi_{ij}$  be a flow (amount of delivery) from  $x_i$  to  $y_j$  and let  $\Xi = [\xi_{ij}]$  be a matrix of these flows (amount of deliveries) such that it transports  $f$  to  $g$  satisfying the following conditions: 1)  $\xi_{ij}$  is nonnegative; 2)  $\sum_{i=1}^N \xi_{ij} \leq 1/M$ ,  $\sum_{j=1}^M \xi_{ij} \leq 1/N$ ; and 3)  $\sum_{i=1}^N \sum_{j=1}^M \xi_{ij} = 1$ . Condition 1 fixes the direction of the delivery from  $i$  to  $j$ . Condition 2 indicates an effective delivery in the sense that it does not take back what has been delivered. Condition 3 indicates that it delivers the entire spike train. The transportation here means that it makes  $f$  equal to  $g$  by moving parts of  $f$ . Then, the EMD between  $f$  and  $g$  is given by

$$\text{EMD}(f, g) = \min\{\sum_{i=1}^N \sum_{j=1}^M d(x_i, y_j) \xi_{ij} : \Xi = [\xi_{ij}] \text{ satisfies conditions above}\}. \quad (2.2.1.2)$$

This concept of spike distance is illustrated in Figure 3A.

When  $N = 0$  (i.e. no spike in the train), the proposed method cannot calculate the distance directly. However, it can deal with such a case indirectly if we consider a spike train with no spike similar to a spike train with spikes everywhere so that the distance of it to any other trains becomes irrelevant to a certain spike timing pattern. Let  $f_0$  be a spike train with no spike and let  $g$  be another spike train to

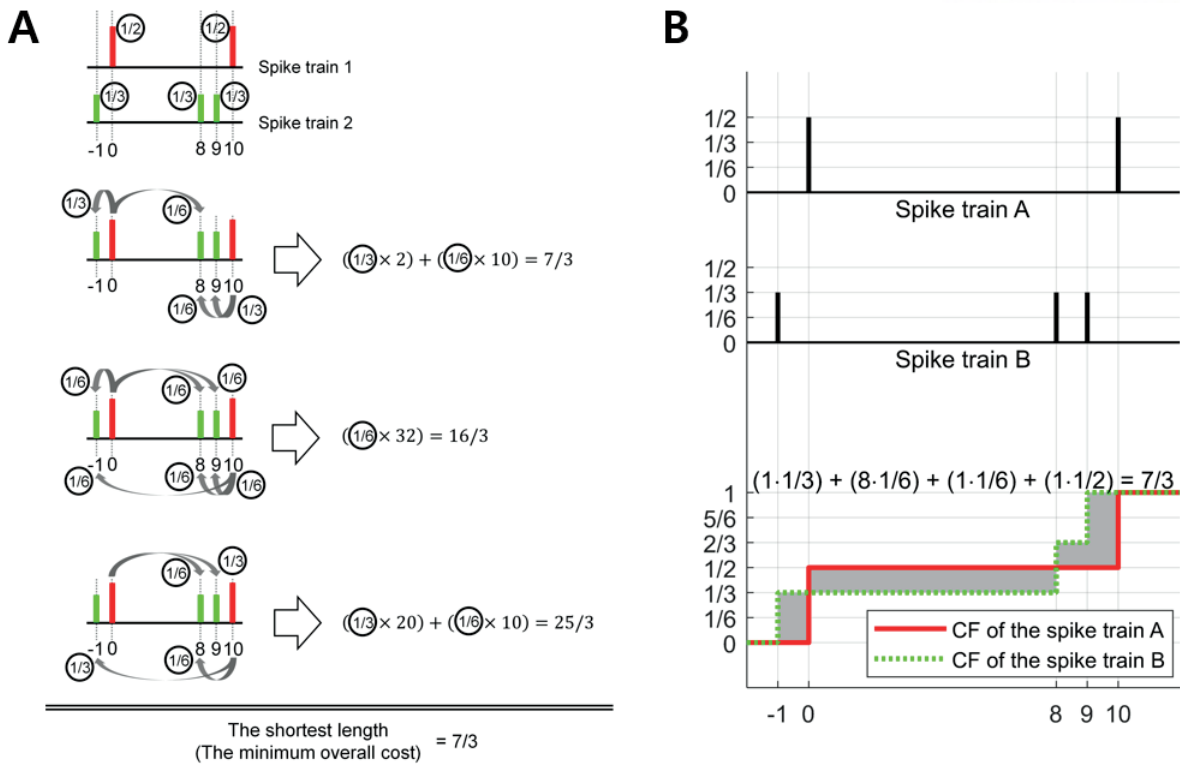
be compared. To calculate  $d(f_0, g)$ , let  $f_n$  be a spike train with  $n$  spikes generated from a uniform probability distribution defined on a certain bounded analysis domain. The bounded analysis domain prevents the distance from increasing to infinity, although the distance measurement depends on how the analysis domain is determined. Then, the EMD calculates  $d(f_0, g) = \lim_{n \rightarrow \infty} E(d(f_n, g))$  where  $E(\cdot)$  indicates an expected value. To deal with an empty spike train in the EMD, we attended to an idea that there was also no information about spike timing if spikes are everywhere, uniformly distributed. It means that a spike train with one spike at a specific location holds more information about spike timing than a spike train with uniformly distributed spikes. In this regard, an empty spike train would be more similar to a spike train with uniformly distributed spikes at every location than a spike train with one spike.

The EMD is a mathematical metric, that is, it satisfies the three conditions: nonnegativity, symmetry and the triangle inequality (Rubner et al., 2000). This property shows that the EMD conforms to our intuition about distance. Moreover, from the fact that the EMD is calculated solely based on spike timing data, it can be seen that the EMD is the shortest length based on spike timing between two spike trains. The EMD is calculated in a non-parametric way so that it produces a unique value. Due to its non-parametric approach, the EMD can avoid the dependency of distance outcomes on parameters.

Moreover, there is an efficient way to calculate the restricted version of the EMD as follows (Cohen, 1999). Let  $F$  and  $G$  be the cumulative functions of  $f$  and  $g$ , respectively. Then, the EMD is given by

$$\text{EMD}(f, g) = \int_{-\infty}^{\infty} |F(t) - G(t)| dt. \quad (2.2.1.3)$$

An example of the calculation procedure above is illustrated in Figure 3B.



**Figure 3. Calculation of the spike distance based on the earth mover's distance.** (A) Illustration of the basic concept of the earth mover's distance (EMD) to measure a distance between spike trains. The objective is to take the smallest value among all possible delivery (flow, terminology in EMD) path lengths between two objects (red and green). In this illustrative example, the red object (e.g., spikes) is delivered to the green object in three possible paths. Assuming the size of the red object is 1, the path length is calculated by delivery distance (in time) multiplied by the amount of delivery (i.e., size of the object). It is also possible to deliver only a fraction of the object, as shown in the third case. If there are multiple deliveries toward the target object, the final delivery path length is the sum of individual delivery path lengths. (B) Illustration of the calculation procedure of the EMD described in Cohen (1999). The distance between two spike trains, A and B, is calculated. Initially, the nonnegative values are assigned to every spike such that the sum of the values in each train is equal to 1 (e.g., 1/2 for each spike in A or 1/3 for each in B). The next step is to produce the cumulative functions (CF) for each spike train (red bold line indicates the CF of spike train A and green dotted line indicates the CF of spike train B). The next step is to integrate the absolute difference between the two CFs (gray shading area). The final result of the calculation procedure is  $7/3$ .

### 2.2.1.B. Evaluation

Our new spike distance was compared to four existing spike distances: 1) the Victor-Purpura distance (Victor and Purpura, 1996) with parameter values,  $q = 0.1, 0.2, \dots, 12.8 \text{ s}^{-1}$ ; 2) the van Rossum distance (van Rossum, 2001) with parameter values,  $\tau = 1, 2, \dots, 16 \text{ s}$ . Note that an alternative calculation method (Houghton and Kreuz, 2012) was used here instead of the original one (van Rossum, 2001); 3) the SPIKE-distance (Kreuz et al., 2013); and 4) the RI-SPIKE-distance (Satuvuori et al., 2017).

The tested time-scale parameters of the Victor-Purpura distance and the van Rossum distance were determined as follows. For the Victor-Purpura distance parameter  $q$ , the time range of a spike train in which we performed the analysis was set to 0 - 10 s. Then we opted for values of  $q$  varying between two opposite cases:  $q = 0.1 \text{ s}^{-1}$  and  $q = 12.8 \text{ s}^{-1}$ . The smallest  $q = 0.1 \text{ s}^{-1}$  in the Victor-Purpura distance made the metric focus on a “spike timing shift” by assigning a cost of 1 to add/delete each spike, whereas it costed at most  $(q = 0.1 \text{ s}^{-1}) \times (10 \text{ s}) = 1$  for time-shifting a spike. Then, the value of  $q$  was increased by a factor of 2 up to the largest  $q = 12.8 \text{ s}^{-1}$ , which turned the algorithm to focus on “spike adding/deleting” by increasing the cost for time-shifting such as  $(q = 12.8 \text{ s}^{-1}) \times (1 \text{ s}) = 12.8$  even for shifting a spike by 1 s.

Similarly, for the van Rossum distance, the smallest value of  $\tau = 1 \text{ s}$  makes the convolved range narrow by setting the width of the exponential function to 1 s. Then, the value of  $\tau$  was increased by a factor of 2 up to the largest value of  $\tau = 16 \text{ s}$ , which makes the convolved range cover the overall spike train by setting the width of the exponential function to 16 s.

Taking spike counts into dissimilarity is a key difference between the EMD and the Victor-Purpura distance or the van Rossum distance. In fact, while the EMD is focused on temporal coding, both the Victor-Purpura distance and the van Rossum distance cover from a mixture of temporal coding and rate coding to pure rate coding by varying the time scale parameter  $q$  or  $\tau$ , as they are so designed originally. We demonstrated such differences between the EMD and the Victor-Purpura distance or the van Rossum distance in the simulations.

A comparison of the five spike distances was conducted to assess how well each distance represented two aspects of similarity between spike trains: spike timing difference, temporal similarity. Furthermore, each distance’s robustness to changes in firing rates was examined for temporal similarity.

To avoid potential errors while replicating the existing distance calculation procedures, we directly utilized the available source code for each distance. The code to calculate the Victor-Purpura distance was obtained from <http://www-users.med.cornell.edu/~jdvicto/spkdm.html>. The code for the van Rossum distance was from <http://wwwold.fi.isc.cnr.it/users/thomas.kreuz/images/vanRossum.m>. The codes for both the SPIKE-distance and the RI-SPIKE-distance were from

<http://wwwold.fi.isc.cnr.it/users/thomas.kreuz/Source-Code/cSPIKE.html>.

For the calculation of the SPIKE-distance and the RI-SPIKE-distance, we always set the time range of the underlying dissimilarity profiles exactly equal to the spike generation interval.

#### 2.2.1.B.1. Spike timing difference

A pair of spike trains with three spikes each was synthesized to test spike timing difference. The locations of the 1<sup>st</sup> and 3<sup>rd</sup> spikes were fixed and matched between the trains. The 2<sup>nd</sup> spike of the first train was fixed close to the 1<sup>st</sup> spike. Then, the location of the 2<sup>nd</sup> spike of the second train was moved toward the 3<sup>rd</sup> spike. This test paradigm was performed in the previous study by Kreuz et al. (2011) to compare several distances. We adopted it here with the inclusion of the van Rossum distance, the RI-SPIKE-distance, and the EMD. In the test, we located the 1<sup>st</sup> spike at 0 s and the 3<sup>rd</sup> at 10 s in the two trains. The 2<sup>nd</sup> spike of the first train was fixed at 1 s. Then, the 2<sup>nd</sup> spike of the second train was moved from 1 s to 9 s in steps of 1 s (see Figure 4A). We measured the distance for each shift of the 2<sup>nd</sup> spike of the second train.

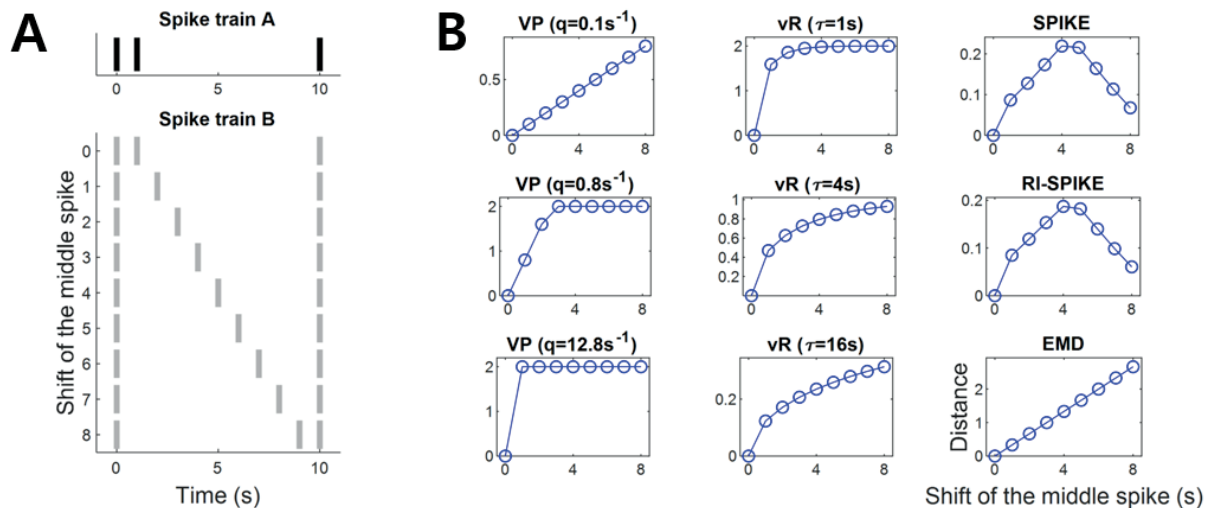
The spike distance measurements exhibited differences among the five spike distances tested in this study. The Victor-Purpura distance linearly increased as the spike timing difference increased with one parameter value ( $q = 0.1$  s), but was saturated with the other parameter values ( $q = 0.8, 12.8$  s). Similarly, the van Rossum distance monotonically increased as the spike timing difference increased with one parameter ( $\tau = 16$  s), but was saturated with another parameter ( $\tau = 1, 4$  s). Both the SPIKE-distance and the RI-SPIKE-distance increased first but later decreased as the spike timing difference increased. The EMD linearly increased as the spike timing difference increased (Figure 4B).

Also, we observed that the SPIKE-distance and the RI-SPIKE-distance consider the spike trains to be more similar if a middle spike is close to the edge spikes than if the middle spike is located at an equal distance from both edge spikes. The reason is that these methods focus on the local dissimilarity between spike trains. Two spike trains are locally similar when the middle spike is close to the edge spikes since then it becomes easier to see it as part of a doublet that together is quite synchronous with the single spike in the other spike train. In contrast, if the middle spike is located at an equal distance from both edge spikes, then the distance to the nearest spike in the other train is maximized, increasing local dissimilarity. The van Rossum distance seems to evaluate a similarity of two spike trains based on synchronization of spike timings within a certain temporal range, where the temporal range was determined by the time-scale parameter  $\tau$ . Then, if two spikes from each spike train occurred within the temporal range, these spikes were deemed to be synchronized. The Victor-Purpura distance with the



parameter  $q = 0.1$  s (i.e. emphasizing temporal differences) and the EMD linearly increase as the difference of middle spikes is linearly increased, because these methods focus on equalizing two spike trains. Hence, for instance, if a difference in the latency of neural responses between spike trains is of interest, the Victor-Purpura distance with a small  $q$  and the EMD can provide an appropriate measure.

The characteristics of distances for small spike timing differences (for example, the middle spike is shifted by 0, 1, or 2 in Figure 4B) can provide information about temporal precision of the spike timing. The Victor-Purpura distance ( $q = 0.1$  s) and the EMD are linearly decreasing when the spike timing difference converges to zero. This linear property allows them to have the information about temporal precision, but with no conclusive answer to whether a timing difference between spike trains is precise or not. On the other hand, the van Rossum distance, the SPIKE-distance and the RI-SPIKE-distance are rapidly decreasing when the spike timing difference is nearing zero so that they can provide precise information whether timing difference falls within some range or not.



**Figure 4. Spike distance results for the measurement of spike timing differences.** (A) Spike train A is fixed whereas spike train B is changed as the location of a middle spike is shifted from left to right so that the spike timing difference between A and B increases linearly. (B) The spike distance results of each of the five distances: the Victor-Purpura distance, the van Rossum distance, the SPIKE-distance, the RI-SPIKE-distance, and the EMD. The horizontal axis represents the amount of the shift of the middle spike in train B. The Victor-Purpura distance showed a linear increase in distance only for a certain parameter (e.g.,  $q = 0.1$  s). The van Rossum distance showed an increase in distance nonlinearly but monotonically. The SPIKE-distance and the RI-SPIKE-distance did not show monotone increases. The EMD showed a linear increase as the spike timing difference increased. VP: Victor-Purpura



distance. vR: van Rossum distance.

### 2.2.1.B.2. Temporal similarity and robustness to firing rate change

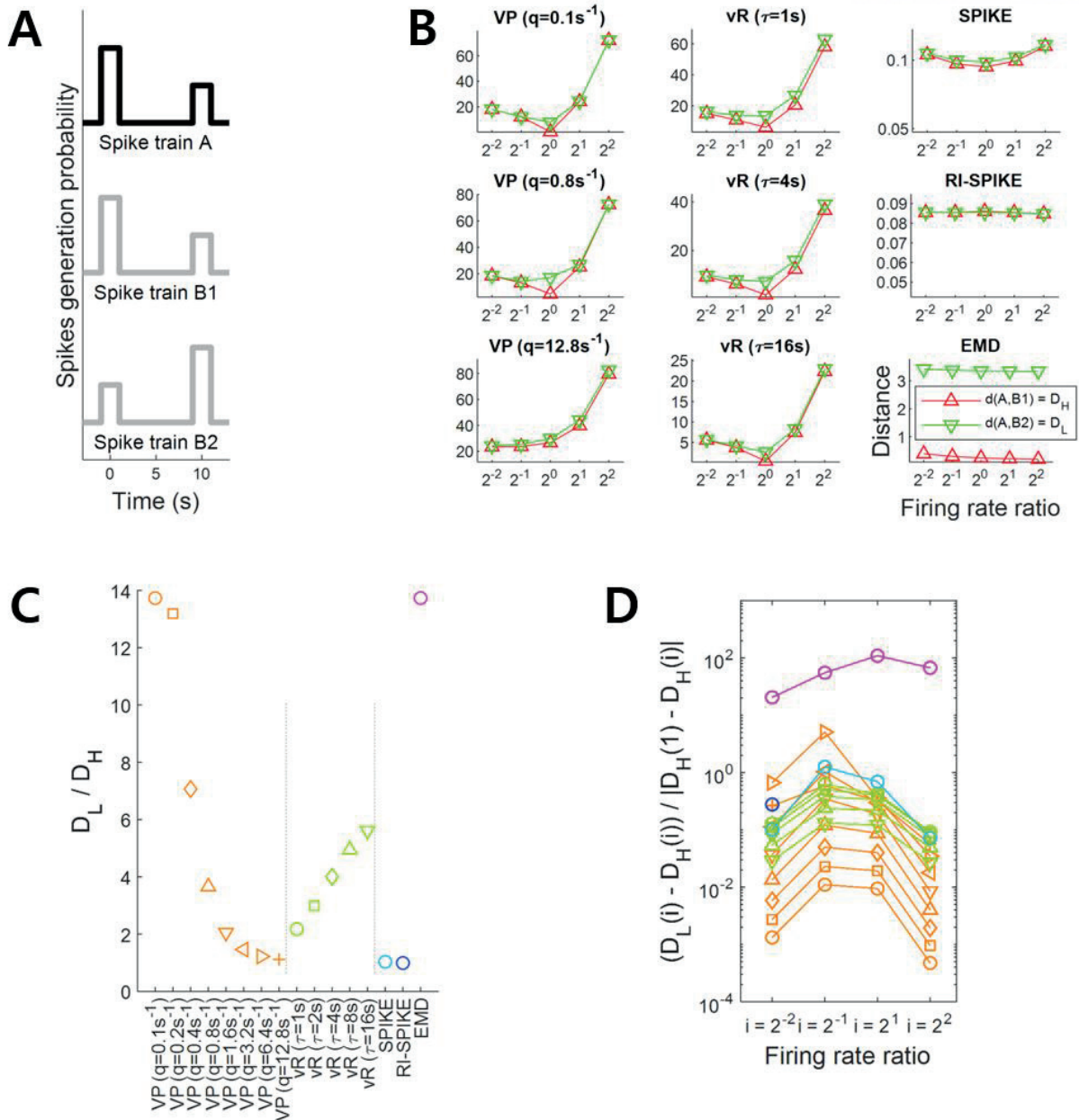
A simulation experiment was performed to test the robustness of each distance against firing rate changes when measuring temporal similarity between spike trains. Spike trains were generated according to a simple probabilistic model. The probabilistic model was built following a certain firing rate profile. Temporal similarity would increase if a pair of spike trains were generated from a probabilistic model sharing a similar profile and decrease if the profiles become more dissimilar. Note that temporal similarity describe here depends only upon firing rate profiles, not firing rates themselves. The probabilistic model used here consisted of two intervals where each interval had a nonzero probability of containing a spike. Spikes in the intervals were randomly generated from a uniform distribution centered at 0 s and 10 s with a halfwidth of 1 s. Then, we built three spike trains denoted as spike trains A, B1, and B2. In the probabilistic model of spike train A, the probability of generating a spike in the 1<sup>st</sup> interval was twice as high as that in the 2<sup>nd</sup> interval. Spike train B1 had the same probabilistic model as spike train A. On the other hand, it was reversed in spike train B2 such that the probability of generating a spike in the 2<sup>nd</sup> interval was twice that in the 1<sup>st</sup> interval (see Figure 5A). Hence, the distance between A and B1 should be smaller than that between A and B2, because temporal patterns would be more similar between A and B1 than between A and B2.

To test the robustness of the distances against firing rate changes, we varied the number of spikes in the trains. We first set the number of spikes in A to  $2^3 \times 3$ , where  $2^3$  spikes were generated three times (twice in the 1<sup>st</sup> interval and once in the 2<sup>nd</sup> interval). Then, five levels of the number of spikes were used to vary the firing rates in B1 or B2. The number of spikes in B1 or B2 was varied as  $2^1 \times 3$ ,  $2^2 \times 3$ ,  $2^3 \times 3$ ,  $2^4 \times 3$ , and  $2^5 \times 3$ , making the spike count ratios of A to B1 or B2  $2^{-2}$ ,  $2^{-1}$ ,  $2^0$ ,  $2^1$ , and  $2^2$ . If a spike distance is robust to firing rate changes, distance variability over all the ratios should be negligible compared to the difference in distance between A to B1 and between A and B2. We calculated the difference in distances between these two pairs (A & B1, A & B2) using each of the five distances by varying the firing rates in B1 or B2.

We evaluated distance measurements between a pair of spike trains with a high or low temporal similarity when the ratio of the firing rates between the trains varied. Let  $D_L$  be a distance with a low temporal similarity (i.e., between A and B2) and  $D_H$  be a distance with a high temporal similarity (i.e., between A and B1). First, we calculated the ratios of  $D_L$  to  $D_H$  from each distance for the case when the firing rates of two spike trains were equal, and the result is summarized in Figure 5C. The Victor-

Purpura distance ( $q = 0.1, 0.8$  s), the van Rossum distance ( $\tau = 4, 16$  s), and the EMD clearly resulted in a smaller distance with a high temporal similarity than with a low temporal similarity (Figure 5B). These low and high temporal similarities reflect the global difference between two spike trains in Figure 5A, not the local difference. The spike trains B1 and B2 in Figure 5A are globally different, but locally similar (near 0 s and 10 s). Since both the SPIKE-distance and the RI-SPIKE-distance focus on the local difference, these distances show less sensitivities for the discrimination between low and high temporal similarity in a global sense. In contrast, the EMD is a global measurement, showing an ability to discriminate global temporal similarity. Although the RI-SPIKE-distance is robust to firing rate changes just as the EMD is, this point indicates a key difference between the RI-SPIKE-distance and the EMD.

Next, to assess the robustness to firing rate changes when unequal firing rates exist between the spike trains, we calculated the ratio  $(D_L(i) - D_H(i)) / |D_H(1) - D_H(i)|$ , where  $i$  denotes the firing rate ratio of spike train B1 (or B2) to that of spike train A for  $i = 1/4, 1/2, 2, 4$  (e.g.,  $D_H(1/2)$  refers to distance measurements when the firing rate ratio is  $1/2$ ). The distance results for each value of  $i$  are given in Figure 5D. As for the robustness to firing rate changes, the Victor-Purpura distance and the van Rossum distance increased as the ratio of the firing rates deviated from 1, which indicates that variability in the distance across the firing rate ratios was larger than the difference in distances between high and low temporal similarities, revealing that the distances were not robust to firing rate changes. This was not the case for the SPIKE-distance and the RI-SPIKE-distance, where the distances remained at similar levels across the ratios of firing rates although variability in the distance across the ratios was larger than the difference in distances between high and low temporal similarities, showing that they were also not robust to firing rate changes. On the other hand, the EMD showed that variability in the distance across the ratios was much smaller than the difference in distances between high and low temporal similarities, demonstrating its robustness to firing rate changes (Figure 5B). The SPIKE-distance, the RI-SPIKE-distance, and the EMD showed the robustness to firing rate changes relative to the Victor-Purpura distance and the van Rossum distance. It implies that those three distances are more suitable for temporal coding.



**Figure 5. Spike distance results for the measurement of temporal similarity.** (A) The probabilistic models of spike generation for spike trains A, B1, and B2 are described. In the simulation, spike trains A and B1 share the same probabilistic model whereas spike trains A and B2 have different probabilistic models. Accordingly, the temporal similarity is high between A and B1, but low between A and B2. (B) The spike distance results of each of the five distances as the ratio of firing rates between the spike trains varies from  $2^{-2}$  to  $2^2$ . The red lines represent distances between the spike trains A and B1 and green lines represent those between A and B2. It is clearly shown that the variability of distances by changes in the ratio is larger than that by changes in the temporal similarity for the four distances, including the Victor-Purpura, the van Rossum, the SPIKE- and the RI-SPIKE- distances. In contrast,

the distances calculated by the EMD remain almost unchanged as the ratio changes, being robust to the firing rate change. (C) Results of spike distance for measuring temporal similarity.  $D_L$  is a distance with a low temporal similarity, and  $D_H$  is a distance with a high temporal similarity. (D) Quantification of robustness as the firing rate changes.  $D_L(i)$  is the distance with a low temporal similarity when the firing rate ratio is  $i$ , and  $D_H(i)$  is the distance with a high temporal similarity when the firing rate ratio is  $i$ . The results of the RI-SPIKE-distance partly disappear because of negative values. VP: Victor-Purpura distance. vR: van Rossum distance.

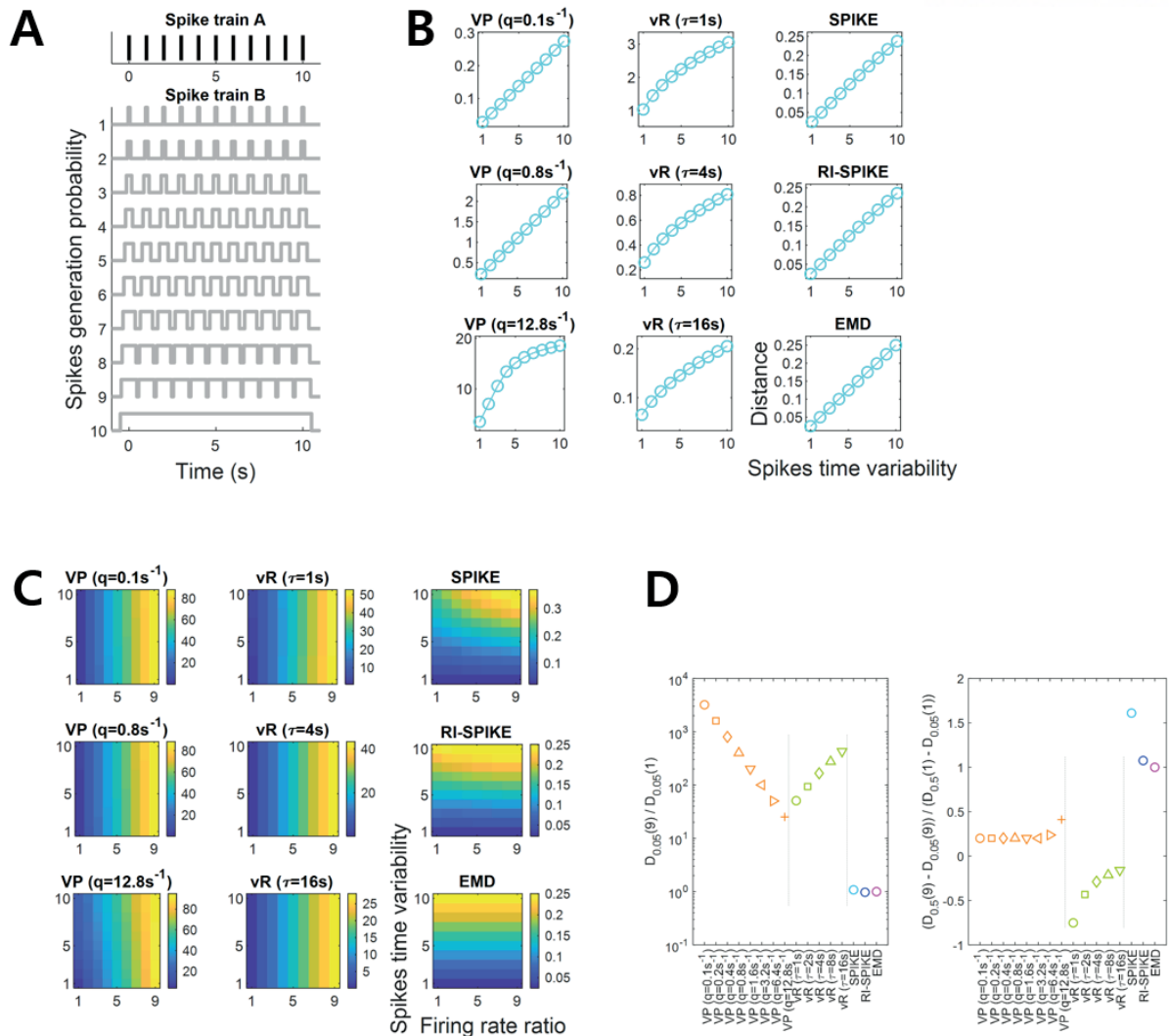
### 2.2.1.B.3. Spike time synchrony and robustness to firing rate change

Another simulation experiment was performed to test the robustness of each distance against firing rate changes when measuring spike time synchrony between spike trains. To this end, a pair of spike trains, denoted as A and B were synthesized. Spike train A was generated to contain eleven equally spaced spikes discharged at 0 s, 1 s, ..., 10 s. Spike train B was generated according to a probabilistic model, consisting of eleven uniform distributions centered at 0 s, 1 s, ..., 10 s. Then, we varied the halfwidth of these uniform distributions across ten levels to manipulate the degrees of spike timing jitter; the halfwidth was set as 0.05 s, 0.1 s, ..., or 0.5 s (see Figure 6A). As the halfwidth was increased, spike timing jitter increased, which was likely to desynchronize spike timing more between A and B. It would then result in an increase in the distance between A and B.

Similar to Section 2.2.1.B.2, we varied the number of spikes in B to test the robustness of the distance to firing rate change. The number of spikes in B varied across nine levels to reflect firing rate changes. It varied as  $1 \times 11$ ,  $2 \times 11$ , ..., and  $9 \times 11$  (the first number in the product indicates the number of spikes randomly generated in each interval of B) so that the ratios of A to B became 1, 2, ..., and 9, respectively. We expected that if the spike distance was robust to firing rate changes, variability in the distance across the ratios should be negligible compared to variability in distance according to different degrees of spike timing jitter. We calculated the distances between A and B for each degree of spike timing jitter for each firing rate level in B.

Spike distances with various synchrony levels were measured using each of the five distances and their robustness to firing rate changes was tested. Every distance clearly showed a similar pattern when the ratio of firing rates was 1 such that the spike distance increased as the degree of spike timing jitter increased (Figure 6B). To assess the robustness to firing rate changes, we quantified the effect of the firing rate ratio on the spike train. Here, let  $D_k(n)$  be the distance when the firing rate ratio of spike train B to spike train A was  $n$ , where  $k$  denotes the halfwidth of the uniform distribution in B. We first

calculated the ratio  $D_{0.05}(9) / D_{0.05}(1)$  using each distance and obtained the results as summarized in the left figure of Figure 6D. Next, we calculated the ratio  $(D_{0.5}(9) - D_{0.05}(9)) / (D_{0.5}(1) - D_{0.05}(1))$  using each distance and obtained results, which are listed in the right figure of Figure 6D. A comparison of these two ratios showed that when the firing rate ratio increased, the Victor-Purpura distance and the van Rossum distance increased rapidly, whereas other distances were almost unchanged. In other words, by using the Victor-Purpura distance and the van Rossum distance, variability in distance across the firing rate ratios was larger than variability in distance due to different degrees of spike timing jitter, showing that the distances were not robust to firing rate changes. The SPIKE-distance, the RI-SPIKE-distance, and the EMD revealed that variability in distance across the ratios was smaller than that among different levels of synchrony, demonstrating that they were robust to firing rate changes. Moreover, the RI-SPIKE-distance and the EMD appeared to be most robust (Figure 6C). These results indicate that the Victor-Purpura distance and the van Rossum distance are suitable to measure the dissimilarity due to both rate difference and temporal synchrony. The SPIKE-distance is also suitable to measure the dissimilarity in both rate difference and temporal synchrony although it seems to be less sensitive to rate difference than the Victor-Purpura distance and the van Rossum distance. On the other hand, the RI-SPIKE-distance and the EMD are suitable to measure temporal synchrony, insensitive to rate differences.



**Figure 6. Spike distance results for the measurement of spike time synchrony.** (A) Spike train A is fixed to have equally spaced eleven spikes. Spike train B is generated by a probabilistic model of spike generation with various spike timing jitter. The spike timing jitter is manipulated by increasing the halfwidth of eleven uniform distributions each centered at the spike timing of train A. Spike time synchrony between A and B decreases as spike timing jitter increases. (B) The spike distance results of the five distances as the ratio of firing rates of B over A are equal to 1. All the distances exhibit approximately linear increases with increases in spike timing jitter. (C) The spike distance results of each of the five distances as ratios of firing rates of B over A increase from 1 to 9. The index of the vertical axis corresponds to the index of the spike trains in (A), where increasing index number indicates increasing spike timing jitter. Distances proposed by Victor-Purpura and van Rossum are significantly affected by the variation in the firing rate ratio, whereas those proposed by the SPIKE-distance, the RI-SPIKE-distance and the EMD are not. (D) Results of spike distance for measuring spike time synchrony.



$D_k(n)$  is the distance when the firing rate ratio of one spike train to another was  $n$ , and  $k$  denotes the halfwidth of the uniform distribution in the spike train. VP: Victor-Purpura distance. vR: van Rossum distance.

#### 2.2.1.B.4. Comparison with Victor and Purpura's distance

The spike distance in the present study is closely related to the Victor-Purpura distance. It is important to compare the properties between the Victor-Purpura distance and the EMD. Satuvuori and Kreuz already discussed the suitability of the Victor-Purpura distance to rate and temporal coding (Satuvuori and Kreuz, 2018). They suggested that the Victor-Purpura distance is suitable to rate coding in general, but suitable to temporal coding only for similar firing rates, even with a wide range of time-scale parameter  $q$ . To verify whether the EMD suffered from a similar issue to the Victor-Purpura distance, we applied the analysis of Satuvuori and Kreuz (Satuvuori and Kreuz, 2018) to the EMD. Three spike trains were generated in the analysis. Spike train A was generated to contain one spike discharged at 5 s. Spike train B was generated according to a probabilistic model of a uniform distribution centered at 5 s with the halfwidth of 1 s. Spike train C was also generated according to a probabilistic model of a uniform distribution centered at 5 s with the halfwidth of 5 s. Spike train B had five levels of the number of spikes;  $2^0$ ,  $2^1$ ,  $2^2$ ,  $2^3$  and  $2^4$ . By comparison, spike train C had only one spike as in spike train A (see Figure 7A). From the point of view of temporal coding, it was expected that the distance between A and B was smaller than the distance between A and C and the distance between B and C, because spike trains A and B had more similar temporal information compared to C. The Victor-Purpura distance was examined for time-scale parameters in the range from 0.01 to 1000.

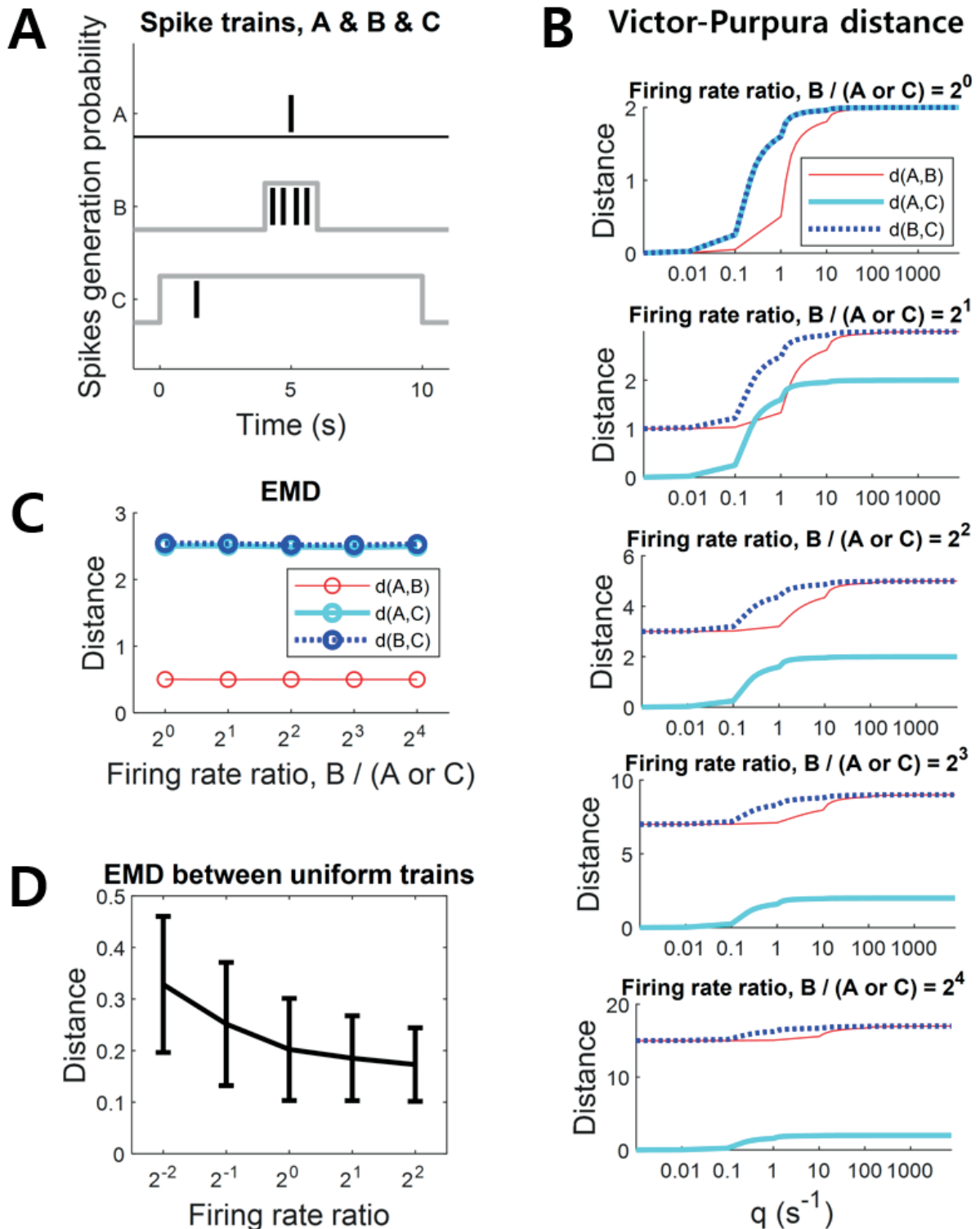
The simulation result for the Victor-Purpura distance in the present study was similar to that in the study by Satuvuori and Kreuz (2018). The expected result was that the distance between the spike trains A and B was smaller than those between A and C and between B and C, because the temporal coding between A and B is more similar than that between other pairs (Figure 7A). When the firing rate ratio of B to A or C was  $2^0$  (i.e. the same firing rates), the Victor-Purpura distance showed the expected result for a wide range of time-scale parameters  $q$  (Figure 7B top). It indicates that the Victor-Purpura distance is suitable for temporal coding if the firing rate ratio is 1. However, as the firing rate ratio of B to A or C increased, the Victor-Purpura distance started to show unexpected results. The distances between A and B and between B and C were increasing for every time-scale parameter  $q$ , reflecting the increased rate difference (Figure 7B). It indicates that the Victor-Purpura distance is not suitable for temporal coding if the firing rate ratio deviates from 1. The smaller value of the time-scale parameter  $q$  emphasizes the temporal coding. However, the result showed that the Victor-Purpura distance is still

rate-sensitive even for a very small value of  $q$ . Therefore, the value of  $q$  apparently changes sensitivity from pure rate coding to combined rate and temporal coding, not to pure temporal coding (Satuvuori and Kreuz, 2018).

On the other hand, the EMD showed the expected results for all tested firing rate ratios. Furthermore, the distances between every pair of spike train remained nearly constant even as the firing rate ratio changed (Figure 7C). It indicates that the EMD is suitable for temporal coding even though the firing rates differ between the spike trains, showing that it does not reflect rate coding. That is, the EMD is sensitive to pure temporal coding in contrast to the Victor-Purpura distance.

Although the EMD is relatively insensitive to firing rate difference than the Victor-Purpura distance, it is uncertain whether the EMD is completely insensitive. In order to test the effect of different rate ratios on the EMD, we calculated the EMD between two Poisson spike trains that were generated uniformly over  $[0, 1]$  s with different rates. The spike trains were generated with firing rates of 1, 2, 4, 8, and 16 Hz. Then, the spike trains with 4 Hz were compared to those with other firing rates (including the identical 4 Hz) so that the firing rate ratios varied over  $2^{-2}$ ,  $2^{-1}$ ,  $2^0$ ,  $2^1$ , and  $2^2$ . The resulting EMD values are provided in Figure 7D. The EMD between trains with the same temporal pattern varied across different firing rate ratios although the EMD variation was much smaller than the firing rate ratios variation.





**Figure 7. Comparison with the Victor-Purpura distance in terms of suitability for temporal coding with different firing rates.** (A) Spike train A has only one spike with fixed timing. Spike train B has five levels of spikes with narrow range spike timing jitters. Spike train C has only on spike with a broad

range spike timing jitter. The desirable expected results are that the distance between spike trains A and B is smaller than the distance between A and C and the distance between B and C. (B) The Victor-Purpura distance with various values of the time-scale parameter  $q$ . The Victor-Purpura distance did not show the desirable result with increases in firing rate ratio. (C) The EMD showed desirable results overall with increases in firing rate ratio, having a nearly constant scale. (D) The EMD between a uniform spike train with different firing rate ratios. It shows that the EMD is not completely insensitive to firing rate differences.

#### 2.2.1.B.5. Application to neural data

We demonstrated the measurement of a temporal similarity between real neuronal spike trains using the EMD. The neural data is publicly available from Flint *et al.* (2012), and can be downloaded from <https://crcns.org/data-sets/movements/dream>. The example of neural spike trains was obtained from the primary motor cortex of a behaving non-human primate (Flint et al. 2012). An example of various levels of temporal similarity measured by the EMD is shown in Figure 8, in which the spike trains observed under the different experimental conditions (i.e., different movement directions of the subject's arm) showed mutually different temporal similarity with the base condition at the arm movement direction of  $45^\circ$  (at which the example neuron fired the most).

A neuron in the primary motor cortex (M1) modulates its firing rates with arm movement directions (Georgopoulos et al., 1982). Arm movements induce a certain temporal pattern such that a spike train of a M1 neuron contains more spikes around movement onset and less spikes before and after movement offset. Also, the firing rate of the neuron is maximal at the preferred direction (PD) of arm movement and decreases gradually when the movement direction deviates farther from the PD (Georgopoulos et al., 1982; Schwartz et al., 1988; Kalaska et al., 1989; Caminiti et al., 1990). Hence, the temporal patterns of spike trains between the PD and other directions are expected to be more dissimilar when the movement direction becomes more different from the PD. We found that the EMD could describe various levels of temporal similarity to the base condition for various directions and specifically showed that distance increased as the angle became orthogonal to the PD. In addition, the EMD on the true data (red lines in the inlet graph of Figure 8A) revealed a clearer difference between the PD and orthogonal angles than that on the surrogate data with randomized spike timings (gray lines of the inlet graph of Figure 8A). Specifically, corresponding to each true spike train, we generated a random spike train by generating spike timings from the uniform distribution while maintaining the number of spikes unchanged. So, if the difference between directions is mainly represented in the number of spikes, the difference between directions should also be maintained in the surrogate data. However, the result

demonstrated that the EMD difference between spike trains of different directions was not merely due to the firing rate difference.

A spike distance shall yield small values between spike trains obtained under similar experimental conditions and large values between spike trains obtained under different conditions. We demonstrated that the EMD satisfied such a criterion using the real neuronal spike data of a non-human primate in Figure 8. In Figure 8, the EMD showed small values when the subject moved the arm in a direction similar to the preferred direction (i.e. similar experimental condition) and large values when the subject moved the arm in a direction dissimilar to the preferred direction (i.e. dissimilar experimental condition) (see red lines in the inset graph of Figure 8A). In particular, the EMD calculated this result based on the temporal pattern rather than on the firing rate difference.

We compared EMD and other spike distances in terms of an ability to distinguish primary motor cortical spike trains with spiking timing information according to the arm movement directions of a nonhuman primate. There were eight equally divided arm movement directions in this 2D center-out arm reaching task. As each spike distance covered a different magnitude scale, each spike distance was normalized by  $D_{new} = (D - D_{min}) / (D_{max} - D_{min})$  so that the distance values ranged between 0 and 1. We selected one of the eight directions as an anchor (e.g. 45°) and measured average pairwise distance using each spike distance measure between a set of spike trains corresponding to the anchor direction and each set of spike trains corresponding to other directions. We found that the EMD well represented differences between spike trains according to movement directions such that the distance is 0 at the PD, 1 at the opposite of the PD, and the intermediate values at other directions (Figure 8B).

We evaluated how the EMD could be used to discriminate the neural spiking patterns of different upper limb movement directions represented in the primary motor cortex (M1) of a non-human primate (Flint et al. 2012). The non-human primate moved the upper limb in eight different directions while spiking timings of the population of M1 neurons were recorded. There were multiple trials of this task in each direction. As the duration of movements varied across trials, we selected an 1-s epoch after the onset of a go cue. Before spike distance computation using various methods including the EMD, we normalized the overall spike count of every spike train in order to assess each method's ability to extract movement-related information only from spiking timing patterns. This normalization was performed based on resampling – i.e. randomly selecting a certain number of spikes from the original spike train. In this manner, every resampled spike train could have the same number of spikes for every direction while retaining the temporal pattern of the original spike trains.

For resampling, we first selected 113 out of 196 M1 neurons, which fired spikes enough to produce spike trains suitable for our distance analysis (a neuron was selected if it fired  $\geq 50$  spikes within the 1-

s epoch on average for each direction). For each selected neuron, we randomly chose  $R$  spikes from the original spike train, repeating this resampling for every spike train of every direction for that neuron. The number of spikes in a resampled spike train,  $R$ , was stochastically determined by generating a random number from the Poisson distribution with the mean rate of 10. The mean rate of 10 was chosen such that the largest number generated from the Poisson distribution with this mean rate was unlikely to exceed the half of 50 (i.e. 25), in order to make resampled spike trains vary over trials. This ensured that the expected number of spikes in every resampled train in every direction was identical, while allowing trial-to-trial variability. Once the resampled spike train was generated, we multiplied 10 to its spike timings to change the spike train range from  $[0, 1]$  s to  $[0, 10]$  s, in order to adjust the range adequate for pre-defined time-scale parameters of the Victor-Purpura distance and the van Rossum distance. Also, as the SPIKE-distance and the RI-SPIKE-distance calculate the distance in a range from the first spike to the last spike, we added two auxiliary spikes at 0 s and 10 s (Figure 9A).

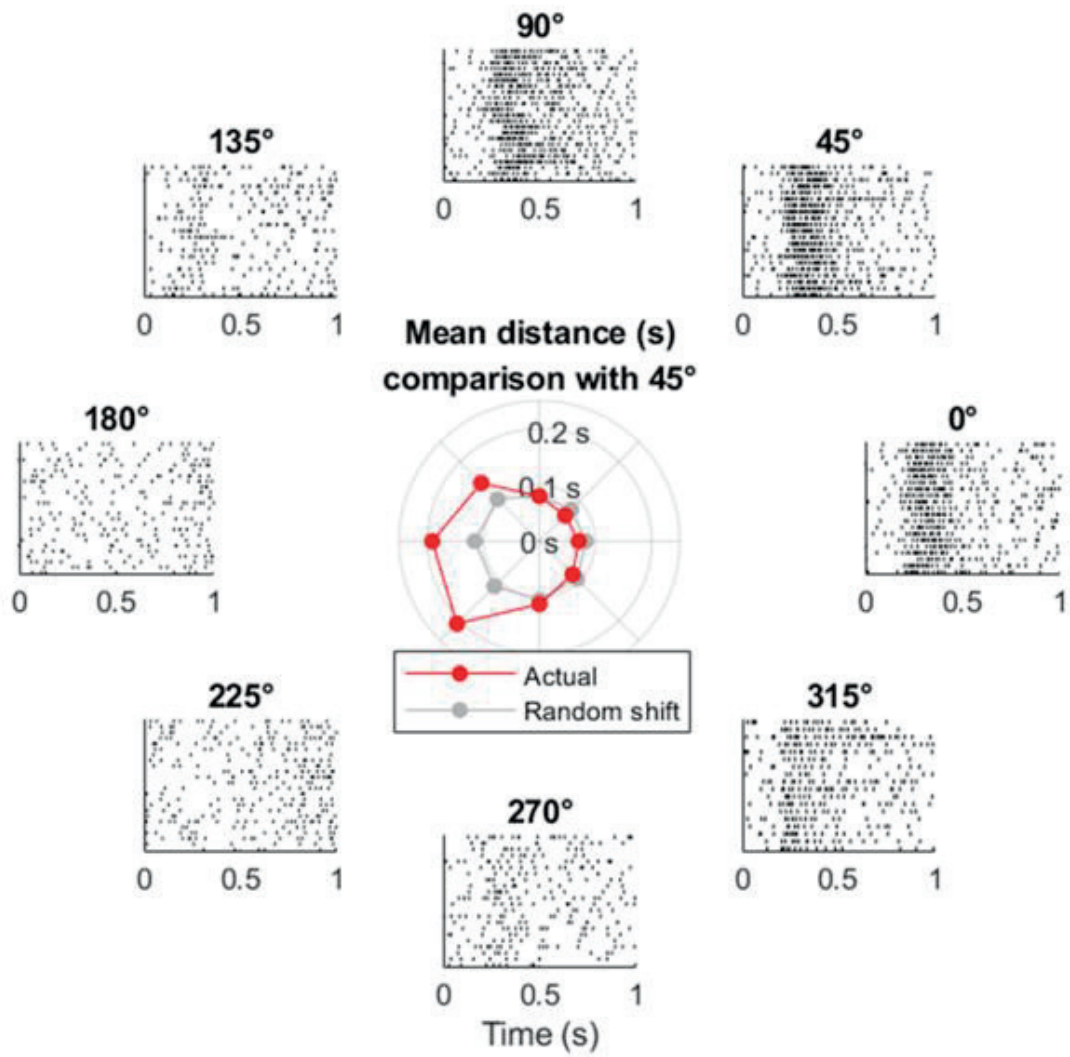
For the assessment of each spike distance method, we randomly selected a single resampled spike train in the  $k$ -th direction and calculated distance between it and every other resampled spike trains using a given spike distance. Those calculated distances were averaged for each direction, yielding the average distances  $d_i$  ( $i = 1, \dots, 8$ ) for each of the 8 directions. The averaged distances were then normalized over direction such that  $\bar{d}_i = (d_i - d_{\min}) / (d_{\max} - d_{\min})$  as above. The shorter normalized distance toward the correct direction,  $\bar{d}_k$ , represented better discrimination of the spiking patterns for the correct direction from other directions. The EMD, as well as the Victor-Purpura distance and the van Rossum distance with specific parameter settings, resulted in shorter distances than others (Figure 9B). Therefore, it demonstrated that the EMD could decode the directional information of upper limb movements in M1 neurons based on spiking timing patterns.

We applied a clustering analysis (Houghton and Victor (2010) and Victor and Purpura (1996)) to the data shown in Figure 9A in order to compare the effect of each distance metrics on decoding the information of movement directions from spike trains. For decoding such directional information based on the shortest distance to the training samples of spike trains, we used the  $k$ -nearest neighbor algorithm (Fix and Hodges, 1951). The decoding performance were measured by the normalized transmitted information proposed in the study by Houghton and Victor (Houghton and Victor (2010)), which ranges from 0 to 1 where a higher value indicates more accurate decoding. The result demonstrated that the EMD produced the best decoding output (with the number of neighbors,  $k=3$ ) (Figure 9C).

We also applied the same clustering analysis to the data shown in Figure 8A without removing firing rate differences, in order to examine the effect of directionally-tuned firing rates on the spike distances. We observed that the Victor-Purpura distance and the van Rossum distance produced larger normalized

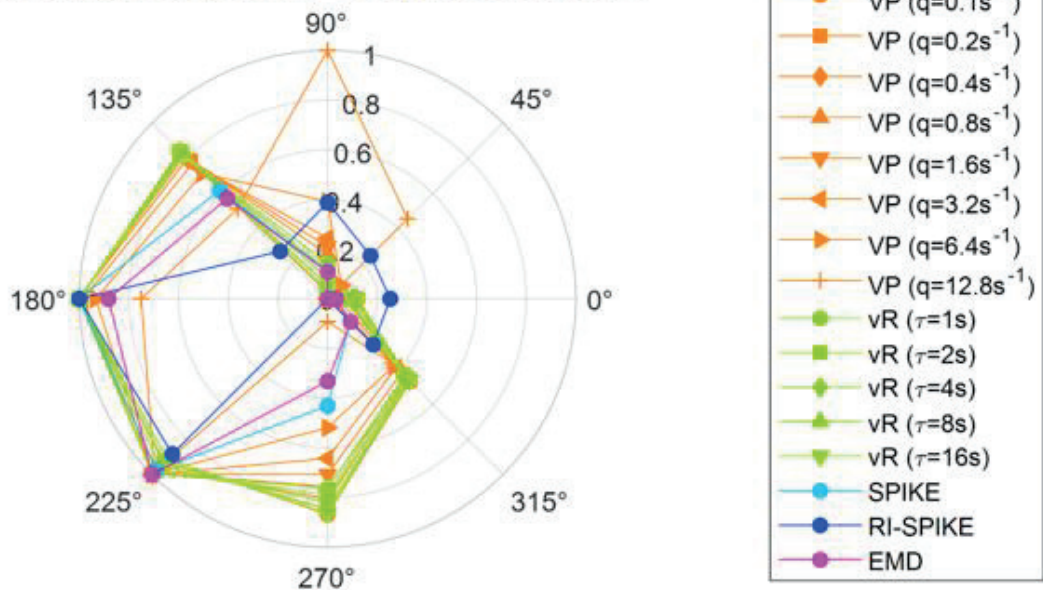
transmitted information than the SPIKE-distance, the RI-SPIKE-distance, and the EMD regardless of the setting of time-scale parameters. It demonstrates that the first two distances are more suitable for rate coding than last three distances, as also shown in the section 2.2.1.B.2 and 2.2.1.B.3.

## A



## B

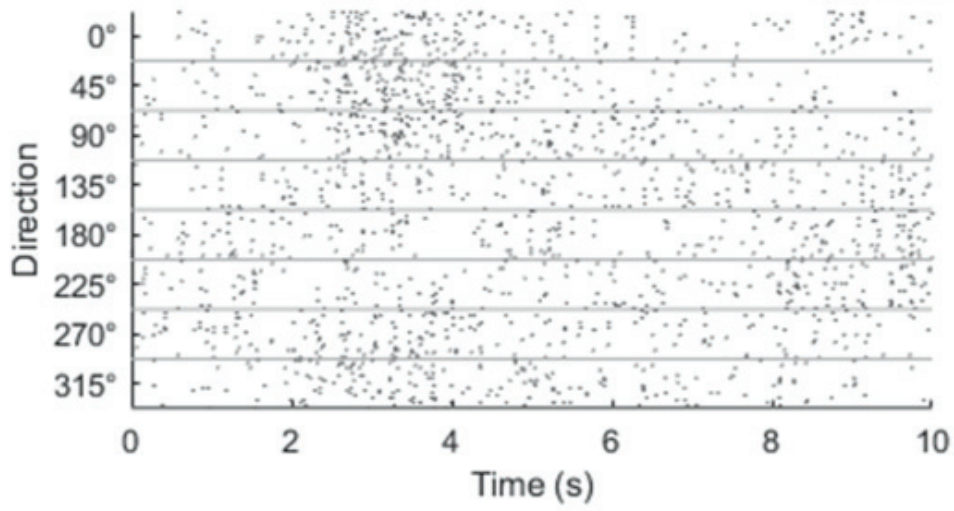
**Normalized distance comparison with 45°**



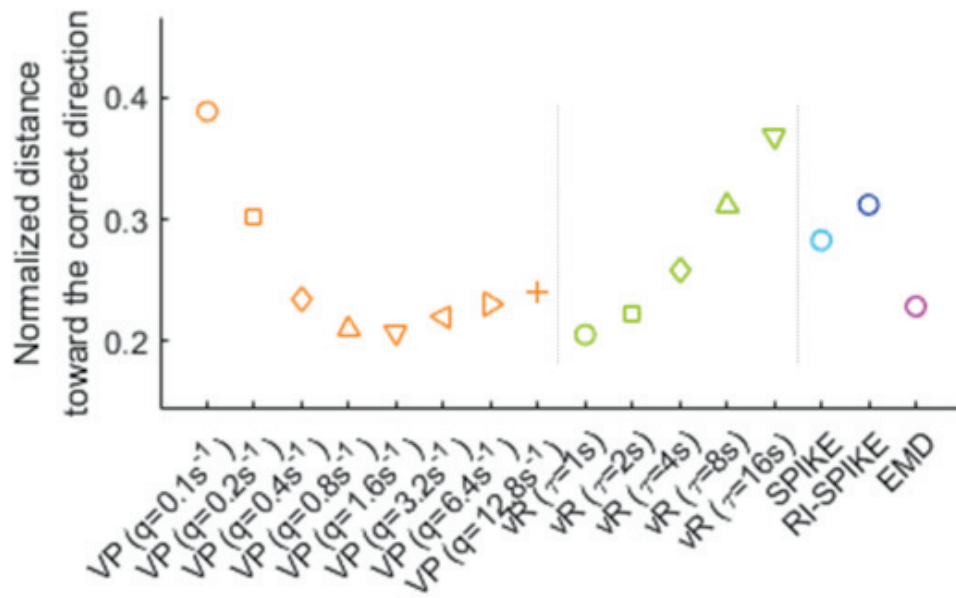


**Figure 8. Application of the spike distance to real neuronal data in the primary motor cortex in a non-human primate.** Data is from Flint et al. (2012). During the data recordings, the subject moved its arm from the central position toward one of the eight target positions and repeated this movement multiple times for each direction. Multiple spike trains of a single neuron for each of the eight target positions are described at each peripheral location, indicated by a directional angle as  $0^\circ$ ,  $45^\circ$ , ..., and  $315^\circ$ . Each spike train is obtained for 1 s after the onset of a movement cue. The spike trains exhibit different temporal patterns for different directions. (A) The direction at  $45^\circ$  is set as the seed direction, where the firing rate is maximum. Then, the spike distance is calculated between the seed direction and each of other seven directions. The mean spike distance between each pair is described using red circles in the center. The spike distance within the seed direction is also calculated for comparison (no calculation between the same spike train). The EMD from actual data (red line of the inner graph) has a clearer difference between the base and orthogonal angles than the EMD from data of randomly shifted spike timing (gray line of the inner graph), which does not exhibit a temporal pattern, showing the EMD difference is not merely due to the firing rate differences. (B) Comparison between the EMD and other spike distances for the data in (A). Each spike distance was normalized such that  $D_{\text{new}} = (D - D_{\text{min}}) / (D_{\text{max}} - D_{\text{min}})$  so that the distance values are filled between 0 and 1, because each spike distance has different magnitude scale. Throughout our spike distance analyses, we have set up a time-scale of spike trains for the Victor-Purpura distance and the van Rossum distance as  $[0, 10]$  s, which makes these distances applicable to both rate coding and temporal coding. To be consistent with such parameter settings of all the analyses done in the study, we also maintained the same time-scale range for the analysis of real neuronal spike data in (A). Since the spike trains of the real neurons we analyzed lasted for 1 s after a task onset, we extended spike trains by multiplying 10 to spike timings, changing the spike train range from  $[0, 1]$  s to  $[0, 10]$  s and used the same parameter settings as other simulation-based analyses for the Victor-Purpura and the van Rossum distances. This extension of the spike train range does not alter the SPIKE-distance, RI-SPIKE-distance and the EMD because they produce time-scale independent distance outcomes. VP: Victor-Purpura distance. vR: van Rossum distance.

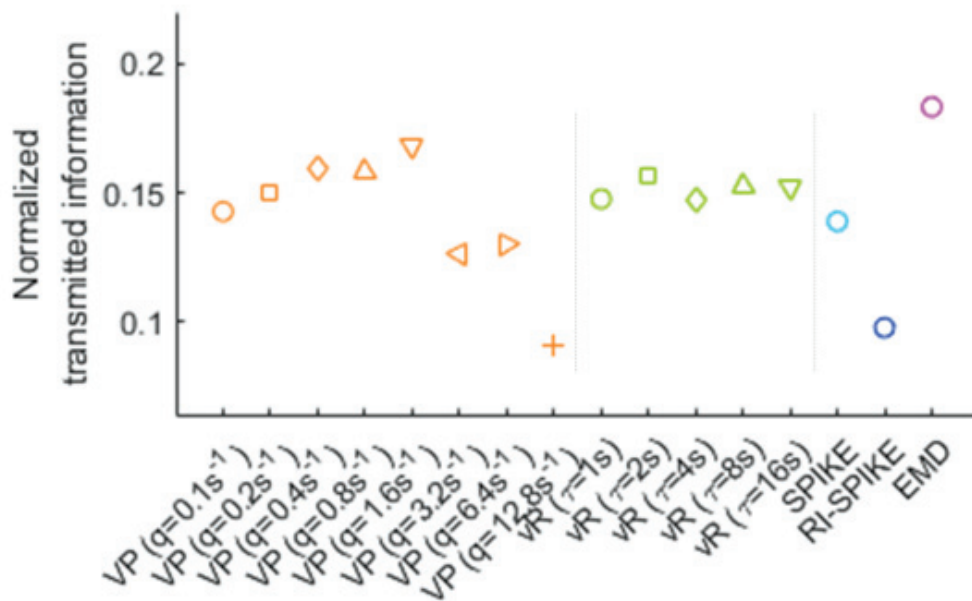
## A



## B



## C





**Figure 9. Application of the spike distance to resampled neuronal data.** (A) Resampled spike trains of the neuron in Figure 8. The spikes in the resampled spike train are randomly chosen from the pool of spike timings in each direction. The spiking patterns in the original spike train is preserved while the number of spikes is controlled. The range of spike trains extended from  $[0, 1]$  s to  $[0, 10]$  s. (B) The directional decoding results using spiking patterns in resampled spike trains. The distance indicates the average distance for each direction. The distance for the direction is normalized by other directions. The small value of the normalized distance toward the correct direction indicates a high magnitude of discrimination of the spiking patterns for the direction from the other directions. (C) Decoding directional information from the spike trains in (A) is performed using the k-nearest neighbor algorithm ( $k = 3$  in our analysis) and evaluated by the normalized transmitted information (see the text). Higher normalized transmitted information indicates better decoding performance. VP: Victor-Purpura distance. vR: van Rossum distance.

### 3. Principle of neural coding

In this section, we will cover the study of which principles enable neural expression. This is an approach to the computational principle and the main topic of the present study.

It has long been accepted that the brain, especially sensory system, has a hierarchical structure (Felleman and Van Essen, 1991; Mesulam, 1998; Harris et al., 2019; Hilgetag and Goulas, 2020). This hierarchical structure is related to gene expression (Burt et al., 2018; Hansen et al., 2021), suggesting that the hierarchical structure is genetically determined and a priori. Then, how can a prior hierarchical brain structure be given the function to represent the external world? This is a neural coding (principle) problem for sensory system;

The hierarchical structure of the sensory system is bidirectional; The hierarchical structure has not only bottom-up pathway but also top-down pathway, even in early sensory (visual) system such as the lateral geniculate nucleus (Murphy and Sillito, 1987; Wang et al., 2006) and the primary visual cortex (Zhang et al., 2014; Muckli et al., 2015; Huh et al., 2018). In this bidirectional hierarchical structure, two types of neuronal noise can be generated, where neuronal noise is defined as the uncertainty of neural responses for given sensory input (Borst and Theunissen, 1999). One of them is noise generated as neural information fluctuates across the hierarchy according to the initial condition of the neural response, even if the external sensory input is static. It is known that this can occur in interconnected structures as a chaotic dynamics sensitive to the initial condition (Rubinov et al., 2009; Tomov et al., 2014). This is neuronal noise when static sensory inputs are given. Another is noise, precisely error, caused by coding different information in each hierarchy because of the transmission delay of information (Berry et al., 1999) when external sensory input is dynamic. Because of the information transmission delay, (lower) hierarchies close to the sensory organ represent relatively recent information, and (upper) hierarchies distant from the sensory organ represent information relatively old, if external inputs are dynamic (changed). By the top-down pathway, old information in the upper hierarchy affects recent information in the lower hierarchy, and this becomes neuronal noise when dynamic sensory inputs are given. Despite these noise (error) problems, it seems that sensory information processing is performed without any major problems in the sensory system of the real brain. For the first noise problem (static sensory input), neural responses in sensory systems are decodable in both neuronal spikes (Berens et al., 2012; Zavitz et al., 2016) and blood-oxygen-level-dependent responses (Kamitani and Tong, 2005; Brouwer and Heeger 2009), indicating that the real brain is robust to noise of this type because this means the uncertainty of neural responses are not very large. For the second noise (error)

problem (dynamic sensory input), the sensory system has information of future sensory input (Palmer et al., 2015; Chen et al., 2017; Sederberg et al., 2018; Liu et al., 2021), minimizing information discrepancy across the hierarchy. This suggests that the real brain is robust to noise (error) of this type because the miss-informed noise (erroneous information) across hierarchy decreases as decreasing of information discrepancy across the hierarchy. Therefore, a neural coding principle that can overcome these noise problems is needed; How can the brain overcome these noise problems?

### 3.1. Efficient coding and predictive coding

Efficient coding, is one of neural coding principles, aims to minimize informational redundancy of neural representations for the external world (Attneave, 1954; Barlow, 1961). Informational redundancy  $R$  is given by (Barlow, 1961)

$$R = 1 - H(X)/C \quad (3.1.1)$$

where  $X$  indicates neural responses,  $H(\cdot)$  is an entropy, and  $C$  is the channel capacity which is the supremum of Shannon mutual information between  $X$  and the stimuli (the external world) variable  $S$ . If  $C$  is fixed, the minimization of the informational redundancy is equal to the maximization of the entropy  $H(X)$ . Hence, the informational redundancy reduction is to use efficiently the space of neural responses subject to the maximal range of neural responses.

Other literature describes efficient coding as maximizing Shannon mutual information between  $X$  and  $S$  (Friston, 2010):

$$I(X; S) = H(X) - H(X|S) \quad (3.1.2)$$

where  $I(X; S)$  is the Shannon mutual information between  $X$  and  $S$ , and  $H(X|S)$  is the conditional entropy of  $X$  given  $S$ . Since  $H(X|S) = 0$  if and only if  $X$  is a function of  $S$ , if  $X$  is deterministic or nearly deterministic (low noise),  $H(X|S)$  is nearly zero. This leads to the result that the maximization of  $I(X; S)$  is nearly equal to the maximization of  $H(X)$  which is the informational redundancy reduction. The biological plausibility of efficient coding has been verified in sensory systems (Laughlin, 1981). Efficient coding has been used to explain neural responses from statistics of the external world (Simoncelli and Olshausen, 2001).

Efficient coding has been applied to several studies on the hierarchical structure: the studies for complex cell property (Karklin and Lewicki, 2009), visual recognition (Hu et al., 2014), and acoustic feature

encoding (Zhang et al., 2019). However, these studies did not take into account the passage of time or studied in a unidirectional hierarchical structure where only bottom-up processing exists. In this environment, the aforementioned noise problems do not occur. These noise problems arise in bidirectional hierarchical structures on the time domain. The real brain environment is a bidirectional hierarchical structure on the time domain.

Predictive coding, is one of neural coding principles, is defined on hierarchical structure (Rao and Ballard, 1999; Spratling, 2017). According to predictive coding higher hierarchy performs top-down predictions on the response of lower hierarchy, and lower hierarchy transmits bottom-up prediction error to higher hierarchy. So, according to predictive coding, there are two different types of subpopulations in the brain: for prediction and for prediction error. Both inference and learning of predictive coding are to minimize the bottom-up prediction errors. Predictive coding has been shown to be able to explain neural responses corresponding to prediction errors, which increases at first and then decreases during inference (Friston, 2005). Predictive coding has also been extended from the explanation of perceptions to actions (Friston, 2010; Clark, 2013). Predictive coding is described as informational redundancy reduction because it only sends the unpredicted information of the stimulus, i.e., the prediction error, to the higher hierarchy (Huang and Rao, 2011). According to this interpretation, predictive coding explains efficient coding.

Prediction error minimization process in predictive coding may reduce the aforementioned neuronal noise; neuronal noise when static and dynamic sensory inputs are given. Nonetheless predictive coding has several theoretical disadvantages. Since inference in predictive coding is to minimize the prediction errors, it seems that the brain should have an additional information processing subsystem to perform the inference. Also, since bottom-up transmitted information is only prediction error, predictive coding requires the error units that are hypothetical entities while in some cases it is even difficult to observe prediction error responses (Solomon et al., 2021). Moreover, existing predictive coding is the problem on real-time information processing (Hogendoorn and Burkitt, 2019).

### **3.2. Spatio-temporally efficient coding**

The goal of the present study is to present a neural coding principle to overcome the aforementioned neuronal noise problem; neuronal noise when static and dynamic sensory inputs are given. To overcome the aforementioned neuronal noise problems, we devised a new neural coding principle that does not

have the disadvantages of predictive coding.

A possible approach to overcome the shortcomings of predictive coding in sensory hierarchical structures is to make bottom-up information transmissions similarly to top-down information transmissions across hierarchies, instead of transmitting bottom-up prediction errors. For example, context-independent bottom-up predictions and context-dependent top-down predictions (Teufel and Fletcher, 2020). Such bidirectional information transmissions eliminate the necessity for hypothetical error units, while presumably elucidating the neural responses of hierarchical structures underlying bottom-up feature integration and top-down predictive coding. A neural coding principle underlying bidirectional information transmissions of hierarchical structures can be found in the theory of efficient coding that draws upon the efficient use of given resources (Laughlin, 2001; Bullmore and Sporns, 2012), which crucially include limited time resources related to processing speed (Griffiths et al., 2015; Lieder and Griffiths, 2020). A possible solution to promote the most efficient use of limited time resources by the bidirectional information transmission system is to render present neural responses similar to future ones before the occurrence of future neural responses. This can be achieved by minimizing the temporal differences between present and future neural responses. Accordingly, we consider this temporal difference minimization as our *learning principle*, referred to as *temporally efficient coding*. Here, *inference* simply refers to a bidirectional information transmission mediated by top-down and bottom-up pathways. Unlike inference in predictive coding, which requires further error minimization, inference in temporally efficient coding involves simple single-step information transmission.

Temporally efficient coding involves a trivial solution: neural responses do not change to changes in external events. This trivial solution is comparable to the dark-room problem of predictive coding or free-energy principle, where an agent stays and is unchanged in a dark room with no surprise or unpredicted parts (Friston et al., 2012; Clark, 2013). We circumvent this issue by adding a complementary neural coding (learning) principle that maximizes the informational entropy of neural responses to alter neural responses to changing external events. It maximizes the neural response space available to represent the external world under the constraints of both the number of neurons and maximum firing rates. Maximal entropy coding indicates that the system uses spatial resources of neural responses efficiently (Attneave, 1954; Barlow, 1961; Laughlin, 1981), referred to as *spatially efficient coding*. By combining spatially efficient coding and temporally efficient coding, we propose a neural coding principle termed *spatio-temporally efficient coding* (Figure 10).

Spatially efficient coding has the same objective as existing efficient coding (Barlow, 1961; Laughlin, 1981), which minimizes informational redundancy because it increases the difference between neural

responses. Temporally efficient coding, on the other hand, can be regarded to increase informational redundancy, as it reduces the difference between neural responses. Two seemingly opposing coding objectives can be reconciled by isolating mechanisms that decrease the differences between consecutive neural responses on time domain and those that increase the differences between neural responses to the apparently different external world. In the real brain, it can be explained that different mechanisms are applied depending on the degree of difference in neural response. In the implementation of the present study, two coding objectives were applied in different ways. As an implementation of temporally efficient coding, we minimize the difference between consecutive neural responses on the time. In spatially efficient coding, the neural responses to the apparently different external stimuli are implemented in a minibatch method that simultaneously learns different images. We maximized the difference between neural responses to different images within each time step (see, Implementation of spatio-temporally efficient coding).

Temporally efficient coding trains the present neural response to be similar to the future neural response in order to efficiently use a given time resource. The temporal trajectory of the neural response is smoothed as the difference between the present and future neural responses is minimized. It thereby minimizes the size of the space represented, when a single stimulus (stimulus in the external world) is represented on the time domain. In other words, it reduces neuronal noise which is defined as the uncertainty of a neural response for given stimulus (Borst and Theunissen, 1999). This is to decrease the conditional entropy of neural response given stimulus,  $H(X|S)$  where  $X$  indicates neural response and  $S$  indicates stimulus. Spatially efficient coding increases  $H(X)$ . Spatio-temporally efficient coding, thus, increases the Shannon mutual information  $I(X; S) = H(X) - H(X|S)$  simultaneously in both terms:  $H(X)$  and  $-H(X|S)$ . This is also the definition of another existing efficient coding (Friston, 2010). Spatio-temporally efficient coding in hierarchical structures, therefore, can also be seen as an extension of existing efficient coding into hierarchical structures on time domain.

By spatio-temporally efficient coding, neural responses change smoothly but dynamically. Those dynamical changes of neural responses for the changing external world differs from the slow feature analysis (Wiskott and Sejnowski, 2002; Berkes and Wiskott, 2005; Creutzig and Sprekeler, 2008) which also minimizes temporal differences of neural responses. Similar to slow feature analysis, there have been studies on the properties of cells in the visual cortex using temporal coherence to obtain slow representations (Hurri and Hyvärinen, 2002; Zou et al., 2011). However, such studies using temporal coherence lack the aspect of dynamically reacting to changes in external input or neural responses of other hierarchies.

Neural system homeostasis is associated with maximization of mutual information between neural

responses and external stimuli (Toyoizumi et al., 2005; Sullivan and de Sa, 2006). Since spatio-temporally efficient coding increases the Shannon mutual information between neural responses and external stimuli, it is related to homeostasis. In particular, temporal difference minimization of neural responses in temporally efficient coding is reminiscent of homeostasis of energy metabolism. Smoothing the temporal trajectory of a neural responses reduces the variance of the neural response distribution so that the neural response stays within a certain range. This is also a consequence of the homeostatic plasticity (Turrigiano and Nelson, 2004) of the brain.

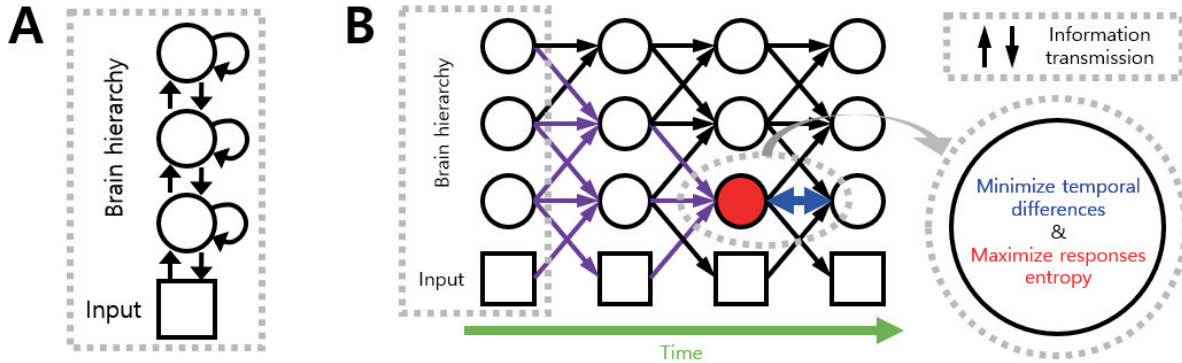
As mentioned earlier, smoothing the temporal trajectories of neural responses reduces the spatial extent of neural responses to static stimuli. This leads to a rapid stabilization of the neural response to the static stimulus; decodable stable neural representations. The rapid stabilization of neural responses shows the characteristic of temporally efficient coding, which makes efficient use of given time resources. Another effect of this smoothing the temporal trajectory of neural response is to render smooth neural representations that locally preserves the structure of the external world. If a stimulus is static or changes smoothly, making the temporal trajectory of the neural response smooth is to render a similar neural response to the similar stimuli. This is smooth neural representations that locally preserves the structure of the external world.

Energy limitation in the brain can be seen as an evolutionary selective pressure (Niven and Laughlin, 2008). Thus the efficient use of given energy is an important neural coding principle (Laughlin, 2001). Temporally efficient coding can be viewed as a reduction in the duration of energy consumption. If the energy consumed per unit time is similar regardless of the duration of energy consumption, temporally efficient coding will lead to a reduction in energy consumption. Spatially efficient coding elicits the maximal information representation under a given energy consumption (the firing rate of neurons). Thus, spatio-temporally efficient coding can be viewed as a plausible principle from an evolutionary point of view.

By spatio-temporally efficient coding (especially temporally efficient coding), bidirectional hierarchical structures learn smooth neural responses. This minimizes temporally differences between neural responses. It renders that neural information less fluctuates across the hierarchy. Therefore, we can expect that spatio-temporally efficient coding can overcome the aforementioned noise problem of first type (case of static sensory inputs). Moreover, to render neural response smooth when sensory input changes smoothly is to make similar neural responses for similar sensory inputs. Since sensory inputs at adjacent time may be similar, neural responses are also similar. These similar neural responses on time domain reduce the discrepancy of represented information across the hierarchy. Therefore, we can expect that spatio-temporally efficient coding can overcome the aforementioned noise problem of



second type (case of dynamic sensory inputs).



**Figure 10. Spatio-temporally efficient coding.** (A) This illustration depicts a hierarchical structure of the brain. Open black circles indicate an ensemble of neuronal units in each hierarchy of the brain. Open black square indicates visual input. Back arrows indicate information transmissions of bottom-up (upward arrow), recurrent (loop arrow), and top-down (downward arrow). (B) This illustration depicts a hierarchical structure of the brain and learning objectives. The hierarchical structure learns to represent input from the external world, which is depicted as black squares (e.g., visual input). Open black circles indicate an ensemble of neuronal units in each hierarchy of the brain. Inference based on spatio-temporally efficient coding is made by neuronal units as bottom-up, recurrent, and top-down information transmissions over time (black arrows). Learning in spatio-temporally efficient coding consists of two objectives: minimizing the temporal differences between present and future neural responses and maximizing the informational entropy of neural responses. For example, information transmissions (purple arrows) are optimised to minimize the temporal differences between present neural responses at the corresponding hierarchy (red filled circle) and future neural responses (circle to the right of the red filled circle) while concurrently maximizing the informational entropy of neural responses at the corresponding hierarchy (red filled circle).

### 3.3. Implementation of spatio-temporally efficient coding

In the present study, sensory (visual) information processing in hierarchical structures was established as biologically inspired temporal processing. Specifically, visual information processing is described as a function  $f_t$  for both image  $X_{image}$  and neural responses  $X_h$  in each hierarchy  $h$  such that it maps



from  $X_{image}$  and  $X_h$  at time  $t - 1$  to those at time  $t$ :

$$f_t: X_{image,t-1} \times X_{h=1,t-1} \times \dots \times X_{h=H,t-1} \rightarrow X_{image,t} \times X_{h=1,t} \times \dots \times X_{h=H,t} \quad (3.3.1)$$

where  $X_{h,t-1}$  and  $X_{h,t}$  are the neural responses  $X_h$  at time  $t - 1$  and  $t$ , respectively, at hierarchy  $h$ . For convenience,  $X_{h=0}$  denotes  $X_{image}$ , in particular  $X_{h=0,t}$  denotes the image presented at time  $t$ . The details of the visual information processing  $f_t$  are as follows: If  $h > 0$ , then

$$f_t|_{X_{h,t}} = \sigma(W_{h+1,h}^T X_{h+1,t-1} + W_{h,h}^T X_{h,t-1} + W_{h-1,h}^T X_{h-1,t-1} + b_h) \quad (3.3.2)$$

where  $X_{h,t}$  is an  $X_h$  value vector at time  $t$ ,  $f_t(\cdot)|_{X_{h,t}}$  indicates restricting the range of the function value  $f_t(\cdot)$  to  $X_{h,t}$ ,  $W_{h+1,h}$  is a synaptic weight matrix from hierarchy  $h + 1$  to  $h$ ,  $T$  is the transpose of a matrix,  $b_h$  is a bias vector at hierarchy  $h$ , and  $\sigma(\cdot)$  is a sigmoid function. The terms of  $W_{h+1,h}^T X_{h+1,t-1}$ ,  $W_{h,h}^T X_{h,t-1}$ , and  $W_{h-1,h}^T X_{h-1,t-1}$  indicate top-down, recurrent, and bottom-up information transmissions, respectively. In case of  $h + 1 > H$  the  $W_{h+1,h}^T X_{h+1,t-1}$  term would be omitted. If  $h = 0$ , then

$$f_t|_{X_{h,t}} = \sigma(W_{h+1,h}^T X_{h+1,t-1} + b_h). \quad (3.3.3)$$

If  $h > 0$ ,  $X_{h,t} = f_t|_{X_{h,t}}$ . On the other hand, when  $h = 0$ ,  $X_{h,t}$  is the image presented at time  $t$ , not  $f_t|_{X_{h,t}}$ . This function  $f_t$  makes *inferences* using spatio-temporally efficient coding. Unlike inference in predictive coding (Rao and Ballard, 1999; Spratling, 2017) that requires additional processes such as minimization of prediction errors, inference in spatio-temporally efficient coding is a function value  $f_t$  itself.

*Learning* in spatio-temporally efficient coding minimizes the ensuing objectives of both temporally and spatially efficient coding. The objective of temporally efficient coding is given by:

$$L_{Temporal} = \sum_{n=1}^N \left( f_t \circ f_{t-1}(x_{h=0,\tilde{t},image=n}) - f_{t+1} \circ f_t \circ f_{t-1}(x_{h=0,\tilde{t},image=n}) \right)^2 \quad (3.3.4)$$

where  $\circ$  is the function composition, squaring is operated component-wise. In the case of static sensory inputs,  $x_{h=0,\tilde{t},image=n}$  indicates that the image is fixed to the  $n$ th sample throughout the temporal processing (i.e.,  $X_{h=0,t-2} = X_{h=0,t-1} = X_{h=0,t}$  is  $n$ th image sample). Otherwise, in the case of dynamic sensory inputs,  $x_{h=0,\tilde{t},image=n}$  indicates that the  $n$ th image is smoothly changed with the passage time  $t$ . In case of  $h = 0$ ,  $f_{t+1} \circ f_t \circ f_{t-1}(X_{h=0,\tilde{t},image=n})|_{X_{h=0}} := X_{h=0,t+1}$  which is the image presented at time  $t + 1$  while  $f_t \circ f_{t-1}(x_{h=0,\tilde{t},image=n})|_{X_{h=0}}$  is the function value as inference.

The function composition  $f_t \circ f_{t-1}$  of two functions (inferences) in the  $L_{Temporal}$  allows the simultaneous learning of all information transmissions across all hierarchies because the depth of hierarchy is  $H = 2$  in the present study (bottom-up, recurrent, and top-down information transmission across the hierarchy of depth  $H = 2$ ). The minimization of the given objective minimizes the temporal differences between present and future neural responses. The objective of temporally efficient coding is to render present neural responses similar to future neural responses that has not yet arrived. By doing so, it uses the given time resources efficiently. This minimization of temporal difference creates a learning effect in which the temporal trajectory of the neural response becomes smooth. The expected effect of this learning effect is to quickly stabilize neural responses when a static external stimulus is given. In that aspect, it is also to use the given time resources efficiently.

The objective of spatially efficient coding is given by

$$L_{Spatial} = \sum_{n=1}^N \log P(f_t(x_{h=0, \tilde{t}, \text{image}=n})|_{X_h}) \quad (3.3.5)$$

where  $h > 0$  in  $X_h$  (so,  $f_t(X_{h=0, \tilde{t}, \text{image}=n})|_{X_h} = X_{h,t}$ ) and  $P(\cdot)$  is a probability. Because the term  $L_{Spatial}$  is an estimation of negative informational entropy, minimizing the objective maximizes the informational entropy of  $f_t$ . This objective increases the entropy of each neuron, and it can be seen that an increase in marginal entropy in each neuron will increase the joint entropy of the entire system. Therefore, spatially efficient coding maximizes the entropy of individual neurons as in Laughlin's study (Laughlin, 1981) and consequently the entropy of the entire system. The objective of spatially efficient coding is to render different neural responses to different inputs from the external world. We, therefore, can overcome the trivial solution of temporally efficient coding, where there is no change in neural response despite changes in the external world. This thereby does not conflict with the learning of temporally efficient coding, which decreases differences in consecutive neural responses in the time domain to inputs of one stream.

To minimize  $L_{Spatial}$ , which is the objective of spatially efficient coding, it is necessary to calculate the probability  $P(\cdot)$  in  $L_{Spatial}$  (equation 3.3.5). Instead of calculating the exact probabilities, we obtained pre-normalised densities in the sense of probabilities without a partition function. As the value of the partition function is fixed, it does not affect the minimization process. Note that, when  $h > 0$ ,  $X_{h,t}$  is the space of all possible neural responses, i.e.,  $X_{h,t} = [0, 1]^{\dim X_{h,t}}$ . Let  $x \in X_{h,t}$  be a value of  $f_t(x_{h=0, \tilde{t}, \text{image}=n})|_{X_h}$ . Kernel density estimation was used to obtain  $P(x)$ . Using a Gaussian kernel with width  $0.1(\dim X_{h,t})^{1/2}$ , the neural response density  $Q(x)$  and compensation density  $Q'(x)$  at

$x \in X_{h,t}$  were obtained. Then, the pre-normalized density of interest is

$$P(x) = \frac{Q(x)}{Q'(x)}. \quad (3.3.6)$$

The neural response density  $Q(x)$  is obtained by kernel density estimation of neural responses on  $X_{h,t}$ . The compensation density  $Q'(x)$  is obtained using pseudo-uniformly generated samples on  $X_{h,t}$  instead of the neural responses. (For implementation of low neural response, i.e., sparse neural responses,  $Q'(x)$  is obtained using pseudo-uniformly generated samples on  $[-1, 1]^{\dim X_{h,t}}$  instead of  $X_{h,t} = [0, 1]^{\dim X_{h,t}}$ , so that  $Q'(x)$  has a fat distribution around zero.) The compensation density is necessary to compensate for the non-uniform intrinsic expectation of  $Q(\cdot)$  resulting from the fact that  $X_{h,t}$  is bounded. At the boundary of  $X_{h,t}$ , the density of neural responses,  $Q(x)$ , measured by kernel density estimation is decrease. This intrinsic decrease corresponds to  $Q'(x)$ . We compensated for the decrease by dividing  $Q(x)$  by  $Q'(x)$ .

Finally, the objective of spatio-temporally efficient coding is a linear combination of those two objectives:

$$L = L_{Temporal} + \lambda L_{Spatial} \quad (3.3.7)$$

where  $\lambda$  is a regularization parameter. A smaller  $\lambda$  indicates a greater emphasis on the temporally efficient coding objective, whereas a larger  $\lambda$  indicates the opposite. As mentioned previously, temporally efficient coding decreases the conditional entropy of neural response given stimulus,  $H(X|S)$  where  $X$  indicates neural response and  $S$  indicates stimulus. Spatially efficient coding increases  $H(X)$ . Hence, the regularisation parameter  $\lambda$  can be seen as controlling the balance between  $H(X)$  and  $-H(X|S)$ . It is a modification of fixed balance of Shannon mutual information  $I(X; S) = H(X) - H(X|S)$  of the existing efficient coding (Friston, 2010).

Temporal trajectories of neural responses are smoothed by temporally efficient coding, but this does not mean just slow neural representations. By spatially efficient coding, different neural responses to different inputs from the external world should be exhibited. Therefore, it should show rapid changes in neural responses to sudden changes in the external world (fast representation). This is the difference from slow feature analysis (Wiskott and Sejnowski, 2002; Berkes and Wiskott, 2005; Creutzig and Sprekeler, 2008) or temporal coherence (Hurri and Hyvärinen, 2002; Zou et al., 2011), which targets slow neural representations. On the one hand, with temporally efficient coding, changes in neural responses should be smoothed out quickly when the external input is not changing (fast stabilization).

Suppose that Gaussian noise is added to a series of temporally correlated external inputs (e.g., static

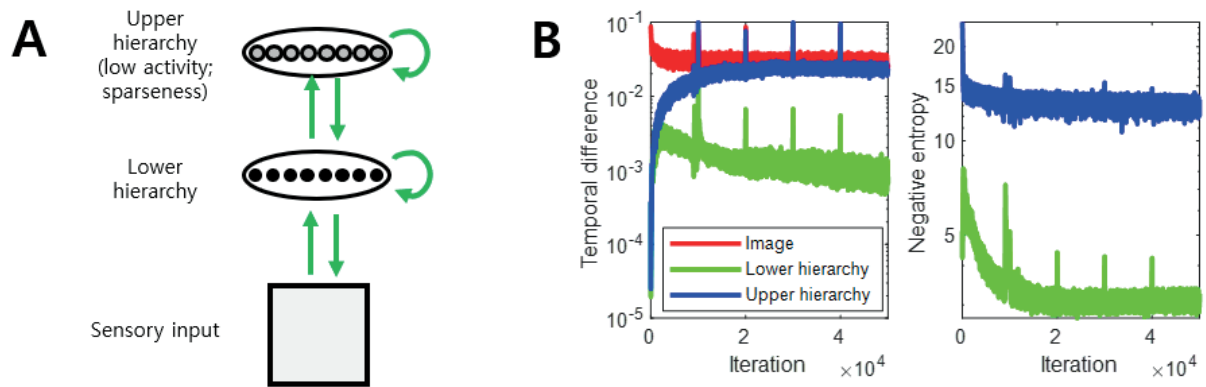
images or smoothly moved images + Gaussian noise). If these noisy external inputs are still temporally correlated, then smoothing the temporal trajectory of neural responses (by temporally efficient coding objective  $L_{Temporal}$ ) is rendering temporally similar neural responses to temporally similar inputs. It is also smooth neural representations that locally preserves the structure of the external world. Moreover, if the noise is provided independently of the input, the effect of the noise on neural representations will be dispelled by multiple independent trials. Therefore, temporally efficient coding objective  $L_{Temporal}$  increases the fidelity of the neural representation with respect to a Gaussian noise.

Smooth neural representation also means making different neural responses to different inputs. In other words, it increases discriminability for different inputs. This is achieved by spatially efficient coding objective  $L_{Spatial}$  that maximizes the entropy of neural responses. Assumed that Gaussian noise is added to external inputs. Let  $D_{true}$  be a binary-valued random variable for the discrimination between two actually different inputs such that  $D_{true} = 1$  means the discrimination that two noised inputs differ and  $D_{true} = 0$  means the discrimination that two noised inputs are same.  $D_{true}$  is probabilistic because of the Gaussian noise mentioned earlier. Let  $D_{response}$  be a binary-valued random variable for the discrimination between two neural responses for inputs such that  $D_{response} = 1$  means the discrimination that two neural responses differ and  $D_{response} = 0$  means the discrimination that two neural responses are same. Let  $P_{true}$  and  $P_{response}$  be the probability mass functions of  $D_{true}$  and  $D_{response}$ , respectively. Spatially efficient coding objective  $L_{Spatial}$  decreases the Kullback–Leibler divergence from  $P_{response}$  to  $P_{true}$ , i.e.,  $D_{KL}(P_{true}||P_{response})$ . Hence it also decreases the binary cross entropy  $H(P_{true}) + D_{KL}(P_{true}||P_{response})$  where  $H(\cdot)$  is an informational entropy.

In the present study, the depth  $H$  of hierarchies was set to 2, the minimum depth to realise both bottom-up and top-down pathways in the same hierarchy. Since neural responses in subcortical sensory (visual) area are regular while neural responses in cortical sensory (visual) area are sparse (Simoncelli, 2003), we implemented low neural responses (sparse neural responses) at upper hierarchy (Figure 11A). Units in adjacent hierarchies are fully connected.

Because  $L = L_{Temporal} + \lambda L_{Spatial}$  is differentiable, the minimization of the objective in spatio-temporally efficient coding was performed with a gradient descent. The Adam optimiser (Kingma and Ba, 2015) was used to perform the stochastic gradient descent with momentum. The parameters of the Adam optimiser used in this study were  $\alpha = 0.001$ ,  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ , and  $\epsilon = 10^{-8}$ . The optimisation lasted  $10^4$  iterations for each repetition and restarted with five repetitions. For each iteration, the duration of temporal processing  $f_t$  was five (i.e.,  $t \in [1,5]$ ), and the minibatch size was 40. In learning, after temporal processing of five (case of static sensory inputs) or nine (case of dynamic

sensory inputs) durations was finished, new temporal processing begins, and the initial values of neural responses of new temporal processing were the last neural responses values of the previous temporal processing. In our simulations, we repeatedly exposed the hierarchical structure to natural scene images, which enabled it to learn the bidirectional information transmissions between top-down and bottom-up hierarchies using spatio-temporally efficient coding with a range of the balancing parameter  $\lambda$ . Successful learning was confirmed by minimizing or stabilising  $L$  during learning (Figure 11B). Further, we verified that the learned hierarchical structure could successfully reconstruct an input image.



**Figure 11. Implementation of spatio-temporally efficient coding.** (A) Architecture of visual hierarchy with depth 2 in simulations. Units in the lower hierarchy set up to have regular neural responses and units in the higher hierarchy set up to have sparse neural responses. These correspond to subcortical and cortical neural responses, respectively. Units in adjacent hierarchies are fully connected. (B) Learning curves of spatio-temporally efficient coding. Left panel: mean temporal difference. Right panel: mean negative entropy. As the negative entropy of the lower and upper hierarchies decreases, it can be observed that the temporal difference increases and then stabilizes or decreases again. Vertical axis has the logarithm scale.

## 4. Spatio-temporally efficient coding: a case of static sensory input

### 4.A. Introduction

In the bidirectional hierarchical structure, two types of neuronal noise can be generated, where neuronal noise is defined as the uncertainty of neural responses for given sensory input (Borst and Theunissen, 1999). One of them is noise generated as neural information fluctuates across the hierarchy according to the initial condition of the neural response, even if the external sensory input is static. It is known that this can occur in interconnected structures as a chaotic dynamics sensitive to the initial condition (Rubinov et al., 2009; Tomov et al., 2014). This is neuronal noise when static sensory inputs are given. Another is noise caused by coding different information in each hierarchy because of the transmission delay of information (Berry et al., 1999) when external sensory input is dynamic. Because of the information transmission delay, (lower) hierarchies close to the sensory organ represent relatively recent information, and (upper) hierarchies distant from the sensory organ represent information relatively old, if external inputs are dynamic (changed). By the top-down pathway, old information in the upper hierarchy affects recent information in the lower hierarchy, and this becomes neuronal noise when dynamic sensory inputs are given. In this section, we will show that spatio-temporally efficient coding can deal with the neuronal noise of first type (case of static sensory inputs).

By spatio-temporally efficient coding (especially temporally efficient coding), bidirectional hierarchical structures learn smooth neural responses. This minimizes temporally differences between neural responses. It renders that neural information less fluctuates across the hierarchy. Therefore, we can expect that spatio-temporally efficient coding can overcome the aforementioned noise problem of first type (case of static sensory inputs). In this section, we will verify this expectation through simulation.

### 4.B. Methods

For the simulations, van Hateren's natural scene image dataset (van Hateren and van der Schaaf, 1998) was used. The dataset was downloaded from <http://bethgelab.org/datasets/vanhateren/>. The images were downsized to  $64 \times 96$  pixels. For the comparison tests, the MNIST handwritten digit dataset (Lecun et al., 1998) was used. The dataset was downloaded from <http://yann.lecun.com/exdb/mnist/>. The images were resized to  $64 \times 96$  pixels to fit the images used in the simulations. All image data were rescaled between 0 and 1.

For a given static image, the duration of temporal processing of  $f_t$  was given as 5 (i.e.,  $t \in [1,5]$ ) in

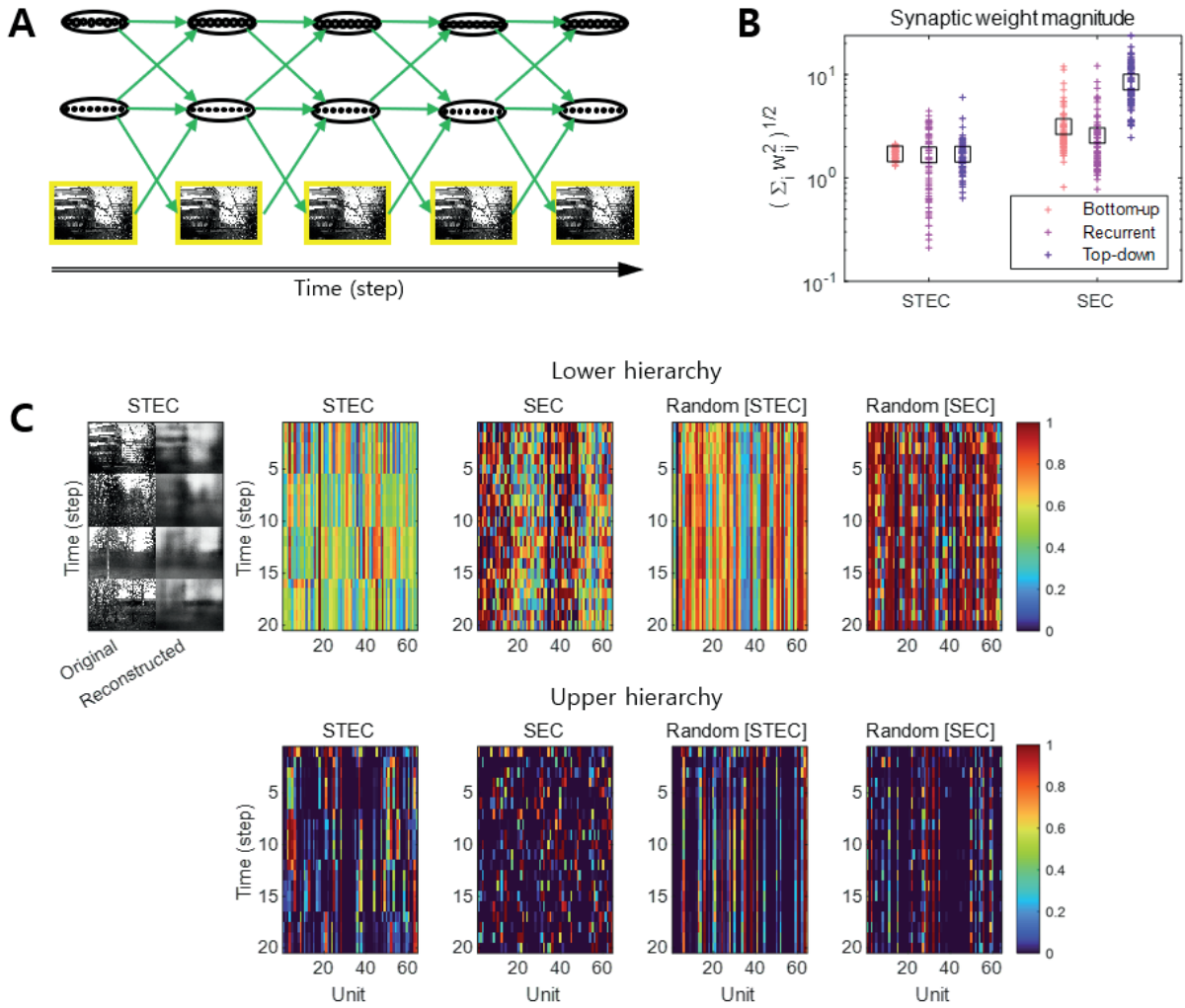
learning (Figure 12A), because four time steps are required for the image information to reach the top hierarchy and return over  $H = 2$  hierarchies, in addition to one time step to obtain future neural responses. The learning via Adam optimizer lasted  $10^4$  iterations for each repetition and restarted with five repetitions. For each iteration, the duration of temporal processing  $f_t$  was five (i.e.,  $t \in [1,5]$ ) as above, and the minibatch size was 40. Learning was deemed to be successful when the value of the objective function ( $L$ , Eq. 3.3.7) was sufficiently stabilized over repetitions.

We set  $\lambda = 5, 100$ , where  $\lambda$  is a regularization parameter in the objective function,  $L = L_{Temporal} + \lambda L_{Spatial}$  (see Eq. 3.3.7). The condition  $\lambda = 5$  was set the balanced condition between spatially efficient coding and temporally efficient coding objectives (STEC). The condition  $\lambda = 1000$  was set one of the control conditions which is overweighted toward spatially efficient coding (SEC). This indicates the strict efficient condition.

A change in balance between the two objectives also altered the relative strengths between bottom-up and top-down synaptic connections. Simulation results demonstrated that top-down synaptic strengths from upper hierarchy to lower hierarchy were increased compared to bottom-up synaptic strengths from images to lower hierarchy and recurrent synaptic strengths from lower hierarchy to lower hierarchy when  $\lambda = 1000$  (SEC), with an emphasis on spatially efficient coding (Figure 12B).

We set important control condition, random networks, based on STEC and SEC: Random [STEC] and Random [SEC]. Random [STEC] was made by random permutations of synaptic weights and bias of STEC at each pathway (bottom-up, top-down, and recurrent). Random [SEC] was similarly made from SEC. Therefore, Random [STEC] and Random [SEC] have same synaptic weight magnitude level with STEC and SEC shown in Figure 12B. Example neural responses trajectories of all conditions are shown in Figure 12C.





**Figure 12. A simulation method for static sensory input.** (A) Learning in the case of static sensory input. (B) The magnitude of synaptic weights on lower hierarchy units is compared for different  $\lambda$ s, where  $\lambda$  is a regularization parameter in the objective function,  $L = L_{Temporal} + \lambda L_{Spatial}$  (see Eq. 3.3.7). Each cross indicates the  $L_2$  norm of synaptic weights for each unit. Black squares indicate the mean. (C) Left: Four different original input images and their reconstructions from the neuronal responses of lower hierarchy. Right: Representative neural responses of the lower (upper row) and upper (lower row) hierarchy units for different conditions. The neural response is the output of the sigmoid function and is therefore normalized to a range between 0 and 1. Note that  $\lambda = 5, 1000$ . The condition  $\lambda = 5$  was set the balanced condition between spatially efficient coding and temporally efficient coding objectives (STEC). The condition  $\lambda = 1000$  was set one of the control conditions which is overweighted toward spatially efficient coding (SEC). Random [STEC] and Random [SEC] were random networks based on STEC and SEC, respectively.



#### 4.C.1. Results: Decodable and rapidly stable neural representations

In this subsection, we will show that spatio-temporally efficient coding can overcome the aforementioned noise problem (case of static sensory inputs).

The objective of temporally efficient coding is to render present neural responses similar to future neural responses that has not yet arrived. This minimization of temporal difference creates a learning effect in which the temporal trajectory of the neural response becomes smooth. It thereby minimizes the size of the space represented, when a single stimulus is represented on the time domain. In other words, it reduces neuronal noise which is defined as the uncertainty of a neural response for given stimulus (Borst and Theunissen, 1999). The expected effect of this learning effect is to quickly stabilize neural responses when a static external stimulus is given. The objective of spatially efficient coding is to render different neural responses to different stimuli. The expected effect of spatio-temporally efficient coding is to render decodable and rapidly stable neural representations which is an appropriate function to hierarchical brain structures on the time domain. If neural responses are decodable and rapidly stable, then neural information less fluctuates across the hierarchy. Hence this indicates overcoming the aforementioned noise problem.

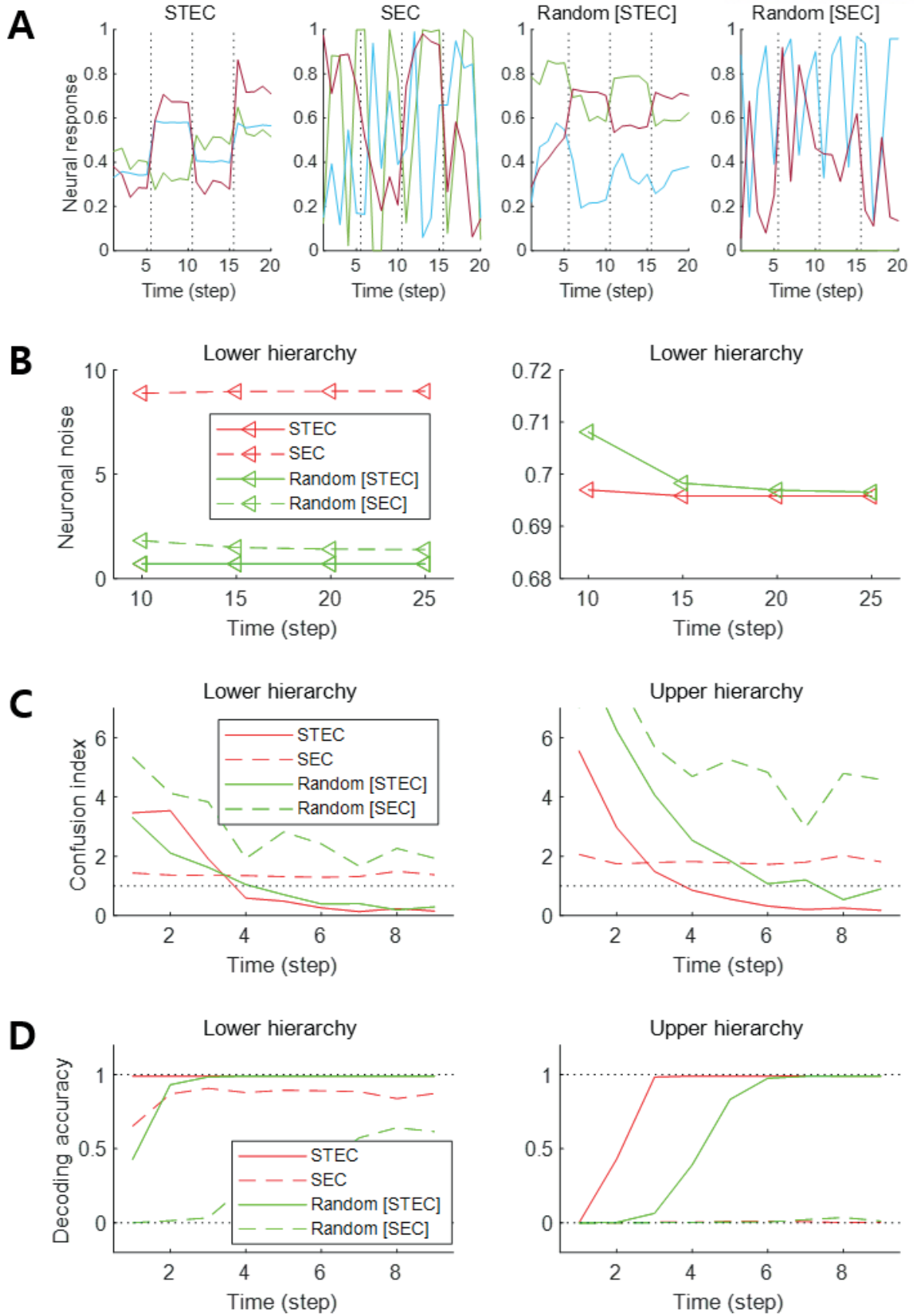
To confirm these expectations, we performed four simulation experiments. The first simulation is to present multiple images for five time step for each image and check the neural responses for those images. When each image was given, it was checked whether the neural response was quickly stabilized (Figure 13A). The second simulation measures the amount of noise in the neural responses when each image is presented during 25 time step. Neuronal noise is defined as the uncertainty of a neural response for given stimulus (Borst and Theunissen, 1999), so we measured the conditional entropy of neural responses during each five time step as neuronal noise. That is, Neuronal noise =  $\frac{1}{|S|} \sum_{s \in S} H(f_t | s, t \in [t' + 1, t' + 5])$  for each  $t' \in \{5, 10, 15, 20\}$  where  $s$  is an image,  $S$  is the set of all images,  $H(\cdot | \cdot)$  is the conditional entropy, and  $f_t$  is the neural response at time  $t$ . Probability was estimated by kernel density estimation without a partition function, and thus shifted conditional entropy was measured. We checked the amount of neuronal noise (Figure 13B). In the third simulation, we checked whether the neural responses to each image stabilized over time. To do this, we measured how confused the neural response over time (9 time step) to one image was with the neural response to a similar image. This is represented by the confusion index. Confusion index for image  $s$  at time  $t = |f_t(s) - f_{10}(s)| / |f_{10}(s) - f_{10}(s')|$  where  $f_t(s)$  is the neural response for image  $s$  at time  $t$  and  $s'$  is the nearest image of  $s$  based on the global feature based distance (Di Gesù and Starovoitov, 1999). If the

confusion index is less than 1, it can be said that the neural response is well stabilized. We checked whether the neural response was stabilized by reducing the confusion index to less than 1 (Figure 13C). In fourth simulation, we checked how decodable the neural responses was. To measure how decodable the neural response is, the neural responses of time steps 1, ..., 9 were decoded using the neural responses of time steps 10 and 11 as a training set. The linear discriminant analysis was selected as the decoder, and each of the 4212 natural scene images was defined as one class (Figure 13D).

As a result of the simulation, different neural responses to different stimuli were shown in the STEC condition, and these neural responses were rapidly stabilized. On the other hand, neural responses were less stabilized under other conditions (Figure 13A). The factor affecting the stabilization of the neural response was the amount of neuronal noise. To quantify the amount of neuronal noise, conditional entropy of neural responses given stimulus was measured. Conditional entropy was measured by collecting neural responses at every five time step. The STEC condition showed lower neuronal noise than the other conditions (Figure 13B).

We checked whether the neural response was well stabilized through the confusion index. Under STEC condition, the confusion index was kept below 1 on average for the majority of time steps at each hierarchy (Figure 13C). This means that the neural representations are rapidly stabilized in STEC condition. On the other hand, the confusion index was not sufficiently reduced in SEC and Random [SEC], and was more slowly reduced in Random [STEC] condition (Figure 13C). This means that sufficient stabilization of neural representations is difficult with only strict efficient coding. Furthermore STEC more rapidly stabilized than random networks. Since initial neural responses were randomly set so that they cause the aforementioned noise (case of static sensory input), the fact that neural responses in STEC more rapidly stabilized than in others suggests that spatio-temporally efficient coding can overcome the aforementioned noise problem.

Finally, we checked whether the neural response is decodable. In the STEC condition, the decoding accuracy gradually increased over time, suggesting decodable stable neural representations. On the other hand, the decoding accuracy was low in SEC and Random [SEC] condition, and slowly increased in Random [STEC] conditions (Figure 13D). Since initial neural responses were randomly set so that they cause the aforementioned noise (case of static sensory input), the fact that decoding accuracy in STEC more rapidly increased than in others suggests that spatio-temporally efficient coding can overcome the aforementioned noise problem.

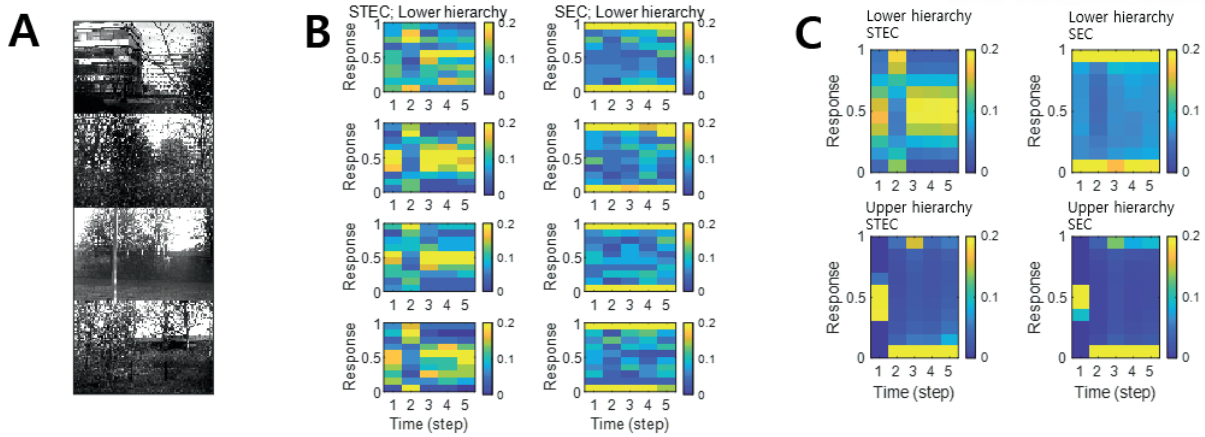


**Figure 13. Decodable and stable neural representations.** (A) Examples of neural responses in the lower hierarchy. These were neural responses of 1st, 17th, and 33th units in Figure 12C. Dotted vertical black lines indicate the presentations of new external input. (B) Neuronal noise that measured as shifted conditional entropy of neural responses given stimulus. Conditional entropy was measured by collecting neural responses at every five time step. Left and right panel showed different vertical axis scales of same data. (C) Confusion index that a measure of how much it confuses neural responses to one stimulus with neural responses to another similar stimulus. The dotted lines denote the confusion index of 1 which indicates the confusion. (D) Decoding accuracy via the naïve Bayes classifier. The upper dotted line denotes the maximum decoding accuracy, 1. The lower dotted line denotes the chance level,  $1/4212$ . Error bars indicate standard deviations.

#### 4.C.2. Results: Relation to homeostasis

The following subsections do not relate to the aforementioned noise problem, but reveal the characteristics of spatio-temporally efficient coding.

Neural system homeostasis is associated with maximisation of mutual information between neural responses and external stimuli (Toyoizumi et al., 2005; Sullivan and de Sa, 2006). Since spatio-temporally efficient coding increases the Shannon mutual information between neural responses and external stimuli, it is related to homeostasis. In particular, temporal difference minimisation of neural responses in temporally efficient coding is reminiscent of homeostasis of energy metabolism. Smoothing the temporal trajectory of a neural responses reduces the variance of the neural response distribution so that the neural response stays within a certain range. We confirmed this through simulation. It was checked whether the neural response had a middle value, not an extreme value such as 0 or 1, which is homeostasis of neural responses. In STEC condition emphasizing temporally efficient coding objective, neural responses were concentrated at the middle value than in SEC condition (Figure 14). Figure 14B showed specific cases corresponding to stimuli in Figure 14A. Figure 14C showed the distribution of overall neural responses, indicating that this is general phenomena of STEC. This is also a consequence of the homeostatic plasticity (Turrigiano and Nelson, 2004) of the brain.



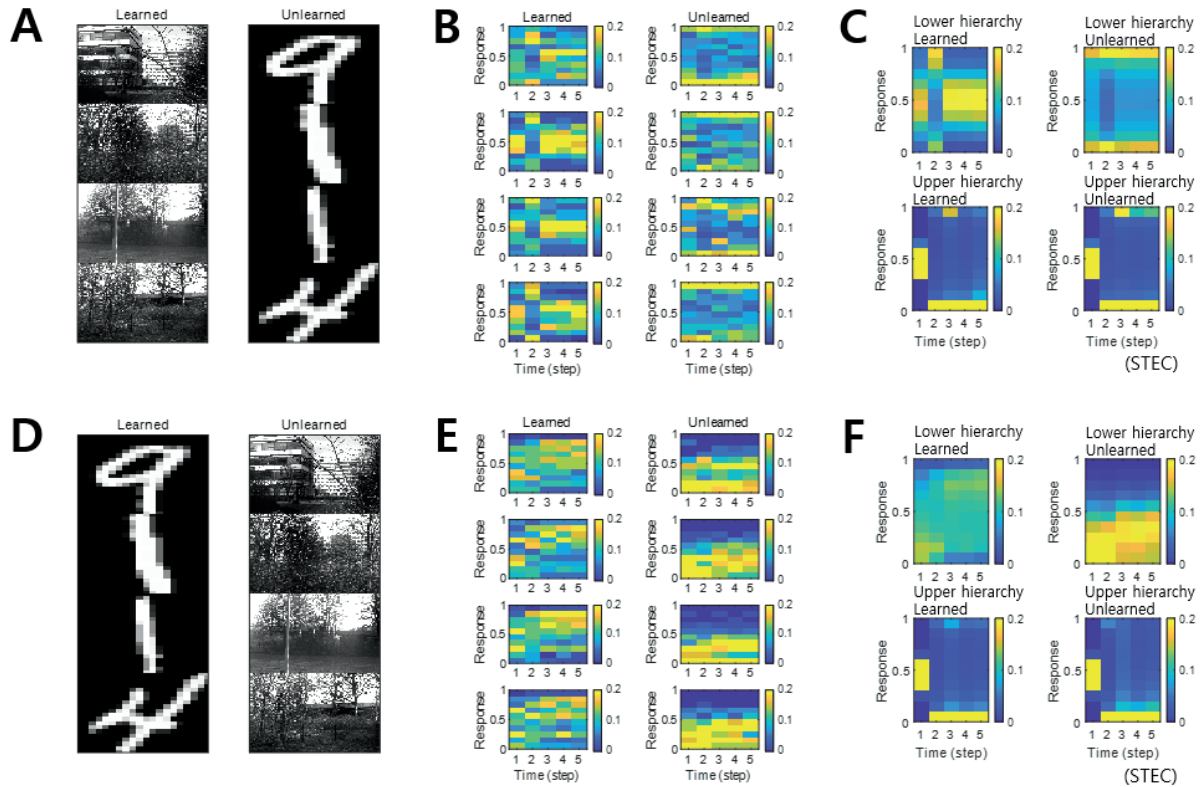
**Figure 14. Neural response distributions.** (A) Examples of natural scene images. (B) Examples of neural response distributions of lower hierarchy units for each of the five time steps of bidirectional information transmissions, corresponding to images in (A). The colour scale indicates the proportion. (C) Overall neural response distributions in response to all input images (either STEC or SEC) at lower and upper hierarchies, respectively. The colour scale indicates the proportion.

#### 4.C.3. Results: Deviant neural responses to unlearned inputs

The visual system often responds selectively to sensory inputs (Margoliash, 1983; Waydo et al., 2006). Even for the type of sensory inputs to which the visual system is responsive, unfamiliar inputs induce larger neural responses compared to familiar inputs (Huang et al., 2018; Issa et al., 2018). These large neural responses to unfamiliar inputs are thought to be due to prediction errors (Issa et al., 2018). On the one hand, spatio-temporally efficient coding renders smooth neural representations, i.e., locally preserving the structure of the external world. Because unlearned inputs differ from learned inputs, if their neural representations are smooth, their neural representations will also differ. By STEC homeostasis, since the neural response to the learned input has a middle value (Figure 14), the neural responses to the unlearned input (if these differ from the neural responses to the learned input) may be closed to extreme values such as 0 or 1. These are deviant neural responses to unlearned inputs. Accordingly, we can expect that spatio-temporally efficient coding could predict the phenomenon of deviant neural responses to unlearned inputs without the introduction of prediction error responses mediated by error units. We checked whether the distribution of neural responses to unlearned input is close to extreme values such as 0 or 1, unlike the distribution of neural responses to learned input.

In the simulations, the visual hierarchical structure learned to be learned with natural scene images, and novel handwritten digit images were used as unlearned visual inputs (Figure 15A). The simulation

results revealed that neural responses were distributed over middle values for learned images and over smaller or larger values for unlearned inputs (Figure 15B and 15C), suggesting that spatio-temporally efficient coding could predict the phenomenon of deviant neural responses to unlearned inputs. The same conclusion could also be reached if the learned images were handwritten digit images and the unlearned images were natural scene images (Figure 15D, 15E, and 15F).



**Figure 15. Neural response distributions for learned and unlearned inputs.** (A) Natural scene images are used as learned inputs (for learning in the visual hierarchical structure) and handwritten digit images are used as unlearned inputs (not used in learning). (B) Examples of neural response distributions of lower hierarchy units for each of the five time steps of bidirectional information transmissions, corresponding to images in (A). The colour scale indicates the proportion. (C) Overall neural response distributions in response to all input images (either learned or unlearned) at lower and upper hierarchies, respectively. (D), (E), (F) are similar to (A), (B), (C), but handwritten digit images are used as learned inputs and natural scene images are used as unlearned inputs.

#### 4.C.4. Results: Preferred orientation biases of receptive fields

Neurons in the visual system prefer horizontal and vertical orientations over oblique orientations (Furmanski and Engel, 2000; Li et al., 2003). Indeed, orientation discrimination is more sensitive to horizontal and vertical orientations than to oblique orientations (Girshick et al., 2011). This is a bias towards cardinals (Girshick et al., 2011). Since smooth neural representations of spatio-temporally efficient coding reflects the structure of the external world well, we predicted that it would also reflect the environmental statistics of natural scenes. We investigated whether units in the visual hierarchical structure that learned by spatio-temporally efficient coding of natural scene images exhibited such biases. We checked whether the neural responses of the units were highest to the horizontal/vertical orientation stimuli. Units in the lower hierarchy had the Gabor-like visual receptive fields, while units in the upper hierarchy had more complex visual receptive fields (Figure 16A). Because the units had an oriented Gabor-like receptive field, we used oriented bar stimuli to measure the unit's preferred orientation. We presented a static bar oriented in one of eight angles that moved in the direction perpendicular to the orientation angle (Figure 16B) and defined the response of each unit to that orientation by the largest response during presentation. We then defined the preferred orientation of each unit as the orientation that elicited the largest response. Example neural responses were shown in Figure 16C. Simulation results revealed that units clearly prefer horizontal and vertical orientations over oblique orientations (Figure 16D), consistent with the orientation bias of visual cortical neurons and in accordance with context-independent bottom-up prediction (Teufel and Fletcher, 2020). This means that smooth neural representations when well-balanced between temporally efficient coding objective and spatially efficient coding objective reflect the environmental statistics of natural scenes.



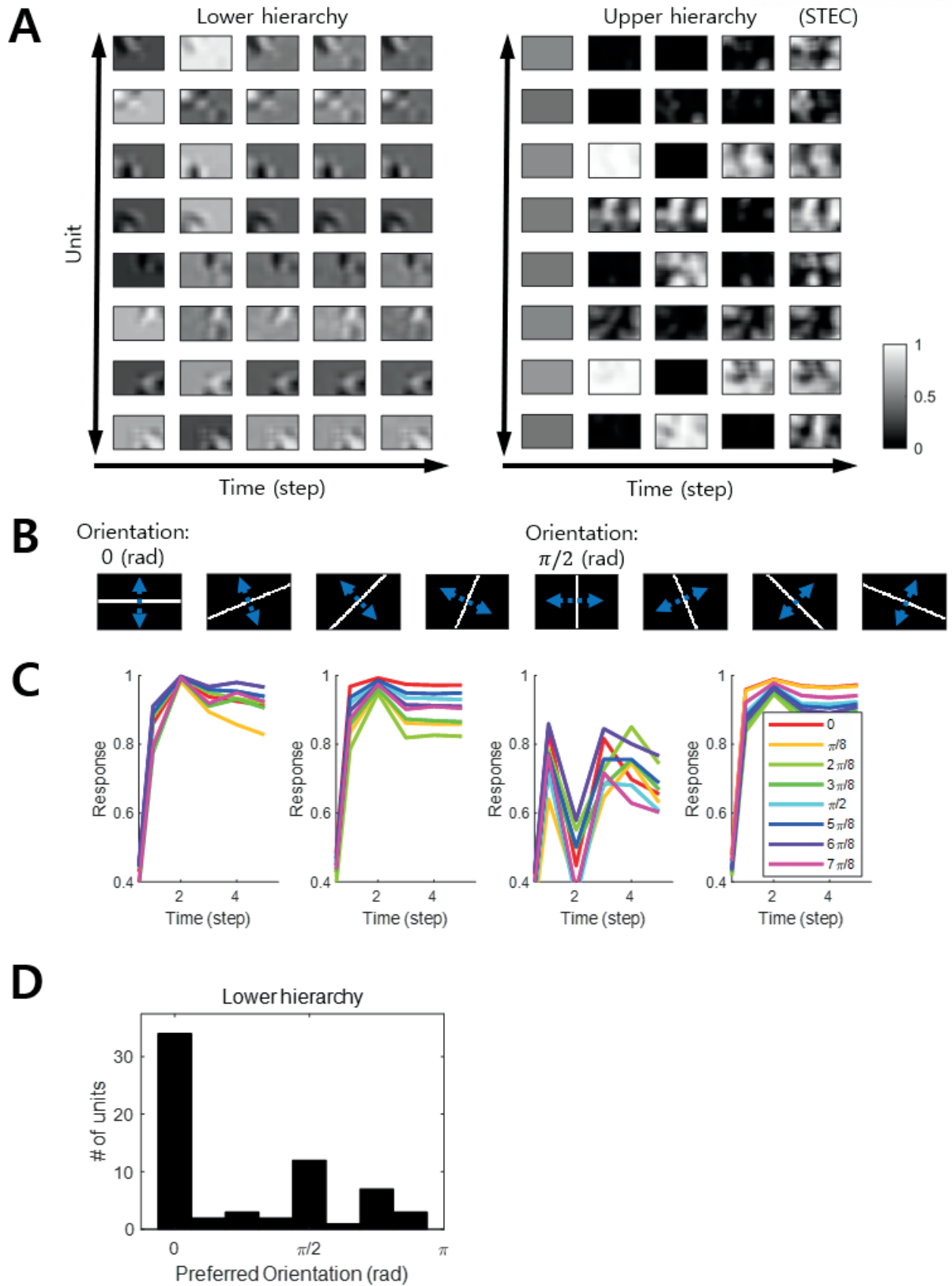


Figure 16. Orientation preference. (A) Examples of visual receptive fields of lower and upper



hierarchy units. (B) Orientation images used to test the visual orientation preference of neuronal units at lower hierarchy in the visual hierarchical structure that learned by the spatio-temporally efficient coding of natural scene images. White bars in each of the eight orientations are moved in the perpendicular directions denoted by blue dotted arrows. (C) Examples of lower hierarchy neural responses for bar stimuli. Each colour indicates each orientation. The neural responses of each orientation were obtained from the bar positions with the largest neural responses at time step 5. These four example units correspond to 1st, 3rd, 5th, and 7th units in (A). (D) Histograms of the orientation preference at time step.

## 5. Spatio-temporally efficient coding: a case of dynamic sensory input

### 5.A. Introduction

In the bidirectional hierarchical structure, two types of neuronal noise can be generated, where neuronal noise is defined as the uncertainty of neural responses for given sensory input (Borst and Theunissen, 1999). One of them is noise generated as neural information fluctuates across the hierarchy according to the initial condition of the neural response, even if the external sensory input is static. It is known that this can occur in interconnected structures as a chaotic dynamics sensitive to the initial condition (Rubinov et al., 2009; Tomov et al., 2014). This is neuronal noise when static sensory inputs are given. Another is noise, precisely error, caused by coding different information in each hierarchy because of the transmission delay of information (Berry et al., 1999) when external sensory input is dynamic. Because of the information transmission delay, (lower) hierarchies close to the sensory organ represent relatively recent information, and (upper) hierarchies distant from the sensory organ represent information relatively old, if external inputs are dynamic (changed). By the top-down pathway, old information in the upper hierarchy affects recent information in the lower hierarchy, and this becomes neuronal noise when dynamic sensory inputs are given. In this section, we will show that spatio-temporally efficient coding can deal with the neuronal noise (error) of second type (case of dynamic sensory inputs).

By spatio-temporally efficient coding (especially temporally efficient coding), bidirectional hierarchical structures learn smooth neural responses. To render neural response smooth when sensory input changes smoothly is to make similar neural responses for similar sensory inputs. Since sensory inputs at adjacent time may be similar, neural responses are also similar. These similar neural responses on time domain reduce the discrepancy of represented information across the hierarchy. Therefore, we can expect that spatio-temporally efficient coding can overcome the aforementioned noise (error) problem of second type (case of dynamic sensory inputs). In this section, we will verify this expectation through simulation.

### 5.B. Methods

The natural scene image dataset created by van Hateren (van Hateren and van der Schaaf, 1998) was used for simulations. The dataset was publicly available at <http://bethgelab.org/datasets/vanhateren/>. The images were downsized to  $128 \times 192$  pixels and rescaled between zero and one.

A sensory input  $X_{image}$  was composed of a series of parts ( $64 \times 64$  pixels) of a natural scene image,

where each part was obtained by moving a gaze at a constant velocity starting at a random location over the whole image ( $128 \times 192$ ). The magnitude of the velocity was randomly selected from an integer between 0 and 4 pixels per time step. If the selected velocity magnitude is too large to move within the whole image, the velocity magnitude was adjusted to be the maximum magnitude that allowed moving within the whole image. The direction of the velocity was randomly selected in the two-dimensional pixel space, and speed was kept as  $L_\infty$  norm (Figure 17A and 17B).

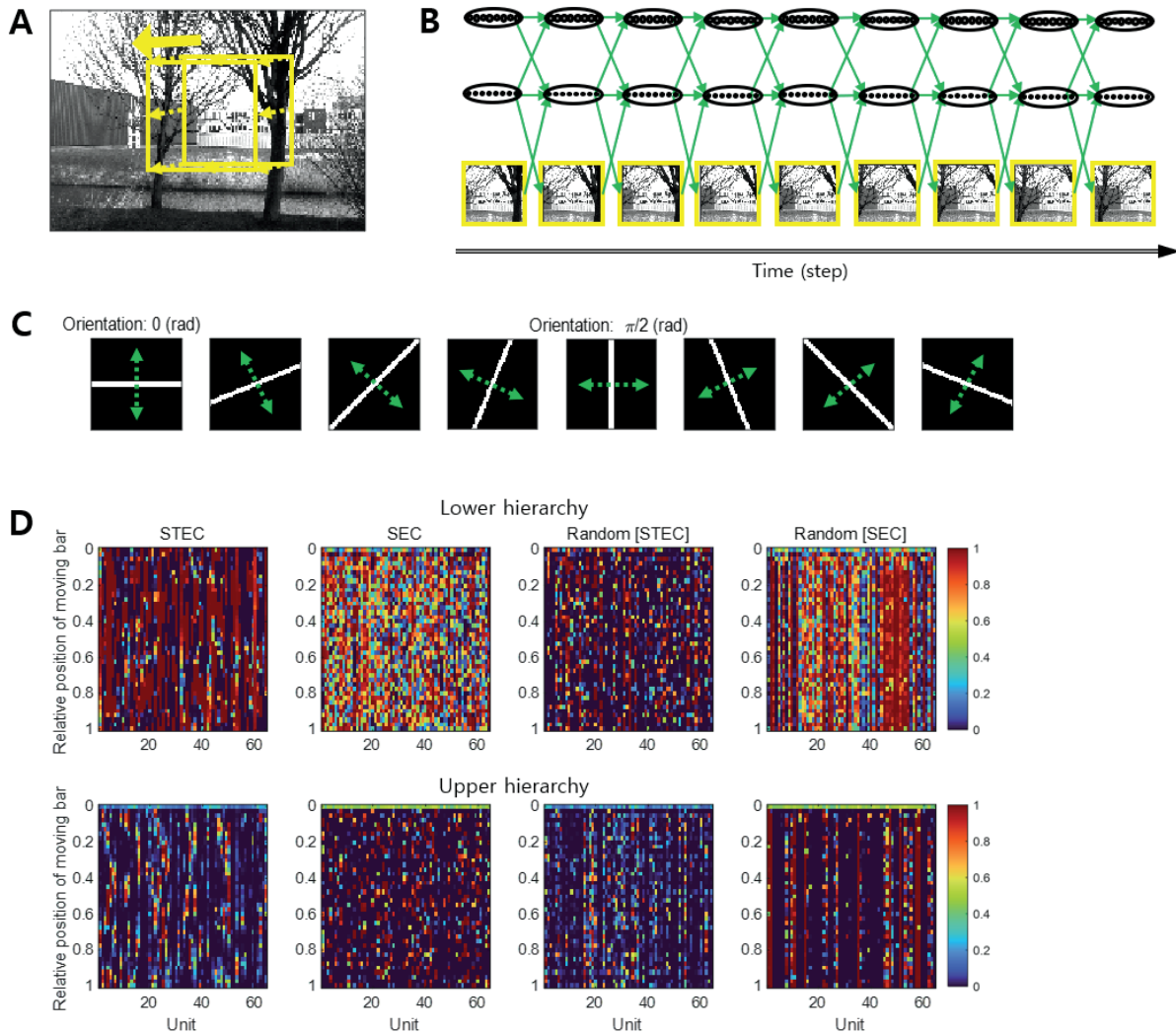
Minimization via Adam optimizer persisted  $10^4$  iterations per one repetition. We restarted the repetition five times. For each minimization iteration above (one of  $10^4$  iterations),  $f_t$  lasted on nine-time steps (i.e.,  $t \in [1,9]$ ), as we assumed that a gaze shifted 9 times in every natural scene image (Figure 17B). The minibatch size was set to 100.

Although smoothly-moving natural scene images were used for learning, bar stimuli that were accurately controlled position and orientation were used for testing. A bar stimulus was created as a  $64 \times 64$  image with a value of one on the bar and zero elsewhere. Bar stimuli had eight orientations:  $0, \pi/8, 2\pi/8, 3\pi/8, 4\pi/8, 5\pi/8, 6\pi/8$ , and  $7\pi/8$  (rad). For each orientation, we created 41 bar stimuli at different positions covering the whole image uniformly. At each position, the length of a bar extended from one edge to the opposite of an image. The width of the bar was kept constant across eight orientations and occupied four pixels in the case of the horizontal orientation (Fig. 17C). Over these 41 positions, we moved the bar stimulus smoothly or randomly. A smoothly-moving bar in a certain orientation sequentially moved from one end position to the other end, taking 41 time-steps to present a smoothly-moving bar. The same bar stimulus was also moved in a reversed order. For convenience, we hereafter denote one of these bidirectional movements as moving forward and the other as moving backward, although there is no specific reason to call one as forward and the other as backward. These bidirectional movements of a bar stimulus generated a total of 16 smoothly-moving bar stimuli. A randomly-moving bar moved from one position to another randomly with no repeated visit to the same position, thus appearing at each of the 41 positions once and taking 41 time-steps, too. We generated 41 randomly-moving bars in each of the eight orientations.

We set  $\lambda = 5, 100$ , where  $\lambda$  is a regularization parameter in the objective function,  $L = L_{Temporal} + \lambda L_{Spatial}$  (see Eq. 3.3.7). The condition  $\lambda = 5$  was set the balanced condition between spatially efficient coding and temporally efficient coding objectives (STEC). The condition  $\lambda = 1000$  was set one of the control conditions which is overweighted toward spatially efficient coding (SEC). This indicates the strict efficient condition.

We set important control condition, random networks, based on STEC and SEC: Random [STEC] and

Random [SEC]. Random [STEC] was made by random permutations of synaptic weights and bias of STEC at each pathway (bottom-up, top-down, and recurrent). Random [SEC] was similarly made from SEC. Therefore, Random [STEC] and Random [SEC] have same synaptic weight magnitude level with STEC and SEC. Example neural responses trajectories of all conditions are shown in Figure 17D.



**Figure 17. A simulation method for dynamic sensory input.** (A) Example dynamic sensory inputs on the natural scene image for learning. Yellow square indicates each visual scene. (B) Spatio-temporally efficient coding for dynamic visual scene. Each yellow square corresponds to one of yellow squares in A. Green arrow indicates information transmission. (C) Bar stimuli with eight orientations, for test. (D) Neural responses of learned sensory hierarchies for smoothly-moving bar stimuli (0 rad orientation).

### 5.C. Results: Consistent neural responses for static and dynamic sensory inputs

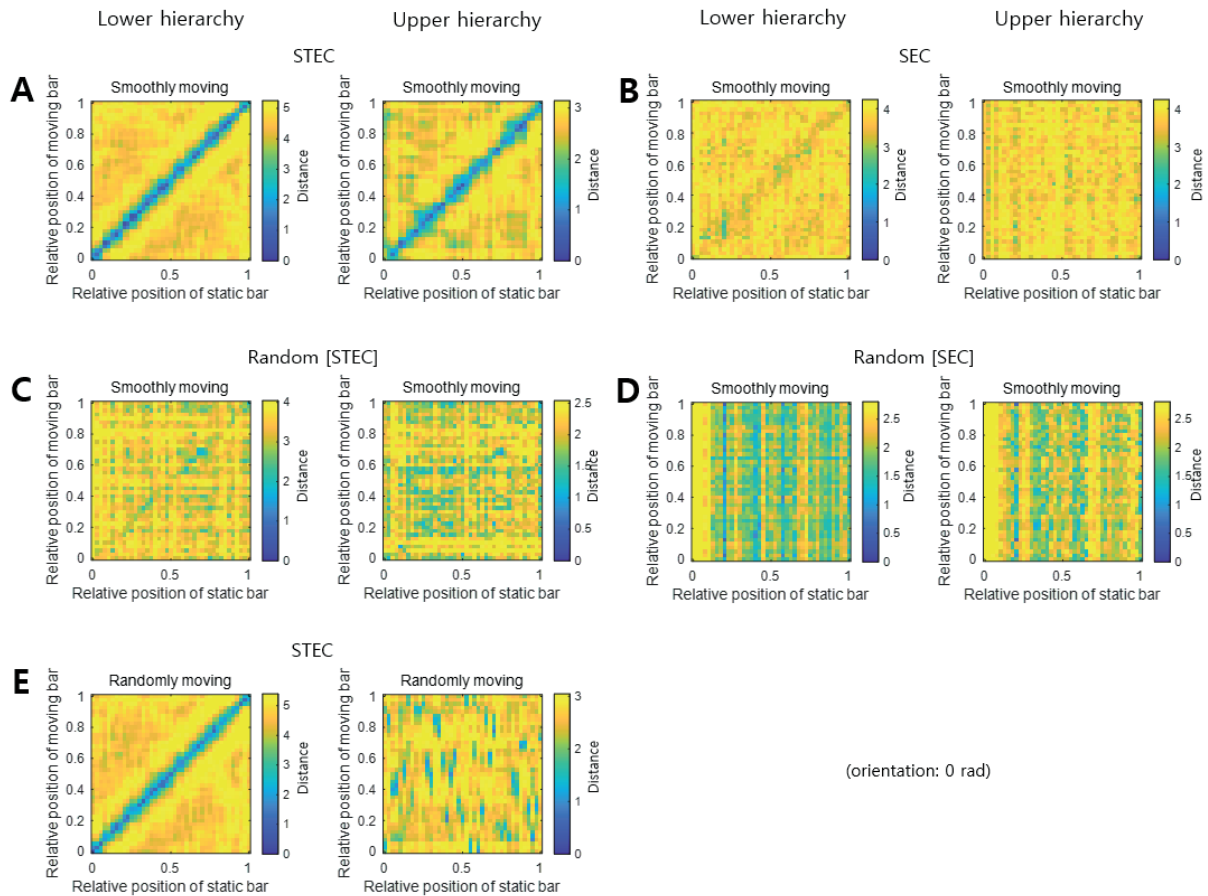
In this subsection, we will show that spatio-temporally efficient coding can overcome the aforementioned noise (error) problem (case of dynamic sensory inputs).

As one method to measure the robustness to the aforementioned noise (error), the distance between the neural responses to static bars and moving bars in various positions and same orientations was measured. Neural responses to moving bars in the brain hierarchy transmit the aforementioned noise (error) to other hierarchies. On the other hand, neural responses to static bars do not have such the aforementioned noise (error). Therefore, it can be said that neural coding overcomes the aforementioned noise (error) if the distance between neural responses for static and moving bars in the same position is small.

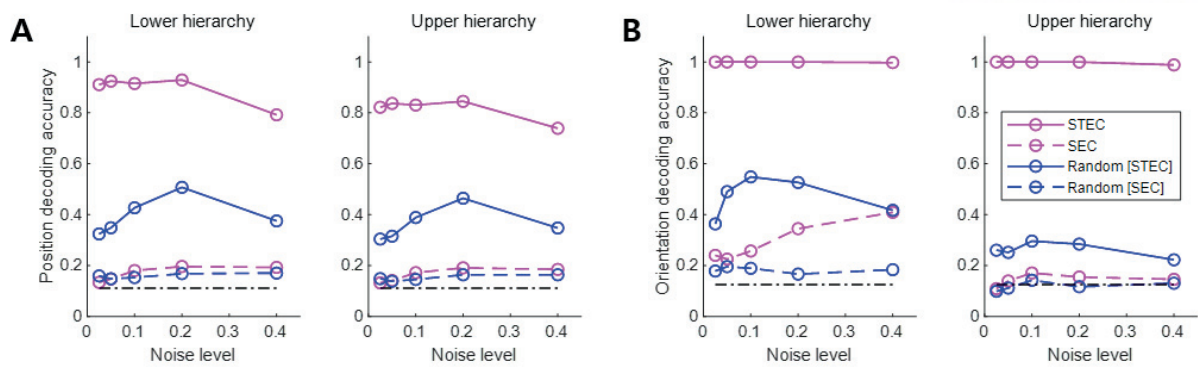
When learned with spatio-temporally efficient coding (STEC), the distance between neural responses for static and moving bars in the same position was small in both upper and lower hierarchies (Fig. 18A), indicating the neural coding is robust to the aforementioned noise (error). On the other hand, the distance between neural responses for static and moving bars in the same position was large in the control conditions (SEC, Random [STEC], Random [SEC]) (Fig. 18B, 18C, and 18D), indicating the neural coding was not robust to the aforementioned noise (error). Distances of off-diagonal entries increase gradually around diagonal entries, indicating that neural responses locally preserve the structure of the external world (Fig. 18A). This robustness in spatio-temporally efficient coding could be achieved by smoothness. The neural response in the cortical hierarchy to a randomly-moving bar is no longer smooth, which leads to the result that the neural coding was not robust to the aforementioned noise (Fig. 18E). This suggests that the robustness to the aforementioned noise (error) is achieved by the smoothness of dynamic sensory inputs.

As another method to measure the robustness of neural coding, the neural response to a static bar was used as a training set and the neural response to a smoothly-moving bar was used as a test set to measure the decoding accuracy for the same position. It can be said that neural coding overcomes the aforementioned noise (error) if information about moving bars can be accurately decoded from neural responses to moving bars using neural response distribution for static bars in the same position.

When learned with spatio-temporally efficient coding (STEC), it was showed higher accuracy than control conditions (SEC, Random [STEC], Random [SEC]) in both upper and lower hierarchies in case of position decoding (Fig.19A). In the case of orientation decoding, it was also showed lower accuracy in the control conditions than when learned with spatio-temporally efficient coding (STEC) (Fig. 19B). These results indicate that smooth neural coding via spatio-temporally efficient coding is robust to miss-informed noise (erroneous information).



**Figure 18. Distance between neural responses denoting consistent neural representations.** (A) Distances between neural responses for static and smoothly-moving bar stimuli with different positions and the 0 rad orientation. STEC. (B) Similar to A, but SEC. (C) Similar to A, but Random [STEC]. (D) Similar to A, but Random [SEC]. (E) Distances between neural responses for static and randomly-moving bar stimuli with different positions and the 0 rad orientation. STEC. A low distance value of diagonal entries indicates consistent neural response between static and moving bar stimuli, denoting consistent neural representations. Gradually increasing distances of off-diagonal entries around diagonal entries indicate that neural responses locally preserve the structure of the external world.



**Figure 19. Decoding of bar stimuli denoting consistent neural representations.** (A) Position decoding accuracy at various noise level for nine-centered-bar positions. Training set is the neural responses for static bar stimuli and test set is the neural responses for smoothly-moving bar stimuli. Black dash-dotted line indicates the chance level. (B) Orientation decoding accuracy at various noise level for eight orientations of three-centered-bar. Training set is the neural responses for static bar stimuli and test set is the neural responses for smoothly-moving bar stimuli. Black dash-dotted line indicates the chance level. A high accuracy value indicates consistent neural response between static and moving bar stimuli, denoting robust neural representations. Noise levels are the standard deviations of the adding random noise to bar stimuli. Decoding is performed with naïve Bayes classifier (Hastie et al., 2009).



## 6. Discussion

The brain sensory systems have bidirectional hierarchical structures. These bidirectional hierarchical structures can cause neuronal noise problems of two kinds. One of them is noise generated as neural information fluctuates across the hierarchy according to the initial condition of the neural response, even if the external sensory input is static. It is known that this can occur in interconnected structures as a chaotic dynamics sensitive to the initial condition (Rubinov et al., 2009; Tomov et al., 2014). This is neuronal noise when static sensory inputs are given. Another is noise, precisely error, caused by coding different information in each hierarchy because of the transmission delay of information (Berry et al., 1999) when external sensory input is dynamic. Because of the information transmission delay, (lower) hierarchies close to the sensory organ represent relatively recent information, and (upper) hierarchies distant from the sensory organ represent information relatively old, if external inputs are dynamic (changed). By the top-down pathway, old information in the upper hierarchy affects recent information in the lower hierarchy, and this becomes neuronal noise (error) when dynamic sensory inputs are given.

In the present study, spatio-temporally efficient coding was devised to overcome two aforementioned noise (error) problems (Section 3). In the section 4, we showed that spatio-temporally efficient coding can overcome the aforementioned noise problem of first kind (case of static sensory inputs) by simulations. In the section 5, we showed that spatio-temporally efficient coding can overcome the aforementioned noise (error) problem of second kind (case of dynamic sensory inputs) by simulations.

Appropriate neural representations in the present study were decodable stable neural representations which is an appropriate function to hierarchical brain structures on the time domain. To demonstrate this, we compared spatio-temporally efficient coding (STEC condition) with other conditions (SEC, Random [STEC], and Random [SEC]) as shown in Figure 13. Only spatio-temporally efficient coding showed decodable stable neural representations. Existing predictive coding is problematic to apply to such real-time information processing (Hogendoorn and Burkitt, 2019). It would be interesting to compare the results with the conceptually improved predictive coding as well. Temporal coherence (Hurri and Hyvärinen, 2002; Zou et al., 2011) or slow feature analysis (Wiskott and Sejnowski, 2002; Berkes and Wiskott, 2005; Creutzig and Sprekeler, 2008) elicits smooth changes in neural responses, similar to temporally efficient coding in the present study. However, they were not used as comparative models because they are not intended to elicit different neural responses to different external inputs and are not suitable for direct application to the bidirectional multiple hierarchical structure of in the present

study. Nevertheless, they can be substituted to the role of temporally efficient coding in the present study.

In the present study, neural responses were rate coded such that they can have continuous value between 0 and 1. However, we can choose spike event coding which has binary value such as 0 and 1. In this case, neural response space is a high-dimensional lattice instead of hypercube, and neural response density function of Equation 3.3.6 should be estimated on a high-dimensional lattice. In this space, implementation of low neural response (sparse neural response) will become sparseness constraint. Such space could treat temporal coding of spike event, which has important meanings.

Since its initial proposal (Attneave, 1954; Barlow, 1961), spatially efficient coding has been validated experimentally (Laughlin, 1981). However, observed correlations between neurons, which maximise entropy to a lesser extent compared to mere spatially efficient coding assuming no inter-neuronal correlations, have yet to be incorporated into the principle of spatially efficient coding. Empirically observed neuronal correlations may drive computational processes of the brain away from strict spatially efficient coding. Recent studies suggest that biological visual systems are intermediate between strict spatially efficient coding and correlated neural responses (Stringer et al., 2019). Therefore, to create biologically plausible computational models, it is necessary to mitigate the spatially efficient coding objective by combining firing-rate-dependent correlations (de la Rocha et al., 2007). This enables more accurate information transmissions of visual perception mediated by visual hierarchical structures. As we focused on integrating spatially efficient coding with temporally efficient coding for computation in hierarchical structures, this study did not incorporate the correlations between neurons in spatially efficient coding, which will be pursued in follow-up studies.

Based on our simulations, we observed that the learning of bidirectional information transmission networks with spatio-temporally efficient coding was hindered when the balancing parameter  $\lambda$  was too small or too large (Figure 12C). Therefore, it was necessary to confine  $\lambda$  within a certain range, in which the magnitude of  $\lambda$  affected neural responses such that a larger  $\lambda$  rendered responses more variable (Figure 12C). Such increased variability is likely to originate from recurrent responses via higher hierarchies. This was confirmed by the observation that top-down synaptic weights become larger than bottom-up synaptic weights when  $\lambda$  increased (Figure 12B). A large  $\lambda$  rendered stronger top-down synaptic connections in lower hierarchy (Figure 12B), which is consistent with the previous finding that top-down synaptic connections are stronger than bottom-up connections in the lateral geniculate nucleus (Sillito et al., 2006). As to why top-down synaptic weights increase with a larger  $\lambda$  value (Figure 12B), we speculate that learning via spatio-temporally efficient coding may increase the range of neural responses to maximise entropy through top-down pathways. While the bottom-up

pathways originating from external inputs are invariant during learning, the top-down pathways originating from higher hierarchy neural responses are more flexible to adjustment to maximise entropy during learning. An increase in these top-down synaptic weights predicts impairment of eye movement tracking partially occluded visual targets in schizophrenic patients (Adams et al., 2012). This may be due to an increased higher hierarchy's influence in patients with schizophrenia. The increased higher hierarchy's influence in our simulations is that they do not stabilize neural responses sufficiently to distinguish a given input from other inputs (Figure 13C).

The deviant neural responses to unlearned inputs observed in this study (Figure 15) arise from smooth neural representations for learned inputs (distributed over the middle value). As such, neural representations for learned inputs extrude neural responses to unlearned inputs into a range of deviant neural responses. We conjecture that the visual system may generate deviant neural responses via a similar mechanism.

Spatio-temporally efficient coding predicted a bias in preferred orientations (Fig 16). In this regard, spatially efficient coding alone has been reported to predict bias in preferred orientations (Ganguli and Simoncelli, 2014). Notably, spatio-temporally efficient coding was able to predict this bias well, even when  $\lambda$  was low, that is, when the spatially efficient coding objective was less weighted (Fig 16D). Therefore, this bias prediction should be viewed as a result of spatio-temporally efficient coding, not as a result of spatially efficient coding alone.

Spatio-temporally efficient coding minimizes temporal differences between the present and future neural responses. It thereby renders the temporal trajectories of neural responses smooth. Smoothing the temporal trajectory of a neural response when external inputs changes smoothly is a coding method that preserves smoothness. In other words, the smoothly changing external world creates smooth neural representations in the brain via spatio-temporally efficient coding. This was also shown in the simulations of the present study using bar stimuli which are frequently selected as experimental stimuli in the early visual system (Figure 17D).

Since there is a transmission delay when information is transmitted between each hierarchy (Berry et al., 1999), when a moving object is represented in the visual system, the upper hierarchy far from the eyes has more past information, and the lower hierarchy close to the eyes has more present information. Inter-hierarchy interconnection conveys this different information to other hierarchies, which becomes noise at each hierarchy. This is the aforementioned noise (error) of second kind (case of dynamic sensory inputs). Neural responses to moving bars in the brain hierarchy transmit the aforementioned noise (error) to other hierarchies. On the other hand, neural responses to static bars do not have such aforementioned noise (error). Therefore, if the neural responses to static bars and moving bars are

similar, it can be said that this neural coding is robust to the aforementioned noise (error). This was shown through the distance between neural responses (Figure 18) and the decoding of bar stimuli (Figure 19) in our simulations, indicating that the smooth coding via spatio-temporally efficient coding is robust to the aforementioned noise (error).

This robustness could be achieved because spatio-temporally efficient coding is smooth coding that preserves structures of the external world locally in the brain. When the external world changes smoothly, smooth coding, which reflects the smooth change in neural response, preserves the difference in appearance of the external world as a difference in neural response. This was also confirmed in the simulations in the present study; gradually increasing distances of off-diagonal entries around diagonal entries indicate that neural responses locally preserve the structure of the external world (Figure 18A). If the structure of the external world is preserved in the neural response and the time difference between the present and the past is not very large, a similar appearance of the external world will be reflected in the similar neural response. This similar neural responses between the present and the past will reduce the difference in information represented in different hierarchies due to transmission delay. This means that the impact of the aforementioned noise (error) is reduced.

Reducing the difference in information represented between the present and the past also means that the difference in information represented between the present and the future is decreasing. This is consistent with the predictive information that the real visual system has information about future stimuli in advance (Palmer et al., 2015; Chen et al., 2017; Sederberg et al., 2018; Liu et al., 2021). It is known that a receptive field similar to reality can be obtained by using such future predictive coding (Singer et al., 2018). The efficient coding principle for future prediction through information bottleneck framework was also presented (Chalk et al., 2018).

The trial-to-trial variability of neural responses can be widely observed in the brain (Malins et al., 2018; Daniel et al., 2019; Nogueira et al., 2020; Li et al., 2021; Zhang et al., 2022). It is observed as both electrophysiological signals (Daniel et al., 2019; Li et al., 2021) and blood-oxygenation-level-dependent signals (Malins et al., 2018; Zhang et al., 2022). These trial-to-trial variability of neural responses are neuronal noise. Assuming strict efficient coding, neurons must code different information to minimise redundancy. This makes neural coding susceptible to neuronal noise (Pryluk et al, 2019). If we allow some redundancy, this coding can be made more robust to neuronal noise (Pryluk et al, 2019). In the real brain, neurons are known to perform robust coding (not strict efficient coding) against neuronal noise by having smooth tuning curves (Stringer et al., 2019). These smooth tuning curves are observed from many experimental data (Chen and Hafed, 2018; Kutter et al., 2018; Chettih and Harvey, 2019; Christensen et al., 2019), supporting the mitigation of strict efficient coding. Spatio-temporally

efficient coding makes the neural responses distributed around the middle value by minimizing the temporal difference of the neural responses (Figure 14). This is deviant from strict efficient coding as it does not reach entropy maximization. Therefore, it may be closer to the coding principle of the real brain.

The present study has several limitations. First, for simplicity, our simulation model contained only two hierarchies. However, it is necessary to explore how spatio-temporally efficient coding operates in models with more hierarchies. We also modelled 64 neuronal units at each hierarchy, as we assumed that this would be sufficient to represent the natural scene images used in this study. Nevertheless, the interactions between the number of neuronal units, levels of hierarchy, and spatio-temporally efficient coding require further investigation. Second, the hierarchical structure used in the present study is too simplistic to explain complex perceptual phenomena such as optical illusions. In the present study, the hierarchical structure has a simple structure that fully connected. However, brain sensory system has more complex hierarchical structure which may be able to explain complex perceptual phenomena. It is necessary to construct a more complex and realistic structure to implement spatio-temporally efficient coding. Third, we assumed that the external sensory input is static or changed smoothly in the present study. This does not include sudden changes in sensory input resulting from saccadic eye movements. Finally, the scope of the present study was limited to the visual system given that its hierarchical structure is well documented, but spatio-temporally efficient coding may be applied to other systems (e.g., somatosensory system) or to movements and planning. Recent efficient coding researches are expanding their scope to perception generalization (Sims, 2018), subjective value (Polanía et al., 2019), and memory (McPherson and McDermott, 2020). It seems that spatio-temporally efficient coding can also broaden its scope.

## References

1. Adams RA, Perrinet LU, Friston K. Smooth pursuit and visual occlusion: Active inference and oculomotor control in schizophrenia. *PLOS ONE* 2012 Oct 26; 7(10):e47502. doi: 10.1371/journal.pone.0047502
2. Adrian E. *The basis of sensation, the action of the sense organs*. New York: Norton; 1928. p 122.
3. Allman J, Miezin F, McGuinness E. Stimulus specific responses from beyond the classical receptive field: Neurophysiological mechanisms for local-global comparisons in visual neurons. *Annual Review of Neuroscience* 1985 Mar 01; 8(1):407-430. doi: 10.1146/annurev.ne.08.030185.002203
4. Attneave F. Some informational aspects of visual perception. *Psychological Review* 1954; 61(3):183–193. doi: 10.1037/h0054663
5. Barlow HB. Possible principles underlying the transformations of sensory messages. In: Rosenblith WA editor. *Sensory communication*, MIT Press; 1961. p 217-234. doi: 10.7551/mitpress/9780262518420.003.0013
6. Berens P, Ecker AS, Cotton RJ, Ma WJ, Bethge M, Tolias AS. A fast and simple population code for orientation in primate V1. *J. Neurosci.* 2012 Aug 01; 32(31):10618-10626. doi: 10.1523/JNEUROSCI.1335-12.2012
7. Berkes P, Wiskott L. Slow feature analysis yields a rich repertoire of complex cell properties. *Journal of Vision* 2005 Jul 20; 5(6):579-602. doi: 10.1167/5.6.9
8. Berry MJ, Brivanlou IH, Jordan TA, Meister M. Anticipation of moving stimuli by the retina. *Nature* 1999 Mar 01; 398(6725):334-338. doi: 10.1038/18678
9. Bolz J, Gilbert CD. Generation of end-inhibition in the visual cortex via interlaminar connections. *Nature* 1986 Mar 01; 320(6060):362–365. doi: 10.1038/320362a0
10. Borst A, Theunissen FE. Information theory and neural coding. *Nature Neuroscience* 1999 Nov 01; 2(11):947-957. doi: 10.1038/14731
11. Brouwer GJ, Heeger DJ. Decoding and reconstructing color from responses in human visual cortex. *J. Neurosci.* 2009 Nov 04; 29(44):13992-14003. doi: 10.1523/JNEUROSCI.3577-

09.2009

12. Bullmore E, Sporns O. The economy of brain network organization. *Nat Rev Neurosci* 2012 May 01; 13(5):336–349. doi: 10.1038/nrn3214
13. Butts DA, Weng C, Jin J, Yeh C-I, Lesica NA, Alonso J-M, Stanley GB. Temporal precision in the neural code and the timescales of natural vision. *Nature* 2007 Sep 01; 449(7158):92-95. doi: 10.1038/nature06105
14. Burt JB, Demirtaş M, Eckner WJ, Navejar NM, Ji JL, Martin WJ, Bernacchia A, Anticevic A, Murray JD. Hierarchy of transcriptomic specialization across human cortex captured by structural neuroimaging topography. *Nat Neurosci* 2018 Sep 01; 21(9):1251–1259. doi: 10.1038/s41593-018-0195-0
15. Caminiti R, Johnson PB, Urbano A. Making arm movements within different parts of space: dynamic aspects in the primate motor cortex. *J. Neurosci.* 1990 Jul 01; 10(7):2039-2058. doi: 10.1523/JNEUROSCI.10-07-02039.1990
16. Chalk M, Marre O, Tkačik G. Toward a unified theory of efficient, predictive, and sparse coding. *Proc Natl Acad Sci USA* 2018 Jan 02; 115(1):186-191. doi: 10.1073/pnas.1711114115
17. Chen KS, Chen C-C, Chan CK. Characterization of predictive behavior of a retina by mutual information. *Front. Comput. Neurosci.* 2017 Jul 20; 11:66. doi: 10.3389/fncom.2017.00066
18. Chen C-Y, Hafed ZM. Orientation and contrast tuning properties and temporal flicker fusion characteristics of primate superior colliculus neurons. *Front. Neural Circuits* 2018; 12:58. doi: 10.3389/fncir.2018.00058
19. Chettih SN, Harvey CD. Single-neuron perturbations reveal feature-specific competition in V1. *Nature* 2019; 567:334–340. doi:10.1038/s41586-019-0997-6
20. Christensen RK, Lindén H, Nakamura M, Barkat TR. White noise background improves tone discrimination by suppressing cortical tuning curves. *Cell Reports* 2019; 29(7):2041-2053.e4. doi: 10.1016/j.celrep.2019.10.049
21. Clark A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences* 2013 May 10; 36(3):181-204. doi: 10.1017/S0140525X12000477
22. Cohen S. Finding color and shape patterns in images. Dissertation, Stanford University; 1999.



23. Creutzig F, Sprekeler H. Predictive coding and the slowness principle: An information-theoretic approach. *Neural Computation* 2008 Apr; 20(4):1026-1041. doi: 10.1162/neco.2008.01-07-455
24. Daniel E, Meindersma T, Arazi A, Donner TH, Dinstein I. The Relationship between trial-by-trial variability and oscillations of cortical population activity. *Scientific Reports* 2019; 9(1):16901. doi: 10.1038/s41598-019-53270-7
25. deCharms RC, Zador A. Neural representation and the cortical code. *Annu. Rev. Neurosci.* 2000 May 01; 23(1):613-647. doi: 10.1146/annurev.neuro.23.1.613
26. de la Rocha J, Doiron B, Shea-Brown E, Josic K, Reyes A. Correlation between neural spike trains increases with firing rate. *Nature* 2007 Aug 01; 448(7155):802–806. doi: 10.1038/nature06028
27. DiCarlo JJ, Zoccolan D, Rust NC. How does the brain solve visual object recognition? *Neuron* 2012 Feb 09; 73(3):415-434. doi: 10.1016/j.neuron.2012.01.010
28. Di Gesù V, Starovoitov V. Distance-based functions for image comparison. *Pattern Recognition Letters* 1999 Feb 01; 20(2):207-214. doi: 10.1016/S0167-8655(98)00115-9
29. Di Lorenzo PM, Victor JD. Taste response variability and temporal coding in the nucleus of the solitary tract of the rat. *Journal of Neurophysiology* 2003 Sep 01; 90(3):1418-1431. doi: 10.1152/jn.00177.2003
30. Felleman DJ, Van Essen DC. Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex*, 1991 Jan 01; 1(1):1-47. doi: 10.1093/cercor/1.1.1
31. Flint RD, Lindberg EW, Jordan LR, Miller LE, Slutzky MW. Accurate decoding of reaching movements from field potentials in the absence of spikes. *Journal of Neural Engineering* 2012 Jun 25; 9(4):046006. doi: 10.1088/1741-2560/9/4/046006
32. Friston K. A theory of cortical responses. *Phil. Trans. R. Soc. B* 2005 Apr 29; 360(1456):815-836. doi: 10.1098/rstb.2005.1622
33. Friston K. The free-energy principle: a unified brain theory? *Nat Rev Neurosci* 2010 Feb 01; 11(2):127–138. doi: 10.1038/nrn2787
34. Friston K, Thornton C, Clark A. Free-energy minimization and the dark-room problem. *Front. Psychol.* 2012 May 08; 3:130. doi: 10.3389/fpsyg.2012.00130

35. Fukushima M, Rauske PL, Margoliash D. Temporal and rate code analysis of responses to low-frequency components in the bird's own song by song system neurons. *J Comp Physiol A* 2015 Dec 01; 201(12):1103–1114. doi: 10.1007/s00359-015-1037-0
36. Furmanski CS, Engel SA. An oblique effect in human primary visual cortex. *Nat Neurosci* 2000 Jun 01; 3(6):535–536. doi: 10.1038/75702
37. Ganguli D, Simoncelli EP. Efficient sensory encoding and bayesian inference with heterogeneous neural populations. *Neural Comput* 2014 Oct 01; 26(10):2103–2134. doi: 10.1162/NECO\_a\_00638
38. Georgopoulos AP, Kalaska JF, Caminiti R, Massey JT. On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *J. Neurosci.* 1982 Nov 01; 2(11):1527-1537. doi: 10.1523/JNEUROSCI.02-11-01527.1982
39. Girshick AR, Landy MS, Simoncelli EP. Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nat Neurosci* 2011 Jul 01; 14(7):926–932. doi: 10.1038/nn.2831
40. Gollisch T, Meister M. Rapid neural coding in the retina with relative spike latencies. *Science* 2008 Feb 22; 319(5866):1108-1111. doi: 10.1126/science.1149639
41. Goodale MA, Milner AD. Separate visual pathways for perception and action. *Trends in Neurosciences* 1992 Jan 01; 15(1):20-25. doi: 10.1016/0166-2236(92)90344-8
42. Griffiths TL, Lieder F, Goodman ND. Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Top Cogn Sci* 2015 Apr 01; 7(2):217-229. doi: 10.1111/tops.12142
43. Hansen JY, Markello RD, Vogel JW, Seidlitz J, Bzdok D, Misic B. Mapping gene transcription and neurocognition across human neocortex. *Nat Hum Behav* 2021 Mar 25. doi: 10.1038/s41562-021-01082-z
44. Harris JA, Mihalas S, Hirokawa KE, Whitesell JD, Choi H, Bernard A, Bohn P, Caldejon S, Casal L, Cho A, Feiner A, Feng D, Gaudreault N, Gerfen CR, Graddis N, Groblewski PA, Henry AM, Ho A, Howard R, Knox JE, Kuan L, Kuang X, Lecoq J, Lesnar P, Li Y, Luviano J, McConoughey S, Mortrud MT, Naemi M, Ng L, Oh SW, Ouellette B, Shen E, Sorensen SA, Wakeman W, Wang Q, Wang Y, Williford A, Phillips JW, Jones AR, Koch C, Zeng H. Hierarchical organization of cortical and thalamic connectivity. *Nature* 2019 Nov 01; 575(7781):195–202. doi: 10.1038/s41586-019-1716-z Hilgetag CC, Goulas A. ‘Hierarchy’ in

the organization of brain networks. *Phil. Trans. R. Soc. B* 2020 Apr 13; 375(1796):20190319.  
doi: 10.1098/rstb.2019.0319

45. Harvey MA, Saal HP, Dammann III JF, Bensmaia SJ. Multiplexing stimulus information through rate and temporal codes in primate somatosensory cortex. *PLOS Biology* 2013 May 07; 11(5):e1001558. doi:10.1371/journal.pbio.1001558
46. Henry CA, Joshi S, Xing D, Shapley RM, Hawken MJ. Functional characterization of the extraclassical receptive field in macaque V1: Contrast, orientation, and temporal dynamics. *J. Neurosci.* 2013 Apr 03; 33(14):6230-6242. doi: 10.1523/JNEUROSCI.4155-12.2013
47. Hastie T, Tibshirani R, Friedman J. *The elements of statistical learning: Data mining, inference, and prediction.* 2nd ed. Springer; 2009. doi: 10.1007/978-0-387-84858-7
48. Hogendoorn H, Burkitt AN. Predictive coding with neural transmission delays: a real-time temporal alignment hypothesis. *eNeuro* 2019 May 06; 6(2). doi: 10.1523/ENEURO.0412-18.2019
49. Houghton C, Kreuz T. On the efficient calculation of van Rossum distances. *Network: Computation in Neural Systems* 2012 Mar 01; 23(1-2):48-58. doi: 10.3109/0954898X.2012.673048
50. Houghton C, Victor JD. Measuring representational distances—the spike-train metrics approach. In: Kriegeskorte N, Kreiman G editors. *Visual population codes: Toward a common multivariate framework for cell recording and functional imaging*, MIT Press; 2010, p 213-250.
51. Hu X, Zhang J, Li J, Zhang B. Sparsity-regularized HMAX for visual recognition. *PLOS ONE* 2014 Jan 02; 9(1):e81813. doi: 10.1371/journal.pone.0081813
52. Huang G, Ramachandran S, Lee TS, Olson CR. Neural correlate of visual familiarity in macaque area V2. *J. Neurosci.* 2018 Oct 17; 38(42):8967-8975. doi: 10.1523/JNEUROSCI.0664-18.2018
53. Huang Y, Rao RPN. Predictive coding, *WIREs Cogn Sci* 2011 Sep 01; 2(5):580-593. doi: 10.1002/wcs.142
54. Hubel DH, Wiesel TN. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology* 1962 Jan 01; 160(1):106-154. doi: 10.1113/jphysiol.1962.sp006837

55. Hubel DH, Wiesel TN. Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology* 1968 Mar 01; 195(1):215-243. doi: 10.1113/jphysiol.1968.sp008455
56. Huh CYL, Peach JP, Bennett C, Vega RM, Hestrin S. Feature-specific organization of feedback pathways in mouse visual cortex. *Current Biology* 2018 Jan 08; 28(1):114-120.e5. doi: 10.1016/j.cub.2017.11.056
57. Hurri J, Hyvärinen A. Temporal coherence, natural image sequences, and the visual cortex. *Proceedings of the Conference on Neural Information Processing Systems*; 2002 Dec 9-12; Vancouver, British Columbia, Canada.
58. Issa EB, Cadieu CF, DiCarlo JJ. Neural dynamics at successive stages of the ventral visual stream are consistent with hierarchical error signals. *eLife* 2018 Nov 28; 7:e42870. doi: 10.7554/eLife.42870
59. Jamali M, Chacron MJ, Cullen KE. Self-motion evokes precise spike timing in the primate vestibular system. *Nature Communications* 2016 Oct 27; 7(1):13229. doi: 10.1038/ncomms13229
60. Johansson RS, Flanagan JR. Coding and use of tactile signals from the fingertips in object manipulation tasks. *Nature Reviews Neuroscience* 2009 May 01; 10(5):345-359. doi: 10.1038/nrn2621
61. Kalaska JF, Cohen DA, Hyde ML, Prud'homme M. A comparison of movement direction-related versus load direction-related activity in primate motor cortex, using a two-dimensional reaching task. *J. Neurosci.* 1989 Jun 01; 9(6):2080-2102. doi: 10.1523/JNEUROSCI.09-06-02080.1989
62. Kamitani Y, Tong F. Decoding the visual and subjective contents of the human brain. *Nature Neuroscience* 2005 Apr 24; 8(5):679-685. doi: 10.1038/nn1444
63. Kantorovich LV. On one effective method of solving certain classes of extremal problems. *Dokl. Akad. Nauk USSR* 1940; 28:212–215.
64. Karklin Y, Lewicki MS. Emergence of complex cell properties by learning to generalize in natural scenes. *Nature* 2009 Jan 01; 457(7225):83-86. doi: 10.1038/nature07481
65. Kingma DP, Ba J. Adam: A method for stochastic optimization. *Proceedings of the 3th International Conference on Learning Representations*; 2015 May 7-9; San Diego, California,

United States.

66. Krause BM, Murphy CA, Uhrich DJ, Banks MI. PV+ cells enhance temporal population codes but not stimulus-related timing in auditory cortex. *Cereb Cortex* 2019 Feb 01; 29(2):627-647. doi: 10.1093/cercor/bhx345
67. Kreuz T, Chicharro D, Greschner M, Andrzejak RG. Time-resolved and time-scale adaptive measures of spike train synchrony. *Journal of Neuroscience Methods* 2011 Jan 30; 195(1):92-106. doi: 10.1016/j.jneumeth.2010.11.020
68. Kreuz T, Chicharro D, Houghton C, Andrzejak RG, Mormann F. Monitoring spike train synchrony. *Journal of Neurophysiology* 2013 Mar 01; 109(5):1457-1472. doi: 10.1152/jn.00873.2012
69. Kreuz T, Haas JS, Morelli A, Abarbanel HDI, Politi A. Measuring spike train synchrony. *Journal of Neuroscience Methods* 2007 Sep 15; 165(1):151-161. doi: 10.1016/j.jneumeth.2007.05.031
70. Kriegeskorte N, Diedrichsen J. Peeling the onion of brain representations. *Annu. Rev. Neurosci.* 2019 Jul 08; 42(1):407-432. doi: 10.1146/annurev-neuro-080317-061906
71. Kutter EF, Bostroem J, Elger CE, Mormann F, Nieder A. Single neurons in the human brain encode numbers. *Neuron* 2018; 100(3):753-761.e4. doi: 10.1016/j.neuron.2018.08.036
72. Laughlin SB. A simple coding procedure enhances a neuron's information capacity. *Z. Naturforsch.* 1981; 36(9-10):910-912. doi: 10.1515/znc-1981-9-1040
73. Laughlin SB. Energy as a constraint on the coding and processing of sensory information. *Current Opinion in Neurobiology* 2001 Aug 01; 11(4):475-480. doi: 10.1016/S0959-4388(00)00237-3
74. Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 1998 Nov; 86(11):2278-2324. doi: 10.1109/5.726791
75. Li D, Constantinidis C, Murray JD. Trial-to-trial variability of spiking delay activity in prefrontal cortex constrains burst-coding models of working memory *J. Neurosci.* 2021; 41(43):8928-8945. doi: 10.1523/JNEUROSCI.0167-21.2021
76. Li B, Peterson MR, Freeman RD. Oblique Effect: A neural basis in the visual cortex. *J. Neurophysiol.* 2003 Jul 01; 90(1):204-217. doi: 10.1152/jn.00954.2002

77. Lieder F, Griffiths TL. Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences* 2020; 43:E1. doi: 10.1017/S0140525X1900061X
78. Liu B, Hong A, Rieke F, Manookin MB. Predictive encoding of motion begins in the primate retina. *Nature Neuroscience* 2021 Aug 02. doi: 10.1038/s41593-021-00899-1
79. Logiaco L, Quilodran R, Procyk E, Arleo A. Spatiotemporal spike coding of behavioral adaptation in the dorsal anterior cingulate cortex. *PLOS Biology* 2015 Aug 12; 13(8):e1002222. doi: 10.1371/journal.pbio.1002222
80. Machens CK, Prinz P, Stemmler MB, Ronacher B, Herz AVM. Discrimination of behaviorally relevant signals by auditory receptor neurons. *Neurocomputing* 2001 Jun 01; 38-40:263-268. doi: 10.1016/S0925-2312(01)00382-4
81. MacLeod K, Bäcker A, Laurent G. Who reads temporal information contained across synchronized and oscillatory spike trains? *Nature* 1998 Oct 01; 395(6703):693-698. doi: 10.1038/27201
82. Malins JG, Pugh KR, Buis B, Frost SJ, Hoeft F, Landi N, Mencl WE, Kurian A, Staples R, Molfese PJ, Sevcik R, Morris R. Individual differences in reading skill are related to trial-by-trial neural activation variability in the reading network. *J. Neurosci.* 2018; 38(12):2981-2989. doi: 10.1523/JNEUROSCI.0907-17.2018
83. Margoliash D. Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow. *J. Neurosci.* 1983 May 01; 3(5):1039-1057. doi: 10.1523/JNEUROSCI.03-05-01039.1983
84. McPherson MJ, McDermott JH. Time-dependent discrimination advantages for harmonic sounds suggest efficient coding for memory. *Proc Natl Acad Sci USA* 2020; 117(50):32169-32180. doi: 10.1073/pnas.2008956117
85. Mechler F, Victor JD, Purpura KP, Shapley R. Robust temporal coding of contrast by V1 neurons for transient but not for steady-state stimuli. *J. Neurosci.* 1998 Aug 15; 18(16):6583-6598. doi: 10.1523/JNEUROSCI.18-16-06583.1998
86. Mesulam MM, From sensation to cognition. *Brain*, 1998 Jun 01; 121(6):1013–1052. doi: 10.1093/brain/121.6.1013
87. Muckli L, De Martino F, Vizioli L, Petro LS, Smith FW, Ugurbil K, et al. Contextual feedback

- to superficial layers of V1. *Current Biology* 2015 Oct 19, 25(20):2690-2695. doi: 10.1016/j.cub.2015.08.057
88. Murphy PC, Sillito AM. Corticofugal feedback influences the generation of length tuning in the visual. *Nature* 1987 Oct 01; 392(6141):727-729. doi: 10.1038/329727a0
  89. Narayan R, Graña G, Sen K. Distinct time scales in cortical discrimination of natural sounds in songbirds. *Journal of Neurophysiology* 2006 Jul 01; 96(1):252-258. doi: 10.1152/jn.01257.2005
  90. Niven JE, Laughlin SB. Energy limitation as a selective pressure on the evolution of sensory systems. *Journal of Experimental Biology* 2008 Jun 01; 211(11):1792-1804. doi: 10.1242/jeb.017574
  91. Nogueira R, Peltier NE, Anzai A, DeAngelis GC, Martínez-Trujillo J, Moreno-Bote R. The effects of population tuning and trial-by-trial variability on information encoding and behavior. *J. Neurosci.* 2020; 40(5):1066-1083. doi: 10.1523/JNEUROSCI.0859-19.2019
  92. Ohzawa I, DeAngelis GC, Freeman RD. Encoding of binocular disparity by simple cells in the cat's visual cortex. *Journal of Neurophysiology* 1996 May 01; 75(5):1779-1805. doi: 10.1152/jn.1996.75.5.1779
  93. Palmer SE, Marre O, Berry MJ, Bialek W. Predictive information in a sensory population. *Proc Natl Acad Sci USA* 2015 Jun 02; 112(22):6908-6913. doi: 10.1073/pnas.1506855112
  94. Pettet MW, Gilbert CD. Dynamic changes in receptive-field size in cat primary visual cortex. *Proc Natl Acad Sci USA* 1992 Sep 01; 89(17):8366-8370. doi: 10.1073/pnas.89.17.8366
  95. Polanía R, Woodford M, Ruff CC. Efficient coding of subjective value. *Nat Neurosci* 2019; 22:134–142. doi: 10.1038/s41593-018-0292-0
  96. Pryluk R, Kfir Y, Gelbard-Sagiv H, Fried I, Paz R. A tradeoff in the neural code across regions and species. *Cell* 2019; 176(3):597-609.e18. doi: 10.1016/j.cell.2018.12.032
  97. Rao RPN, Ballard DH. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci* 1999 Jan 01; 2(1):79-87. doi: 10.1038/4580
  98. Reich DS, Mechler F, Victor JD. Temporal coding of contrast in primary visual cortex: When, what, and why. *Journal of Neurophysiology* 2001 Mar 01; 85(3):1039-1050. doi: 10.1152/jn.2001.85.3.1039



99. Riesenhuber M, Poggio T. Hierarchical models of object recognition in cortex. *Nat Neurosci* 1999 Nov 01; 2(11):1019-1025. doi: 10.1038/14819
100. Riesenhuber M, Poggio T. Models of object recognition. *Nat Neurosci* 2000 Nov 01; 3(11):1199-1204. doi: 10.1038/81479
101. Rubinov M, Sporns O, van Leeuwen C, Breakspear M. Symbiotic relationship between brain structure and dynamics. *BMC Neuroscience* 2009 Jun 02; 10(1):55. doi: 10.1186/1471-2202-10-55
102. Rubner Y, Tomasi C, Guibas LJ. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision* 2000 Nov 01; 40(2):99-121. doi: 10.1023/A:1026543900054
103. Satuavuori E, Kreuz T. Which spike train distance is most suitable for distinguishing rate and temporal coding? *Journal of Neuroscience Methods* 2018 Apr 01; 299:22-33. doi: 10.1016/j.jneumeth.2018.02.009
104. Satuavuori E, Mulansky M, Bozanic N, Malvestio I, Zeldenrust F, Lenk K, Kreuz T. Measures of spike train synchrony for data with multiple time scales. *Journal of Neuroscience Methods* 2017 Aug 01; 287:25-38. doi: 10.1016/j.jneumeth.2017.05.028
105. Schneider GE. Two Visual Systems. *Science* 1969 Feb 28; 163(3870):895-902. doi: 10.1126/science.163.3870.895
106. Schwartz AB, Kettner RE, Georgopoulos AP. Primate motor cortex and free arm movements to visual targets in three-dimensional space. I. Relations between single cell discharge and direction of movement. *J. Neurosci.* 1988 Aug 01; 8(8):2913-2927. doi: 10.1523/JNEUROSCI.08-08-02913.1988
107. Sederberg AJ, MacLean JN, Palmer SE. Learning to make external sensory stimulus predictions using internal correlations in populations of neurons. *Proc Natl Acad Sci USA* 2018 Jan 30; 115(5):1105-1110. doi: 10.1073/pnas.1710779115
108. Semedo JD, Zandvakili A, Machens CK, Yu BM, Kohn A. Cortical Areas Interact through a Communication Subspace. *Neuron* 2019 Apr 03; 102(1):249-259.e4. doi: 10.1016/j.neuron.2019.01.026
109. Serre T, Oliva A, Poggio T. A feedforward architecture accounts for rapid categorization. *Proc Natl Acad Sci USA* 2007 Apr 10; 104(15):6424-6429. doi: 10.1073/pnas.0700622104

110. Sihn D, Kim S-P. A spike train distance robust to firing rate changes based on the earth mover's distance. *Front. Comput. Neurosci.* 2019 Dec 10; 13:82. doi: 10.3389/fncom.2019.00082
111. Sihn D, Kim S-P. Spatio-temporally efficient coding assigns functions to hierarchical structures of the visual system. *bioRxiv* 2021 Aug 13; 456321. doi: 10.1101/2021.08.13.456321
112. Sihn D, Kim S-P. Spatio-temporally efficient coding assigns functions to hierarchical structures of the visual system. *Front. Comput. Neurosci.* 2022 May 27; 16:890447. doi: 10.3389/fncom.2022.890447
113. Sillito AM, Cudeiro J, J HE. Always returning: feedback and sensory processing in visual cortex and thalamus. *Trends Neurosci.* 2006 Jun 01; 29(6):307-316. doi: 10.1016/j.tins.2006.05.001
114. Simoncelli EP. Vision and the statistics of the visual environment. *Current Opinion in Neurobiology* 2003 Apr 01; 13(2):144-149. doi: 10.1016/S0959-4388(03)00047-3
115. Simoncelli EP, Olshausen BA. Natural image statistics and neural representation. *Annu. Rev. Neurosci.* 2001 Mar 01; 24(1):1193-1216. doi: 10.1146/annurev.neuro.24.1.1193
116. Sims CR. Efficient coding explains the universal law of generalization in human perception. *Science* 2018; 360(6389):652-656. doi: 10.1126/science.aag1118
117. Singer Y, Teramoto Y, Willmore BDB, Schnupp JWH, King AJ, Harper NS, Gallant JL, Kastner S. Sensory cortex is optimized for prediction of future input. *eLife* 2018 Jun 18; 7:e31557. doi: 10.7554/eLife.31557
118. Solomon SS, Tang H, Sussman E, Kohn A. Limited evidence for sensory prediction error responses in visual cortex of macaques and humans. *Cereb Cortex* 2021 Jun 01; 31(6):3136-3152. doi: 10.1093/cercor/bhab014
119. Spratling MW. A review of predictive coding algorithms. *Brain and Cognition* 2017 Mar 01; 112:92-97. doi: 10.1016/j.bandc.2015.11.003
120. Stringer C, Pachitariu M, Steinmetz N, Carandini M, Harris KD. High-dimensional geometry of population responses in visual cortex. *Nature* 2019 Jul 01; 571(7765):361-365. doi: 10.1038/s41586-019-1346-5
121. Sullivan TJ, de Sa VR. Homeostatic synaptic scaling in self-organizing maps. *Neural Networks* 2006 Jul; 19(6):734-743. doi: 10.1016/j.neunet.2006.05.006

122. Tabuchi M, Monaco JD, Duan G, Bell B, Liu S, Liu Q, Zhang K, Wu MN. Clock-generated temporal codes determine synaptic plasticity to control sleep. *Cell* 2018 Nov 15; 175(5):1213-1227.e18. doi: 10.1016/j.cell.2018.09.016
123. Teufel C, Fletcher PC. Forms of prediction in the nervous system. *Nat Rev Neurosci* 2020 Apr 01; 21(4):231–242. doi: 10.1038/s41583-020-0275-5
124. Tomov P, Pena RFO, Zaks MA, Roque AC. Sustained oscillations, irregular firing, and chaotic dynamics in hierarchical modular networks with mixtures of electrophysiological cell types. *Front. Comput. Neurosci.* 2014 Sep 02; 8:103 doi: 10.3389/fncom.2014.00103
125. Toyozumi T, Pfister J-P, Aihara K, Gerstner W. Generalized Bienenstock–Cooper–Munro rule for spiking neurons that maximizes information transmission. *Proc Natl Acad Sci USA* 2005 Mar 28; 102(14):5239-5244. doi: 10.1073/pnas.0500495102
126. Turrigiano GG, Nelson SB. Homeostatic plasticity in the developing nervous system. *Nature Reviews Neuroscience* 2004 Feb 01; 5(2):97-107. doi: 10.1038/nrn1327
127. van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res* 2008 Nov; 9(86):2579–2605.
128. van Hateren JH, van der Schaaf A. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. Royal Soc. B* 1998 Mar 07; 265(1394):359-366. doi: 10.1098/rspb.1998.0303
129. van Rossum MCW. A novel spike distance. *Neural Computation* 2001 Apr 01; 13(4):751-763. doi: 10.1162/089976601300014321
130. Vargas-Irwin CE, Franquemont L, Black MJ, Donoghue JP. Linking objects to actions: Encoding of target object and grasping strategy in primate ventral premotor cortex. *J. Neurosci.* 2015 Jul 29; 35(30):10888-10897. doi: 10.1523/JNEUROSCI.1574-15.2015
131. Vaserstein LN. Markov processes over denumerable products of spaces, describing large systems of automata. *Probl. Peredachi Inf.* 1969; 5(3):64–72.
132. Victor JD, Purpura KP. Nature and precision of temporal coding in visual cortex: a metric-space analysis. *Journal of Neurophysiology* 1996 Aug 01; 76(2):1310-1326. doi: 10.1152/jn.1996.76.2.1310
133. Victor JD, Purpura KP. Spatial phase and the temporal structure of the response to gratings in V1. *Journal of Neurophysiology* 1998 Aug 01; 80(2):554-571. doi: 10.1152/jn.1998.80.2.554

134. Wang L, Narayan R, Graña G, Shamir M, Sen K. Cortical discrimination of complex natural stimuli: Can single neurons match behavior? *J. Neurosci.* 2007 Jan 17; 27(3):582-589. doi: 10.1523/JNEUROSCI.3699-06.2007
135. Wang W, Jones HE, Andolina IM, Salt TE, Sillito AM. Functional alignment of feedback effects from visual cortex to thalamus. *Nat Neurosci* 2006 Oct 01; 9(10):1330-1336. doi: 10.1038/mn1768
136. Waydo S, Kraskov A, Quian Quiroga R, Fried I, Koch C. Sparse representation in the human medial temporal lobe. *J. Neurosci.* 2006 Oct 04; 26(40):10232-10234. doi: 10.1523/JNEUROSCI.2101-06.2006
137. Wiskott L, Sejnowski TJ. Slow feature analysis: Unsupervised learning of invariances. *Neural Computation* 2002 Apr 01; 14(4):715-770. doi: 10.1162/089976602317318938
138. Yamins DLK, Hong H, Cadieu CF, Solomon EA, Seibert D, DiCarlo JJ. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc Natl Acad Sci USA* 2014 Jun 10; 111(23):8619-8624. doi: 10.1073/pnas.1403112111
139. Zandvakili A, Kohn A. Coordinated neuronal activity enhances corticocortical communication. *Neuron* 2015 Aug 19; 87(4):827-839. doi: 10.1016/j.neuron.2015.07.026
140. Zandvakili A, Kohn A. Simultaneous V1-V2 neuronal population recordings in anesthetized macaque monkeys. *CRCNS.org* 2019. doi: 10.6080/K0B27SHN
141. Zavitz E, Yu H-H, Rowe EG, Rosa MGP, Price NSC. Rapid adaptation induces persistent biases in population codes for visual motion. *J. Neurosci.* 2016 Apr 20; 36(16):4579-4590. doi: 10.1523/JNEUROSCI.4563-15.2016
142. Zhang Q, Cramer SR, Ma Z, Turner KL, Gheres KW, Liu Y, Drew PJ, Zhang N. Brain-wide ongoing activity is responsible for significant cross-trial BOLD variability. *Cerebral Cortex* 2022; bhac016, doi: 10.1093/cercor/bhac016
143. Zhang Q, Hu X, Hong B, Zhang B. A hierarchical sparse coding model predicts acoustic feature encoding in both auditory midbrain and cortex. *PLOS Computational Biology* 2019 Feb 11; 15(2):e1006766. doi: 10.1371/journal.pcbi.1006766
144. Zhang S, Xu M, Kamigaki T, Hoang Do JP, Chang W, Jenvay S, et al. Long-range and local circuits for top-down modulation of visual cortex processing. *Science* 2014 Aug 08; 345(6197):660-665. doi: 10.1126/science.1254126

145. Zou WY, Ng AY, Yu K. Unsupervised learning of visual invariance with temporal coherence.  
Proceedings of the Conference on Neural Information Processing Systems; 2011 Dec 12-17;  
Granada, Spain.

## **Acknowledgment**

Contents of the present study were based on the author's previous studies (Sihn and Kim, 2019; Sihn and Kim, 2021; Sihn and Kim, 2022) and descriptions in the studies.

I thank the reviewers for reviewing this dissertation.