

## Article

# A Fast and Efficient Ensemble Transfer Entropy and Applications in Neural Signals

Junyao Zhu <sup>1,2</sup>, Mingming Chen <sup>1,2</sup>, Junfeng Lu <sup>1,2</sup>, Kun Zhao <sup>3</sup>, Enze Cui <sup>1,2</sup>, Zhiheng Zhang <sup>1,2</sup> and Hong Wan <sup>1,2,\*</sup><sup>1</sup> School of Electrical Engineering, Zhengzhou University, Zhengzhou 450001, China<sup>2</sup> Henan Key Laboratory of Brain Science and Brain-Computer Interface Technology, Zhengzhou 450001, China<sup>3</sup> School of Intelligent Engineering, Zhengzhou University of Aeronautics, Zhengzhou 450001, China

\* Correspondence: wanhong@zzu.edu.cn; Tel.: +86-037-167-781-421

**Abstract:** The ensemble transfer entropy ( $TE_{ensemble}$ ) refers to the transfer entropy estimated from an ensemble of realizations. Due to its time-resolved analysis, it is adapted to analyze the dynamic interaction between brain regions. However, in the traditional  $TE_{ensemble}$ , multiple sets of surrogate data should be used to construct the null hypothesis distribution, which dramatically increases the computational complexity. To reduce the computational cost, a fast, efficient  $TE_{ensemble}$  with a simple statistical test method is proposed here, in which just one set of surrogate data is involved. To validate the improved efficiency, the simulated neural signals are used to compare the characteristics of the novel  $TE_{ensemble}$  with those of the traditional  $TE_{ensemble}$ . The results show that the time consumption is reduced by two or three magnitudes in the novel  $TE_{ensemble}$ . Importantly, the proposed  $TE_{ensemble}$  could accurately track the dynamic interaction process and detect the strength and the direction of interaction robustly even in the presence of moderate noises. The novel  $TE_{ensemble}$  reaches its steady state with the increased samples, which is slower than the traditional method. Furthermore, the effectiveness of the novel  $TE_{ensemble}$  was verified in the actual neural signals. Accordingly, the  $TE_{ensemble}$  proposed in this work may provide a suitable way to investigate the dynamic interactions between brain regions.

**Keywords:** ensemble transfer entropy ( $TE_{ensemble}$ ); ensemble local transfer entropy ( $te_{ensemble}$ ); dynamic interaction; statistical test method



**Citation:** Zhu, J.; Chen, M.; Lu, J.; Zhao, K.; Cui, E.; Zhang, Z.; Wan, H. A Fast and Efficient Ensemble Transfer Entropy and Applications in Neural Signals. *Entropy* **2022**, *24*, 1118. <https://doi.org/10.3390/e24081118>

Received: 25 May 2022

Accepted: 9 August 2022

Published: 13 August 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The brain is a highly complex system [1,2]. Multiple interconnected brain regions with specific information processing capabilities interact to support the cognitive tasks [3,4], and the strength and the direction of interactions change dynamically [5,6]. For example, the dynamic interactions between the hippocampus (Hp) and posterior parietal cortex (PPC) have been detected in mental arithmetic tasks [5]. The strength of the information flow from the Hp to the dorsal PPC reaches the maximum during mental arithmetic. The maximum value from the Hp to the central PPC is found in verbal memory recall. In rodent spatial associative tasks, the information flows from Hp to the prefrontal cortex (PFC), but the direction reverses in the sampling period [6]. Therefore, a complete description of these interactions, in terms of both strength and directionality, is necessary to reveal the function and the cooperative work of brain regions.

As a measurement of the information interaction between two signals, transfer entropy (TE) is model-free and does not assume any signal or interaction structure [7,8]. Therefore, it has been widely used in neuroscience [9–12]. However, TE is the average of information transfer over time. The application of a sliding window is the most common way to explore the dynamic interaction process within and between brain regions. The neural signals should be divided into continuous (non)overlapping segments [13] or be separated into different epochs according to the task [14,15]; then, TE for each segment (epoch) is

calculated. To ensure enough samples for TE estimation, the choice of window length is usually a compromise between the estimation precision and the temporal resolution of the dynamic process. The larger the time window, the lower the resolution.

To improve the temporal resolution, a technique (called the ensemble method) takes advantage of multiple realizations of the dynamic process [16], such as numerous recordings of evoked or event-related potentials/fields [17,18]. By estimating TE from the ensemble members instead of individual trials [19], it allows for a time-resolved analysis of the interaction dynamics. Based on this technique, Gómez-Herrero and his colleagues proposed a data-efficient estimator of probability densities to calculate TE from an ensemble of realizations [20]. Additionally, Wollstadt [21] combined the ensemble method with the TE self-prediction optimality ( $TE_{SPO}$ ) estimator, which was introduced by Wibral [22] to achieve the optimal estimation of delayed information transfer in an ensemble of independent repetition trials. In the following, we use the ensemble transfer entropy ( $TE_{ensemble}$ ) to indicate the transfer entropy estimated from an ensemble of realizations with the  $TE_{SPO}$  estimator. Instead of TE estimation for each trial, a single  $TE_{ensemble}$  metric value can be accurately estimated from the ensemble members.  $TE_{ensemble}$  is not only suitable for short-time data but also for non-stationary signals, which are commonly observed in neuroscience. So,  $TE_{ensemble}$  can be used for analyzing the dynamic interaction processes between neural signals.

However, the  $TE_{ensemble}$  algorithm is still at the theoretical level and cannot be applied to the actual neural signals due to its enormous amount of calculation [21]. Firstly, to estimate the  $TE_{ensemble}$  metric value accurately, a mutual information estimation method proposed by Kraskov, Stogbauer, and Grassberger (KSG) is used [23]. The KSG estimator spends most of the CPU time searching for neighbors, especially in high-dimensional spaces. The complexity of this algorithm is  $\mathcal{O}((n \times N)^2)$ , where  $n$  and  $N$  are the number of independent repetitions and the sample size in a trial, respectively. The computational complexity is much larger than the methods based on partitioning the observation space (the complexity is  $\mathcal{O}(n \times N)$ ) [24]. Secondly, constructing the null hypothesis distribution in  $TE_{ensemble}$  leads to the calculation increasing several orders of magnitude. In TE, the null hypothesis distribution can be constructed by one set of surrogate data ( $n$  trials) [9]. However, only a single metric value can be obtained from  $n$  trials in the ensemble method, so multiple sets (usually more than 500) of surrogate data are needed to construct the null hypothesis distribution [21]. The computational complexity of TE with KSG estimator is  $\mathcal{O}(2n \times N^2)$ , but for  $TE_{ensemble}$ , the complexity is  $\mathcal{O}((1 + m) \times (n \times N)^2)$ , where  $m$  is the number of surrogate data sets. So,  $TE_{ensemble}$  is much more complex than TE. For a neuroscience experiment with channel pairs (100)  $\times$  the number of surrogate data sets (1000)  $\times$  stimulus conditions (4)  $\times$  subjects (15), the elapsed time of  $TE_{ensemble}$  is 240 weeks [21].

One approach to reducing the time consumption for  $TE_{ensemble}$  is to use faster hardware devices such as FPGA [25], graphic processing unit (GPU) [21], and computer cluster [26].  $TE_{ensemble}$  working on GPU is one of the most effective methods to alleviate the time-consuming problem. However,  $TE_{ensemble}$  still requires extended running time even with GPU when enormous data is involved. The experimental data in neuroscience takes about 4.8 weeks on a single GPU (NVIDIA GTX Tian) and will take longer when it searches for information transfer delay. However, the use of multiple GPUs puts forward a higher requirement on computer performance. Another approach is to use the simple estimation method. In the phase transfer entropy proposed by Lobier [27], mutual information is estimated for phase time series using a simple binning method. This approach effectively reduces the running time and can calculate the strength and the direction of interaction [28–30]. However, simple discretization by partitioning the observation space ignores the neighborhood relationship in continuous data, which may cause the loss of important information [31], leading to the failure of mutual information estimation in real-valued data.

Hence, to reduce the computational cost, a fast, efficient  $TE_{ensemble}$  with a simple statistical test method is proposed here. Based on the characteristic that  $TE_{ensemble}$  is the

average value of its local transfer entropy ( $te_{ensemble}$ ), we use a simple  $t$ -test for the  $te_{ensemble}$  of the raw data against the  $te_{ensemble}$  from the surrogate data (one set) as the statistical test method in the novel  $TE_{ensemble}$ . Because just one set of surrogate data is used, the time consumption of the novel  $TE_{ensemble}$  is significantly reduced. Then, we employ a widely used neural mass model (NMM) to produce neural signals through which the characteristics of the novel  $TE_{ensemble}$  are compared with those of the traditional method. The results show that the time consumption is reduced by two or three magnitudes in the novel  $TE_{ensemble}$ . Importantly, the proposed  $TE_{ensemble}$  robustly detects the strength and the direction of interaction, and it reaches stability with the increase in the sample size, which is slower than the traditional  $TE_{ensemble}$ . Furthermore, the novel  $TE_{ensemble}$  can track the dynamic interaction processes between signal pairs and the effectiveness of the novel  $TE_{ensemble}$  has also been verified in the realistic neural signals recorded from pigeons.

This paper is organized as follows: Section 2 introduces the novel  $TE_{ensemble}$  and the NMM we used. Section 3 investigates the characteristics of the novel  $TE_{ensemble}$  on the simulated signal pairs and the actual neural signals and compares the performance of the novel  $TE_{ensemble}$  with that of the traditional method. Section 4 discusses the results, and Section 5 is a conclusion.

## 2. Materials and Methods

In information-theoretic framework, Shannon entropy defines the measurement of information uncertainty. For a random variable  $X$  probability distribution  $p(x)$ , its Shannon entropy is:

$$H(X) = - \sum_x p(x) \log_2 p(x) \quad (1)$$

Shannon entropy can be extended to two random variables. For the  $X$  and  $Y$  probability distribution  $p(x)$  and  $p(y)$ , the joint entropy can be defined as in Equation (2):

$$H(X, Y) = - \sum_{x,y} p(x, y) \log_2 p(x, y) \quad (2)$$

The conditional entropy in Equation (3) is the average uncertainty about  $x$  that remains when the value of  $y$  is known:

$$H(X|Y) = - \sum_{x,y} p(x, y) \log_2 p(x|y) \quad (3)$$

The mutual information between  $X$  and  $Y$  measures the reduction of one variable's uncertainty by the knowledge of another one:

$$I(X; Y) = H(X) + H(Y) - H(X, Y) = - \sum_{x,y} p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} \quad (4)$$

By assuming a third random variable  $Z$ , the conditional mutual information of  $X$  and  $Y$  is:

$$I(X; Y|Z) = H(X|Z) + H(Y|Z) - H(X, Y|Z) \quad (5)$$

Mutual information has been widely used in neuroscience [32]. However, the major problem is that mutual information contains no directionality. Transfer entropy, which describes the uncertainty reduction in predicting the target variable by adding the historical information of a new variable [33], is proposed to solve this deficiency. For  $X$  and  $Y$ ,  $TE_{X \rightarrow Y}$  defines the conditional mutual information between  $Y$  and  $X^-$  (historical information of  $X$ ) under  $Y^-$  (historical information of  $Y$ ):

$$TE_{X \rightarrow Y} = I(Y; X^- | Y^-) = H(Y|Y^-) - H(Y|X^-, Y^-) \quad (6)$$

Suppose (1) there is an interaction delay  $u$  between  $X$  and  $Y$ ;

(2)  $X$  and  $Y$  can be approximated by a Markov process of order  $k$  and  $l$ , respectively.

With these assumptions, TE can be rewritten in a more general form as in Equation (7):

$$\begin{aligned}
 TE_{X \rightarrow Y}^{(k,l)}(t, u) &= I(Y_t; \mathbf{X}_{t-u}^{(k)} | \mathbf{Y}_{t-1}^{(l)}) = H(\mathbf{X}_{t-u}^{(k)}, \mathbf{Y}_{t-1}^{(l)}) - H(Y_t, \mathbf{X}_{t-u}^{(k)}, \mathbf{Y}_{t-1}^{(l)}) + H(Y_t, \mathbf{Y}_{t-1}^{(l)}) - H(\mathbf{Y}_{t-1}^{(l)}) \\
 &= \sum_{y_t, \mathbf{x}_{t-u}^{(k)}, \mathbf{y}_{t-1}^{(l)}} p(y_t, \mathbf{x}_{t-u}^{(k)}, \mathbf{y}_{t-1}^{(l)}) \log_2 \frac{p(y_t | \mathbf{y}_{t-1}^{(l)}, \mathbf{x}_{t-u}^{(k)})}{p(y_t | \mathbf{y}_{t-1}^{(l)})}
 \end{aligned} \tag{7}$$

Equation (7) can be viewed as the average of information transfer over time. Based on this, the local transfer entropy (te) is proposed [34], realizing the local or pointwise interaction (Equation (8)):

$$te_{X \rightarrow Y}^{(k,l)}(t, u) = i(Y_t; \mathbf{X}_{t-u}^{(k)} | \mathbf{Y}_{t-1}^{(l)}) = \log_2 \frac{p(y_t | \mathbf{y}_{t-1}^{(l)}, \mathbf{x}_{t-u}^{(k)})}{p(y_t | \mathbf{y}_{t-1}^{(l)})} \tag{8}$$

From Equations (7) and (8), we know that the transfer entropy is the average value of the local transfer entropy.

### 2.1. Ensemble Transfer Entropy ( $TE_{ensemble}$ ) and Ensemble Local Transfer Entropy ( $te_{ensemble}$ )

When independent repetition trials of an experimental condition meet the cyclostationarity, these trials are taken as an ensemble of realizations, and various probability density functions (PDFs) can be accurately estimated from the ensemble members. In this paper, we use the subscript *ensemble* to indicate the ensemble transfer entropy with the TE self-prediction optimality estimator:

$$\begin{aligned}
 TE_{ensemble}(X \rightarrow Y, t, u) &= I(Y_t; \mathbf{X}_{t-u}^{(k)} | \mathbf{Y}_{t-1}^{(l)}) \\
 &= \sum_{y_t(r), \mathbf{x}_{t-u}^{(k)}(r), \mathbf{y}_{t-1}^{(l)}(r)} p(y_t(r), \mathbf{x}_{t-u}^{(k)}(r), \mathbf{y}_{t-1}^{(l)}(r)) \log_2 \frac{p(y_t(r) | \mathbf{y}_{t-1}^{(l)}(r), \mathbf{x}_{t-u}^{(k)}(r))}{p(y_t(r) | \mathbf{y}_{t-1}^{(l)}(r))}
 \end{aligned} \tag{9}$$

where  $r$  is the number of independent repetition trials.

When the number of repetitions is sufficient to provide the necessary amount of data to estimate various PDFs in the time window  $t' \in [t^-, t^+]$  reliably, the  $TE_{ensemble}$  in  $t'$  can be estimated:

$$\begin{aligned}
 TE_{ensemble}(X \rightarrow Y, t', u) &= I(Y_{t'}; \mathbf{X}_{t'-u}^{(k)} | \mathbf{Y}_{t'-1}^{(l)}) \\
 &= \sum_{y_{t'}(r), \mathbf{x}_{t'-u}^{(k)}(r), \mathbf{y}_{t'-1}^{(l)}(r)} p(y_{t'}(r), \mathbf{x}_{t'-u}^{(k)}(r), \mathbf{y}_{t'-1}^{(l)}(r)) \log_2 \frac{p(y_{t'}(r) | \mathbf{y}_{t'-1}^{(l)}(r), \mathbf{x}_{t'-u}^{(k)}(r))}{p(y_{t'}(r) | \mathbf{y}_{t'-1}^{(l)}(r))}
 \end{aligned} \tag{10}$$

With these definitions in place, we can obtain the ensemble local transfer entropy ( $te_{ensemble}$ ):

$$te_{ensemble}(X \rightarrow Y, t, u) = i(Y_t; \mathbf{X}_{t-u}^{(k)} | \mathbf{Y}_{t-1}^{(l)}) = \log_2 \frac{p(y_t(r) | \mathbf{y}_{t-1}^{(l)}(r), \mathbf{x}_{t-u}^{(k)}(r))}{p(y_t(r) | \mathbf{y}_{t-1}^{(l)}(r))} \tag{11}$$

### 2.2. Estimating Ensemble Transfer Entropy

A TE estimator KSG with less bias has been widely used [24]. This method is based on the nearest neighbor estimator of Kozachenko and Leonenko [35]. The distance of  $k$ -th nearest neighbor in the high-dimensional spaces are projected to the low-dimensional spaces so that the deviations caused by the different spatial scales in low-dimensional spaces are significantly reduced. In this paper, the KSG estimator is applied to  $TE_{ensemble}$  and  $te_{ensemble}$ . Instead of searching for the nearest neighbors in the state space constructed by the individual trial, we proceed in all repetitions [36]:

$$TE_{ensemble}(X \rightarrow Y, t, u) = \psi(k) + \langle \psi\left(n_{\mathbf{y}_{t-1}^{(l)}(r)} + 1\right) - \psi\left(n_{y_t(r)\mathbf{y}_{t-1}^{(l)}(r)} + 1\right) - \psi\left(n_{\mathbf{y}_{t-1}^{(l)}(r)\mathbf{x}_{t-u}^{(k)}(r)} + 1\right) \rangle \quad (12)$$

where  $\psi$  is the Digamma function,  $\psi(x) = \Gamma(x)^{-1} \frac{d\Gamma(x)}{dx}$ ;  $\langle \cdot \rangle$  means average;  $n_{\mathbf{y}_{t-1}^{(l)}(r)}$ ,  $n_{y_t(r)\mathbf{y}_{t-1}^{(l)}(r)}$  and  $n_{\mathbf{y}_{t-1}^{(l)}(r)\mathbf{x}_{t-u}^{(k)}(r)}$  are the number of samples falling into the strip of the marginal space  $\mathbf{y}_{t-1}^{(l)}(r)$ ,  $y_t(r)\mathbf{y}_{t-1}^{(l)}(r)$ ,  $\mathbf{y}_{t-1}^{(l)}(r)\mathbf{x}_{t-u}^{(k)}(r)$ , respectively. The strip is defined by the distance to its  $k$ -th nearest neighbors. In general,  $k$  is 4 [23]:

$$te_{ensemble}(X \rightarrow Y, t, u) = \psi(k) + \psi\left(n_{\mathbf{y}_{t-1}^{(l)}(r)} + 1\right) - \psi\left(n_{y_t(r)\mathbf{y}_{t-1}^{(l)}(r)} + 1\right) - \psi\left(n_{\mathbf{y}_{t-1}^{(l)}(r)\mathbf{x}_{t-u}^{(k)}(r)} + 1\right) \quad (13)$$

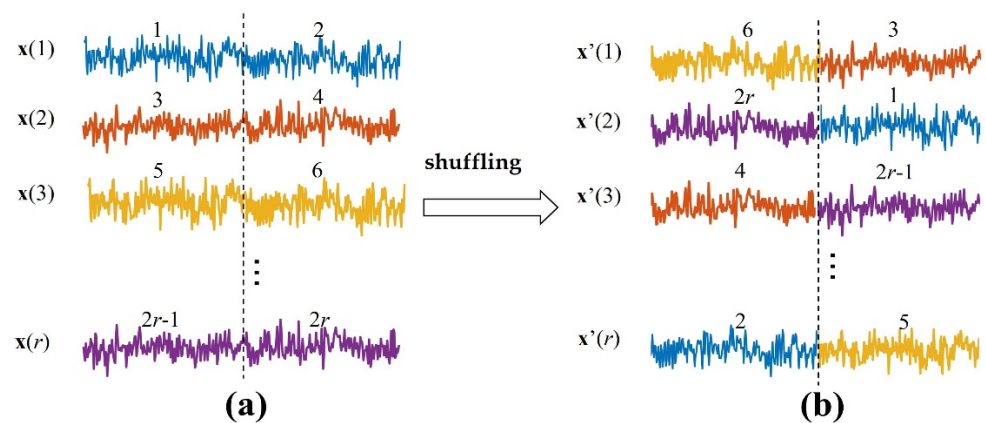
### 2.3. Parameter Selection

The information transfer delay  $u$  between  $X$  and  $Y$ , the embedded dimension (Markov approximation order  $k$  and  $l$ ) and embedded delay  $\tau$ , have a significant impact on  $TE_{ensemble}$  estimation. We use the TE self-prediction optimality estimator to obtain the transfer delay [21]. When the  $TE_{ensemble}(X \rightarrow Y, t, u)$  is maximal, the assumed delay  $u$  is equal to the true information transfer delay  $\delta$  (Equation (14)) [22].  $k$ ,  $l$ , and  $\tau$  are calculated by using the Rawditz criterion [37]:

$$\delta = \operatorname{argmax}_u(TE_{ensemble}(X \rightarrow Y, t, u)) \quad (14)$$

### 2.4. Surrogate Data and The Improved Statistical Test Method

$TE_{ensemble}$  is a biased estimation with no upper bound [9], so it is necessary to generate surrogate data and construct the null hypothesis distribution to test the statistical significance of the  $TE_{ensemble}$  metric value. In the surrogate data, it is assumed that there is no information transfer between the source variable  $X$  and the target variable  $Y$ . The commonly used method is to shuffle  $X$ , which destroys the dependence between  $X$  and  $Y$  while retaining the probability distribution of the variables [38]. Here, the source signals of each independent repetition trial are separated into two segments (Figure 1). Then, these segments are shuffled to ensure each segment is not in the same position as before. We can obtain the surrogate data ( $X', Y$ ).



**Figure 1.** Generating the surrogate data. (a)  $x(1), x(2), \dots, x(r)$  are the source signals of independent repetition trials. Each trial is separated into two segments, these segments are shuffled in the ensemble members to ensure each segment is not in the same position as before, and then the surrogate data  $x'(1), x'(2), \dots, x'(r)$  in (b) are generated.

In the traditional  $TE_{ensemble}$  method, at least 500 sets of surrogate data are required to generate and then  $TE_{ensemble}$  metrics are estimated to construct the null hypothesis distribution. The null hypothesis can now be rejected or retained by comparing the  $TE_{ensemble}$

metric value of the raw data to the null hypothesis distribution at the 1% (5%) level of significance [21].

Here, we modify the statistical test method in the traditional  $TE_{ensemble}$ . The  $t$ -test is a parametrical statistical significance test, which is used to test whether there is any difference in the mean values of two groups. Based on the characteristic that  $TE_{ensemble}$  is the average value of  $te_{ensemble}$  [39], a  $t$ -test is performed on the  $te_{ensemble}$  values of the raw and the surrogate data. If the null hypothesis is rejected, it indicates that there is a significant difference between the  $TE_{ensemble}$  of the raw data and the  $TE_{ensemble}$  from the surrogate data.

Due to its high power, the  $t$ -test has been widely used to measure the difference in the mean values from two groups. In small samples, the  $t$ -test is valid only for data that is normally distributed [40]. However, because of the central limit theorem, the  $t$ -statistic is normally distributed with unit variance when the sample size is large, no matter what distribution the data has. Thus, the  $t$ -test will always be appropriate for large enough samples [41,42]. However, how large is large? The sample size relates to the difference in variance and the prevalence of extreme outliers. A large body of literatures indicate that "sufficiently large" is often less than 500 in extremely non-normally distributed data [41].

In the ensemble method, although the distributions of  $te_{ensemble}$  values (which are estimated from the raw and the surrogate data, respectively) are non-normally distributed (Figure S1), the samples of  $te_{ensemble}$  are often substantially larger than 500. Therefore, the  $t$ -test is applicable to the  $te_{ensemble}$  values. We also compare the  $t$ -test and the Wilcoxon rank sum test (Figure S2). The results of the two methods are almost the same.

### 2.5. Neural Mass Model

Signal pairs are generated by NMM described in [43], which simulates the connectivity between multiple regions of interesting (ROIs) through long-range excitatory connections. In the NMM, the average spike density of pyramidal neurons of the presynaptic area ( $Z^X$ ) affects the target region by a weight factor  $\omega$  and a time delay  $u$  (Equation (15)):

$$u^Y(t) = n^Y(t) + \omega^{X \rightarrow Y} Z^X(t - u) \quad (15)$$

where  $n(t)$  is a Gaussian white noise. The superscripts  $X$  and  $Y$  are represented by the presynaptic and target region, respectively.

Signal pairs generated by the NMM are nonlinear and have significant  $\beta$  (about 20 Hz) activity. By changing the information transfer delay  $u$  and weight factor  $\omega^{X \rightarrow Y}$ ,  $\omega^{Y \rightarrow X}$ , we obtain the simulated signal pairs with directional interaction.

We use the Trentool Matlab toolbox to estimate  $TE_{ensemble}$  and  $te_{ensemble}$  ([https://github.com/trentool/TRENTOOL3\\_manual](https://github.com/trentool/TRENTOOL3_manual) (accessed on 15 November 2021)). The Matlab codes of the NMM are available in ModelDB (<http://modeldb.yale.edu/263637> (accessed on 1 January 2022)).

## 3. Results

In this section, we first use the simulated signal pairs to evaluate the characteristics of the novel  $TE_{ensemble}$ , and compare the performance and time consumption of the novel  $TE_{ensemble}$  with that of the traditional method. Then, we verified the effectiveness of the novel  $TE_{ensemble}$  in the actual neural signals.

### 3.1. Results on NMM

#### 3.1.1. The Novel $TE_{ensemble}$ Measures the Strength and the Direction of Interaction Robustly

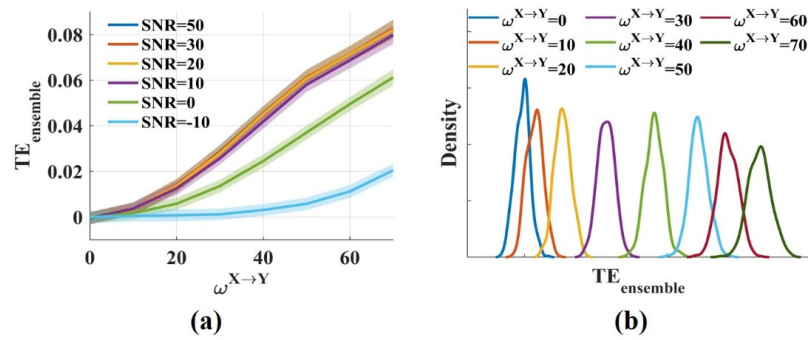
The simulated signal pairs ( $\omega^{X \rightarrow Y} = 0, 10, 20, 30, 40, 50, 60, 70$   $\omega^{Y \rightarrow X} = 0$ ,  $u = 20$  ms, sampling frequency = 100, trial length = 2 s) were used to investigate whether the novel  $TE_{ensemble}$  can reliably detect the strength and the direction of interaction between brain regions. We also explored the impact of the  $t$ -test significance  $p$  level on the results. For each  $\omega^{X \rightarrow Y}$ , 1000 simulated signal pairs were pooled together, then 100 pairs were drawn

randomly to estimate  $TE_{ensemble}$ . This procedure was repeated 500 times and resulted in a distribution of  $TE_{ensemble}$  values for each  $\omega^{X \rightarrow Y}$ . The  $TE_{ensemble}$  metric value measured the strength of interaction. By computing the false positive rate and the sensitivity, the accuracy of the direction estimation for the novel  $TE_{ensemble}$  was obtained. We used the same method that was introduced in [27] to calculate the false positive rate and the sensitivity. By comparing the  $t$ -test values for  $te_{ensemble}$  of the raw and the surrogate data with the threshold value (the  $t$ -test value at the significance  $p$  level with the freedom degree greater than 1000), we could calculate the proportion of false positive at  $\omega^{X \rightarrow Y} = 0$  and the sensitivity (proportion of true positive) for  $\omega^{X \rightarrow Y} \neq 0$ . The threshold values were 3.0902, 2.807, 2.5756, and 2.3263, respectively, corresponding to  $p = 0.002, 0.005, 0.01,$  and  $0.02$ . Then, we obtained the  $TE_{ensemble}$  sensitivity values as a function of  $\omega^{X \rightarrow Y}$  and the coupling detection threshold (CDT) for 0.8 sensitivity were computed by linear interpolation. CDT represented the smallest coupling value for which  $TE_{ensemble}$  detected 80% of the directed interactions. Therefore, low CDT indicated that the significant interactions were detected even for weak coupling while high CDT meant that the information transfer could be detected only for solid coupling. We compared the false positive rate and the sensitivity (CDT) of the novel  $TE_{ensemble}$  with those of the traditional  $TE_{ensemble}$ . In the traditional method, 500 sets of surrogate data were used to construct the null hypothesis distribution for each  $\omega^{X \rightarrow Y}$ . The false positive rate and the sensitivity (CDT) values were obtained by comparing the  $TE_{ensemble}$  values with the null hypothesis distribution at the significance  $\alpha$  level ( $\alpha = 0.01, 0.05$ ).

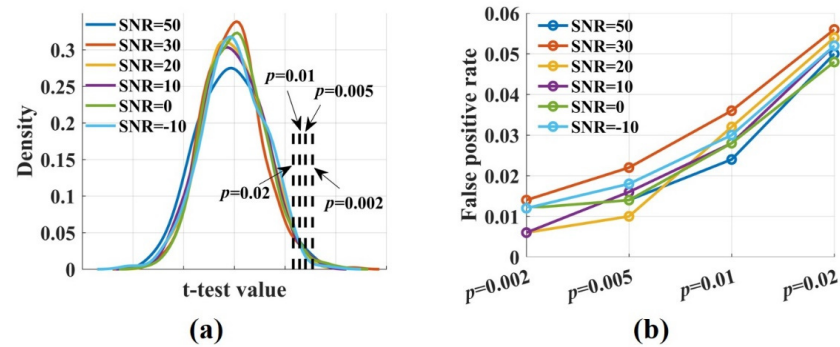
Neural signals (EEG/MEG/LFP) are corrupted by both environmental and biological noise [44–46]. Therefore, the analysis method we apply should be robust to noise. To explore the influence of noise on the novel  $TE_{ensemble}$ , we added gaussian white noise with different energies to the simulated signals. Then, the false positive rate, sensitivity, and CDT values for coupled signal pairs with a signal-to-noise ratio (SNR) = 50, 30, 20, 10, 0, and  $-10$  dB were calculated (the signal pairs with SNR = 50 dB were considered noise free). Finally, we compared the performance of the novel  $TE_{ensemble}$  with that of the traditional method.

The results showed that the  $TE_{ensemble}$  values fluctuated around 0 for  $\omega^{X \rightarrow Y} = 0$  and increased monotonically with  $\omega^{X \rightarrow Y}$  from 0 to 70 (Figure 2a). Low, realistic noise (SNR = 50, 30, 20, 10 dB) had little effect on the  $TE_{ensemble}$  values. The  $TE_{ensemble}$  values decreased moderately when SNR was 0 dB, but for strong noise (SNR =  $-10$  dB), the  $TE_{ensemble}$  values reduced greatly. Figure 2b shows the distributions of the  $TE_{ensemble}$  values for different  $\omega^{X \rightarrow Y}$  when SNR was 20 dB. The distributions had some overlap, but the overlap areas were minimal. Therefore, different  $\omega^{X \rightarrow Y}$  could be distinguished from the  $TE_{ensemble}$  values. The novel  $TE_{ensemble}$  in this work was consistent with the traditional method for measuring the interaction strength due to the same  $TE_{ensemble}$  estimator they used. Above all, it suggests that the novel  $TE_{ensemble}$  effectively measures the interaction strength even in the presence of noise.

To calculate the false positive rate and the sensitivity of the novel  $TE_{ensemble}$ , the  $t$ -test values were computed by performing a  $t$ -test for the  $te_{ensemble}$  values, which were estimated from the raw and the surrogate data. Figure 3a shows the densities of the  $t$ -test values for signal pairs ( $\omega^{X \rightarrow Y} = 0$ ) with varied SNRs. They were broadly similar and normally distributed. The proportion of false positive was calculated by comparing the  $t$ -test values with the threshold value. It increased with the significance  $p$  level from 0.002 to 0.02. The false positive rate fluctuated around 0.01 when  $p$  was 0.02 and increased to 0.05 when  $p$  was 0.02 (Figure 3b). Noise had no impact on the proportion of false positive. Even if SNR was  $-10$  dB, the false positive rate remained at a low level.



**Figure 2.** The novel  $TE_{ensemble}$  measures the interaction strength robustly. (a)  $TE_{ensemble}$  values fluctuated around 0 when  $\omega^{X \rightarrow Y}$  was 0 and increased monotonically with  $\omega^{X \rightarrow Y}$  from 0 to 70. The solid lines and shaded areas represent the mean and the variance of the  $TE_{ensemble}$  values, respectively; (b) The densities of the  $TE_{ensemble}$  values for varied  $\omega^{X \rightarrow Y}$  (SNR = 20 dB) were computed. The distributions had some overlap, but the overlap areas were minimal.



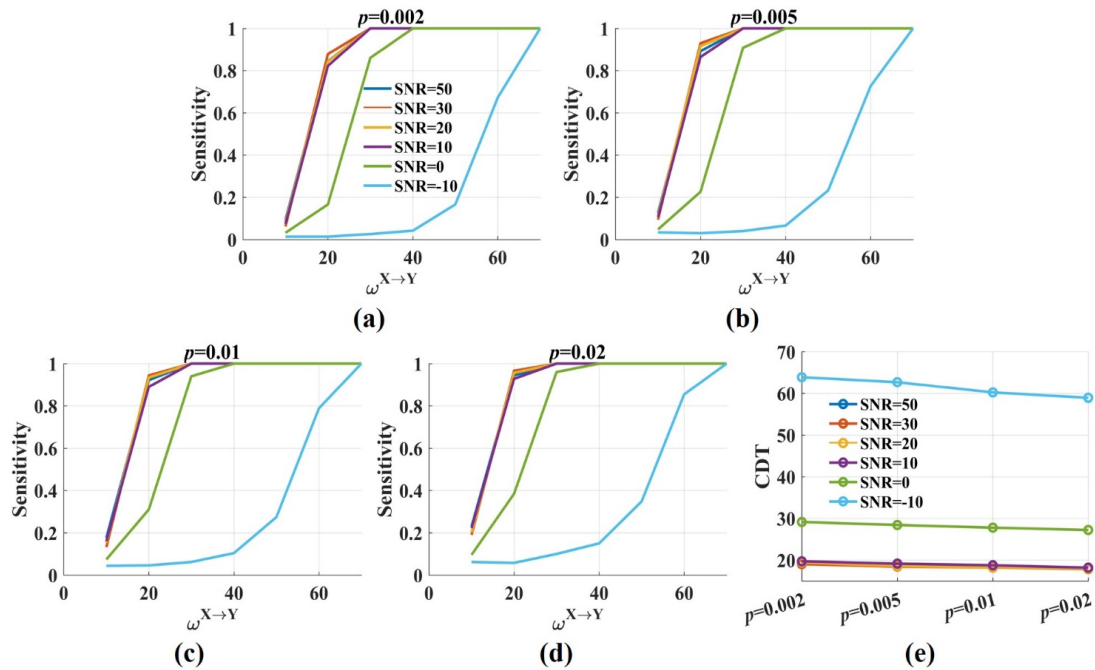
**Figure 3.** The densities of the  $t$ -test values and the false positive rate at different significance  $p$  levels. (a) The densities of the  $t$ -test values with  $\omega^{X \rightarrow Y} = 0$ , SNR = 50, 30, 20, 10, 0, -10 dB were broadly similar and normally distributed; (b) The false positive rate increased with the significance level- $p$  from 0.002 to 0.02. It fluctuated around 0.01 when  $p$  was 0.002 and increased to 0.05 when  $p$  was 0.02. Noise had no impact on the proportion of false positive.

Figure 4 shows the sensitivity and the CDT values against the significance  $p$  levels with  $\omega^{X \rightarrow Y}$  from 10 to 70. The results were twofold. First, when SNR was constant, the sensitivity values improved gradually with the increase in the significance  $p$  levels. The CDT value was 19.4 when  $p$  was 0.002 and reduced to 17.9 when  $p$  was 0.02. However, the decrease was minimal, and the CDT values were almost the same for varied  $p$  values (Figure 4e). The results indicated that the significance  $p$  level had a limited effect on the sensitivity and the CDT values. Second, moderate noise (SNR = 50, 30, 20, 10 dB) had little impact on the sensitivity and the CDT values. However, when SNR was 0 dB, the sensitivity values reduced, and the CDT values increased by nearly 10. Even more, the sensitivity values dramatically decreased when SNR was -10 dB and only when  $\omega^{X \rightarrow Y}$  was 60, about 80% of directed interactions were detected by the novel  $TE_{ensemble}$ .

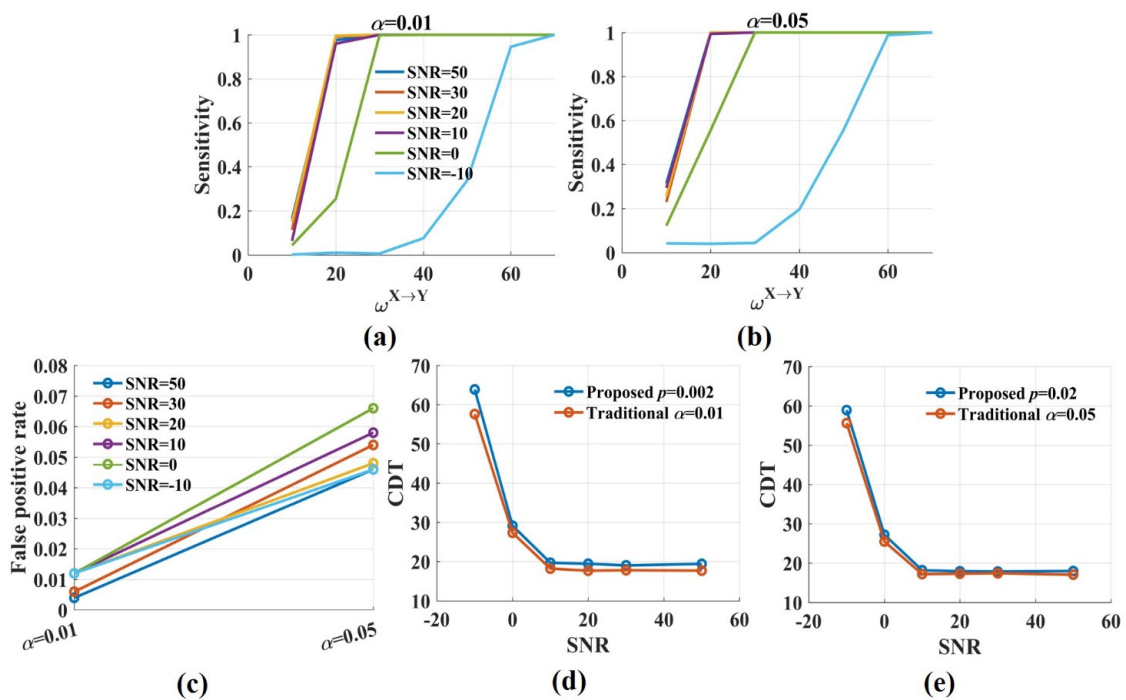
The sensitivity values of the novel  $TE_{ensemble}$  and the traditional  $TE_{ensemble}$  were compared in two cases (Figure 5). For case 1, the significance  $p$  level was 0.002 for the novel  $TE_{ensemble}$  and the significance  $\alpha$  level was 0.01 in the traditional method. For case 2,  $p$  and  $\alpha$  were 0.02 and 0.05, respectively. In case 1, the two methods had the same false positive rate. The sensitivity values of the novel  $TE_{ensemble}$  were somewhat lower than those of the traditional method. For SNR = 50, 30, 20, 10, 0 dB, the differences in the CDT values for the two methods were minimal, about 1.5. The difference increased to 6 when SNR was -10 dB. In case 2, the false positive rate of the traditional method was higher than that of the novel  $TE_{ensemble}$ . The two methods had the almost same sensitivity values (the CDT



values) regardless of SNRs. The results indicated that the proposed  $TE_{ensemble}$  almost had the same performance as the traditional method in detecting direction with realistic noise.



**Figure 4.** The sensitivity and the CDT values of the novel  $TE_{ensemble}$  with varied significance  $p$  levels. (a–d) The sensitivity values of  $TE_{ensemble}$  were plotted against the weight factor  $\omega^{X \rightarrow Y}$  for different SNRs with the significance  $p$  level = 0.002, 0.005, 0.01, 0.02. The sensitivity values improved with  $p$  from 0.002 to 0.02. Moderate noise had a limited effect on the sensitivity. However, the sensitivity values were reduced for low SNR; (e) The CDT values (sensitivity = 0.8) were plotted against the significance  $p$  levels with varied SNRs. Moderate noises and  $p$  values had a limited effect on the CDT values, and low SNR increased them.



**Figure 5.** The sensitivity and the CDT values of the traditional  $TE_{ensemble}$  when the significance  $\alpha$  level was 0.05 and 0.01, respectively. (a,b) The sensitivity values in the traditional method were increased

with  $\alpha$  from 0.01 to 0.05. Moderate noises had a limited effect on sensitivity, but the sensitivity values dramatically reduced when SNR was  $-10$  dB; (c) The false positive rate in the traditional  $TE_{ensemble}$  fluctuated around 0.01 when the significance  $\alpha$  level was 0.01 and increased to 0.05 or higher when  $\alpha$  was 0.05; (d,e) The CDT values of the novel  $TE_{ensemble}$  were compared with those of the traditional method for varied SNRs. For SNR = 50, 30, 20, 10, 0 dB, the differences in the CDT values for the two methods were tiny. The differences increased when SNR was  $-10$  dB.

Therefore, the novel  $TE_{ensemble}$  in this paper could robustly detect the strength and the direction of the interaction when SNR was above 0 dB, but this ability was reduced when SNR was lower than  $-10$  dB.

Above all, the significance  $p$  level had a limited effect on the sensitivity and the CDT values, but it changed the proportion of false positive dramatically. Therefore, in the following,  $p = 0.002$  was used as the significance level for the  $t$ -test in the novel  $TE_{ensemble}$ .

### 3.1.2. Window Length and the Number of Trials Affect the Stability of the Novel $TE_{ensemble}$

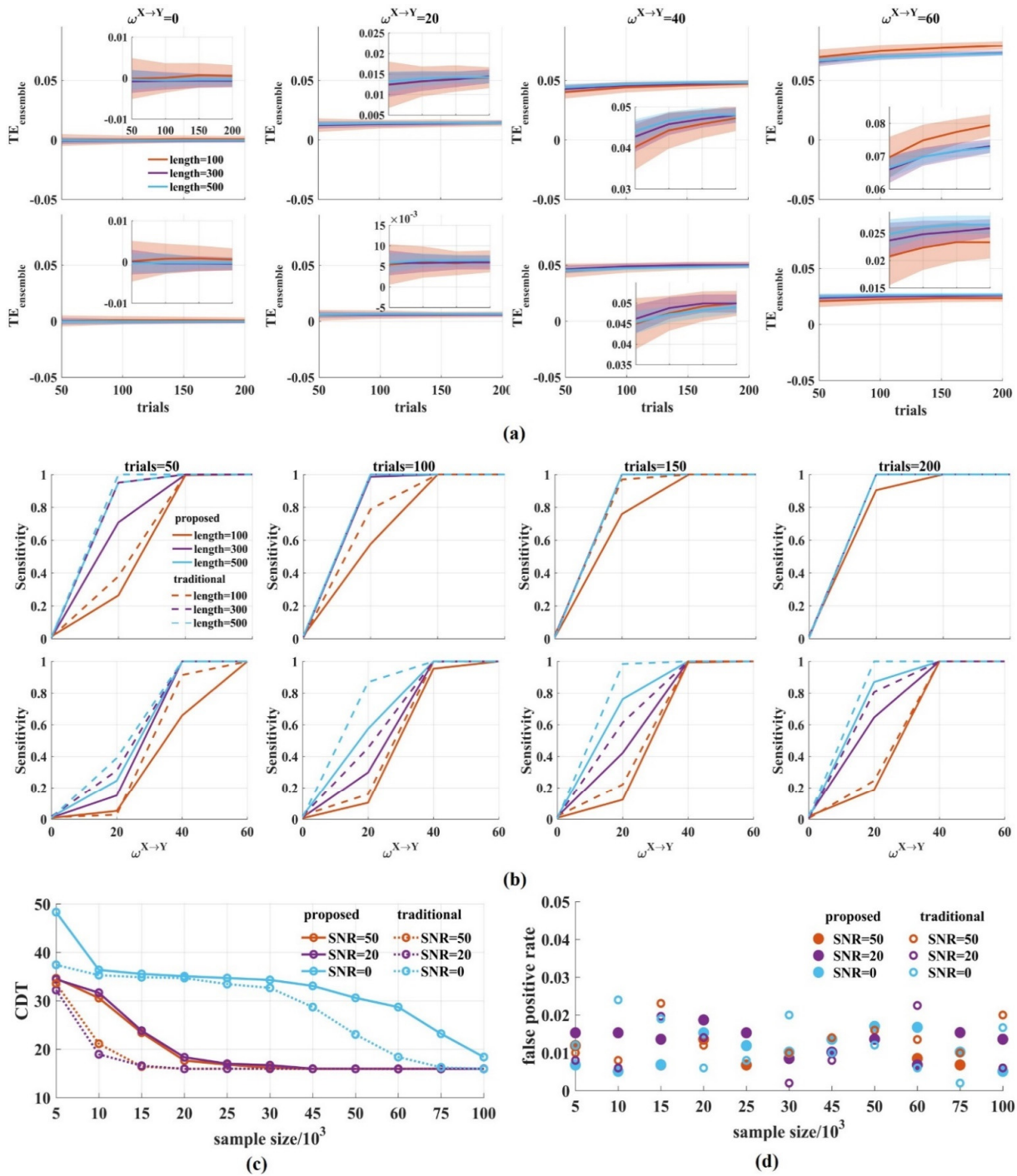
In order to explore the interaction dynamics, a short time window is used to improve the temporal resolution. However, accurate  $TE_{ensemble}$  estimation requires enough samples. We, therefore, investigated the effect of the window length and the number of trials on the performance of the novel  $TE_{ensemble}$ . Then, the results were compared with those of the traditional method. The samples in a time window were 100, 300, and 500, respectively. The number of trials were 50, 100, 150, and 200 with SNR = 0, 20 dB.

Figure 6a shows that when the SNR was 20 dB, the  $TE_{ensemble}$  values clustered around 0 for  $\omega^{X \rightarrow Y} = 0$  and increased monotonically with  $\omega^{X \rightarrow Y}$  from 20 to 60, regardless of the sample size in the ensemble members (Figure 6a top). The same results occurred for low SNR (Figure 6a bottom). As expected, by either increasing the window length or the number of trials, the variance of the  $TE_{ensemble}$  values was reduced with an increased sample size.

Figure 6b shows the sensitivity of the novel  $TE_{ensemble}$  with varied window lengths and the number of trials when SNR was 0, 20 dB. The longer the window length and the more trials used, the higher the sensitivity  $TE_{ensemble}$  obtained. Under the same conditions, the sensitivity for SNR = 20 dB is higher than that of SNR = 0 dB.

To quantify the sensitivity, we calculated the CDT values for a varied amount of simple sizes with SNR = 50, 20, 0 dB and compared the CDT values of the novel  $TE_{ensemble}$  with those of the traditional method. The results showed that the CDT values for the two methods gradually reduced with the increase in the sample size, regardless of the SNR values, and finally reached the same stable state (Figure 6c). For SNR = 50, 20 dB, the CDT values in the traditional  $TE_{ensemble}$  reached stability when the sample size was 15,000 while it was 20,000 in the novel  $TE_{ensemble}$ . The same results were obtained when SNR was 0 dB. The traditional method reached stability with 75,000 samples, which was faster than that of the novel  $TE_{ensemble}$ .

Finally, we calculated the false positive rate for different sample sizes. Figure 6d shows that the false positive rate fluctuated between 0 and 0.02 for the two methods regardless of the samples and noise. There were several values above 0.02 in the traditional method. The false positive rate of the novel  $TE_{ensemble}$  seems to be more stable. Therefore, the window length and the number of trials affect the stability of the novel  $TE_{ensemble}$ . With the increase in sample size, the variance of the  $TE_{ensemble}$  values was reduced and the sensitivity of the novel  $TE_{ensemble}$  reached its steady state, but it was slower than that of the traditional method.

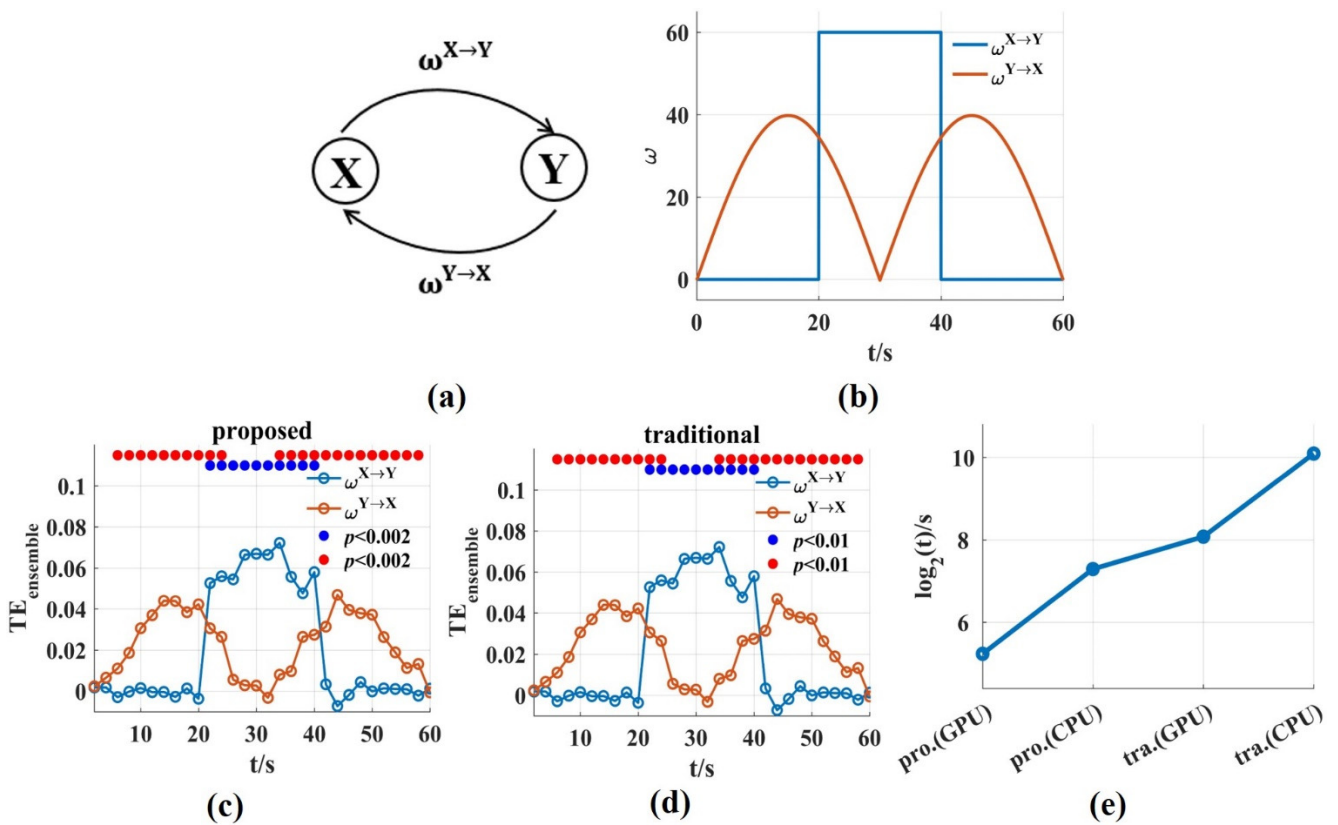


**Figure 6.** The performance of the novel  $TE_{ensemble}$  with varied window lengths and number of trials. (a) We calculated the  $TE_{ensemble}$  values for varied window lengths (100, 300, 500) and number of trials (50, 100, 150, 200) (top: SNR = 20 dB; bottom: SNR = 0 dB). In the upper (or bottom) right-hand corner is a detailed drawing. The  $TE_{ensemble}$  values clustered around 0 for  $\omega^{X \rightarrow Y} = 0$  and increased monotonically for  $\omega^{X \rightarrow Y}$  from 20 to 60, regardless of the SNR values. The variance of the  $TE_{ensemble}$  values was reduced with the increase in the sample size. The solid lines and shaded areas represent the mean and the variance of the  $TE_{ensemble}$  values, which were calculated 500 times by drawing a certain number of pairs from 1000 signal pairs; (b) The sensitivity of the novel  $TE_{ensemble}$  was calculated. The solid lines are the sensitivity values of the novel  $TE_{ensemble}$  and the dotted lines are

those of the traditional method. With the increase in the sample size, the sensitivity improved (top: SNR = 20 dB; bottom: SNR = 0 dB); (c) The CDT values against the sample size with SNR = 50, 20, 0 dB. With the increase in samples, the CDT values reached their stable state in the novel  $TE_{ensemble}$ , which was slower than that of the traditional  $TE_{ensemble}$ ; (d) We obtained the false positive rates for different sample sizes with SNR = 50, 20, 0 dB. They fluctuated between 0 and 0.02 for the two methods regardless of the samples and noise. There were several values above 0.02 in the traditional method.

### 3.1.3. The Novel $TE_{ensemble}$ Requires Less Computation Time to Track the Dynamic Interaction Process

To evaluate the ability and computation time of the novel  $TE_{ensemble}$  in this work in tracking the dynamic interaction process, signal pairs were generated with varying coupling strength ( $u = 20$  ms, trial length = 60 s, number of trials = 100, sampling frequency = 100 Hz).  $\omega^{X \rightarrow Y}$  was on/off boxcar and  $\omega^{Y \rightarrow X}$  varied with the absolute value of a sinusoid (Figure 7b). Then, we used a scanning approach to reconstruct  $TE_{ensemble}$  values and the corresponding interaction delay  $u$ . We scanned assumed delays in the interval  $u = [0, 30]$  ms with 10 ms steps and a window length  $\Delta t = 2$  s was used for the novel and the traditional  $TE_{ensemble}$ . Because they used same estimation method, the  $TE_{ensemble}$  values for both could track the dynamic process of  $\omega^{X \rightarrow Y}$  and  $\omega^{Y \rightarrow X}$ , and bidirectional coupling had no effect on each other (Figure 7c,d).

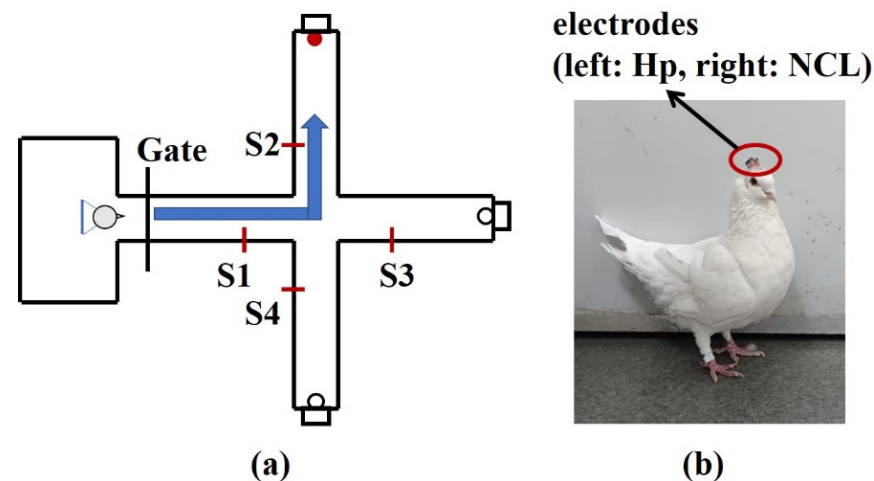


**Figure 7.** The ability and the time consumption of the novel  $TE_{ensemble}$  in tracking the dynamic interaction process. (a,b) Signal pairs were generated with varying coupling strength.  $\omega^{X \rightarrow Y}$  is the on/off boxcar and  $\omega^{Y \rightarrow X}$  varies with the absolute value of a sinusoid; (c,d) The  $TE_{ensemble}$  values for the novel and the traditional method could track the dynamic process of  $\omega^{X \rightarrow Y}$  and  $\omega^{Y \rightarrow X}$ . The solid blue and red circles indicated the significant interaction. (e) The time consumptions of the novel and the traditional  $TE_{ensemble}$  on GPU and CPU, respectively. The time consumption was reduced by about two or three orders of magnitude compared with that of the traditional method. The pro. and tra. are the abbreviations of proposed and traditional, respectively.

The time consumptions were compared for the two methods on GPU and CPU, respectively. For the above bidirectional interaction, the computation time of the traditional method on CPU (CPU-Intel(R) Core (TM) i5-10210U) was 1,237,322.12 s (500 surrogate data sets were used), and the running time on GPU was 37,494.6 s while the time consumption of the novel  $TE_{ensemble}$  was 9634.99 s on CPU, and just 291.97 s (4.87 min) on GPU (Figure 7e). The time consumption on GPU was calculated to be 33 times faster than that on CPU. Due to only one set of surrogate data used, the time consumption of the novel  $TE_{ensemble}$  was significantly lower than that of the traditional method, and the novel  $TE_{ensemble}$  solved the problem of computational complexity fundamentally.

### 3.2. Applying the Novel $TE_{ensemble}$ on the Actual Neural Signals

To explore the applicability of  $TE_{ensemble}$  in this paper for the actual neural signals, the local field potentials (LFPs) of hippocampal (Hp) and nidopallium caudolaterale (NCL) were recorded on two pigeons (P087, P089) when they were performing a goal-directed decision-making task. The pigeons were trained to start from the waiting area, pass through the straight area, and turn left, forward, or right as the goal location light instructed. If the pigeons choose the correct direction, they could obtain food (Figure 8). Details of the experimental process are described in [47].

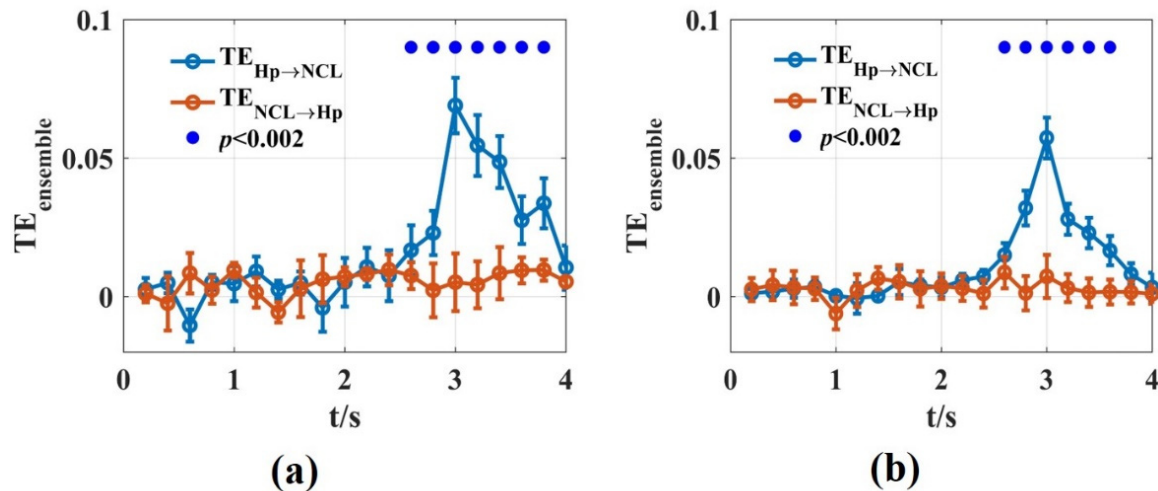


**Figure 8.** Diagrammatic sketch of a plus-maze and pigeon with implanted arrays. (a) The pigeons were trained to start from the waiting area, pass through the straight area, and turn left, forward, or right as the goal location light instructed. After a reward was consumed, they returned to the waiting area. S1, S2, S3, S4 are infrared sensors. (b) The microelectrode arrays were implanted at Hp and NCL.

Data processing was performed using custom scripts written in MATLAB. First, we removed the trials containing strong motion artifact. The adaptive common average reference was used for all channels of the remaining trials [46] and 50 Hz line noise was suppressed using an adaptive notch filter. Then, the signals were resampled to 1000 Hz and the slow gamma frequency band (40–60 Hz) of LFPs recorded from Hp and NCL were extracted for analysis. Finally, from the waiting to turning period, a total of 4-s signals were used. The number of independent repetition trials for each pigeon was 500.

We used the novel  $TE_{ensemble}$  that was introduced in this work to calculate the dynamic interaction between Hp and NCL of the pigeon in the goal-directed decision-making tasks. A time window of  $\Delta t = 200$  ms was used. For each non-overlapping time window, we scanned the assumed interaction delays in the interval  $u = [10, 50]$  ms with 5 ms steps. In total, 100 trials were used by randomly drawing from 500 trials for  $TE_{ensemble}$  estimation. This procedure was repeated 10 times. The results of the novel  $TE_{ensemble}$  were consistent with Zhao's, using the functional network and partial directional coherence [47]. The information flowed from Hp to NCL in the turning period with a 30 ms time delay. The novel  $TE_{ensemble}$  reported the dynamic interaction process in detail and found that the

interaction between the two brain regions reached the maximum at about 1 s after the animal entered the turning area (Figure 9). This may be related to the fact that the animal saw the light stimulation after entering the turning area, and then transferred the spatial location information formed in Hp to NCL. Therefore, the novel  $TE_{ensemble}$  in this paper is suitable for analyzing the dynamic interaction process between the actual neural signals.



**Figure 9.** The  $TE_{ensemble}$  values of the Hp and NCL brain regions in pigeons when they were performing the goal-directed decision-making tasks. **(a,b)**  $TE_{ensemble}$  values were calculated using the LFPs recorded from Hp and NCL for pigeon P087 **(a)** and P089 **(b)**. The blue lines are the  $TE_{ensemble}$  values from Hp to NCL and the red lines are the opposite. The solid blue circles represented the significant interaction from Hp to NCL when the significance  $p$  level was 0.002.

#### 4. Discussion

To reduce the time consumption of the ensemble transfer entropy ( $TE_{ensemble}$ ) and explore the dynamic interaction process in neuroscience, we proposed a fast, efficient  $TE_{ensemble}$  in which we modified the traditional statistical method. A  $t$ -test for the  $te_{ensemble}$  values that were estimated from the raw and the surrogate data was performed to test whether there was a significant difference in their mean values- $TE_{ensemble}$ . Because just one set of surrogate data was used, the time consumption of the novel  $TE_{ensemble}$  was significantly reduced. To validate the improved efficiency, the coupled signal pairs generated by a neural mass model were used. First, the novel  $TE_{ensemble}$  in this paper robustly detected the strength and the direction of the interaction between signal pairs with moderate noises (SNR was above 0 dB) and its performance decreased dramatically when SNR was  $-10$  dB. It yielded almost the same false positive rate and sensitivity as those of the traditional  $TE_{ensemble}$ . Second, with the increase in the window length and the number of trials, the novel  $TE_{ensemble}$  reached its stable state, but it was slower than that the traditional method. Third, the novel  $TE_{ensemble}$  could accurately track the dynamic interaction process and its computation time was reduced by two to three orders of magnitude compared with the traditional method. Finally, the applicability of the novel  $TE_{ensemble}$  in the realistic neural signals was verified on the LFP signals of Hp and NCL when pigeons performed goal-directed decision-making tasks. Therefore, the novel  $TE_{ensemble}$  in this paper may be a suitable way to investigate the dynamic interaction process between brain regions.

$TE_{ensemble}$  is a biased estimation and does not have a meaningful upper bound [9], so it is necessary to construct the null hypothesis distribution to test the statistical significance of  $TE_{ensemble}$ , which is an essential part in  $TE_{ensemble}$  [21]. In TE, the null hypothesis distribution can be constructed by one set of surrogate data ( $n$  trials) [9]. However, in the ensemble method, only a single metric value can be obtained from  $n$  trials [20], so the null hypothesis distribution is built by multiple sets (usually more than 500) of surrogate data, which dramatically increases the amount of calculation [21]. In this paper, we introduced a simple statistical method in the novel  $TE_{ensemble}$  by performing a  $t$ -test for  $te_{ensemble}$  of the

raw and the surrogate data (one set). Because just one set of surrogate data was needed, the computation time was reduced significantly and the computational complexity in  $TE_{ensemble}$  was fundamentally solved. However, there is still a large amount of calculation in the novel  $TE_{ensemble}$  with the KSG estimator. For the construction of a multi-brain dynamic interaction network, one workaround is that the novel  $TE_{ensemble}$  runs on GPU. Another way is to use the TE estimator with a small amount of calculation, for instance, the symbolic version of TE based on ordinal pattern symbolization, kernel-based transfer entropy, and the transfer entropy rate through Lempel–Ziv complexity. The next step is to generalize these estimators to the ensemble method and compare their performance in an ensemble of realizations.

One may wonder whether the service conditions of the  $t$ -test are met in the novel  $TE_{ensemble}$ . In fact, the  $te_{ensemble}$  of the raw data and the surrogate data are not normally distributed (Figure S1a,b). However, based on the central limit theorem, the  $t$ -test is always appropriate for large enough samples, regardless of the distribution of the data [41,42,48]. In the ensemble method, the sample size of  $te_{ensemble}$  is larger (generally more than 5000). So, it is stable to use the  $t$ -test to detect whether there is any significant difference between  $TE_{ensemble}$  of the raw data and  $TE_{ensemble}$  from the surrogate data. Meanwhile, we compared the false positive rate and the sensitivity of the novel  $TE_{ensemble}$  with the  $t$ -test and Wilcoxon rank-sum test. The two methods obtained the same false positive rate and CDT values (Figure S2a,b). Therefore, in this paper, it is possible to use a  $t$ -test of  $te_{ensemble}$  values from the raw data against  $te_{ensemble}$  values from the surrogate data to detect the significant difference between the  $TE_{ensemble}$  values of the raw and the surrogate data.

In the novel  $TE_{ensemble}$ , we chose  $p = 0.002$  as the significance level for the  $t$ -test. The false positive rate fluctuated around 0.01 when  $p$  was 0.002 and increased to 0.05 when  $p$  was 0.02. Someone may doubt whether the novel  $TE_{ensemble}$  is able to control the false positive rate at the desired level and they believe there is just a 1% chance of their result being a false alarm when  $p$  is 0.01. However, the false positive rate is not only related to  $p$  value but also intimately connected to the sample size. When the sample size is large, the sensitivity of the statistical test method is very high. The result is positive even for the two groups with a small difference [49]. In the ensemble method, the samples of  $te_{ensemble}$  are usually larger than 5000. If we expect the false positive rate to be 0.01, we should reduce the  $p$  value instead of  $p = 0.01$ . Meanwhile, the results in Section 3.1.1 also confirmed this conclusion. Only when  $p$  is 0.002, the false positive rate is around 0.01, and it is 0.03 instead of 0.01 when  $p$  is 0.01. So, the  $p$ -value does not measure the probability that the studied hypothesis is true. It reflects our level of tolerance for the false positive rate [50,51].

One of the major challenges in brain science is that the neural signals are corrupted by technical noises (power line interference, impedance fluctuation, motion artifacts, etc.) and biological artifacts (volume conductor, eye movement, eyeblink, muscular, etc.) [52–55]. The presence of noise can mask the features of the neural signals and affect the analysis of the interactions between brain regions. Various methods have been proposed to eliminate noise. For instance, the elimination of noise at the source by standardizing the experimental operation [45], reduction in the power line interference by the adaptive notch filter [56], removal of muscle artifact by ensemble empirical mode decomposition and multiset canonical correlation analysis [57], and so on. However, some filters that are obtained by convolution of the input with their impulse responses may blur the temporal or causal relations between signal and external events [45]. We should be cautious when using them. If signals are recorded on multiple channels, spatial filters may be applied to remove noise [46]. However, some noises are sufficiently complex so that we cannot disentangle them completely from neural processing, which we really need. Therefore, the analysis methods we use should be robust to noise. In Section 3.1.1, we investigated the robustness of the novel  $TE_{ensemble}$  to noise. The results show that the novel  $TE_{ensemble}$  can measure the strength and the direction of interaction robustly when the SNR is above 0 dB. Therefore, the novel  $TE_{ensemble}$  we propose is valid in the presence of moderate noises.

$TE_{ensemble}$  based on the dependent repetition trials detects interaction within a short time window. It has high temporal resolution and is suitable for analyzing the dynamic interaction process between neural signals [20]. However, the selection of the time window length needs to pay attention to the following points: First, we used a scanning method to obtain the interaction time delay  $u$ . However, the time delay can only be estimated accurately when the sample size is greater than 10,000 [21]. Therefore, in order to estimate the time delay accurately in collapse trials, the length of the time window should be selected to ensure that there are enough samples in the ensemble members. Second, in the novel  $TE_{ensemble}$ , to obtain an accurate estimation of future information, enough historical information should be involved. We used the Ragwitz criterion to calculate the embedded window (embedded dimension  $k, l$  \* embedded delay  $\tau$ ), which includes the past information of the source and the target signals and has the ability to predict the future of the target signal [37]. So, the larger the embedded window, the longer the time window that should be picked. Finally, the selection of the time window length is limited by the number of independent repetition trials. Based on the analysis in Section 3.1.2, the performance of the novel  $TE_{ensemble}$  is affected by the sample size. The larger the sample size, the better the stability of the novel  $TE_{ensemble}$ . The sensitivity and the CDT values reached stability when the sample size was more than 20,000 with moderate noise. Therefore, we can select a small window to improve the temporal resolution with more independent repetitions. When the number of independent repetitions is less, a large time window should be used to ensure the stability of the performance.

In TE, the KSG estimator requires the signal to be stationary to obtain accurate results [32]. However, this is difficult to realize in neuroscience and most neural signals are non-stationary.  $TE_{ensemble}$  solves this problem by estimating using independent repetitions in which equivalent events (or equivalent brain activity) occur periodically. The neural signals recorded from these trials are assumed to be cyclo-stationary [58]. In general, independent repetitions meet this hypothesis [21]. However, the brain activity in the learning tasks changes gradually and its neural signals are not cyclo-stationary, so in this case, caution should be exercised when using the novel  $TE_{SPO}$ .

Local transfer entropy is a time-varying version of TE, which was proposed by Lizier to realize the local or pointwise interaction estimation, and TE can be expressed as an average of local transfer entropy [39]. In recent years, local transfer entropy has been used in neuroscience [59]. Ramón demonstrated that the local mutual information is suitable for measuring the dynamics of cross-frequency coupling in brain electrophysiological signals [32]. Local transfer entropy has been applied to explore the dynamic coupling of the phase–amplitude during seizures [59]. However, Sezen questions the local causality measure of local transfer entropy because the causal nature does not necessarily remain in each part [60]. In addition, there is high-frequency leakage into the local transfer entropy, and it is difficult to explain this phenomenon theoretically. The use of local transfer entropy to measure dynamic interaction needs further research. In this paper, local transfer entropy is not directly used to investigate the dynamic interaction between signal pairs. We used the characteristic that TE is the average value of the local transfer entropy for the statistical analysis in the novel  $TE_{ensemble}$ .

Research in rodents and avian has shown that goal-directed behavior involves multiple brain regions. Among them, Hp and PFC/NCL play important roles. Hp participates in goal-directed behavior by recognizing key locations in space [61]. PFC/NCL is involved in weighing conflicting, then making a decision [62]. Hp and PFC/NCL have very close functional interactions that contribute to goal-directed behavior's successful execution [63]. In a previous work, we investigated the interaction between Hp and NCL of pigeons in a goal-directed task using the local function network and partial directional coherence. The results show that during the turning area, the functional interaction of the Hp–NCL increases significantly, and the information flows from Hp to NCL [48], which was also detected using the novel  $TE_{ensemble}$ . However, the whole decision-making period is as long as 2 s, in which the dynamics of Hp–NCL interaction is unknown. The novel  $TE_{ensemble}$



solves this problem and it was found that the  $TE_{ensemble}$  metric value reaches the maximum at about 1 s after the animal entered the turning area. This may be due to the spatial location information forming in Hp when the animals saw the light stimulation in the turning area, and then it being transferred from Hp to NCL for decision-making. The results show that the novel  $TE_{ensemble}$  is suitable for investigating the dynamic interaction between the actual neural signals.

## 5. Conclusions

We introduced a fast and efficient ensemble transfer entropy ( $TE_{ensemble}$ ) to detect the dynamic interaction process between neural signals. It uses a  $t$ -test for the ensemble local transfer entropy ( $te_{ensemble}$ ) from the raw data against the  $te_{ensemble}$  from the surrogate data as the statistical method in the novel  $TE_{ensemble}$ . Due to just one set of surrogate data being used, the time consumption is significantly reduced. We compared the performance of the novel  $TE_{ensemble}$  with that of the traditional  $TE_{ensemble}$ . The novel  $TE_{ensemble}$  exhibited many characteristics. First, the time consumption was reduced by two or three magnitudes. Second, the novel  $TE_{ensemble}$  reliably measures the strength and the direction of the interaction in the presence of moderate noise. Compared with the traditional  $TE_{ensemble}$ , the novel  $TE_{ensemble}$  was slower to reach its steady state with the increase in the sample size. Third, the novel  $TE_{ensemble}$  could track the dynamic interaction process accurately. Finally, the novel  $TE_{ensemble}$  was applicable to the actual neural signals. Taken together, the novel  $TE_{ensemble}$  may be a suitable method for quantifying the dynamic interactions in neuroscience.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/e24081118/s1>, Figure S1: The distributions of  $te_{ensemble}$  values which were estimated from the raw and the surrogate data with varied  $\omega^{X \rightarrow Y}$ ; Figure S2: The false positive rate and the CDT values of the novel  $TE_{ensemble}$  with the  $t$ -test and the Wilcoxon rank sum test (left:  $t$ -test, right: Wilcoxon rank sum test).

**Author Contributions:** Conceptualization, J.Z., M.C., J.L. and H.W.; methodology, J.Z., M.C. and J.L.; software, J.Z., E.C. and Z.Z.; validation, J.Z. and Z.Z.; formal analysis, J.Z. and E.C.; investigation, J.Z.; resources, H.W.; data curation, K.Z. and Z.Z.; writing—original draft preparation, J.Z.; writing—review and editing, J.Z., M.C. and H.W.; visualization, J.Z., M.C., J.L. and E.C.; supervision, H.W.; project administration, H.W.; funding acquisition, H.W. and M.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by National Natural Science Foundation of China, grant number 61673353 and 62173310.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Life Science Ethical Review Committee of Zhengzhou University (No SYXK 2019-0002).

**Data Availability Statement:** The Trentool Matlab toolbox for  $TE_{SPO}$  estimation is open source and can be download at [https://github.com/trentool/TRENTOOL3\\_manual](https://github.com/trentool/TRENTOOL3_manual) (accessed on 15 November 2021). Matlab codes of the neural mass model are available in ModelDB (A neural mass model for critical assessment of brain connectivity; Ursino et al., 2020) at <http://modeldb.yale.edu/263637> (accessed on 1 January 2022). The actual neural data presented in this study are available on request from the corresponding author.

**Acknowledgments:** This study is completed during the period when COVID-19 is rampant around the world. Thanks to the medical staff and people from all walks of life who have made contributions to the fight against the epidemic.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Tononi, G.; Edelman, G.M.; Sporns, O. Complexity and coherency: Integrating information in the brain. *Trends Cogn. Sci.* **1998**, *2*, 474. [[CrossRef](#)]
2. Bullmore, E.; Sporns, O. The economy of brain network organization. *Nat. Rev. Neurosci.* **2012**, *13*, 336–349. [[CrossRef](#)] [[PubMed](#)]
3. Eichenbaum, H. Prefrontal-hippocampal interactions in episodic memory. *Nat. Rev. Neurosci.* **2017**, *18*, 547–558. [[CrossRef](#)] [[PubMed](#)]
4. Kanwisher, N. Functional specificity in the human brain: A window into the functional architecture of the mind. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 11163–11170. [[CrossRef](#)]
5. Ad, A.; Vmab, C. Causal dynamics and information flow in parietaltemporal-hippocampal circuits during mental arithmetic revealed by high-temporal resolution human intracranial EEG. *Cortex* **2022**, *147*, 24–40. [[CrossRef](#)]
6. Place, R.; Farovik, A.; Brockmann, M.; Eichenbaum, H. Bidirectional prefrontal-hippocampal interactions support context-guided memory. *Nat. Neurosci.* **2016**, *19*, 992–994. [[CrossRef](#)]
7. Bossomaier, T.; Barnett, L.; Harré, M.; Lizier, J.T. *An Introduction to Transfer Entropy*, 1st ed.; Springer International Publishing: Cham, Switzerland, 2016; pp. 65–95. [[CrossRef](#)]
8. Schreiber, T. Measuring Information Transfer. *Phys. Rev. Lett.* **2000**, *85*, 461–464. [[CrossRef](#)]
9. Lindner, M.; Vicente, R.; Priesemann, V.; Wibral, M. TRENTOOL: A Matlab open source toolbox to analyse information flow in time series data with transfer entropy. *BMC Neurosci.* **2011**, *12*, 119. [[CrossRef](#)]
10. Vicente, R.; Wibral, M.; Lindner, M.; Pipa, G. Transfer entropy—A model-free measure of effective connectivity for the neurosciences. *J. Comput. Neurosci.* **2011**, *30*, 45–67. [[CrossRef](#)]
11. Vakorin, V.A.; Kovacevic, N.; McIntosh, A.R. Exploring transient transfer entropy based on a group-wise ICA decomposition of EEG data. *NeuroImage* **2010**, *49*, 1593–1600. [[CrossRef](#)]
12. Barnett, L.; Barrett, A.B.; Seth, A.K. Granger causality and transfer entropy are equivalent for Gaussian variables. *Phys. Rev. Lett.* **2009**, *103*, 238701. [[CrossRef](#)] [[PubMed](#)]
13. Lehnertz, K.; Geier, C.; Rings, T.; Stahn, K. Capturing time-varying brain dynamics. *EPJ Nonlinear Biomed. Phys.* **2017**, *5*, 2. [[CrossRef](#)]
14. Hillebrand, A.; Tewarie, P.; van Dellen, E.; Yu, M.; Carbo, E.W.S.; Douw, L.; Gouw, A.A.; van Straaten, E.C.W.; Stam, C.J. Direction of information flow in large-scale resting-state networks is frequency-dependent. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 3867–3872. [[CrossRef](#)] [[PubMed](#)]
15. Tewarie, P.; Liuzzi, L.; O'Neill, G.C.; Quinn, A.J.; Brookes, M.J. Tracking dynamic brain networks using high temporal resolution MEG measures of functional connectivity. *NeuroImage* **2019**, *200*, 38–50. [[CrossRef](#)]
16. Voytek, B.; D'Esposito, M.; Crone, N.; Knight, R.T. A method for event-related phase/amplitude coupling. *NeuroImage* **2013**, *64*, 416–424. [[CrossRef](#)]
17. Andrzejak, R.G.; Ledberg, A.; Deco, G. Detecting event-related time-dependent directional couplings. *New J. Phys.* **2006**, *8*, 6. [[CrossRef](#)]
18. Leski, S.; Wojcik, D.K. Inferring coupling strength from event-related dynamics. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* **2008**, *78* (Suppl. S1), 41918. [[CrossRef](#)]
19. Martini, M.; Kranz, T.A.; Wagner, T.; Lehnertz, K. Inferring directional interactions from transient signals with symbolic transfer entropy. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* **2011**, *83*, 11919. [[CrossRef](#)]
20. Gómez-Herrero, G.; Wu, W.; Rutanen, K.; Soriano, M.; Pipa, G.; Vicente, R. Assessing coupling dynamics from an ensemble of time series. *Entropy* **2015**, *17*, 1958–1970. [[CrossRef](#)]
21. Wollstadt, P.; Martinez-Zarzuela, M.; Vicente, R.; Diaz-Pernas, F.J.; Wibral, M. Efficient transfer entropy analysis of non-stationary neural time series. *PLoS ONE* **2014**, *9*, e102833. [[CrossRef](#)]
22. Wibral, M.; Pampu, N.; Priesemann, V.; Siebenhühner, F.; Seiwert, H.; Lindner, M.; Lizier, J.T.; Vicente, R. Measuring information-transfer delays. *PLoS ONE* **2013**, *8*, e55809. [[CrossRef](#)]
23. Kraskov, A.; Stögbauer, H.; Grassberger, P. Estimating mutual information. *Phys. Rev. E* **2004**, *64*, 66138. [[CrossRef](#)] [[PubMed](#)]
24. Hlaváková-Schindler, K.; Palu, M.; Vejmelka, M.; Bhattacharya, J. Causality detection based on information-theoretic approaches in time series analysis. *Phys. Rep.* **2007**, *441*, 1–46. [[CrossRef](#)]
25. Shao, S.; Guo, C.; Luk, W.; Weston, S. Accelerating transfer entropy computation. In Proceedings of the IEEE 2014 International Conference on Field-Programmable Technology (FPT), Shanghai, China, 10–12 December 2014. [[CrossRef](#)]
26. Dourado, J.R.; Júnior, J.; Maciel, C. Parallelism Strategies for Big Data Delayed Transfer Entropy Evaluation. *Algorithms* **2019**, *12*, 190. [[CrossRef](#)]
27. Lobier, M.; Siebenhühner, F.; Palva, S.; Palva, J.M. Phase transfer entropy: A novel phase-based measure for directed connectivity in networks coupled by oscillatory interactions. *NeuroImage* **2014**, *85*, 853–872. [[CrossRef](#)]
28. Bastos, A.M.; Jan-Mathijs, S. A Tutorial Review of Functional Connectivity Analysis Methods and Their Interpretational Pitfalls. *Front. Syst. Neurosci.* **2016**, *9*, 175. [[CrossRef](#)]
29. Avena-Koenigsberger, A.; Misic, B.; Sporns, O. Communication dynamics in complex brain networks. *Nat. Rev. Neurosci.* **2018**, *19*, 17–33. [[CrossRef](#)]
30. Mona, F.; Gholam-Ali, H.Z.; Hamid, S.Z. Nonlinear effective connectivity measure based on adaptive Neuro Fuzzy Inference System and Granger Causality. *NeuroImage* **2018**, *181*, 382–394. [[CrossRef](#)]
31. Patricia, W.; Sellers, K.K.; Lucas, R.; Viola, P.; Axel, H.; Flavio, F.; Michael, W.; Hilgetag, C.C. Breakdown of local information processing may underlie isoflurane anesthesia effects. *PLoS Comput. Biol.* **2016**, *13*, e1005511. [[CrossRef](#)]

32. Mca, B.; Jha, C.; Ad, A.; Kd, B.; Rcs, D.; Sm, A. Measuring transient phase-amplitude coupling using local mutual information. *NeuroImage* **2019**, *185*, 361–378. [[CrossRef](#)]
33. Kaiser, A.; Schreiber, T. Information transfer in continuous processes. *Phys. D Nonlinear Phenom.* **2002**, *166*, 43–62. [[CrossRef](#)]
34. Lizier, J.T.; Prokopenko, M.; Zomaya, A.Y. Local information transfer as a spatiotemporal filter for complex systems. *Phys. Rev. E Stat. Nonlin. Soft. Matter Phys.* **2008**, *77*, 26110. [[CrossRef](#)] [[PubMed](#)]
35. Kozachenko, L.F.; Leonenko, N.N. A statistical estimate for the entropy of a random vector. *Probl. Inf. Transm.* **1987**, *23*, 9–16.
36. Frenzel, S.; Pompe, B. Partial mutual information for coupling analysis of multivariate time series. *Phys. Rev. Lett.* **2007**, *99*, 204101. [[CrossRef](#)] [[PubMed](#)]
37. Ragwitz, M.; Kantz, H. Markov models from data by simple nonlinear time series predictors in delay embedding spaces. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* **2002**, *65*, 56201. [[CrossRef](#)]
38. Runge, J. Conditional independence testing based on a nearest-neighbor estimator of conditional mutual information. In Proceedings of the 21st International Conference on Artificial Intelligence and Statistics, Lanzarote, Spain, 9–11 April 2018. [[CrossRef](#)]
39. Lizier, J.T. Measuring the Dynamics of Information Processing on a Local Scale in Time and Space. In *Directed Information Measures in Neuroscience. Understanding Complex Systems*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 161–193. [[CrossRef](#)]
40. Surhone, L.M.; Timpledon, M.T.; Marseken, S.F. *Student's T-Test*, 1st ed.; Betascript Publishing: Whitefish, MT, USA, 2010; pp. 36–52.
41. Lumley, T.; Diehr, P.; Emerson, S.; Chen, L. The Importance of the Normality Assumption in Large Public Health Data Sets. *Annu. Rev. Public Health* **2002**, *23*, 151–169. [[CrossRef](#)]
42. Poncet, A.; Courvoisier, D.S.; Combesure, C.; Perneger, T.V. Normality and Sample Size Do Not Matter for the Selection of an Appropriate Statistical Test for Two-Group Comparisons. *Methodology* **2016**, *12*, 61–71. [[CrossRef](#)]
43. Ursino, M.; Ricci, G.; Magosso, E. Transfer Entropy as a Measure of Brain Connectivity: A Critical Analysis With the Help of Neural Mass Models. *Front. Comput. Neurosci.* **2020**, *14*, 45. [[CrossRef](#)]
44. de Cheveigné, A.; Parra, L.C. Joint decorrelation, a versatile tool for multichannel data analysis. *NeuroImage* **2014**, *98*, 487–505. [[CrossRef](#)]
45. de Cheveigné, A.; Nelken, I. Filters: When, Why, and How (Not) to Use Them. *Neuron* **2019**, *102*, 280–293. [[CrossRef](#)]
46. Liu, X.; Hong, W.; Shan, L.; Yan, C.; Li, S. Adaptive common average reference for in vivo multichannel local field potentials. *Biomed. Eng. Lett.* **2017**, *7*, 7–15. [[CrossRef](#)]
47. Zhao, K.; Nie, J.; Yang, L.; Liu, X.; Shang, Z.; Wan, H. Hippocampus-nidopallium caudolaterale interactions exist in the goal-directed behavior of pigeon. *Brain Res. Bull.* **2019**, *153*, 257–265. [[CrossRef](#)] [[PubMed](#)]
48. Stonehouse, J.M.; Forrester, G.J. Robustness of the t and U tests under combined assumption violations. *J. Appl. Stat.* **1998**, *25*, 63–74. [[CrossRef](#)]
49. Hair, J.F.; Black, B.; Babin, B.J.; Anderson, R. *Multivariate Data Analysis*, 7th ed.; Prentice Hall: Hoboken, NJ, USA, 2014; pp. 4–12.
50. Wasserstein, R.L.; Lazar, N.A. The ASA Statement on p-Values: Context, Process, and Purpose. *Am. Stat.* **2016**, *70*, 129–133. [[CrossRef](#)]
51. Nuzzo, R. Scientific method: Statistical errors. *Nature* **2014**, *506*, 150–152. [[CrossRef](#)]
52. Kanitscheider, I.; Coen-Cagli, R.; Pouget, A. Origin of information-limiting noise correlations. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E6973–E6982. [[CrossRef](#)]
53. Pregowska, A. Signal Fluctuations and the Information Transmission Rates in Binary Communication Channels. *Preprints* **2020**, *2020*, 2020070297. [[CrossRef](#)]
54. Zhang, M.; Qu, H.; Xie, X.; Kurths, J. Supervised learning in spiking neural networks with noise-threshold. *Neurocomputing* **2017**, *219*, 333–349. [[CrossRef](#)]
55. Wang, Y.; Xu, H.; Li, D.; Wang, R.; Jin, C.; Yin, X.; Gao, S.; Mu, Q.; Xuan, L.; Cao, Z. Performance analysis of an adaptive optics system for free-space optics communication through atmospheric turbulence. *Sci. Rep.* **2018**, *8*, 1124. [[CrossRef](#)]
56. Keshtkaran, M.R.; Zhi, Y. A fast, robust algorithm for power line interference cancellation in neural recording. *J. Neural Eng.* **2014**, *11*, 26017. [[CrossRef](#)]
57. Chen, X.; He, C.; Peng, H. Removal of Muscle Artifacts from Single-Channel EEG Based on Ensemble Empirical Mode Decomposition and Multiset Canonical Correlation Analysis. *J. Appl. Math.* **2014**, *2014*, 261347. [[CrossRef](#)]
58. Gardner, W.A.; Napolitano, A.; Paura, L. Cyclostationarity: Half a century of research. *Signal Process.* **2006**, *86*, 639–697. [[CrossRef](#)]
59. Martínez-Cancino, R.; Delorme, A.; Wagner, J.; Kreutz-Delgado, K.; Makeig, S. What Can Local Transfer Entropy Tell Us about Phase-Amplitude Coupling in Electrophysiological Signals? *Entropy* **2020**, *22*, 1262. [[CrossRef](#)] [[PubMed](#)]
60. Cekić, S. Time, frequency, and time-varying Granger-causality measures in neuroscience. *Stat. Med.* **2018**, *37*, 1910–1931. [[CrossRef](#)]
61. Nardi, D.; Bingman, V.P. Asymmetrical participation of the left and right hippocampus for representing environmental geometry in homing pigeons. *Behav. Brain Res.* **2007**, *178*, 160–171. [[CrossRef](#)]
62. Güntürkün, O. The Avian “Prefrontal Cortex” and Cognition. *Curr. Opin. Neurobiol.* **2006**, *15*, 686–693. [[CrossRef](#)]
63. Liu, T.; Bai, W.; Xia, M.; Tian, X. Directional hippocampal-prefrontal interactions during working memory. *Behav. Brain Res.* **2018**, *338*, 1–8. [[CrossRef](#)]