

**Нажесткин Иван Андреевич**

**Сложность и вариативность как характеристики процесса  
самоорганизации нейронов головного мозга в процессе обучения**

**Специальность: 1.5.2 – Биофизика**

**АВТОРЕФЕРАТ**

**диссертации на соискание ученой степени  
кандидата физико-математических наук**

**Москва – 2022**

Работа прошла апробацию в федеральном государственном автономном образовательном учреждении высшего образования «Московский физико-технический технический институт (национальный исследовательский университет)»

**Научный руководитель:** Сварник Ольга Евгеньевна, кандидат психологических наук, ведущий научный сотрудник, лаборатория психофизиологии им. В.Б. Швыркова, Институт психологии Российской академии наук

**Ведущая организация:** Федеральное государственное бюджетное научное учреждение «Федеральный исследовательский центр Институт прикладной физики Российской академии наук»

Защита состоится **«23» декабря 2022 года в 13 часов 00 минут** на заседании диссертационного совета **ФБМФ.1.5.2.01**, созданного на базе федерального государственного автономного образовательного учреждения высшего образования «Московский физико-технический технический институт (национальный исследовательский университет)» (МФТИ, Физтех)

**по адресу:** 14170, Московская область, г. Долгопрудный, Институтский переулок, д. 9.

С диссертацией можно ознакомиться в библиотеке МФТИ и на сайте организации <https://mipt.ru>

Автореферат разослан **«23» ноября 2022 г.**

Ученый секретарь  
диссертационного совета

Зубарев Илья Владимирович

## **ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ**

### **Актуальность темы исследования**

Головной мозг представляет собой систему из десятков миллиардов нейронов, объединяющихся в сложные сети. Подобные сети способны обеспечивать сложное и разнообразное поведение организма в окружающей среде. Исследование процессов самоорганизации нейронов является актуальной задачей на настоящее время. Понимание подобных процессов позволит создавать самообучающиеся нейроморфные системы, способные выполнять широкий спектр задач, обучаясь им самостоятельно. Также это позволит улучшать работу живого мозга, улучшать его способности и лечить различные заболевания.

Для изучения процессов самоорганизации проводятся эксперименты с нахождением различных паттернов активности нейронов и сопоставлением их с выполняемым поведением. Наиболее перспективным является подход с регистрацией нейронной активности мозга лабораторных животных во время процесса обучения новому навыку – это позволяет непосредственно отследить происходящие в мозге изменения и непосредственно наблюдать за процессом самоорганизации. Более подробные результаты можно получить в экспериментах с регистрацией нейронной активности во время последовательного обучения двум последовательным навыкам – это позволяет оценить вовлечение нейронного субстрата получаемого опыта в уже имеющуюся нейронную сеть и использование уже имеющихся элементов для построения новых сетей.

В данной работе производилось исследование самоорганизации нейронов во время приобретения нового навыка с использованием двух различных подходов. Классические методы, основанные на анализе вариативности активности единичных нейронов, были применены к ретроспленальной коре – участку мозга, связанному с запоминанием ключевых объектов в среде и построением связей между ними. Новый подход, основанный на применении теории интегрированной информации, был впервые применён для исследования самоорганизации в течение всего процесса обучения. В головном мозге нейроны объединяются в тесные

сообщества, связанные между собой. В подобных сообществах выполняется узкоспециализированная обработка информации, а результаты строятся путём интеграции информации из сообществ. В теории интегрированной информации предложен показатель  $\Phi$ , который количественно показывает качество совместной работы таких сообществ. По мнению исследователей, этот показатель непосредственно описывает внутренние свойства системы, а не предоставляет внешнее описание [Niizato et al., 2020] и способен предсказывать успешность работы простых систем [Edlund et al., 2011; Engel and Malone, 2018], а также головного мозга [Isler et al., 2018; Fujii et al., 2019]. Однако, применение ИТ к данным нейронной активности значительно затруднено, поскольку вычислительная сложность алгоритма растёт с увеличением числа нейронов быстрее экспоненты. Было показано, что показатель интегрированной информации, вычисленный с использованием различных приближённых методов, описывает степень адаптации мозга к условиям окружающей среды.

### Цель и задачи

Целью работы является исследование процессов самоорганизации нейронов в головном мозге во время приобретения новых навыков и разработка методов оценки изменений спайковой активности (генерации потенциалов действия), непосредственно связанных с процессом самоорганизации.

Для достижения заявленной цели были поставлены следующие задачи:

1. Получить данные нейронной активности ретроспленальной коры, гиппокампа и миндалины у лабораторных животных в процессе приобретения нового поведения.
2. Проанализировать вариативность частот и межспайковых интервалов нейронов при выполнении поведенческих актов и вне их, разработать программное обеспечение для выполнения данного анализа.
3. Выбрать и применить целесообразный математический метод для вычисления коэффициента интегрированной информации  $\Phi$  на имеющихся данных с учётом их природы с целью анализа сложности нейронной сети, выбрать и

проверить методы аппроксимации для вычисления коэффициента интегрированной информации  $\Phi$  для систем большого размера.

### **Научная новизна работы**

В работе обнаружены новые закономерности динамики формирования нейронных паттернов ретроспленциальной коры головного мозга течение всего процесса приобретения навыка, начиная с самых ранних этапов. Впервые исследован процесс добавления представлений новых объектов в ретроспленциальной коре при изменении требуемого действия, которое необходимо для успешного функционирования в среде. Исследовано изменение вариативности нейронной активности, являющейся индикатором процессов поиска необходимых паттернов, необходимых для успешного выполнения задачи, в течение всего процесса обучения. Впервые показана модуляция вариативности в зависимости от успешности выполняемой задачи для нейронов, кодирующих критически важные для обучения объекты.

Впервые теория интегрированной информации (ИТ) была применена для оценки сложности организации нейронной активности в головном мозге на всех этапах приобретения навыков. Показано, что показатель интегрированной информации  $\Phi$  имеет корреляцию с классической метрикой успешности в данной среде (количеством вознаграждений). Также показано, что изменения, описываемые показателем  $\Phi$  в малой подвыборке нейронов исследуемой области мозга, отражают работу всей области.

Впервые на многодневных спайковых данных нейронной активности продемонстрировано применение нескольких приближённых подходов к вычислению показателя интегрированной информации, позволяющих значительно ускорить расчёты. Показана возможность применения трёх различных методов приближения.

### **Теоретическая и практическая значимость работы**

Полученные в настоящей работе закономерности являются основой для постановки новых гипотез и для более глубокого понимания принципов работы головного мозга. Проведённое исследование уточняет функциональное значение ретроспленциальной коры, гиппокампа и миндалины в головном мозге и задаёт вектор дальнейших исследований в данной области. Полученные закономерности могут быть использованы для воздействий на нейронные сети (манипуляций) в процессе научения для оптимизации приобретения новых навыков.

### **Положения, выносимые на защиту**

1. Степень вариативности активности у нейронов ретроспленциальной коры различалась в зависимости от объектов среды, которые они кодировали, и от успешности выполнения навыка.
2. Показано, что сложность, оцененная с помощью показателя интегрированной информации  $\Phi$ , имеет положительную корреляцию с успешностью приобретаемого навыка.
3. С использованием коэффициента интегрированной информации  $\Phi$  показана положительная корреляция между уровнями сложности нейронных сетей двух взаимодействующих отделов головного мозга при обучении.
4. Предложены три приближённых метода для быстрого вычисления коэффициента интегрированной информации  $\Phi$ : два с упрощением искомого разбиения системы и один с ограничением количества нейронов. При этом показано, что методы с упрощением искомого разбиения позволяют получить  $\Phi$ , имеющий более высокий уровень корреляции с успешностью выполнения приобретаемого навыка.

### **Степень достоверности и апробации результатов**

По материалам диссертации опубликовано 3 статьи, индексируемых системами Scopus и Web of Science, а также 4 тезисов в сборниках конференций.

## **Структура и объём работы**

Работа состоит из введения, 5 глав основного содержания, заключения, выводов; и литературы. Работа изложена на 138 страницах, включая 28 рисунков. Список литературы включает 348 ссылок.

## **ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ**

**Глава 1** представляет собой обзор литературы, посвящённой существующим математическим методам анализа нейронной активности. **Глава 2** описывает эксперимент с классическими методами, связанными с исследованием вариативности нейронной активности. **Глава 3** описывает применение теории интегрированной информации (ИИ) для исследования степени адаптации к среде в одном отделе головного мозга. **Глава 4** описывает применение ИИ для исследования адаптации к окружающей среде в двух отделах головного мозга при переобучении на связанный навык. **Глава 5** описывает различные методы приближённого вычисления показателя интегрированной информации.

### **2.1. Постановка и проведение эксперимента с анализом вариативности нейронной активности. Обработка данных.**

Для исследования самоорганизации нейронных сетей классическими методами был проведён эксперимент с многоэтапным последовательным обучением сложному навыку. Обучение производилось в специальной клетке с механизированной кормушкой и двумя педалями на противоположной стороне (рис. 1). Финальный навык заключался в том, что крысы должны были нажать на одну из двух педалей для получения вознаграждения из кормушки. Обучение проводилось поэтапно: сначала крысам было необходимо привыкнуть к кормушке (научиться получать из неё пищу) (1), затем вознаграждение выдавалось за отворот от кормушки и последующий возврат к ней (2), далее – за выход на середину клетки

(3), затем – за подход к педали (4), и, наконец, за нажатие на педаль (5). Каждый этап в данной последовательности, фактически, был отдельным навыком, который приобретали животные, однако рассматривались только два последних навыка, как наиболее содержательные. Регистрация нейронной активности осуществлялась с помощью массива из четырёх микроэлектродов (NeuroNexus, США) с 16 сайтами для регистрации. Каждая экспериментальная сессия записывалась на видео, и после регистрации на видеозаписи отмечались ключевые поведенческие акты, как связанные с выполнением навыка (нажатие на педаль, движение от педали к кормушке, нахождение у кормушки и т.д.), так и не связанные (стойки, груминг).

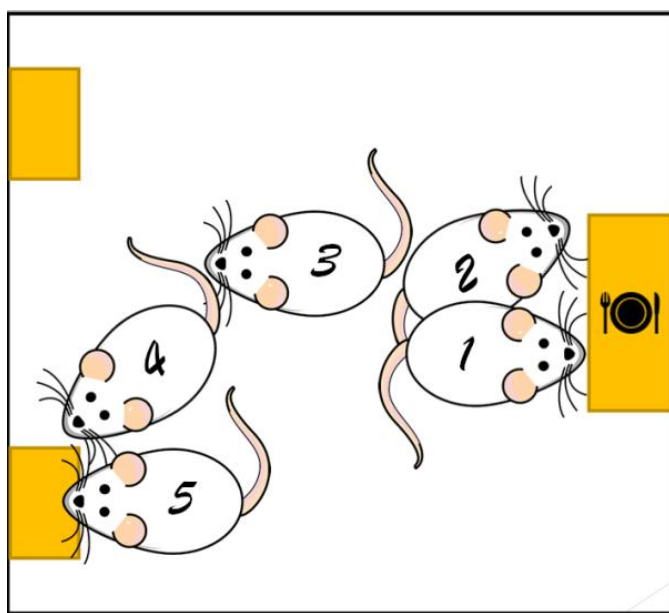


Рис. 1. Схема искусственной среды для обучения животных) с этапами обучения.

Среди зарегистрированных нейронов выделялись специфически активирующиеся к поведенческим актам. Нейрон считался специфическим по отношению к поведенческому акту, если было зарегистрировано превышение его активности в этом акте более чем на 50% относительно средней частоты данного нейрона на всем протяжении периода его регистрации.

Каждая экспериментальная сессия была разбита на 6 равных периодов длиной по 5 минут каждый. Для исследования пластичности нейронов в ходе обучения была рассчитана вариативность активности нейронов. Известно [Churchland et al., 2006; Mandelblat-Cerf et al., 2009; Athalye et al., 2017], что вариативность является признаком дезорганизованности нейронной активности, поиска требуемого паттерна, обеспечивающего выполнение задачи. Рассчитывалась глобальная вариативность (разность между межспайковыми интервалами в смежных периодах, усреднённая по всем экспериментальным сессиям) и локальная вариативность – разность между максимальным и минимальным межспайковыми интервалами в каждом из периодов, усреднённая



по всем экспериментальным сессиям. Также исследовалась вариативность активности нейронов между основными поведенческими актами по алгоритму из [Karlsson et al., 2012]. Каждому акту получения пищи в соответствие была поставлена точка в трёхмерном пространстве, у которой координатами были частоты исследуемого нейрона в интервале  $\pm 1$  сек вокруг начала отхода от педали, начала нахождения у кормушки и завершения нахождения у кормушки. Мерой вариативности было среднее арифметическое всех попарных расстояний между точками.

Для имплантации микроэлектродов была выбрана ретроспленальная кора (РСК), поскольку она используется в пространственном обучении. РСК задействована в большом количестве задач, таких как рабочая память в пространстве [Vann and Aggleton, 2002; St-Laurent et al., 2009], память о страхе в определённой обстановке [Keene and Bucci, 2008; Katche et al., 2013; Tanaka et al., 2014; Kwaris et al., 2015; Sigwald et al., 2016], память о времени событий [Bowers et al., 1988; Todd et al., 2015], построение траекторий и маршрутов [Cho and Sharp, 2001], пространственная навигация [Cooper and Mizumori, 1999; Cooper et al., 2001; Pothuizen et al., 2008; Alexander and Nitz, 2015]. РСК также выполняет запоминание мест и объектов, критически важных для выполнения задачи [Smith, 2002; Vann and Aggleton, 2002; Veddler et al., 2017; de Landeta et al., 2020].

## 2.2. Результаты

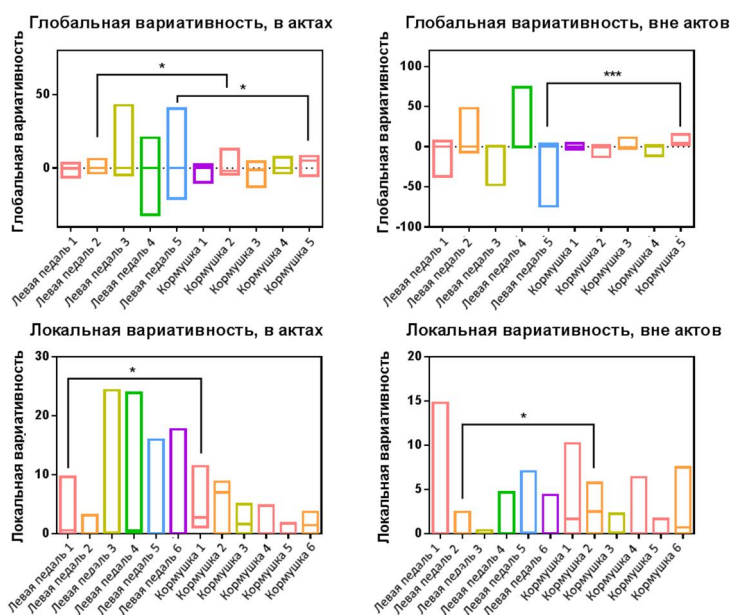


Рис. 2. Локальная и глобальная вариативность нейронов, специфических относительно левой педали и кормушки в зарегистрированных поведенческих актах и вне их в течение сессии обучения. Прямоугольником обозначены минимальные и максимальные значения, горизонтальная черта – медиана.

Были обнаружены периоды, в которых вариативность нейронов, специфических по отношению к кормушке, была достоверно выше, чем у нейронов, специфических по отношению к педали. Подобная закономерность наблюдалась и для локальной, и для глобальной вариативности (рис. 2). При этом вариативность нейронов, специфических к кормушке, была выше вариативности нейронов, специфических по отношению к левой педали. Кормушка участвовала в приобретении нескольких

различных навыков, а педаль была введена в процесс обучения позже, и связанное с ней задание изменялось. Нейроны, связанные с новым объектом, таким образом, имели меньшую вариативность, чем нейроны, связанные со старым.

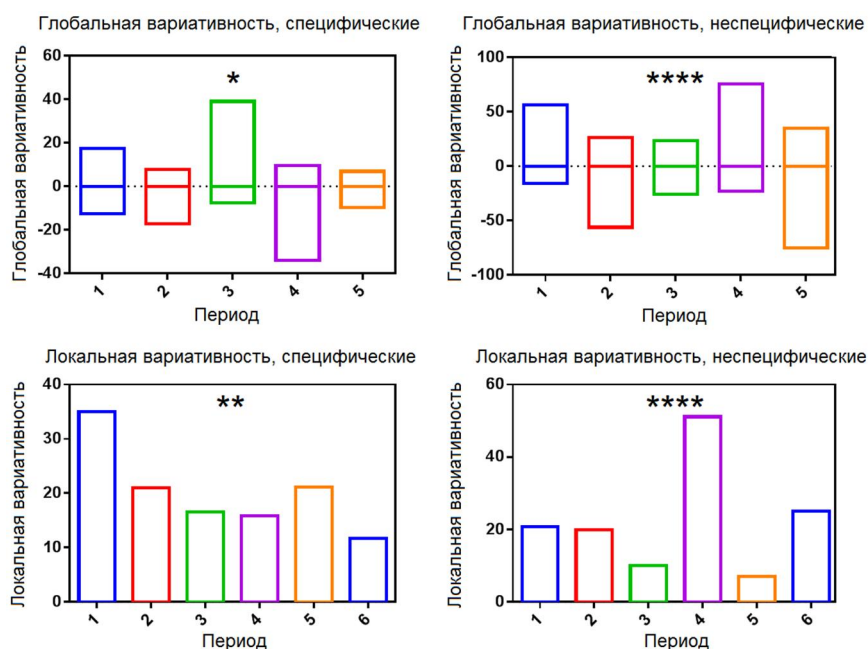


Рис. 3. Изменение локальной и глобальной вариативности специфических и неспецифических нейронов в течение сессии обучения.

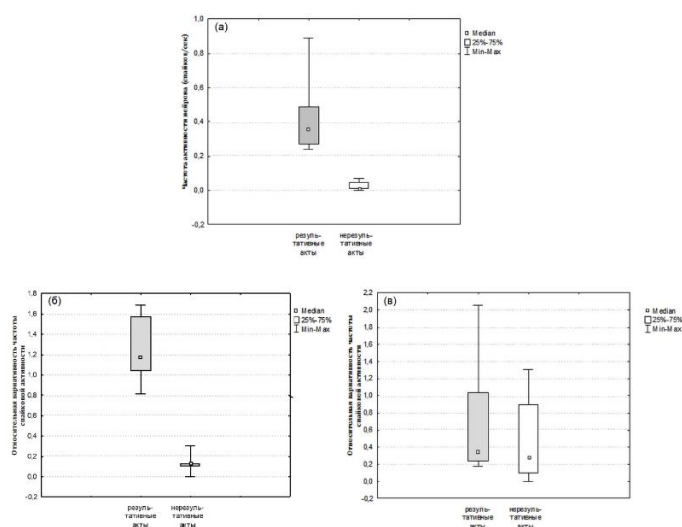


Рис.4. (а) Частоты активности специфических нейронов в процессе обучения в успешных (вознаграждаемых) и неуспешных поведенческих актах. Критерий Вилкоксона,  $p = 0.04$ ,  $Z = 2.02$ . (б). Относительная вариативность активности специфических нейронов в успешных и неуспешных поведенческих актах. Критерий Вилкоксона,  $p = 0.04$ ,  $Z = 2.02$ . (в). Относительная вариативность активности неспецифических нейронов в успешных и неуспешных поведенческих актах. Критерий Вилкоксона,  $p=0,21$ ,  $z=1,24$ .

И для специфических, и для неспецифических нейронов, вариативность изменялась в течение всей сессии (рис. 3). Во время выполнения успешных поведенческих актов специфические нейроны демонстрируют достоверно более высокую частоту активности, чем во время выполнения неуспешных (рис. 4, а). У специфических нейронов вариативность от акта к акту имела прямую взаимосвязь с успешностью выполнения задачи: вариативность во время выполнения успешных актов была выше, чем во время неуспешных (критерий Вилкоксона,  $p=0.04$ ,  $Z=2.02$ ) (рис. 4, б). Для нейронов, не имевших специфических активаций, такой зависимости обнаружено не было ( $p=0.21$ ,  $Z=1.24$ ) (рис. 4, в).

## 2.3. Выводы и обсуждение.

Выявлено, что все зарегистрированные нейроны в течение каждой сессии обучения изменяют свою вариативность. Предположительно, локальный успех, когда животному удаётся выполнить требуемую задачу, вызывает увеличение степени организованности, что отражается снижением вариативности. Дальнейшие неудачи снова вызывают рост вариативности. Кроме того, у неспецифических

нейронов изменение вариативности было более значимым, чем у специфических. Предположительно, нейроны, ранее не использованные в выполнении данного навыка, сильно изменяют свою активность, чтобы методом перебора определить такой паттерн, который будет полезен для достижения результата и будет включён в уже имеющуюся группу, обеспечивающую выполнение навыка.

Исследована динамика активаций специфических нейронов при ситуации рассогласования, когда предыдущее поведение уже не способно обеспечить адаптацию к среде. Вариативность активности нейронов, кодировавших кормушку (старый объект), в некоторых периодах была больше, чем у нейронов, кодировавших педаль (новый объект). Это можно объяснить тем, что в кормушке крысы получали пищу, получение которой было целью выполняемого навыка. Поэтому именно с кормушкой связана ситуация рассогласования. При этом включается механизм поиска нового решения, характеризующийся повышением вариативности.

Также исследована вариативность специфических нейронов непосредственно во время выполнения основных этапов поведенческого акта. Во время выполнения успешных актов вариативность была значимо выше, чем во время выполнения неуспешных. При этом, такая закономерность наблюдалась только для специфических нейронов. Таким образом, процесс поиска требуемого паттерна осуществлялся только специфическими нейронами, то есть теми, которые уже вовлечены в выполнение навыка. Включение остальных нейронов происходит, по-видимому, вне поведенческих актов, в перерывах между ними.

### **3. Применение теории интегрированной информации к исследованию самоорганизации нейронов в процессе адаптации к окружающей среде.**

В данной главе исследовалась возможность применения нового подхода – теории интегрированной информации (Integrated Information Theory, ИТ) – к оценке изменений, происходящих в нейронной сети головного мозга в процессе обучения. Параметр интегрированной информации  $\Phi$  способен предсказывать степень успешности выполняемой сложной системой задачи [Edlund et al., 2011; Engel and Malone, 2018]. Поэтому была выдвинута гипотеза о взаимосвязи  $\Phi$  и степени

адаптации головного мозга к условиям окружающей среды. Таким образом,  $\Phi$  может стать новым маркером, количественно оценивающим степень функциональной связанности для нейронной сети головного мозга. Это может быть полезным в исследовании вовлечения различных областей мозга в процесс обучения, для исследования путей передачи и обработки информации в мозге.

В основе ИТ лежит оценка двух свойств, которыми должна обладать сложная система: способностью к дифференциации (сегрегации) и интеграции информации. Сегрегация обозначает, что система должна иметь возможность находиться в как можно большем числе состояний, чтобы иметь возможность кодировать как можно большее число стимулов. При этом крайне важна специализация – различная информация (например, стимулы разного типа) кодируются различными частями системы. Интеграция же заключается в том, что части системы должны быть взаимосвязанными, их состояния должны быть взаимозависимыми. Это крайне важное условие, поскольку без него система не сможет работать как единое целое, не сможет использовать всю доступную информацию для принятия решения.

Попытки применения ИТ к данным нейронной активности головного мозга уже производились. Для данных ЭЭГ в вычислительной модели мозга  $\Phi$  увеличивался с ходом обучения [Fujii et al., 2019], а в многодневной записи ЭЭГ мозга новорождённых детей наблюдалось увеличение  $\Phi$  с возрастом, то есть с ходом развития головного мозга [Isler et al., 2018]. Однако, в этих исследованиях использовались электроэнцефалографические данные, записанные с большого числа нейронов. Расчёт  $\Phi$  по спайковым данным до настоящего времени был значительно затруднен в связи с большой вычислительной сложностью определения  $\Phi$ . В связи с этим были применены приближённые методы вычисления  $\Phi$ . Таким образом, в настоящей работе впервые было произведено вычисление параметра интегрированной информации на основе спайковых данных в течение всего процесса обучения. Была выявлена взаимосвязь  $\Phi$  с количеством вознаграждений, полученным во время обучения.

### **3.1. Постановка и проведение эксперимента**

Для исследований был использован набор открытых данных [Jadhav et al., 2016; Tang et al., 2017; Jadhav and Frank, 2020]. Крысы (N=3) были обучены навыку ориентирования в трёхрукавном W-образном лабиринте. Вознаграждение выдавалось за проход из конца одного рукава к концу другого по определённым правилам, описанным в оригинальной статье [Jadhav et al., 2016]. Обучение проводилось в течение 8 дней, в каждый из которых проводилась одна или две сессии обучения продолжительностью 20-30 минут. В каждой сессии было зарегистрировано от 15 до 56 нейронов.

Вычисление  $\Phi$  осуществлялось для каждого из 8 равных периодов, на которые разделялась сессия. Для сопоставления полученных данных с внешним маркером поведения животного (количеством вознаграждений) вычислялась корреляция Спирмена. Нулевая гипотеза об отсутствии корреляции проверялась по критерию Стьюдента. Поскольку экспериментальные сессии имели разную длину, количество вознаграждений нормировалось на длину периода.

### 3.2. Вычисление показателя интегрированной информации для спайковых данных нейронной активности

Для расчётов использовался известный алгоритм авторегрессионного  $\Phi$ , поскольку вычисление по определению существенно затруднено вследствие нестабильности результата на практически получаемых наборах данных. Вычисление осуществляется по формуле:

$$\Phi_{AR} = \frac{\min_M \left( \frac{1}{2} \ln \frac{\det(\Sigma(X))}{\det(\Sigma(E^X))} - \sum_{k=1}^2 \frac{1}{2} \ln \frac{\det(\Sigma(M_k))}{\det(\Sigma(E^{M_k}))} \right)}{\frac{1}{2} \ln \left[ \min_k \{ (2\pi e)^{|M_k|} \det(\Sigma(M_k)) \} \right]} \quad (1)$$

Здесь  $\Sigma(X)$  – ковариационная матрица для переменной  $X$ ,  $E^X$  – остатки линейной регрессии, предсказывающей значение переменной  $X$  в момент времени  $t + \Delta t$  на основании значения в момент времени  $t$ ,  $M_k$  – подвыборки (две части) системы,  $|X|$  – мощность множества  $X$ . Подвыборки по определению выбираются таковыми, чтобы значение  $\Phi$  было минимальным. Таким образом оценивается наиболее слабое место системы, в котором имеется недостаток связей между

элементами, что приводит к увеличению средних путей передачи информации и, таким образом, к снижению производительности системы.

Для вычисления  $\Phi$  системы необходимо указать, что является узлами и определить состояния узлов в каждый момент времени. В данном случае, узлами являются нейроны, а состояния узлов определяются следующим образом. Время разбивается на отрезки (бины) равного размера. В случае, если в бине  $i$  нейрон произвёл спайк, то считается, что в момент времени  $i$  узел (нейрон) имел состояние 1, а если спайка не было, то считается, что узел имел состояние 0. Выбор размера бина осуществляется на основе компромисса между сохранением тонких паттернов спайковой активности нейронов и избыточными данными. В данном исследовании был выбран размер бина 0.03 сек.

В некоторых случаях  $\Phi$  вычислить не удавалось. Такое поведение происходит, когда узлы сети меняют свою активность слишком редко, и регрессии, фигурирующие в (1), не могут быть вычислены. В данном случае это происходило, когда нейрон был почти всегда не активен, или, наоборот, активировался почти в каждом бине. В таком случае заново отбиралось на единицу меньшее число нейронов и вычисление повторялось. После вычисления полученное значение нормировалось на количество нейронов, фактически участвовавших в расчёте. Это было сделано для обеспечения возможности сравнения результатов, полученных на различном числе нейронов.

Одним из параметров является сдвиг по количеству временных шагов  $\Delta t$ . При вычислении  $\Phi$  рассматривается информация, генерируемая системой за некоторое количество шагов (последовательных состояний). Имеются все основания полагать, что в головном мозге обработка информации происходит периодически [VanRullen, 2016; Fiebelkorn et al., 2018; Helfrich et al., 2018] и длина этого периода по разным оценкам варьируется от 50 до 500 мс [Isler et al., 2018]. Как и в [Engel and Malone, 2018], при вычислениях использовалось такое значение  $\Delta t$ , которое приводило к максимальному среднему значению  $\Phi$ .

### 3.3. Результаты

Все животные успешно приобрели требуемый навык. Для каждого животного было определено значение характерного временного масштаба обработки информации  $\Delta t$ :  $\Delta t = 0.42$  сек для первого животного,  $\Delta t = 0.45$  сек для второго и  $\Delta t = 0.45$  для третьего. Значения приблизительно равны 0.5 сек, что соответствует предыдущим результатам [].

Показатель интегрированной информации  $\Phi$  был рассчитан при данных оптимальных значениях  $\Delta t$  для всех периодов длиной 1/8 сессии. При этом для  $\Phi$  и количества вознаграждений была обнаружена статистически значимая корреляция (рис. 5). Для первого животного коэффициент корреляции Спирмена  $\rho = 0.4697$ , вероятность нулевой гипотезы  $p = 0.0001$ ; для второго животного  $\rho = 0.2791$ ,  $p = 0.0041$ , для третьего  $\rho = 0.3196$ ,  $p = 0.0009$ . С учётом поправки Бонферрони на множественные сравнения с числом проверяемых гипотез  $m = 3$ , когда минимальный уровень статистической значимости  $\alpha/m = 0.05/3 = 0.017$ , для всех трёх животных наблюдалась статистически значимая корреляция.

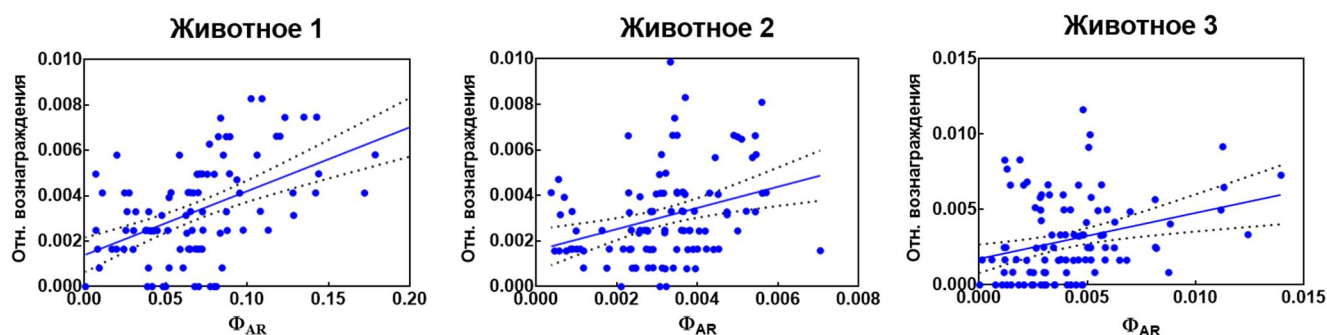


Рис. 5. Взаимосвязь показателя интегрированной информации с относительным количеством вознаграждений (количеством вознаграждений, нормированным на длину сессии)..

### 3.4. Выводы и обсуждение

Была продемонстрирована способность показателя интегрированной информации  $\Phi$  отражать степень адаптации системы к условиям. Предыдущие результаты других исследователей, описанные ранее, показывали, что  $\Phi$  может описывать непосредственно степень сложности системы изнутри, схватывая её структуру, приводящую к определённому результату. При этом подобное описание описывает саму систему, а не её метрики поведения, зависящие от окружающей



среды [Edlund et al., 2011]. Полученные в настоящей работе выводы находятся в полном соответствии с данными результатами и показывают возможность применения  $\Phi$  для анализа нейронов в процессе обучения. Этот кардинально новый подход позволит более точно оценивать степень вовлечённости определённых нейронных групп в процесс обучения, что позволит на более подробном уровне наблюдать за нейропластическими изменениями, происходящими в мозге. Также можно делать предположения относительно того, какой должна быть структура синаптических связей, чтобы показатель интегрированной информации, а значит и успешность адаптации сети, был максимальным. В работе [Barrett and Seth, 2011] приводятся размышления относительно того, каковой должна быть структура сети для получения максимального значения  $\Phi$ . Подобное моделирование, выполненное с учётом предположений, сделанных с использованием генеративных моделей, позволит узнать больше о конкретных механизмах самоорганизации, приводящих к адаптации нейронной сети к окружающим условиям.

## **4. Исследование процесса самоорганизации в двух взаимодействующих отделах головного мозга в процессе приобретения новых навыков**

### **4.1. Введение**

В предыдущей главе исследовалась самоорганизация нейронов в процессе обучения пространственному навыку в гиппокампе. Во время обучения показатель интегрированной информации изменялся в соответствии с выполнением животным требуемого навыка. Однако возникает вопрос, каким образом это изменение согласуется с предыдущим гиппокамп-зависимым опытом. Ранее описывались работы, в которых в процессе переобучения в области, связанной с выполнением старого навыка, наблюдались пластические изменения [Сварник и др., 2014]. В этой главе аналогичное исследование было выполнено для показателя интегрированной информации  $\Phi$ .

### **4.2. Постановка и проведение эксперимента**

Для проведения этого исследования также был использован набор открытых данных [Girardeau et al., 2017]. Крыс последовательно обучали двум навыкам:

простому гиппокамп-зависимому навыку пробежки по линейному лабиринту, и аверзивному (связанному со страхом), когда требовалось пробежать по тому же лабиринту, не обращая внимания на сильные потоки сжатого воздуха. Микроэлектроды были имплантированы в гиппокамп и в миндалину. Вычисление  $\Phi$  в каждой области мозга проводилось по такому же алгоритму, как и в предыдущей главе.

### 4.3. Результаты

Все животные успешно приобрели второй навык. При этом у всех животных характерный прогресс в обучении наблюдался до определённого дня, после чего количество получаемых в день вознаграждений переставало расти.

Для  $\Phi$  и в гиппокампе, и в миндалине у всех животных была обнаружена статистически значимая корреляция с относительным количеством вознаграждений (рис. 6). Также исследовалась совместная работа гиппокампа и миндалины в приобретении навыка. Статистически значимая корреляция между  $\Phi$  в гиппокампе и миндалине была обнаружена в те дни, когда наблюдался прогресс в обучении; после данного момента корреляции обнаружено не было (рис. 7).

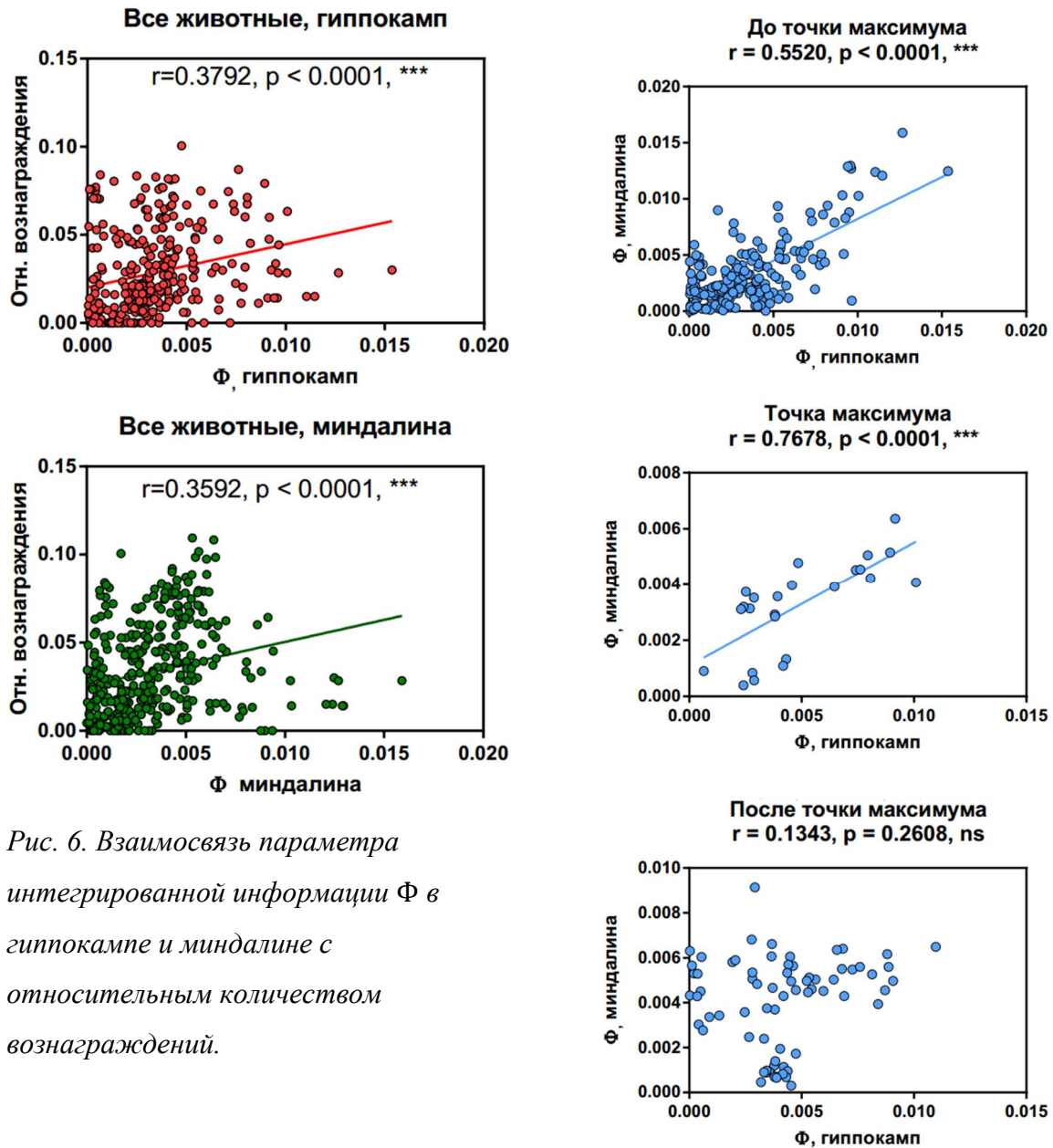


Рис. 6. Взаимосвязь параметра интегрированной информации  $\Phi$  в гиппокампе и миндалине с относительным количеством вознаграждений.

Рис. 7. Взаимосвязь параметра интегрированной информации в гиппокампе и миндалине до дня с максимальным количеством вознаграждений, в этот день и после него.

#### 4.4. Выводы и обсуждение

Была показана способность  $\Phi$  к оценке взаимосвязи между отдельными областями мозга при обучении. Ранее для оценки взаимодействия между областями использовались подходы, основанные на математической теории информации, например [Kirst et al., 2016; Palmigiano et al., 2017; Avena-Koenigsberger et al., 2018].

Новый подход позволяет оценить интеграцию областей друг между другом вместо простой оценки информации, передаваемой между областями. В данном случае показано взаимодействие гиппокампа с миндалиной, имеющее место только во время фактического обучения (во время роста количества вознаграждений). Также полученные результаты находятся в полном соответствии с полученными ранее результатами, показывающими, что область мозга, задействованная в реализации предыдущего навыка, связана и с реализацией нового, связанного с ним (например, [Сварник и др., 2014]).

## **5. Приближённые подходы к вычислению показателя интегрированной информации.**

### **5.1. Введение**

Вычисление  $\Phi$  предложенным методом (по 15 нейронам) может вызвать проблемы при рассмотрении больших сетей, в которых такое количество может стать чрезвычайно малым по сравнению с общим количеством нейронов, попадающих в расчёт. Поэтому необходимы приближённые методы, которые позволяют произвести расчёт по полной выборке, получая  $\Phi$ , не сильно отличающееся от точного значения.

Подобные временные затраты происходят из определения показателя  $\Phi$ , которое основано на поиске наиболее слабого места системы – разбиения, которое минимизирует  $\Phi$ . Поиск осуществляется только полным перебором. Приближённые подходы основаны на приближённом нахождении такого разбиения без полного перебора. Одним из первых предложений было использование так называемого атомарного разбиения (Atomic partition) [Edlund, 2011; Toker and Sommer, 2019, Mediano et al., 2019], которое представляет собой простое деление системы из  $N$  узлов на  $N$  частей, где каждый узел оказывается в одной части. Это частично решает задачу, но при этом структурой сети, обеспечивающей передачу информации, пренебрегается.

Нейронные сети в мозге не однородны, а состоит из отдельных тесно связанных кластеров. Каждый из этих кластеров является в некоторой степени самостоятельной единицей, информация из которых потом интегрируется []. По

определению в теории интегрированной информации именно это и рассматривается, поэтому такое распределение близко к искомому. Для выявления подобных кластеров используются методы статистического анализа нейронной активности. Сила синаптической связи определяется по величине корреляции спайковых последовательностей нейронов. После этого используются широко известные методы нахождения тесных кластеров во взвешенном графе, например, метод Ловена [Blondel et al., 2008]. В настоящей работе сравнивались значения  $\Phi$ , рассчитанные с помощью предыдущего способа (по выборке из 15 нейронов), и на полном наборе нейронов с атомарным разбиением и разбиением, найденным по алгоритму Ловена.

Для вычислений был использован тот же набор данных с гиппокамп-зависимым обучением, что и в первом исследовании [Jadhav et al., 2016; Jadhav and Frank, 2020]. В статистическом анализе использовалась поправка Бонферрони для 15 проверяемых гипотез. Вычисление  $\Phi$  каждым методом производилось в периодах в  $1/8$  сессии, как и ранее. Все остальные вычислительные операции также осуществлялись аналогично.

## 5.2. Результаты и обсуждение

Значение характерного временного масштаба  $\Delta t$ , оптимальные для всех используемых методов расчёта, представлены в Таблице 1. Для различных методов значения практически не отличались и были приблизительно равны 500 мс, что, как и в Главе 2, согласуется с ранее полученными оценками. Статистически значимая корреляция между  $\Phi$  и относительным количеством вознаграждений наблюдалась для всех животных в случае использования метода с использованием 15 нейронов. При использовании приближённых методов для второго животного статистически значимой корреляции не наблюдалось (рис. 8).

	15 нейронов	Разбиение Ловена	Атомарное разбиение
Животное 1	0.42 сек	0.45 сек	0.45 сек
Животное 2	0.48 сек	0.56 сек	0.39 сек
Животное 3	0.39 сек	0.36 сек	0.36 сек

Таблица 1. Значения характерного временного масштаба  $\Delta t$ , оптимальное для всех использованных методов вычисления  $\Phi$ .

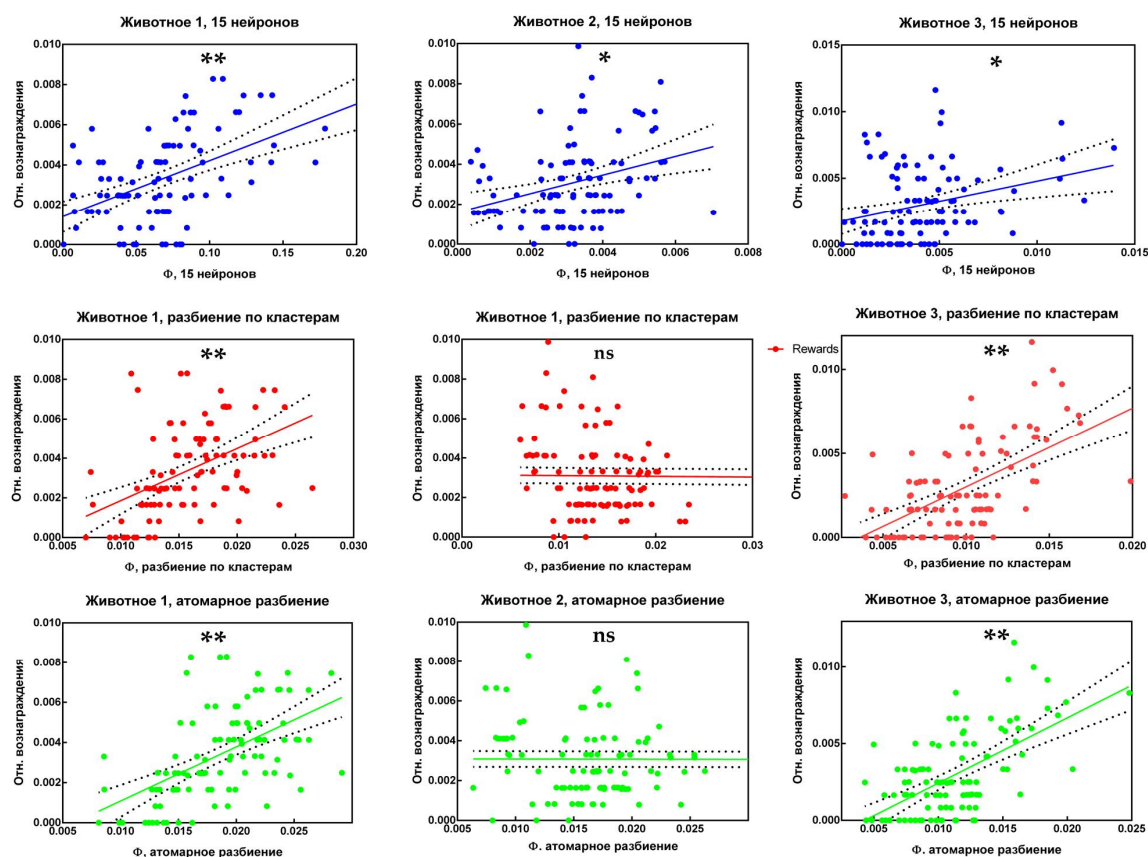


Рис. 8. Взаимосвязь показателя интегрированной информации и относительного количества вознаграждений в периодах для всех животных и всех методов вычисления.

Как видно,  $\Phi$ , вычисленный приближёнными методами, не всегда описывает адаптацию головного мозга к окружающей среде у второй крысы. В то же время, у второго животного во все дни, кроме последних было зарегистрировано максимальное количество нейронов из всех. Можно предположить, что дополнительные нейроны, которые были учтены при вычислениях приближёнными методами, не влияют на изменение  $\Phi$ , поскольку не участвуют в обучении. Для проверки этой гипотезы в каждой экспериментальной сессии были взяты только 15 нейронов, и к ним были применены приближённые методы расчёта  $\Phi$ . Отбор осуществлялся тем же способом, что и ранее. При этом для всех животных наблюдалась статистически значимая корреляция  $\Phi$  и количества вознаграждений (рис. 9).

Таким образом, видно, что только часть нейронов участвует в процессах, связанных с обучением и соответствующим изменением структуры нейронной сети. Удаление подобных нейронов привело к росту корреляции между показателем  $\Phi$  и относительным количеством вознаграждений.

### 5.3. Выводы

В данной главе были рассмотрены приближённые подходы, позволяющие значительно расширить область применения нового разработанного подхода – исследования пластичности в нейронных сетях с помощью теории

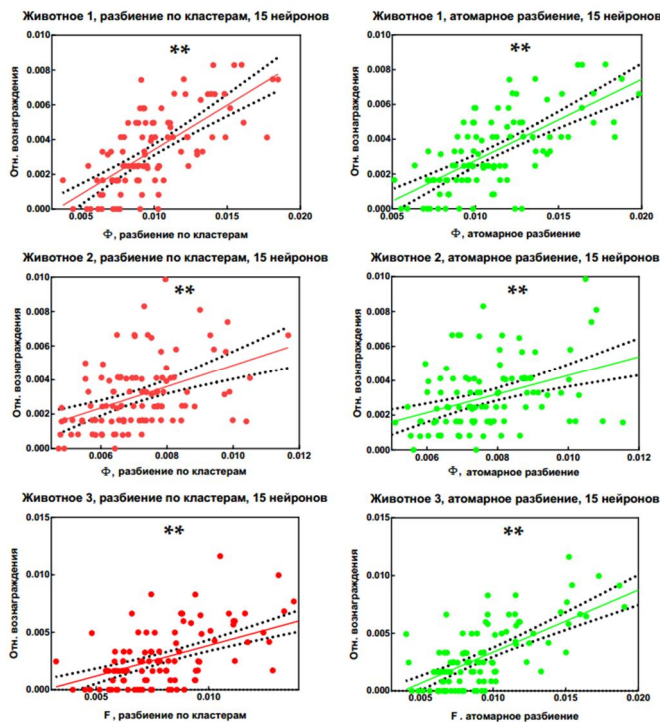


Рис. 9. Взаимосвязь параметра интегрированной информации  $\Phi$ , вычисленного с использованием приближённых методов на ограниченной выборке, с относительным количеством вознаграждений.

интегрированной информации. Новые результаты расширяют область его применения, позволяя увеличить количество рассматриваемых нейронов и значительно сократить время вычислений. Было подтверждено, что показатель  $\Phi$ , рассчитанный с использованием приближённых методов, также описывает степень приспособления нейронной системы к окружающей среде. Полученная корреляция  $\Phi$  с классическим показателем успешности сильнее корреляции, полученной с помощью точного расчёта. Это можно объяснить тем,

что точный расчёт получен с рассмотрением только разбиений системы на 2 части, в то время как метод Ловена позволяет обнаружить разбиение на несколько частей. Такое разбиение лучше соответствует классическому определению  $\Phi$  и позволяют оценить все слабые места системы, в которых обмен информацией между компонентами сведён к минимуму.

## ЗАКЛЮЧЕНИЕ

В настоящей работе исследовалась самоорганизация нейронов в головном мозге на разных этапах процесса приобретения нового навыка. Исследование самоорганизации проводилось двумя различными способами: с помощью анализа вариативности и с помощью анализа сложности, для чего впервые к последовательностям нейронных потенциалов действия была применена теория интегрированной информации.

В первой части работы оценивалась самоорганизация нейронов ретроспленциальной коры. Было установлено, что в ходе процесса обучения сложному навыку нейроны постоянно изменяют свою вариативность, которая, как известно, является мерой дезорганизованности и поиска решения. Изменения происходили в течение 5 минут, что показывает, что изменения в нейронной сети мозга происходят достаточно быстро. Также исследовались нейроны, кодирующие критически важный объект пространства, при изменении навыка, который было необходимо выполнить животному в пространстве. Нейроны, кодирующие этот объект, показывали большую вариативность активности (большую степень дезорганизованности), чем нейроны, кодирующие новый объект (связанный со вторым навыком). Наконец, исследовалось изменение частоты нейронов между поведенческими актами, и только нейроны, кодирующие критически важные объекты, изменяли свою частоту в зависимости от успешности или неуспешности реализуемого поведения. Таким образом, можно сделать выводы, что во время обучения нейрон много раз теряет организованность своей активности, что требуется для поиска решения выполняемой задачи, после чего организованность восстанавливается. За счёт этого, предположительно, наблюдается гибкость поведения и способность адаптироваться к новым условиям. Кроме того, из всего объёма нейронов только некоторые участвуют в выполнении задачи и претерпевают пластические изменения. Подобные закономерности впервые были обнаружены в ретроспленциальной коре.

Во второй части работы исследовалась сложность нейронных сетей с помощью теории интегрированной информации. Показатель интегрированной



информации  $\Phi$ , изначально предназначавшийся для оценки уровня сознания в головном мозге, оценивает преимущества совместной работы элементов системы, возникающие при росте её сложности. Однако, к головному мозгу его вычисление применялось только в ограниченном объёме вследствие серьёзной вычислительной сложности алгоритма. В работе использовались три способа, с помощью которых удалось обойти данное ограничение. Было подтверждено, что показатель  $\Phi$ , вычисленный всеми тремя способами, способен предсказывать степень успешности реализации задачи головным мозгом. Кроме того, была также выявлена возможность его применения для оценки скоординированной работы нескольких областей в мозге. Полученные выводы предоставляют новую возможность оценки вовлечения нейронов в выполнение задачи и оценки из взаимодействия.

### **Список работ, опубликованных автором по теме диссертации**

1. [Применение теории интегрированной информации к нейронной активности гиппокампа] Нажёсткин И. А., Сварник О. Е. Изменение показателя интегрированной информации для активности нейронов гиппокампа в процессе формирования оперантных навыков у крыс //Биофизика. – 2022. – Т. 67. – №. 1. – С. 73-81. DOI: 10.31857/S0006302921060069
2. [Применение теории интегрированной информации к исследованию двух взаимодействующих отделов мозга] Nazhestkin I. A., Svarnik O. E. Integrated Information Coefficient Estimated from Neuronal Activity in Hippocampus-Amygdala Complex of Rats as a Measure of Learning Success //Journal of Integrative Neuroscience. – 2022. – Т. 21. – №. 5. – С. 128. DOI: 10.31083/j.jin2105128
3. [Приближённые подходы к вычислению показателя интегрированной информации] Nazhestkin I., Svarnik O. Different Approximation Methods for Calculation of Integrated Information Coefficient in the Brain during Instrumental Learning //Brain Sciences. – 2022. – Т. 12. – №. 5. – С. 596. DOI: 10.3390/brainsci12050596

### **Список тезисов докладов по теме диссертации**

1. Нажесткин И.А. Пространственные активации нейронов ретроспленальной коры в процессе формирования нового инструментального поведения // Международный молодежный научный форум «ЛОМОНОСОВ-2019»
2. Fomin E., Nazhestkin I., and Svarnik O. Dynamics of neural activity variability in the process of acquiring new experience // National Congress on Cognitive Research, Artificial Intelligence and Neuroinformatics (CAICS-2020)
3. Нажесткин И.А. Интегрированная информация как метрика успешности обучения // Международный молодежный научный форум «ЛОМОНОСОВ-2021»
4. Нажесткин И.А. Применение теории интегрированной информации для исследования пластичности нейронных сетей головного мозга при обучении // XXIII Международная научно-техническая конференция "Нейроинформатика-2021"