Оптический журнал

Нейроиконика

УДК 612.8

Исследование инвариантных к масштабным преобразованиям механизмов классификации изображений

© 2019 г. Г. А. Моисеенко*, канд. биол. наук; С. В. Пронин*; Ю. Е. Шелепин*,**, доктор мед. наук

*Институт физиологии им. И.П. Павлова Российской академии наук, Санкт-Петербург
**Санкт-Петербургский государственный университет, Санкт-Петербург

E-mail: galina_pbox@mail.ru

Поступила в редакцию 20.08.2019

DOI:10.17586/1023-5086-2019-86-11-66-71

Цель исследования — изучение маркеров работы механизма классификации, инвариантного к масштабным преобразованиям изображений тестовых объектов. Наблюдателям предлагалось классифицировать изображения по критерию «живой/неживой» объект. Методом когнитивных вызванных потенциалов проведено 2 серии исследований. Угловые размеры изображений объектов, предъявляемых наблюдателю, 3° в одной серии и 0,4° в другой. Установлено, что классификация изображений стимулов различных угловых размеров по семантическим признакам (живой/неживой) вызывает различие амплитуд компонентов P200 (в отведениях F7, F8, фронтальной области). Выявлена роль компонентов вызванных потенциалов P200 в лобных областях мозга как маркера процесса классификации. Показано, что нейронные сети фронтальной коры для обеспечения классификации объектов используют инвариантное описание изображений, осуществленное на предыдущих этапах их обработки.

Ключевые слова: *классификация объектов, принятие решений, инвариантность восприятия, когнитивные вызванные потенциалы.*

Коды OCIS: *330.1070, 330.4270*

ВВЕДЕНИЕ

Зрительная классификация изображений — результат обобщенного описания воспринимаемых объектов и их детализации. Возможность восприятия индивидуальных особенностей изображений наблюдаемого объекта сочетается со способностью инвариантного восприятия при изменении освещенности объекта, изменении его размера, ориентации, разворота по осям и ракурса. Формирование инвариантного описания является важнейшим свойством восприятия [1–3]. Установлена способность к инвариантному опознанию у животных и человека [2–4]. Инвариантные свойства изображения объекта определяют его принадлежность к тому или иному классу. Они не зависят от размера и ориентации объекта, активируют один и тот же нейронный ансамбль, вызывая один и тот же отклик на выходе зрительной системы. Восприятие во многом базируется на возможности мозга находить в окружающем мире инвариантные характеристики [3]. На основании инвариантного описания во фронтальной области мозга происходит принятие решений, например, в задачах классификации изображений. Для осуществления этапа классификации нейронные сети на предыдущих уровнях, а именно в затылочно-нижневисочной коре, должны обеспечить инвариантное описание изображения. Исследование механизмов, обеспечивающих инвариантное распознавание образов в «зрительном мозге», является ключевым направлением исследований сенсорной физиологии [5–12]. В последние годы вновь возрос интерес к изучению механизмов инвариантности в связи с работами по созданию систем искусственного интеллекта [11–15].

В предыдущих работах нами исследовалось влияние инструкций наблюдателю на процессы классификации в мозге человека по наличию или отсутствию деталей в изображении. Для этого испытуемым предъявлялись отфильтрованные

на разных пространственных частотах изображения, проецируемые в область фовеа, и давались две инструкции: классифицировать изображения по физическим (неразмытое/размытое) и семантическим (живой/неживой объект) признакам. В результате проведенных исследований выявлены компоненты когнитивных вызванных потенциалов, связанные только с семантикой изображений в височных (в компоненте N170) и в лобных (в компоненте P200) отведениях. Различия в амплитудах компонентов вызванных потенциалов, связанные с семантикой изображений (живой/неживой), имели статистически значимую величину даже в тех условиях, когда задача испытуемого заключалась в классификации физических свойств (неразмытый/размытый) изображений объектов [7].

Целью данного исследования являлось исследование влияния размеров стимулов тестовых изображений на характеристики когнитивных вызванных потенциалов в задачах классификации.

МЕТОДИКА

Проведено 2 серии исследований с помощью метода когнитивных вызванных потенциалов.

В первой серии на экране монитора наблюдателям предъявлялись исходные (без цифровой фильтрации) контурные изображения объектов живой и неживой природы. Испытуемым давалась инструкция классифицировать изображения по критерию «живой/неживой», т.е. нажимать на одну клавишу мыши, если на экране был объект живой природы, и на другую — если объект относился к неживой природе. Для проведения первой серии исследований использовался набор из 90 черно-белых контурных изображений (45 изображений живых объектов, 45 — неживых) (рис. 1а). На экране монитора *Sony Multiscan* 200PS с кадровой частотой 100 Гц и разрешением экрана 1024×768 сделано 90 предъявлений каждого класса изображений. Средняя яркость и контраст всех стимулов были одинаковы. Расстояние между монитором и глазами испытуемого составляло 1,5 м. Угловой размер изображений в первой серии исследований — 3°. Изображения предъявлялись в случайном порядке, продолжительность демонстрации — 100 мс, интервал между предъявлениями — 1 с, кроме этого было предусмотрено время (около 300–500 мс) для нажатия испытуемым на клавишу при решении задачи.

Во второй серии инструкция испытуемым была аналогичной (классифицировать изображения по признакам «живой/неживой»). Но угловые размеры стимулов были меньше — 0,4° (рис. 1б).

В обеих электрофизиологических сериях исследований для регистрации вызванных потенциалов использовались энцефалограф фирмы «Мицар-ЭЭГ-201» с частотой дискретизации 500 Гц и программа WinEEG. Электроды располагались по системе 10–20 с добавочным затылочным отведением Oz с референтными ушными электродами. По́лосы пропускания фильтра высоких частот — 70 Гц, фильтра низких частот — 0,53 Гц, полосового фильтра 45–55 Гц. Глазодвигательные артефакты удалялись с помощью метода независимых компонентов, двигательные, глотательные и другие артефакты — вручную. Перед анализом данных удалялись и высокочастотные помехи с помощью фильтра высоких частот 50 Гц. Вызванные потенциалы усреднялись на разные классы стимулов на протяжении каждого исследования отдельно.

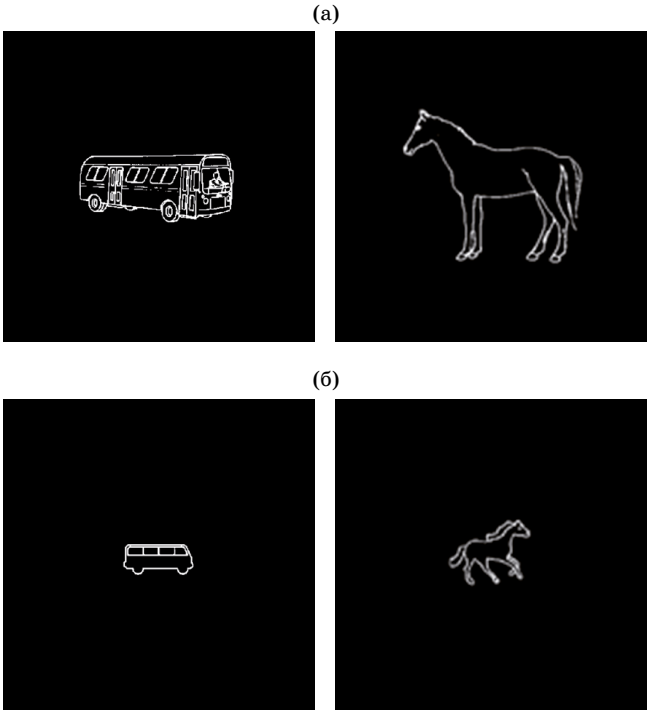

Рис. 1. Примеры больших (а) и малых (б) черно-белых контурных изображений объектов живой и неживой природы.

ОБРАБОТКА ДАННЫХ

Время реакции и процент правильных ответов регистрировались программой, предъявляющей тестовые изображения.

Для анализа вызванных потенциалов использовался пространственно-временно́й анализ сенсорных и когнитивных составляющих зрительных вызванных потенциалов [7]. Для измерения амплитуды компонентов выбирались временные окна, соответствующие колебаниям латентных периодов пиков основных компонентов вызванных потенциалов, обобщенных по всем испытуемым в разных отведениях (*Grand Average*). Эти окна, начиная от момента предъявления стимула, соответствовали компонентам в первой серии

исследований — P100 и N100 (80–150 мс), P170 и N170 (150–220 мс), P250 и N250 (220–300 мс), P300-500 (300–600). Амплитуды основных компонентов вызванных потенциалов во второй серии исследований анализировались во временных окнах — P100 и N100 (110–180 мс), P170 и N170 (180–250 мс), P250 и N250 (250–350 мс), P300-500 (350–700 мс). В пределах указанных окон фактическое временное окно могло немного изменяться, поэтому для каждой области мозга выбирался более точный временной интервал.

Для оценки статистической значимости различий амплитуд вызванных потенциалов в выбранных временных интервалах использовались парный двухвыборочный *t*-тест для средних различий (*t*-критерий для зависимых выборок) и *T*-критерий Вилкоксона. Результаты считали значимыми при достоверных различиях $p < 0{,}05$.

ХАРАКТЕРИСТИКА ИССЛЕДУЕМОЙ ГРУППЫ

В группу испытуемых 1-го исследования входили 14 добровольцев (4 мужчин и 10 женщин) в возрасте от 19 до 39 лет, праворукие, без неврологических патологий, с остротой зрения не менее 1.

В группу испытуемых 2-го исследования входили 14 добровольцев (8 мужчин и 6 женщин) в возрасте от 20 до 38 лет, праворукие, без неврологических патологий, с остротой зрения не менее 1.

РЕЗУЛЬТАТЫ

Результаты 1-й серии исследований (изображения большого размера)

В результате анализа электроэнцефалограммы (ЭЭГ), регистрируемой во время выполнения задачи, выявлены достоверные различия амплитуд различных компонентов вызванных потенциалов в зависимости от семантических признаков изображений.

На рис. 2 изображены карты достоверных различий амплитуд компонентов вызванных потенциалов по семантическим признакам во всех отведениях в разные интервалы времени при инструкции испытуемым отличать контурные изображения по критерию «живой/неживой». Данные рассчитаны по результатам измерений у 14-ти испытуемых. Средняя вероятность правильности ответов составила 97 ± 2%, порог статистической достоверности, рассчитанный с помощью *T*-критерия Вилкоксона, — 5% ($p < 0{,}05$). Из рисунка видно, что во временном интервале 80–150 мс нет различий реакций на разные типы стимулов, а в интервале 150–220 мс выявлены различия в амплитуде на изображения объектов по признаку «живой/неживой» в лобных, центральных и нижневисочных областях; в интервале 220–300 мс — в теменных областях мозга, а в интервале 300–600 мс — во всех областях мозга. В затылочных областях различия амплитуд были на все четыре категории стимулов (живые/неживые, размытые/неразмытые стимулы) в двух временных интервалах.

Результаты 2-й серии исследований (изображения малого размера)

Время реакции испытуемых составило 1210 ± 108 мс, вероятность правильных ответов — 87 ± 2%. На рис. 3 изображены карты достоверных различий амплитуд компонентов вызванных потенциалов по семантическим признакам во всех отведениях в разные интервалы времени при инструкции испытуемым отличать контурные изображения по признаку «живой/неживой». Из рисунка видно, что во временном интервале 110–180 мс нет различий в реакции наблюдателей на разные типы стимулов. В интервале 180–250 мс выявлены различия амплитуд компонентов

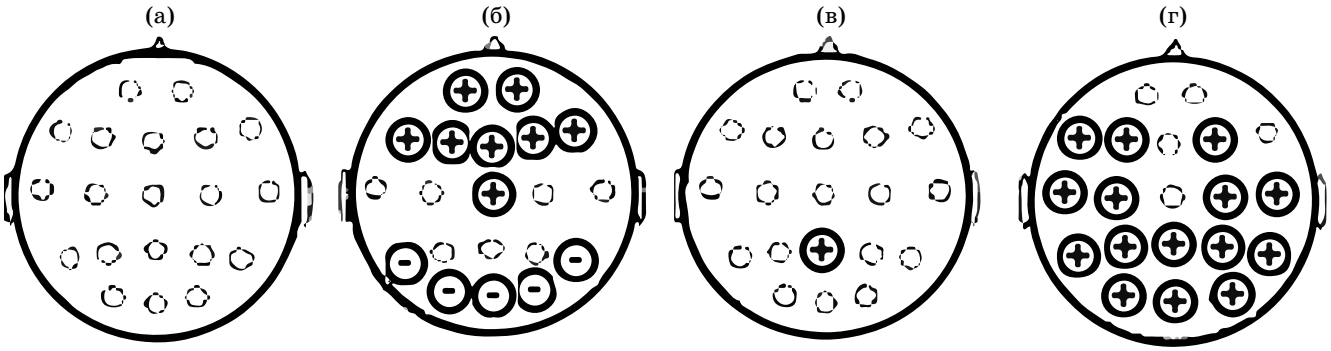

Рис. 2. Локализация достоверных различий амплитуд анализируемых компонентов вызванных потенциалов в задачах классификации объектов живой и неживой природы (стимулы — большие изображения) для всех стандартных отведений ЭЭГ от поверхности кожи головы в разные интервалы времени (а — 80–150 мс, б — 150–220 мс, в — 220–300 мс, г — 300–600 мс). На рисунках обозначены положения электродов на голове испытуемых для отведений ЭЭГ, в которых были выявлены достоверные различия ($p < 0{,}05$) амплитуд компонентов с положительной (кружок с плюсом), с отрицательной (кружок с минусом) полярностью и положения электродов, где не были выявлены достоверные различия (пустой кружок).

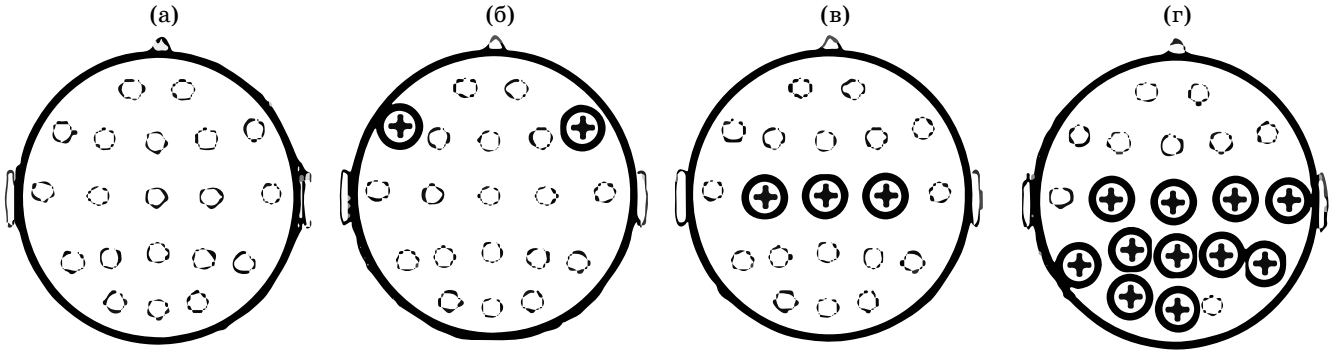

Рис. 3. Локализация достоверных различий амплитуд анализируемых компонентов вызванных потенциалов в задачах классификации объектов живой и неживой природы (стимулы — малые изображения) во всех областях мозга в разные интервалы времени (а — 110–180 мс, б — 180–250 мс, в — 250–350 мс, г — 350–700 мс). На рисунках обозначены положения электродов на голове испытуемых для отведений ЭЭГ, в которых были выявлены достоверные различия ($p < 0{,}05$) амплитуд компонентов с положительной (кружок с плюсом), с отрицательной (кружок с минусом) полярностью и положения электродов, где не были выявлены достоверные различия (пустой кружок).

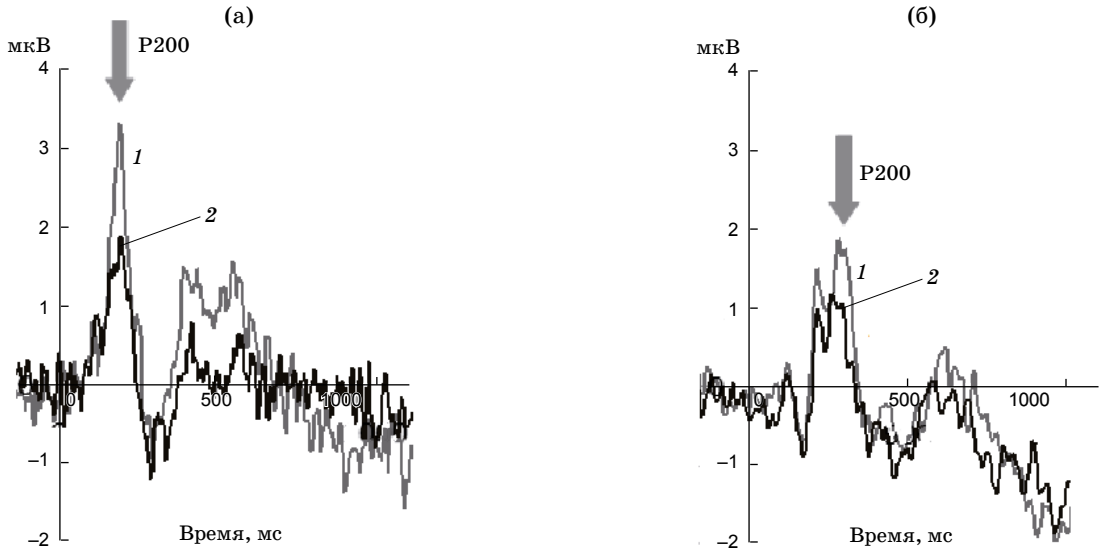

Рис. 4. Примеры усредненных вызванных потенциалов (отведение F7). Достоверные различия амплитуд вызванных потенциалов по семантическим признакам (живой (*1*)/неживой (*2*)) на большие (а) и малые изображения (б).

вызванных потенциалов на изображения объектов живой и неживой природы в лобных областях, в интервале 250–350 мс — в центральных областях мозга, а в интервале 350–700 мс — в центральных, теменных и височных областях.

В обеих сериях исследований различные по семантическим признакам (живой/неживой) тестовые изображения вызывают различные по амплитуде компоненты P200 во фронтальных отведениях F7, F8 (рис. 4). Компонент P200 при этом является маркером классификации по критерию «живой/неживой» во фронтальных областях.

ОБСУЖДЕНИЕ

В наших предыдущих исследованиях была изучена зависимость электрофизиологических маркеров работы механизма классификации в нейронных сетях головного мозга от степени детализации и выраженности обобщенного описания за счет предварительной пространственно-частотной фильтрации тестовых изображений. В данном исследовании представлены результаты исследования зависимости работы механизмов классификации от преобразования размеров изображений.

Инвариантность восприятия как фундаментальное свойство работы зрительной системы человека всегда зависит от противоположного свойства восприятия — конкретизации. Поэтому исследование инвариантности к размеру определяется целью исследования и условиями проведения исследований [4, 6]. Инвариантность восприятия к масштабным преобразованиям зависит от выбранного алфавита стимулов и тренировки. Показано,

что в случае алфавита, составленного из сильно различающихся изображений, их опознание возможно даже при значительном изменении таких параметров, как ориентация и размер [6].

Повышение порогов инвариантного восприятия фрагментированных контурных изображений наступает тогда, когда размер изображения очень мал (менее 1°) и приближается к размерам фовеолы, или очень велик (более 50°) и уходит далеко на периферию сетчатки [3, 8]. Алфавит изображений в работах [3, 8] был использован тот же, который использован в настоящей работе. При наблюдении изображений больших угловых размеров (более 50°) пороги инвариантного восприятия нарушаются вследствие искажений ретинотопического представления и ограничений прослеживающих движений глаз, а при малых угловых размерах (менее 1°) ограничения вносит шум дискретизации.

Благодаря проведенным ранее исследованиям [3, 8] подобраны следующие угловые размеры изображений: в первой серии исследований они составляли 3°, а во второй серии — 0,4°. Размеры больших изображений выходили за пределы фовеа, а изображения малых размеров соответствовали размерам фовеолиты (центральной части фовеолы). Острота зрения наблюдателей была равной 1, соответственно вероятность правильных ответов составила 95%, т.е. они правильно распознавали тестовые изображения. Таким образом, при размерах изображений, равных фовеолите, восприятие сохраняется, но пороги инвариантного восприятия повышаются, так как сказывается шум дискретизации изображения.

Хотя пороги восприятия изменяются, но правильное распознавание сохраняется даже при малых размерах стимулов, ограниченных разрешающей способностью и шумом дискретизации. Поэтому можно предположить, что амплитуды определенных компонентов вызванных потенциалов изображений объектов живой и неживой природы могут быть различными. Полученные результаты наглядно демонстрируют правильность данного утверждения. Действительно, в проведенных исследованиях были установлены локализации достоверных различий амплитуд вызванных потенциалов в нижнелобных отведениях. Именно здесь (отведения F7 и F8) и при больших, и при малых размерах стимулов есть различия реакций на объекты живой и неживой природы. На изображения больших размеров, превышающие размеры фовеа, различия этих компонентов выражены практических во всех лобных отведениях. Малая разница вызванных потенциалов на объекты небольших размеров выявляет различия в задаче категоризации, вероятно, лишь в самых «важных» для контроля работы мозга отведениях (лобные F7 и F8). Эти отведения соответствуют областям принятия решений [17, 18]. Компоненты вызванных потенциалов, по которым можно судить о классификации по критерию «живой/неживой» объект, приходящиеся на время около 200 мс, соответствуют моменту принятия решения.

Различия по критерию «живой/неживой» объект в более поздних компонентах вызванного потенциала (250–350 мс) в области центральной извилины говорят о передаче принятого решения в области мозга, обеспечивающие организацию двигательного процесса.

Более поздние компоненты вызванных потенциалов (350–700 мс) — период работы механизма контроля принятого решения [21].

Можно предположить, что во фронтальной области, которая обеспечивает принятие решений при выполнении задач классификации изображений, мозг использует результат анализа изображения на предыдущих уровнях (в задних отделах нижневисочной коры), который выделяет из него признаки, инвариантные к размеру изображения. Проведены измерения амплитуды вызванных потенциалов, согласно которым у хищников, не имеющих выраженной лобной области, инвариантное описание осуществляется в зоне Клэра–Бишопа — теменно-височно-затылочном участке коры [18], удаление которой у приматов приводит к нарушениям распознавания [19].

В нейронные сети лобной коры приходит информация, инвариантная к масштабным преобразованиям. Поэтому возникает возможность классификации изображений по семантическим признакам независимо от размера изображения.

Работа выполнена в рамках финансирования Программы фундаментальных научных исследований государственных академий на 2013–2020 гг. (ГП-14, раздел 63), Институт физиологии им. И.П. Павлова.

ЛИТЕРАТУРА

1. *Bohm D.* The special theory of relativity. NY.: Benjamin Inc., 1965 (*Бом Д.* Физика и восприятие, роль инвариантов в восприятии / в кн. Специальная теория относительности / М.: Мир, 1967. С. 239–281.)

2. *Глезер В.Д., Дудкин К.Н., Подвигин Н.Ф., Невская А.А., Праздникова Н.В.* Зрительное опознание и его нейрофизиологические механизмы. М.: Наука, 1975. 272 с.

3. *Шелепин Ю.Е., Чихман В.Н., Вахрамеева О.А., Пронин С.В., Фореман Н., Пэсмор П.* Инвариантность зрительного восприятия // Экспериментальная психология. 2008. Т. 1. № 1. С. 7–33.

4. *Бондарко В.М., Семенов Л.А., Солнушкин С.Д., Чихман В.Н.* Инвариантность к размеру при восприятии иллюзорных и фрагментированных контуров в зависимости от возраста наблюдателей // Физиология человека. 2008. Т. 36. № 6. С. 41–47.

5. Нейротехнологии / под ред. Шелепина Ю.Е., Чихмана В.Н. / Коллективная монография. СПб.: изд. ВВМ, 2018. 398 с.

6. *Глезер В.Д.* Зрение и мышление. Изд. 2, испр. и доп. СПб.: Наука, 1993. 284 с.

7. *Моисеенко Г.А., Шелепин Ю.Е., Хараузов А.К., Пронин С.В., Чихман В.Н., Вахрамеева О.А.* Классификация и распознавание изображений живой и неживой природы // Оптический журнал. 2015. Т. 82. № 10. С. 53–64.

8. *Вахрамеева О.А., Шелепин Ю.Е., Мезенцев А.Ю., Пронин С.В.* Изучение восприятия неполных контурных изображений различного размера // Российский физиологический журнал им. И.П. Сеченова. 2008. Т. 94. № 10. С. 1158–1170.

9. *Leek E.C., Roberts M., Oliver Z.J., Cristino F., Pegna A.J.* Early differential sensitivity of evoked-potentials to local and global shape during the perception of three-dimensional objects // Neuropsychologia. 2016. V. 89. P. 495–509.

10. *Dzhelyova M., Rossion B.* The effect of parametric stimulus size variation on individual face discrimination indexed by fast periodic visual stimulation // Neuroscience. 2014. V. 15. № 87. P. 1471–2202.

11. *Cecotti H.* Toward shift invariant detection of event-related potential is non-invasive brain-computer interface // Pattern Recognition Lett. 2015. V. 66. P. 127–134.

12. *Wiskott L.* How does our visual system achieve shift and size invariance? // In Problems in Systems Neuroscience / eds. by van Hemmen J.L. and Sejnowski T.J. Oxford: Oxford University Press, 2003.

13. *Баранов Р.П., Белоконь А.В.* Определение и приоритет признаков объектов на изображении в системах распознавания // Актуальные проблемы авиации и космонавтики. 2012. С. 328–329.

14. *Wiskott L., Sejnowski T.* Slow feature analysis: Unsupervised learning of invariances // Neural Computation. 2002. V. 14. P. 715–770.

15. *Lowe D.G.* Object recognition from local scale-invariant features // Proc. Internat. Conf. Computer Vision. 1999. P. 1150–1157.

16. *Шелепин Ю.Е., Фокин В.А., Хараузов А.К., Пронин С.В., Чихман В.Н.* Локализация центра принятия решения и восприятие формы зрительных стимулов // Доклады АН. 2009. Т. 429. № 6. С. 835–837.

17. *Шелепин Ю.Е.* Введение в нейроиконику. СПб.: изд. «Троицкий мост», 2017. 350 с.

18. *Шелепин Ю.Е.* О зависимости параметров вызванных потенциалов в стриарной коре кошки от размера изображения // Физиолог. журн. СССР. 1973. Т. 59. № 5. С. 688–695.

19. *Шелепин Ю.Е.* Локализация областей зрительной коры кошки, дающих инвариантный ответ при изменении размера изображения // Нейрофизиология. 1973. Т. 5. № 2. С. 115–121.

20. *Яковлев В.В.* Различия в описании зрительного образа на уровне заднетеменной и нижневисочной коры обезьян // Физиология. 1982. С. 754–757.

21. *Шелепин Ю.Е., Фокин В.А., Хараузов А.К., Фореман Н., Пронин С.В., Вахрамеева О.А., Чихман В.Н.* Нейроиконика — пространственно-временное картирование активных областей головного мозга при распознавании изображений // Оптический журнал. 2011. Т. 78. № 12.