

УДК 612.8; 528.854

Общность алгоритмов движений глаз, обеспечивающих распознавание жанровых сцен в текстах и в изображениях

© 2019 г. **Е. Ю. Шелепин***, аспирант; **О. В. Жукова***, канд. психол. наук;
С. В. Пронин*; **О. В. Защиринская****, доктор психол. наук;
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: yshelopin@yandex.ru

Поступила в редакцию 01.09.2019

DOI:10.17586/1023-5086-2019-86-11-79-89

Исследованы характеристики движений глаз в задачах распознавания текстов и комиксов с описаниями динамических жанровых сцен. Анализ взаимосвязи мысленно построенного наблюдателем смыслового пространства комиксов и смыслового пространства тестовых изображений, построенного исследователем, проведен с помощью измерений характеристик движений глаз, выполнявших роль своеобразного маркера, проецирующего на объект исследования субъективный алгоритм анализа контента. Выявлены общие черты, присущие распределению саккад и фиксаций взора в пространстве изображения при восприятии комиксов и текстовых изображений, вероятно, отражающие алгоритм выявления семантической, смысловой структуры изображений, общий для анализа разных способов передачи информации в мультимедиа. Этот алгоритм нарушается при различных заболеваниях головного мозга. Сделано предположение, что независимо от различия представления информации в текстах и в комиксах понимание обеспечивают общие механизмы построения внутреннего, воображаемого контента, изменяющегося во времени.

Ключевые слова: контент, распознавание сцен, движение глаз, чтение, текст, изображения сцен, мультимедиа.

Коды OCIS: 100.4996, 330.2210, 330.6110

ВВЕДЕНИЕ

Понимание работы механизмов «зрительного мозга» человека, осуществляющих оценку развития ситуации во взаимоотношениях участников жанровой сцены или боевой обстановки, поиск и положение основных и второстепенных участников представляют значительный интерес для обучения современных интеллектуальных систем. Проблема состоит в том, что количество информации, содержащейся даже в одном кадре современных камер высокого разрешения, может быть очень велико. В существующих искусственных распознающих системах реального времени анализ изображения разбивается на два канала. Первый канал осуществляет собственно круговой обзор и обеспечивает обнаружение объектов, целей, представляющих интерес. Второй канал обеспечивает наведение системы высокого разрешения для детального

анализа объекта и проводит классификацию, например, по критерию свой–чужой [1]. Аналогии современных инженерных решений многоканальной пространственно-частотной предварительной фильтрации имеются в зрительной системе пауков, кошачьих и приматов, осуществляющих и глобальную (пространственно низкочастотную) оценку, и изучение деталей (пространственно высокочастотных) объекта, выбранного в качестве цели [2]. Один из традиционных подходов к классификации изображений — использование согласованной фильтрации [3]. В зрительной системе человека есть системы обзора полусферы (периферия сетчатки) и детального анализа выделенного объекта (центральная часть сетчатки). Для наведения центральной части глаза (фовеа) на цель и понимания сцены в деталях служит система движения глаз. Понимание семантики в задачах

классификации отдельного объекта и в задачах понимания содержания жанровой сцены определяют принятие решений. Жанровые сцены можно описать и представить в виде изображений текстов (письменной речи), либо в виде серий рисунков, комиксов. В связи с этим представляет интерес анализ общности и различий в характере движений глаз, обеспечивающих восприятие при предъявлении этих двух групп изображений.

Целью настоящей работы явилось исследование характеристик перемещений взора наблюдателя, необходимых для понимания семантики при восприятии текста с описанием жанровых сцен и при восприятии комиксов с изображениями жанровых сцен.

МЕТОДИКА

Оборудование и процедура

Для проведения экспериментов использовался программно-аппаратный комплекс, состоящий из ноутбука *Asus* (разрешение экрана 1680/1050) и системы *iView XRed 250* (SMI, Германия) с частотой дискретизации в 250 Гц. Исследовались следующие параметры: количество и длительность саккад и фиксаций, время просмотра/прочтения, общий паттерн движений глаз.

Стимулы

В исследовании использовались по 5 вербальных и невербальных стимулов. Вербальные стимулы были представлены письменными текстами — жанровыми сценами русских народных сказок, изображенными черными буквами на сером фоне. Фрагменты текста в среднем содержали 700 знаков (15 строк), включая пробелы и знаки препинания. В качестве невербальных стимулов использовались комиксы Херлуфа Бидструпа — датского художника-карикатуриста (1912–1988 гг.). Каждый комикс содержал 8 кадров сюжета.

Движениями глаз управляют две системы (крупно- и мелкоклеточные, или магно- и парво-) зри-

тельного анализатора, передающие информацию преимущественно в диапазоне низких и высоких пространственных частот [4–6]. Поэтому все стимулы предварительно подвергались цифровой фильтрации с помощью свертки исходного изображения с DoG-функциями (DoG — *difference of Gaussians*) — вейвлетами, представляющими собой разность двух двумерных функций Гаусса [4–6]. Фильтрация проводилась в спектральной области путем умножения спектра исходного изображения на функцию, представляющую собой разность двух гауссоидов с различающимися в два раза стандартными отклонениями. Функция осесимметричная, профиль ее сечения описывается следующим выражением:

$$f(v) = \exp(-v^2/(2\sigma_1^2)) - \exp(-v^2/(2\sigma_2^2)),$$

$$\sigma_2 = 0,5\sigma_1, \quad (1)$$

где v — пространственная частота, σ_1 и σ_2 — стандартные отклонения гауссоидов.

Выбирая параметр σ_1 , можно перемещать максимум функции (1) по оси частот, выделяя тем самым из спектра различные частотные диапазоны. В данной работе для обработки изображений с комиксами использовались фильтры со следующими положениями максимума частотной характеристики: 3,75, 2,81, 2,11, 1,58, 1,19 цикл/град, т.е. отношение соседних частот соответствовало 0,75 (табл. 1).

Для обработки изображений с текстами использовался другой набор фильтров — 3,2, 2,56, 2,05, 1,64, 1,31 цикл/град, с отношением соседних частот, равным 0,8 (табл. 2).

Фильтрации текстов и комиксов выполнены разными наборами фильтров, так как спектры исходных изображений (комиксов и текстов) не совпадают. Тем не менее, в обоих случаях первый фильтр давал изображение, на котором видны все детали, последний — размытое, на котором сложно распознать объект (см. рис. 1).

Таблица 1. Размеры DoG-функции, использованной для фильтрации изображений с комиксами

Положение максимума частотной характеристики, цикл/град	Размер DoG-функции, пиксели	Угловой размер DoG-функции, град
3,75	12	0,24
2,81	16	0,32
2,11	20	0,40
1,58	32	0,64
1,19	44	0,88

Таблица 2. Размеры DoG-функции, использованной для фильтрации изображений текстов

Положение максимума частотной характеристики, цикл/град	Размер DoG-функции, пиксели	Угловой размер DoG-функции, град
3,2	14	0,28
2,56	18	0,36
2,05	20	0,40
1,64	26	0,52
1,31	34	0,68

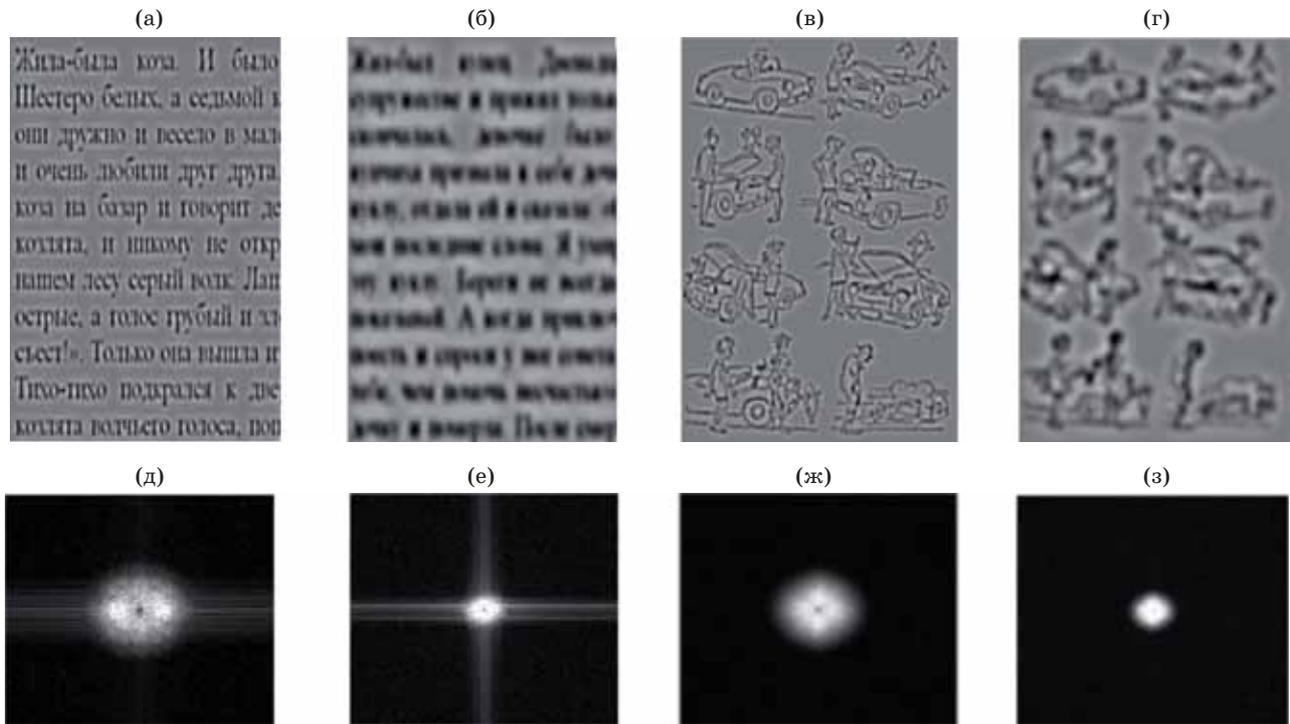


Рис. 1. Примеры тестовых изображений текстов (а, б) и комиксов (в, г), пропущенных через фильтры с различными пространственно-частотными полосами пропускания, и соответствующие им спектры (д–з).

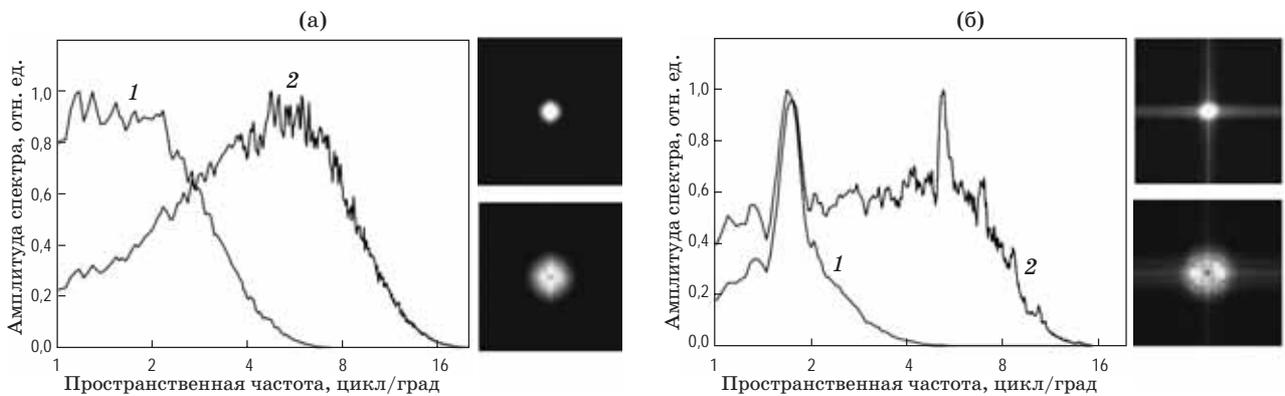


Рис. 2. Графики усредненных по разным ориентациям сечений спектров изображений комиксов (а), текстов (б) и их спектры. Графики получены путем усреднения 100 сечений двумерного пространственно-частотного спектра. Ориентации сечений охватывали угол 360° (с шагом в $3,6^\circ$). Усредненные сечения спектров низкочастотного изображения (максимум спектра на частоте 1,7 цикл/град) — 1, высокочастотного изображения (максимум спектра на частоте 5 цикл/град) — 2.

Размер букв оставался постоянным и в среднем составлял 5 мм ($0,52^\circ$) в ширину, 7 мм ($0,73^\circ$) в высоту, толщина ножки буквенного знака — 0,9 мм ($0,09^\circ$). Стимулы предъявлялись в случайном порядке. На рис. 2 представлены графики усредненных сечений спектров.

На рис. 2б видно, что в спектре текстов после высокочастотной фильтрации (кривая 2) есть два пика, соответствующие строкам и буквам, после

низкочастотной фильтрации высокочастотный пик, соответствующий буквам, исчезает (кривая 1). В спектре комиксов (рис. 2а) это соответствует деталям изображения, но не так наглядно.

Инструкция

Всем испытуемым давалась инструкция понять сюжет изображения. Время просмотра не было ограничено. После восприятия каждого стимула

испытуемым предлагалось заполнить анкету, направленную на изучение понимания увиденного, где они оценивали свое понимание стимула по 6-балльной шкале (0 — ничего не понял, 6 — понял отлично) и то, насколько интересными для них были предъявляемые тексты и изображения, а также предлагалось коротко изложить смысл увиденного. Процедура исследования проходила в индивидуальном порядке.

Статистическая значимость эффектов, связанных с параметрами движения глаз, сравнивалась с помощью дисперсионного анализа с повторными измерениями. Оценка статистической значимости (p) выполнялась с учетом поправки Грингауза–Гайссера при чтении текстов в связи с небольшим отклонением от сферичности. Также использовались t -критерии парных выборок и регрессионный анализ. Для устранения ложно позитивных ошибок при множественных сравнениях использована коррекция Бонферрони. Проведено преобразование данных (квадратный корень) для нормализации распределений полученных значений [7].

Испытуемые

Основную экспериментальную выборку составляли 25 испытуемых в возрасте 18–25 лет (студенты и сотрудники СПбГУ — 13 женщин, 12 мужчин). Все испытуемые являлись носителями русского языка, имели нормальное зрение и не страдали неврологическими заболеваниями. Каждый испытуемый подписывал письменное согласие на участие в исследовании.

Дополнительную выборку составили 60 испытуемых в возрасте от 15 до 19 лет — старшеклассники с разным уровнем интеллектуального развития. Были сформированы три подгруппы по уровню интеллектуального развития: нормально развивающиеся школьники (20 человек), старшеклассники с диагнозами «смешанные специфические расстройства психического развития» (F83 по международной классификации болезней МКБ-10, 20 человек) и старшеклассники с легкой умственной отсталостью (F70 по МКБ-10, 20 человек). Необходимо отметить, что старшеклассники участвовали в пилотной серии исследования особенностей восприятия и понимания жанровых сцен при нарушениях интеллектуального развития. Им предъявлялись только два стимула: текст с размером вейвлетного элемента $0,40^\circ$ (20 пикселей) и комикс с размером вейвлетного элемента $0,32^\circ$ (16 пикселей).

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Обработка изображений в основной экспериментальной выборке

Результаты зрительного анализа движений глаз — фиксаций и саккад — при восприятии комиксов показали, что глаза двигаются по стандартному для европейцев чтению текста — «Z-пути», т.е. слева направо и сверху вниз. Если комикс содержал неожиданное развитие сюжета, как правило, в конце повествования, то фиксаций на данных кадрах было больше, чем на других кадрах (рис. 3).

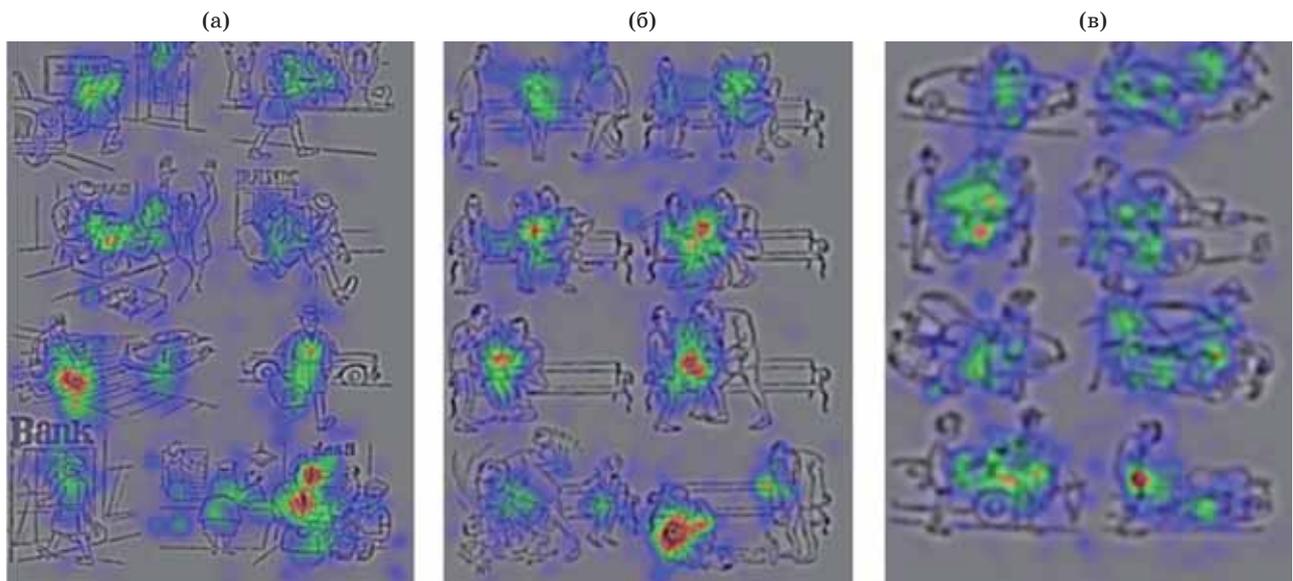


Рис. 3. Примеры тепловых карт распределения фиксаций при восприятии комиксов Херлуфа Бидструпа. «Ограбление» (размер вейвлетного элемента $0,24^\circ$) — (а), «Свидание» (размер вейвлетного элемента $0,32^\circ$) — (б), «Амур прочитался» (размер вейвлетного элемента $0,64^\circ$) — (в).

Подобная стратегия рассматривания комиксов характерна вплоть до последнего из примененных в работе уровней размытия изображения — с размером вейвлетного элемента $0,88^\circ$ (44 пиксела). При этих уровнях размытия сюжет жанровой сцены понять сложно, однако глобальная структура изображения определяла последовательность движений глаз — почти как при чтении. Даже максимальный уровень размытия позволял понимать комикс, а структура фиксаций сохранялась.

Тепловые карты распределения фиксаций при восприятии текстов показали, что движения глаз, так же как и при восприятии комиксов, осуществляются слева направо и сверху вниз, традиционно для испытуемых. Плотность распределения длительности фиксаций также неравномерна. Как в комиксах, так и текстах она максимальна в наиболее интересных для испытуемого областях.

При небольших уровнях размытия (с размером вейвлета от $0,28$ до $0,40^\circ$, или от 14 до 20 пикселей

соответственно) большая длительность фиксаций зафиксирована в верхней части изображения. В данном случае, вероятно, было достаточно нескольких первых предложений, чтобы выполнить инструкцию — понять смысл текста. С увеличением уровня размытия (размер вейвлета от $0,52$ до $0,68^\circ$, или от 26 до 34 пикселей соответственно) длительность и количество фиксаций уменьшаются, взор может достаточно долго фиксироваться в какой-то одной точке (рис. 4). Рассматривание сцены становится дезорганизованным. Такое распределение стратегий рассматривания текстов с различной степенью размытия было характерно для большинства испытуемых.

Так как в эксперименте применялись именно те угловые размеры, которые использует наблюдатель и при чтении текста, и при рассматривании комикса, в результатах наблюдения видны существенные различия. Совершенно очевидно, что при тех же параметрах наблюдения стандартных размеров букв текста и картинок на странице книги или в кадре

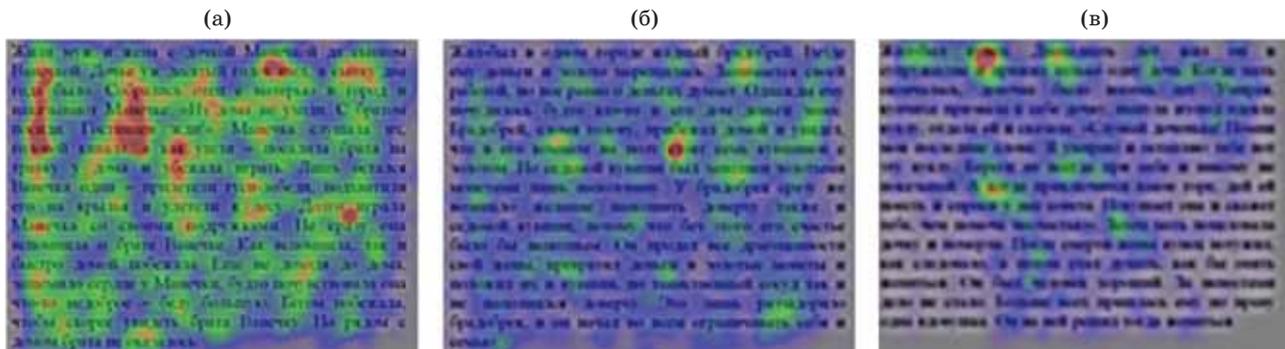


Рис. 4. Тепловые карты распределения фиксаций при восприятии текстов с размерами вейвлетного элемента $0,40$ (а), $0,52$ (б), $0,68^\circ$ (в). Данные получены в основной группе испытуемых (25 человек).

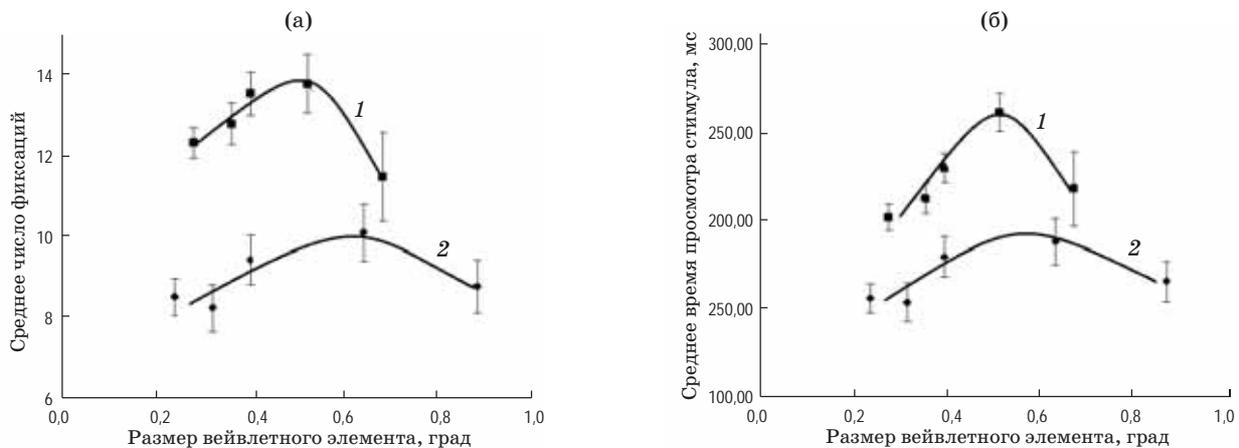


Рис. 5. Зависимости количества фиксаций (а) и времени просмотра стимулов (б) от размера вейвлетного элемента, 1 — тексты, 2 — комиксы. Данные получены в основной группе испытуемых (25 человек). Разброс представлен средней ошибкой. Применено преобразование данных (квадратный корень) для нормализации распределений полученных значений.

на экране дисплея компьютера угловые размеры элементов в комиксе больше (пространственная частота ниже) и он устойчивее к размытию. Данные по числу и времени фиксации в зависимости от пространственной частоты вейвлетного элемента разложения представлены на рис. 5.

Динамика кривых при восприятии и комиксов, и текстов похожа. При небольших уровнях размытия стимулов наблюдаются небольшое количество фиксации и короткая длительность просмотра. С увеличением сложности, начиная с размера вейвлетного элемента $0,40^\circ$ (20 пикселей) в случае комикса и $0,36^\circ$ (18 пикселей) в случае текста, происходит постепенное увеличение показателей. Пик значений приходится на стимулы с размерами вейвлетного элемента $0,64^\circ$ (32 пикселя) для комикса и $0,52^\circ$ (26 пикселей) для текста. Далее происходит спад всех параметров движения глаз.

По результатам проведенного дисперсионного анализа повторных измерений показано, что существуют статистически значимые различия в параметрах движения глаз по различным уровням фильтрации комиксов. Это справедливо в отношении количества фиксации ($F_{4,96} = 5,27, p < 0,05$), времени просмотра ($F_{4,96} = 5,22, p < 0,05$) и длины саккад ($F_{3,68} = 4,04, p < 0,05$). Апостериорные сравнения с поправкой Бонферрони показывают наличие статистически значимых отличий в основном между пиком ($0,64^\circ$, или 32 пикселя) и крайними точками кривой (размеры вейвлетного элемента $0,24$ и $0,32^\circ$) на уровне $p < 0,05$. Глобальный F -критерий показывает статистически значимые различия в параметрах движения глаз также по различным уровням фильтрации текстов. Но если по анализу времени просмотра ($F_{2,37} = 4,97, p < 0,05$) и длине саккад ($F_{4,96} = 27,37, p < 0,001$) результаты, с поправкой на сферичность Грингауза-Гайссера, статистически значимы, то по количеству фиксации — только на уровне тенденции ($F_{2,45} = 2,52, p = 0,094$). Апостериорные сравнения также показывают наличие статистически значимых отличий между пиком ($0,52^\circ$, или 26 пикселей соответственно) и крайними точками кривой на уровне $p < 0,05$.

В табл. 3 приведены результаты парного сравнения основных параметров движения глаз на одинаковом уровне размытия стимулов — при размере вейвлетного элемента $0,4^\circ$ (20 пикселей). Показано, что значения параметров движения глаз при восприятии текстов и комиксов статистически значимо отличаются. При этом количество фиксации больше при чтении текста. Это характерно и для времени «чтения», и для средней длины саккад.

Испытуемые не только осуществляли большее количество фиксации на текстах, но и лучше понимали семантику стимулов на небольших уровнях размытия: при размере вейвлетного элемента от $0,28$ до $0,4^\circ$ (от 14 до 20 пикселей соответственно) (рис. 6).

Далее происходит резкий спад субъективных оценок понимания смысла, достигая на максимальном уровне фильтрации — $0,68^\circ$ (34 пикселя) — достаточно низких показателей ($0,49 \pm 0,13$). Понимание комиксов ухудшается раньше, начиная со второго уровня фильтрации — размер вейвлетного элемента $0,4^\circ$ (20 пикселей), но ухудшается

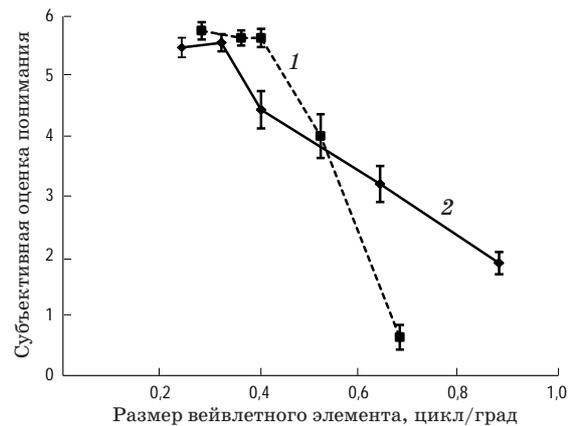


Рис. 6. Зависимости субъективной оценки понимания стимулов от размера вейвлетного элемента, 1 — тексты, 2 — комиксы. Данные получены в основной группе испытуемых (25 человек). Разброс представлен средней ошибкой. Применено преобразование данных (квадратный корень) для нормализации распределений полученных значений.

Таблица 3. Параметры движения глаз при чтении текста и комикса с размером вейвлетного элемента $0,4^\circ$ (20 пикселей)

Параметр	Комикс	Текст	p
Количество фиксации	$9,4 \pm 0,61$	$13,515 \pm 0,55$	$p < 0,001$
Время просмотра, мс	$179,61 \pm 11,53$	$229,67 \pm 8,21$	$p < 0,001$
Длина саккад	$13,26 \pm 0,14$	$14,02 \pm 0,17$	$p < 0,05$

Примечание. Данные получены в основной группе испытуемых (25 человек). Использовано преобразование данных (квадратный корень) для нормализации распределений полученных значений. Разброс представлен ошибкой средней. Сравнение средних значений осуществлялось с помощью парного t -критерия Стьюдента.

постепенно таким образом, что даже на максимальном уровне размытия размер вейвлета $0,88^\circ$ (44 пиксела) понимание остается на достаточно высоком уровне ($1,31 \pm 0,08$).

На рис. 7 представлены графики зависимостей основных параметров движений глаз при восприятии комиксов и чтении текстов.

Видно, что существует прямая зависимость количества фиксаций и времени просмотра жанровых сцен. Так, коэффициент детерминации при восприятии комиксов показывает (R квадрат = 0,93), что 93% вариабильности фиксаций обусловлено временем просмотра. Проверка значимости модели показала, что нулевую гипотезу об отсутствии взаимосвязи можно отвергнуть ($F = 1533$, $p < 0,001$). Коэффициент детерминации при чтении текстов ниже ($F = 0,79$). Так, только 79% вариабильности фиксаций зависит от времени просмотра ($F = 455$, $p < 0,001$). Связь прямая: чем больше количество фиксаций, тем дольше время просмотра. Можно предположить, что пока испытуемый может понять смысл представленной жанровой сцены, не смотря на уровень размытия, он продолжает выполнять предложенную инструкцию. Соответственно количество фиксаций и время просмотра стимула увеличивается. Когда размытие настолько велико, что понять сюжет комикса не представляется возможным, увеличение степени размытия, наоборот, снижает время просмотра

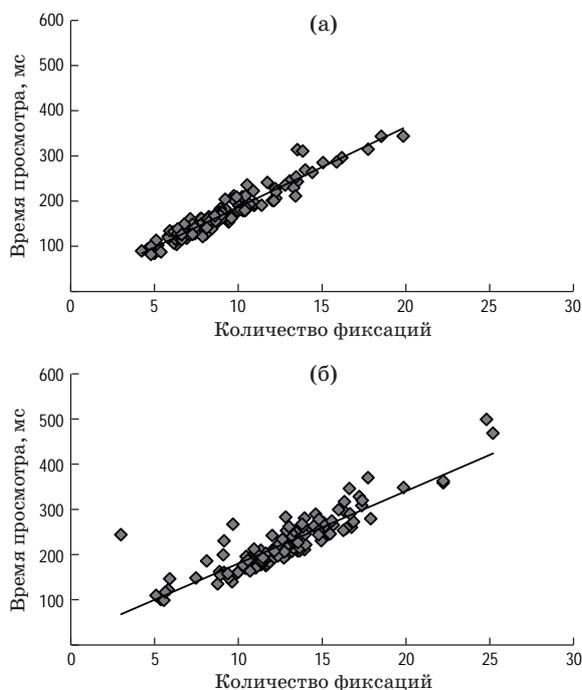


Рис. 7. Зависимости времени просмотра комиксов (а) и текстов (б) от количества фиксаций. Графики построены на основе генеральной совокупности данных (25 испытуемых и 5 стимулов).

и количество фиксаций. Испытуемые стараются сфокусировать взор только на последнем, ключевом для понимания сюжета кадре комикса.

При анализе зависимостей количества фиксаций и времени просмотра от длины саккад получаем более слабую связь между откликом и регрессорами (предикторами). При этом коэффициенты корреляции принимают отрицательные значения. То есть чем больше количество фиксаций и дольше время просмотра, тем меньше длина саккад.

Обработка изображений в дополнительной экспериментальной выборке

В дополнительной выборке получены схожие с основной выборкой результаты. Так, количество фиксаций у всех старшеклассников при чтении текста больше, чем при просмотре комикса. Так же, во всех группах получены одинаковые зависимости между изучаемыми параметрами движения глаз (прямая зависимость количества фиксаций и времени просмотра стимулов и обратная между количеством фиксаций (времени просмотра) и длиной саккад).

Статистически значимых зависимостей между субъективными оценками понимания сюжета у испытуемых и характеристиками движения глаз не выявлено ($p > 0,05$) ни в одной из изучаемых групп.

При просмотре комикса с размером вейвлетного элемента $0,32^\circ$ (16 пикселей) однофакторный дисперсионный анализ не выявил статистически значимых отличий в параметрах движения глаз между тремя группами школьников ($p > 0,05$). Отличия на уровне тенденции показаны только при анализе времени просмотра стимула ($F_{2, 59} = 3,149$, $p = 0,05$). Апостериорные сравнения с поправкой Бонферони показали, что данный уровень значимости достигнут при сравнении нормально развивающихся школьников со школьниками с «легкой умственной отсталостью» ($p < 0,05$). В то же время на основе общего анализа паттерна движений глаз при просмотре комикса показано, что существуют межгрупповые отличия в том, на что дети с разным интеллектуальным развитием обращают внимание.

При чтении текстов глобальный F -критерий дисперсионного анализа показал статистически значимые отличия между тремя группами школьников в основном при анализе времени просмотра стимула и длины саккад (табл. 4).

Отличия во времени просмотра, вероятно, связаны с отличиями в длительности фиксаций. Дети с «легкой умственной отсталостью» распознают тексты хуже в сравнении с другими экспериментальными группами, что находит отражение в характеристиках движения глаз (уменьшении количества фиксаций и времени просмотра стимула). Апостериорные сравнения с поправкой

Таблица 4. Параметры движения глаз при просмотре текста с размером вейвлетного элемента 0,40° (20 пикселей) школьниками с разным уровнем интеллектуального развития

Параметры движения глаз	Нормально развивающиеся школьники (20)	Школьники со «смешанными специфическими расстройством психического развития» (20)	Школьники с «легкой умственной отсталостью» (20)	<i>p</i>
Количество фиксаций	12,20 ± 0,53	13,89 ± 1,12	12,08 ± 0,94	<i>p</i> > 0,05
Время просмотра, мс	254,27 ± 11,96	326,56 ± 24,88	264,19 ± 17,77	<i>p</i> < 0,05
Длина саккад	14,17 ± 0,17	13,37 ± 0,17	13,36 ± 0,23	<i>p</i> < 0,05

Примечание. Данные представлены по группе испытуемых (60 чел.). Использовано преобразование данных (квадратный корень) для нормализации распределений полученных значений. Разброс представлен ошибкой средней. *p* — уровень значимости различий в данных между тремя группами испытуемых.

Бонфферони показали наличие статистически значимых отличий в основном между нормально развивающимися школьниками и учениками с «легкой умственной отсталостью» ($p < 0,05$). Таким образом, можно сказать, что существуют статистически значимые отличия при восприятии стимулов между различными группами школьников, но в большей степени это касается чтения текстов, а не просмотра комиксов.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Анализ организации движения глаз — это своеобразный физиологический «проективный тест», так как он позволяет по траектории их движения судить о некоторых «алгоритмах» работы мозга человека. В отличие от субъективных психологических проективных тестов методы анализа движения глаз работают как «объективные проективные тесты».

С помощью метода регистрации движений глаз исследованы особенности распознавания семантики текстов и последовательностей изображений, подвергнутых цифровой фильтрации. Большинство исследований ранее проводилось при одновременном предъявлении и текста, и изображения [8–9], причем информация, представленная разными способами, дополняла друг друга. Траекторию движений глаз определяет стоящая перед испытуемым цель, в психофизиологических исследованиях — это инструкция [8–10].

В данном исследовании предъявляемый комикс не содержал элементов текста. Хотя в оригинале автор давал краткое, из нескольких слов, его название. Испытуемые должны были самостоятельно определить последовательность восприятия комикса, чтобы понять семантику каждого кадра и всего сюжета в целом. Зрительный анализ движения глаз как по отдельным испытуемым, так и по всем группам, показал, что восприятие комиксов осуществляется слева направо и сверху вниз, т.е. по классическому для письменного тек-

ста «Z-пути». Данный результат вполне ожидаем, так как, чтобы выполнить инструкцию — понять изображенное на рисунке, испытуемые должны были, по сути, «прочитать» комикс. Был выявлен схожий паттерн распределения фиксаций при восприятии тестовых изображений, вероятно отсележивающий «смысловую композицию, грамматическую структуру» изображений обоих типов. Можно предположить, что при восприятии комикса задействованы те же механизмы, что и при чтении письменного текста. При построении комикса этот же стиль движений глаз и руки происходил, вероятно, и у художника при проецировании воображаемой композиции на создаваемый рисунок. Чтение у современного человека столь значимо в жизни и занимает столь много времени, что и при просмотре изображения лица наблюдается чаще всего траектория, напоминающая укороченное чтение. Собеседник как бы читает ключевые элементы лица — выражение от глаза к глазу и на губы. Левое–правое направление от глаза к глазу зависит от того, какой текст обычно читает испытуемый — русский, английский или иврит, арабский [11].

Согласно полученным экспериментальным данным, количество фиксаций, а также общее время просмотра и средняя длина саккад больше при чтении текстов, а не при рассматривании комиксов [4–6]. Просмотр рисунков — более экономичный и менее затратный тип зрительной работы. При чтении текста человек совершает большое количество фиксаций и саккад. Это наиболее затратный вид работы глазодвигательной системы, как было показано в специальных исследованиях Айно Ламминпи с соавторами [12]. Необходимость саккад объясняется свойствами сетчатки: основную зрительную информацию человек получает, обрабатывая сигналы из фовеа и ее центра — фовеолиты — области максимально высокой остроты зрения. В ее пределах острота зрения не меняется. При фиксации взора на точке глаз колеблется либо в пределах $\pm 12'$ — это радиус фовеолиты, либо в пределах $\pm 2,5'$ — это радиус рецептивного поля [13].

Следовательно, при чтении требуется перемещать глаза так, чтобы в фовеолиту попадали все новые и новые слова.

При восприятии же комикса в ряде исследований показано, что взор наблюдателя сосредоточен преимущественно в центральной области сцены [13]. Цель подобного поведения, с одной стороны, связана с расчетом центра тяжести фигур, а с другой стороны — с меньшими затратами. Картина воспринимается как бы в общем, но при этом взор легко может быть перемещен в любую точку периферии, и человеку необходимо значительно меньшее количество саккад для получения общей информации, чем при работе с текстом.

Таким образом, можно сказать, что и письменный текст, и комикс содержат как локальные, так и глобальные признаки, но соотношение этих признаков в стимулах разное, что и объясняет значительную разницу в параметрах движений глаз. Именно поэтому контент страницы в интернете содержит все большее количество рисунков.

Принято считать, что запуск саккад обеспечивает магносистема, а остановку взора в конце саккады — парвосистема, которая считывает информацию во время фиксации взора. Магноклеточная система обеспечивает восприятие преимущественно низкочастотной информации, парвоклеточная — преимущественно высокочастотной информации в зрительном анализаторе. Для получения стимулов, избирательно воздействующих на магно- или парво- системы, применяется цифровая фильтрация изображений в области высоких или низких пространственных частот. Пространственно-частотная фильтрация позволяет удалить или, наоборот, подчеркнуть локальные признаки, содержащиеся в изображениях, и избирательно исследовать вклад каждой из систем в обеспечение организации движения глаз [2, 6]. Поэтому и в наших исследованиях использовалась пространственная высоко- и низкочастотная фильтрация тестовых изображений. Фильтрация осуществлялась для двух типов тестовых изображений, несущих наблюдателю информацию о жанровых сценах. Первый тип — это тексты, требующие для распознавания (прочтения) упорядоченного просмотра — «сканирования», — комиксы — ряд небольших изображений, связанных по смыслу друг с другом, который надо понять. В обоих случаях имеется упорядоченная последовательность элементов сцены, определяемая в тексте грамматикой, а в комиксах композицией. Для понимания смысла тестовых изображений (тексты или комиксы) требуется распознать, по крайней мере, большую часть фрагментов, из которых они построены, затем понять смысл изображения [2, 4]. За время одной фиксации взора человек может в фовеа захватить одновременно от 4 до 8 букв текста, соответствующих слову или смысловому

элементу текста, а в комиксах — это законченный элемент изображения сцены: голова, глаз, фигура, стул, скамейка, цветок, букет, автомобиль и проч. [2, 6]. Периферия поля зрения захватывает соседние области менее четко.

В условиях размытия стимулов качество изображения существенно ухудшается. Испытуемым все сложнее понять сюжет изображения. Вероятно, именно с этим связано увеличение количества фиксаций и времени просмотра.

Важно, что для обработки комиксов и текстов были использованы разные наборы фильтров. В случае комикса отношение соседних частот соответствовало 0,75, а в случае текста — 0,8. Можно предположить, что при использовании одинаковых для разного типа стимулов фильтров динамика кривых была бы практически идентичной. Полученные данные согласуются с результатами, полученными в исследованиях особенностей движения глаз при чтении текстов с разным уровнем размытия. Так, в исследованиях [4–6] показано, что пиковые значения количества и длительности фиксаций приходятся на размер вейвлетного элемента, равный 32 пикселям. Далее происходит уменьшение количества фиксаций и времени просмотра. При анализе зависимостей количества фиксаций и времени просмотра от длины саккад найдена хоть и слабая, но достоверная связь. При этом коэффициенты корреляции принимают отрицательные значения. То есть, чем больше количество фиксаций и дольше время просмотра, тем меньше длина саккад. Испытуемые как бы «останавливались» взором на сложном участке и достаточно долго его рассматривали, пытаясь понять, что изображено на картинке. Возврат к сложным участкам изображения и длительность остановки увеличивались при патологии.

Интерес представляет тот факт, что при чтении размытого текста, в сравнении с размытым изображением комикса, происходит более резкое снижение показателей количества фиксаций и длительности просмотра стимула. Об этом также говорят и субъективные оценки испытуемых. Вероятно, при сильных уровнях размытия недостаточно информации для эффективного управления работой глазодвигательной системы в процессе сегментации текста. Но если при чтении текста сильное размытие имеет критические последствия — наблюдатель перестает выполнять инструкцию, то при восприятии комиксов даже на сильных уровнях фильтрации он продолжает работать. В таком случае изображение комикса превращается в причудливо разбросанные по кадру пятна, и возникшее абстрактное изображение служит стимулом для свободных ассоциаций [2].

У испытуемых с нарушением интеллектуального развития наблюдается ухудшение распознавания текстов в сравнении с группой «норма»,

но не комиксов, которое проявляется в уменьшении количества фиксаций и времени просмотра стимула. Можно предположить, что дети с «легкой умственной отсталостью» уставали быстрее при выполнении задания, что нашло отражение в характеристиках движений глаз. При просмотре комикса статистических значимых отличий в числе саккад и длительности фиксаций между группами не выявлено. Но дети с разным интеллектуальным развитием по-разному используют ключевые элементы изображений в пространстве контента. Так, у учеников с «легкой умственной отсталостью» при просмотре комиксов и текстов отмечена большая длительность фиксаций на элементах, раскрывающих эмоциональную функцию изображения. У нормально развивающихся школьников и школьников со «смешанными специфическими расстройствами психического развития» длительные фиксации отмечены на словах, передающих динамику событий и их участников. Нормально развивающиеся подростки при просмотре текстов и комиксов ориентируются больше на логический анализ событий, а учащиеся с «легкой умственной отсталостью» лучше понимают текст и комикс при его эмоциональном осмыслении.

Общность реакций движений глаз при просмотре комиксов и текстов связаны как с опытом чтения, так и с наличием «грамматики» и тезауруса как в тексте, так и в комиксе. Тем самым были получены экспериментальные результаты, подтверждающие справедливость гипотезы о наличии общих мозговых механизмов для построения вербальной и зрительной картины мира. Иными словами, наш результат дает экспериментальное подтверждение гипотезе, согласно которой речь как абстрактное мышление строится на основе конкретного зрительного мышления [14]. В зрительном конкретном описании сцены существуют глобальное целостное описание структуры наблюдаемого изображения (гештальт) и локальное детальное описание, построенное на выделении отдельных признаков. Важность глобального описания в том, что оно строится как в отдельном канале, передающем низкочастотные составляющие двумерного пространственно-частотного спектра, так и на основе низкочастотной огибающей, построенной на основе вычленения отдельных признаков, выделяемых высокочастотным каналом [2, 4–6]. Представленные результаты подтверждают эти утверждения. Они, как мы видим, показывают общность движения глаз и при сохранении деталей, и после низкочастотной фильтрации, когда эти мелкие детали отсутствуют. Траектории движений и остановок глаз соответствуют строкам и смыслу текста и композиции и смыслу рисунка. Важно подчеркнуть, что параметры движения глаз при чтении являются косвенным, но

объективным средством, отражающим аналитический способ мышления и речевые процессы. Представляет также интерес исследование особенностей распознавания жанровых сцен у испытуемых с нарушением интеллектуального развития.

Поэтому анализ движения глаз и может работать как проективный тест, обеспечивающий выявление алгоритмов управления глазодвигательной системой в условиях, когда изображение невозможно охватить и распознать в момент одной фиксации взора. Заметим, что термин «фиксация взора» весьма условный, так как в этот момент сохраняются высокочастотные микродвижения глаз [12], характер которых отражает отношение наблюдателя к воспринимаемой информации. Прослеживающие движения глаз связаны с композицией сцены и с грамматикой текста после фильтрации — удаления высоких пространственных частот и сохранения низких частот в спектре изображения. Это говорит о том, что в низкочастотной части изображений может находиться существенная структурная информация, содержащаяся в контенте, своеобразная «квазиграмматика», сохраняющаяся даже в условиях потери многих деталей и важных признаков в области высоких пространственных частот. Тем самым мы обращаем внимание читателя на общие принципы, которыми руководствуется мозг в пространстве контента и при проецировании их на пространство наблюдаемого изображения.

При исследовании особенностей восприятия и распознавания текстов и комиксов, подвергнутых высокочастотной и низкочастотной фильтрации, с помощью метода регистрации движений глаз было показано, что последовательность изображений, связанных по сюжету, может выступать в роли своеобразного «текста». При наличии инструкции «понять сюжет повествования» взор движется по классическому для письменного текста «Z-пути», что говорит о том, что испытуемые не просто смотрят, но «читают» невербальный «текст».

ЗАКЛЮЧЕНИЕ

На основе проведенного анализа можно предположить, что общность семантики (содержания) независимо от разных средств передачи информации (текст, рисунок) вызывает у читателя помимо явных различий и некоторые важные общие механизмы представления жанровых сцен, обеспечивающие общность воображаемого контента в нейронных сетях головного мозга человека. Полученные данные переключаются со структурными методами решения задач распознавания образов [15, 16]. Представляет интерес сравнение обработки зрительной системой изображений сюжетов (комиксов) и описания этих динамических сюжетов в речи, в тексте с обработкой в слуховой систе-

ме речевых сигналов [17–21]. Ранее было показано, что характеристики движения глаз изменяются при изменении соотношения прилагательных и глаголов в тексте [5].

Независимо от различия представления информации в текстах и в комиксах понимание контента обеспечивают общие механизмы построения внутреннего, воображаемого контента, изменяющегося во времени. Описание изображений, их распознавание и мысленное название распознанных объектов и целостных сцен может происходить в расширенных «речевых зонах Брока и Вернике», вероятно, еще и при участии симметричных им

зон в правом полушарии головного мозга, активирующихся при возникновении зрительно-лексических ассоциаций [19–23].

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. Исследование старшеклассников с разным уровнем интеллектуального развития выполнено в рамках проекта Российского научного фонда № 14-18-02135 «Психофизиологические и нейролингвистические аспекты процесса распознавания вербальных и невербальных паттернов».

ЛИТЕРАТУРА

1. *Московченко Л.В., Тушиков В.А., Лысенко Э.Л.* Концепция построения перспективных систем оптико-электронной пассивной локации для надводных кораблей // Национальная оборона. 2012. № 9. С. 60–61.
2. *Шелепин Ю.Е.* Введение в нейрокогнику. СПб.: Троицкий мост, 2017. 350 с.
3. *Красильников Н.Н.* Цифровая обработка 2D- и 3D-изображений. СПб.: БХВ-Петербург, 2011. 608 с. ISBN 978-5-9775-0700-4.
4. *Ламминия А.М., Пронин С.В., Шелепин Ю.Е.* Пространственно-частотная фильтрация текста для локального и глобального анализа // Оптический журнал. 2018. Т. 85. № 8. С. 38–45.
5. *Ламминия А.М., Малахова Е.Ю.* О восприятии движения реальной и воображаемой сцены // Нейротехнологии / под ред. Шелепина Ю.Е., Чихмана В.Н. СПб., 2018. С. 257–268. ISBN 978-5-9651-1198-5.
6. *Ламминия А.М., Вахрамеева О.А., Райт Д.В. и др.* Влияние вейвлетной фильтрации изображений текстов на характеристики движений глаз при чтении // Сенсорные системы. 2013. Т. 27. № 1. С. 3–9.
7. *Шипунов А.Б., Балдин Е.М., Волкова П.А. и др.* Наглядная статистика. Используем R! М.: ДМК Пресс, 2014. 298 с.
8. *Cohn N.* The visual language of comics: Introduction to the structure and cognition of sequential images. London, UK: Bloomsbury, 2013. 240 p.
9. *Hegarty M., Just M.A.* Constructing mental models of machines from text and diagrams // J. Memory and Language. 1993. V. 32. № 6. P. 717–742.
10. *Rayner K., Rotello C.M., Stewart A.J., et al.* Integrating text and pictorial information: Eye movements when looking at print advertisements // J. Experimental Psychology. 2001. V. 7. № 3. P. 219–226.
11. *Барабанищев В.А.* Экспрессии лица и их восприятие. М.: Институт психологии РАН, 2012. 341 с.
12. *Ламминия А.М., Шелепин Ю.Е., Бауэр С.М., Вахрамеева О.А., Райт Д.В.* Влияние оптических свойств текста на работу каналов зрительной системы человека // Сб. тезисов докладов Всерос. молодежной конф. «Нейробиология интегративных функций мозга». СПб., 2013. С. 43.
13. *Гассовский Л.Н., Никольская Н.А.* Движения глаз в процессе непрерывной фиксации точки // Тр. ГОИ. 1941. Т. 15. С. 112–120.
14. *Глезер В.Д.* Зрение и мышление. Л.: Наука, 1993.
15. *Фу К.* Структурные методы в распознавании образов. М.: Мир, 1977. 320 с.
16. *Тушиков В.А., Павлова В.А., Крюков С.Н. и др.* Лингвистические методы в задачах распознавания изображений // Известия ЮФУ (технические науки). 2015. Т. 1. № 162. С. 142–151.
17. *Shannon R.V., Zeng F.G., Kamath V., et al.* Speech recognition with primarily temporal cues // Science. 1995. V. 270. P. 303–304.
18. *Lublinskaya V.V., Ross J., Ogorodnikova E.A.* Auditory perception and processing of amplitude modulation in speech-like signals: Legacy of the Chistovich–Kozhevnikov group // NATO Science Series. Ser. 1: Life and Behavioural Sciences (Dynamics of Speech Production and Perception). 2006. V. 374. P. 87–103.
19. *Шелепин К.Ю., Пронин С.В., Шелепин Ю.Е.* Распознавание фрагментированных изображений и возникновение «инсайта» // Оптический журнал. 2015. Т. 82. № 10. С. 70–78.
20. *Шелепин К.Ю., Труфанов Г.Е., Фокин В.А. и др.* Активность нейронных сетей головного мозга человека до, во время и после инсайта при распознавании изображений // Нейротехнологии / под ред. Шелепина Ю.Е., Чихмана В.Н. СПб., 2018. С. 207–243. ISBN 978-5-9651-1198-5.
21. *Shelepin K.Y., Trufanov G.E., Fokin V.A., et al.* Digital visualization of the activity of neural networks of the human brain before, during, and after insight when images are being recognized // JOT. 2018. V. 85. № 8. P. 468–475.
22. *Ardila A., Bernal B., Rosselli M.* Language and visual perception associations: Meta-analytic connectivity modeling of Brodmann area 37 // Behavioral Neurology. 2015. V. 2015. P. 2–14.
23. *Ardila A., Bernal B., Rosselli M.* How localized are language Brain areas? A review of Brodmann areas involvement in oral language // Archives of Clinical Neuropsychology. 2016. № 31. P. 112–122.