

УДК 159.9.07

О.В. Левашов

Аттракторы внимания на ранней (preattentive) стадии зрительного анализа сцен и новая модель узнавания

Аннотация:

Описаны результаты экспериментальной проверки гипотезы Ноттона и Старка о «жестких» траекториях осмотра на более сложных изображениях пространственных сцен (репродукции картин, реклама, фото реальных сцен). Гипотеза Ноттона и Старка не подтвердилась. Однако обнаружено, что для зрительного внимания человека существует определенная иерархия визуальных «аттракторов», куда внимание и, соответственно, движения глаз, направляются в первую очередь. Наиболее «сильными» аттракторами оказались участки изображений, включающие сложные необычные формы, например, отражения в воде, искаженные реальные предметы (как на картинах сюрреалистов). Траектории осмотра изображения проходят через аттракторы в последовательности, зависящей от соотношения их перцептивных «весов» и поэтому могут быть разными.

Ключевые слова: модели зрения, раннее зрение (preattentive vision), анализ зрительных сцен, траектории осмотра изображений, М- и Р-системы в зрении, аттракторы внимания

Об авторе: Левашов Олег В., кандидат биологических наук, Научный центр неврологии (Москва), научный сотрудник Лаборатории возрастной физиологии мозга; эл. почта: olevashov@gmail.com.

Введение

При рассматривании реальных изображений глаз человека движется по сложной траектории, выполняя быстрые скачки, которые чередуются с фиксациями, когда глаз останавливается, и происходит считывание информации в окрестности этой точки фиксации [9; 1]. При этом выделяют так называемую «довнимательную» (preattentive) стадию, когда скачки программируются автоматически (бессознательно), в зависимости от визуального контекста [2]. Осознание увиденного наступает позднее. И тогда характер движения глаз задается сознательно, в соответствии с задачей (примеры см. [9]).

Изучая механизм управления движений глаз в «довнимательный» период, Ноттон и Старк [8] предложили собственную методику для того, чтобы «развернуть» процесс узнавания во времени. Для этого они использовали стимулы в виде очень крупных и неярких изображений. Как оказалось, первые 2-3 скачка в таких условиях зависят только от визуального контекста изображения. Они обнаружили, что для конкретного человека траектория первых скачков глаз повторяется каждый раз при показе одного и того же изображения. Для другого изображения эта траектория иная, но тоже в целом повторяющаяся. Такие повторяющиеся траектории они назвали «путями сканирования» (scan-path). На основе полученных результатов авторы предположили, что на изображениях есть какие-то ключевые признаки, которые зрительный мозг моментально распознает, а в процессе скачков и фиксаций эти признаки «собираются» вместе и сопоставляются с эталонными цепочками признаков (с «кольцами признаков»). Однако авторам не удалось понять главное – как мозг за столь короткое время (200-250 мс, латентное время скачка) – выделяет главные «точки интереса» на изображении и направляет туда движения глаз.

Попытку ответить на этот вопрос предприняли Н. Завалишин и И. Мучник [3]. Они ввели понятие «функции информативности» изображения, в максимумах или минимумах которой стремится останавливаться глаз. Для простых линейных изображений типа цифр и углов они смогли математически описать функцию информативности. В дальнейшем они получили удовлетворительное экспериментальное подтверждение своей модели. Однако

открытым остался вопрос – как может выглядеть функция информативности для реальных многоградационных изображений пространственных сцен.

В данной статье описаны результаты экспериментальной проверки гипотезы Ноттона и Старка о «жестких» траекториях осмотра на сложных изображениях пространственных сцен (репродукции картин, реклама, фото реальных сцен). Гипотеза Ноттона и Старка не подтвердилась, однако было обнаружено, что для зрительного внимания человека существует система визуальных «аттракторов», куда внимание и, соответственно, движения глаз, направляются в первую очередь. Наиболее сильными аттракторами оказались участки изображений, включающие сложные необычные формы, например, отражения в воде, искаженные реальные предметы (как на картинах сюрреалистов). Первые полученные результаты настоящего исследования были представлены на Европейской конференции по зрительному восприятию [11].

Метод

Испытуемые. Участниками экспериментов были 52 человека разных профессий, почти все с высшим образованием, с нормальным или скорректированным до нормального зрением в возрасте от 19 до 75 лет.

Стимулы. Использовали 9 наборов изображений (Из), по 5 Из в каждой экспериментальной серии, – цветные Из реальных сцен, интерьера, картины известных художников (реалистов и сюрреалистов) и визуальную рекламу. Этими наборами были: 1. Метеоры (Греция). 2. Бассейн. 3. Реклама. 4. 3-D Реклама (показывающая расположение предметов в пространстве). 5. «Невнятная» (плохая с точки зрения понимания) реклама. 6. Третьяковка (известные картины). 7. Картины С. Дали, 8. Картины Р. Магритта и 9. Картины М. Эрнста. Размеры и положение стимулов на экране было нормализовано по величине и положению на экране, но все же варьировали по величине в небольших пределах – от 20° до 24°. Точка фиксации вспыхивала в центре предполагаемого изображения на 1 сек, но исчезала перед экспозицией тестового Из.

Аппаратура. Стимулы предъявлялись на экране монитора с помощью специальной компьютерной программы, позволяющей варьировать длительность экспозиции и порядок предъявления стимулов в серии [5].

Процедура. Испытуемый сидел перед экраном, на котором в случайном порядке по очереди экспонировались пять Из одного набора. Каждый раз после показа стимула в том же месте экрана на 300-600 мс экспонировалась «постэкспозиционная матрица» такого же размера, как и стимул, имеющая вид поля из 9 цветных прямоугольников, в каждом из которых помещалось 9 матриц из мелких цифр (размер цифр 20-25' и расстояние между ними 30-40') от 1 до 9. В каждой такой матрице все 9 цифр были одинаковы (см. рисунок 1). Первая серия из 5 Из экспонировалась на 300-400 мс (в зависимости от быстроты зрительного восприятия Ис), что позволяло определить локализацию первой фиксации глаз после начала экспозиции Из. Во второй серии показывали те же Из, но в другом порядке, а время экспозиции составляло 600-800 мс, в третьей экспозиция составляла 900-1200 мс. Время экспозиции подбиралось для каждого испытуемого на этапе тренировки.

Задача. Испытуемый должен был назвать экспериментатору (или набрать сам на клавиатуре) цвет квадранта и цифру, которую он успевал увидеть. В силу малого размера цифр испытуемый видел только те цифры, которые попадали в область его фовеального зрения. Таким образом, точность определения местоположения очередной фиксации глаза составляла 2-3°. После всех трех серий можно было последовательно определить локализацию первых фиксаций глаз и получить траекторию смещения зрительного внимания по данному изображению. Каждый испытуемый выполнял, по крайней мере, 4 серии, на 4 разных наборах изображений. Каждая серия по времени занимала не более 5 мин.

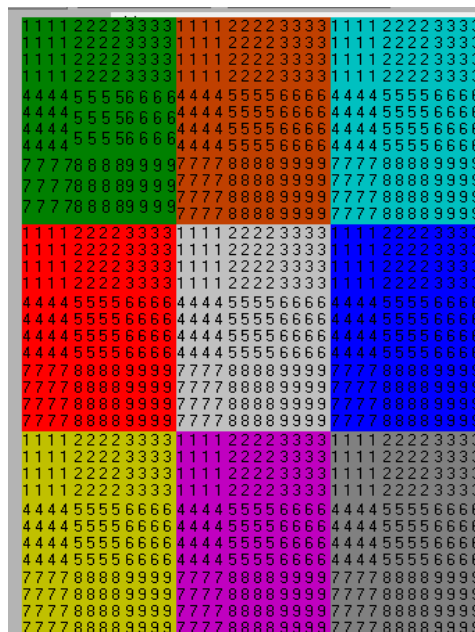


Рисунок 1. Постэкспозиционная матрица для отчета испытуемых

Результаты и обсуждение

Все Ис легко справлялись со своей задачей после предварительной тренировки. Она включала 5-7 минут знакомства с цветами матрицы и с самой процедурой, а также показ тестовой серии (это был первый набор «Греция, Метеоры»).

Некоторые Ис участвовали только в одной сессии из 4 серий с разными наборами изображений в силу своей занятости, но многие участвовали во множестве сессий, в разные дни (это были, в основном, коллеги). Общее число наблюдений и измерений за несколько лет проведения эксперимента составило порядка 5-7 тысяч.

Оценивали: 1. Характер участков изображений, в которых наблюдались фиксации внимания. 2. Если эти участки совпадали с характерными частями предметов на изображениях (например, если это были части тела людей, контуры зданий, крупные текстуры, предметы интерьера и т.п.), то они считались «точками внимания» или зрительными «аттракторами» (ЗА). Как оказалось, очень редко внимание Ис фиксировалось на незначимых фрагментах. 3. Подсчитывалось количество ЗА, зафиксированных испытуемым за время всей серии. 4. Подсчитывалась «длина» траектории сканирования каждого Из (в условных единицах, единицей считалась матрица размером 4x4 знаков, т.е. маленькая матрица одинаковых цифр).

Основные результаты

1. Мы не наблюдали повторяющихся «траекторий осмотра» (scan-pathes), декларированных в работе [8].

2. Зрительная система произвольно выбирает одни и те же характерные аттракторы, но траектории осмотра Из могут быть разные, даже при повторном показе того же самого изображения тому же испытуемому. Главное – это наличие определенных «популярных» (наиболее привлекательных) участков изображений, которые оказались общими практически для всех испытуемых независимо от пола и возраста. Именно поэтому мы и назвали эти участки сцен «зрительными аттракторами» (ЗА).

3. Анализ количества фиксаций внимания на аттракторах позволил составить примерную «иерархию» аттракторов по степени их важности для зрительного внимания и узнавания.

Наиболее важными оказались следующие ЗА (в порядке предпочтения):

1. Искаженные предметы и необычные объекты, напоминающие по форме знакомые объекты (например, предметы на картинах С. Дали и М. Эрнста).

2. Отражения предметов в воде или в других отражающих средах, особенно в средах сложной формы (см. рисунки 2 и 3).

3. Лица людей, фигуры человека или животных, части тела.
4. Надписи (в рекламе).
5. Крупные текстуры (например, на рисунке 2)

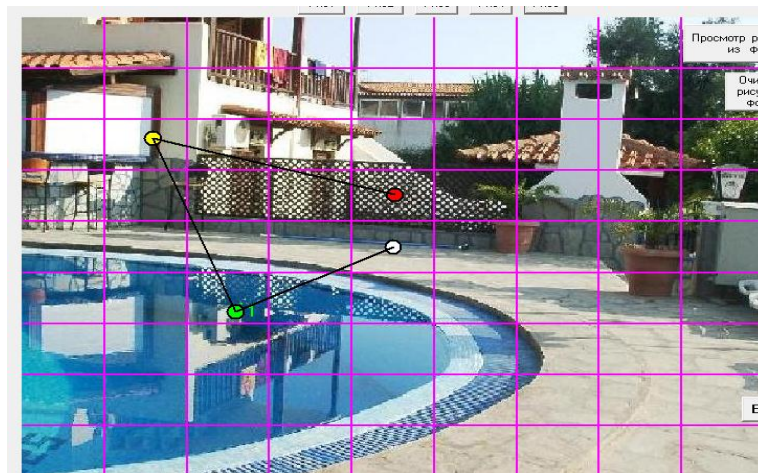


Рисунок 2. Отражения в воде архитектурных объектов – сильный «аттрактор». На всех приведенных примерах белым цветом отмечена точка фиксации как начальная точка траектории осмотра



Рисунок 3. Сложное отражение в трехмерной текучей среде – наиболее сильный аттрактор (вторая и третья фиксации внизу рекламы). Обращает на себя внимание, что испытуемый не обратил внимания на значимый по мысли авторов рекламы объект – торговую марку HEAD (справа внизу).

Стратегии осмотра у разных испытуемых

По характеру распределения зрительного внимания всех испытуемых можно приблизительно разбить на две группы – условно назовем их «аналитики» и «визуалы». Визуалы сразу могут оценить общую трехмерную конфигурацию сцены, их больше привлекают формы и предметы, чем надписи. Их внимание привлекают признаки по всему Из, так что их траектории осмотра Из протяженнее, чем у «аналитиков» (см. рисунок 4).

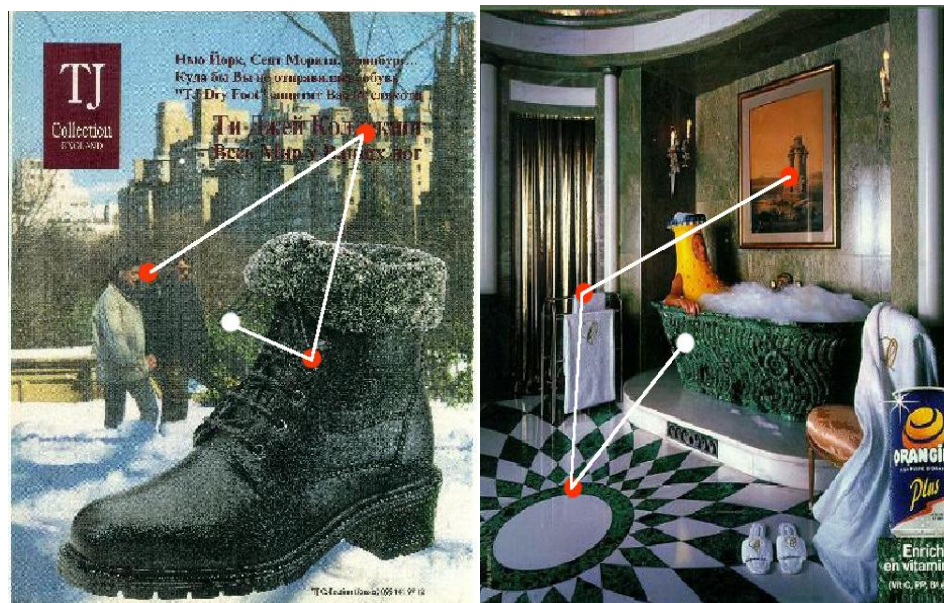


Рисунок 4. Типичные траектории смещения внимания у «визуалов». Видно, что испытуемые улавливают виртуальную трехмерную структуру сцены, смещая внимание по всей сцене

Аналитики обычно не могут оценивать общую трехмерную конфигурацию сцены, их внимание ограничено центральной частью Из, но лучше «схватывают» надписи (см. рисунок 5).

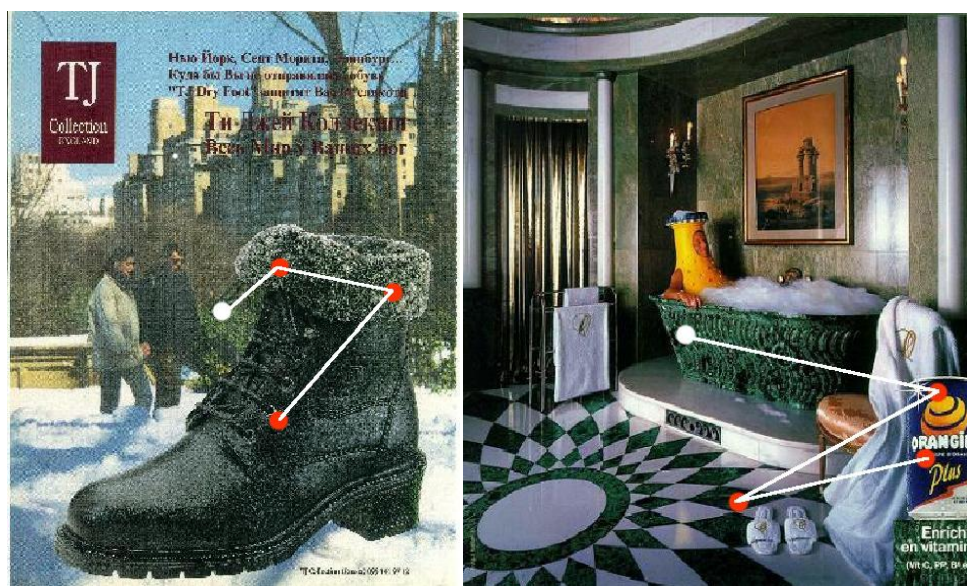


Рисунок 5. Типичные траектории смещения внимания у «аналитиков». Заметно, что зрительное внимание испытуемых ограничено частью всей сцены

В единичных случаях отмечались аномальные траектории осмотра (всего 2 Ис), когда отчеты о фиксации совпадали с незначимыми местами (например, пустая стена интерьера, небо), и в то же время эти Ис пропускали важные аттракторы. Один из них обследовался на предмет болезни Паркинсона, другой принимал в этот период по рецепту врача транквилизаторы. В одном случае мы наблюдали «зависание» в окрестности центральной точки фиксации с отсутствием смещения внимания. Этот неожиданный результат объяснил сам испытуемый – опытный художник: «А мне и так вся картина видна, мне не нужно переводить взгляд и внимание на детали».

Некоторые рекламные изображения с эротическими мотивами позволили выявить определенные гендерные различия в траекториях осмотра.

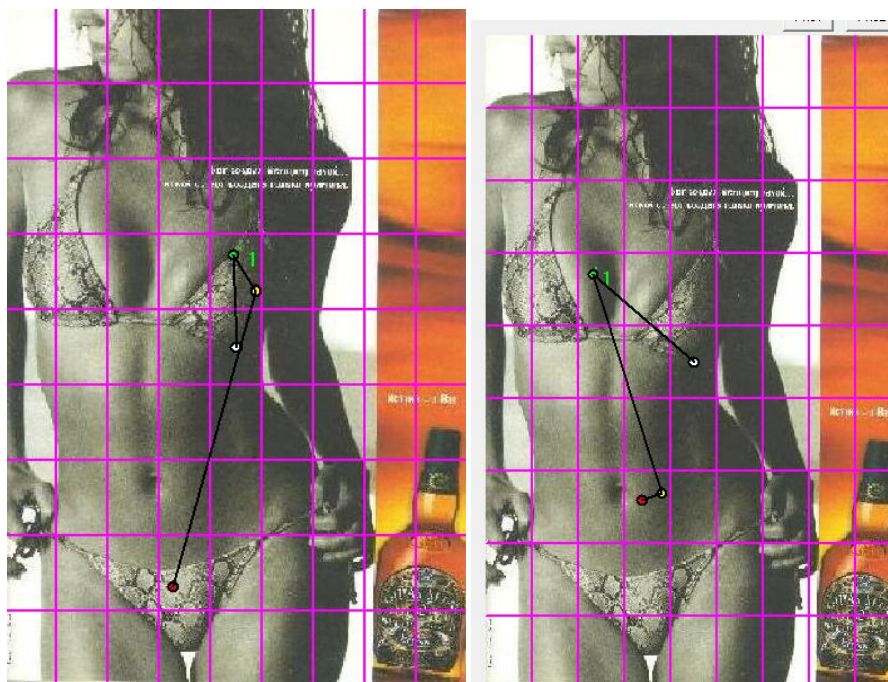


Рисунок 6. Траектории осмотра рекламы виски с добавленным рекламистами эротическим объектом (характер распределения внимания типичен для испытуемых мужского пола)

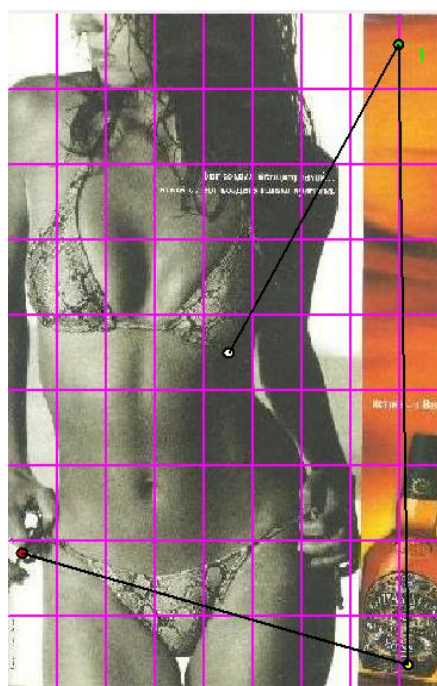


Рисунок 7. Траектории осмотра рекламы виски с эротическим объектом (типично для испытуемых женского пола)

Длина траекторий осмотра и число замеченных аттракторов

На рисунке 8 показана гистограмма распределения длины траектории осмотра и количество просмотренных за это время аттракторов для 12 испытуемых (каждый столбик гистограммы – это данные одного Ис). Видно, что среди этих Ис пятеро (столбики слева)

имеют небольшую по сравнению с остальными длину траектории осмотра и что эти пятеро «схватывают» в среднем в два раза меньше аттракторов, чем остальная группа Ис.

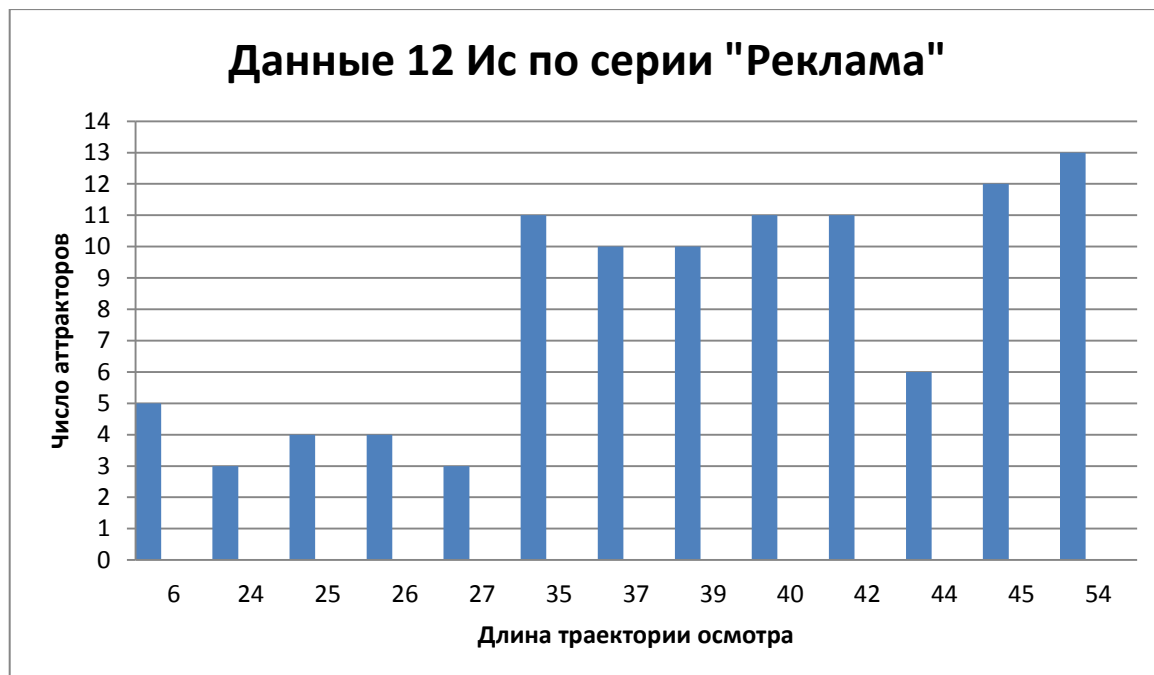


Рисунок 8. Различия в длине траектории осмотра рекламных изображений и в количестве замеченных «аттракторов» (сумма найденных аттракторов и длин траекторий по 5 изображениям в серии)

На рисунке 9 показано аналогичное распределение для 15 Ис, рассматривавших другие 5 сложных изображения (эти изображения были сложнее в том смысле, что там были представлены более сложные трехмерные сцены). Участвовали все 12 Ис из предыдущего эксперимента и еще трое других.

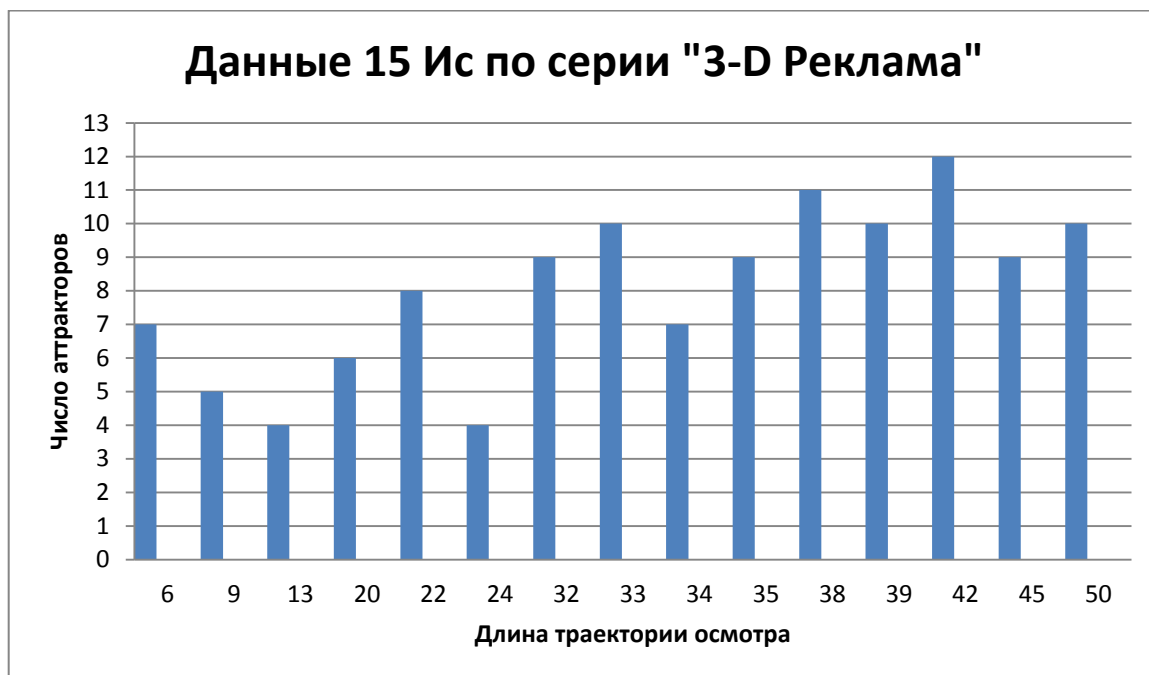


Рисунок 9. Различия в длине траектории осмотра рекламных изображений и в количестве замеченных «аттракторов» (сумма найденных аттракторов и длин траекторий по 5 изображениям в серии)

Видно, что шестеро первых Ис на рисунке 9, как и на рисунке 8, в среднем «схватывают» примерно в два раза меньше аттракторов, чем остальные Ис.

Опрос и анализ характера деятельности этой группы Ис показал, что группа «визуалов» включает больше лиц творческих профессий (музыканты, дизайнеры и художники), а также людей, занимающихся спортом.

Статистические оценки

Первая группа (набор Реклама):

Среднее значение количества найденных ЗА: для «аналитиков» (первые 5 Ис слева на рисунке 8) – 3,8, для «визуалов» (остальные 7 Ис) – 9,9. Различие достоверно на уровне $p < 0,01$ (критерий Стьюдента).

Средняя длина траектории осмотра: для аналитиков 21,6 условных единиц, для визуалов – 36,2. Различие достоверно на уровне $p < 0,01$.

Вторая группа (набор Трехмерная реклама):

Среднее значение количества найденных ЗА: для «аналитиков» (первые 6 Ис слева на рисунке 9) – 5,6, для «визуалов» (остальные 9 Ис) – 9,7. Различие достоверно на уровне $p < 0,05$.
Средняя длина траектории осмотра: для аналитиков 15,6 условных единиц, для визуалов – 39. Различие достоверно на уровне $p < 0,01$.

Общее обсуждение и модель параллельного зрительного узнавания

Полученные нами результаты могут говорить о том, что на самом раннем этапе зрительного анализа не происходит ничего такого, о чем обычно говорят при построении моделей зрительного узнавания, а именно, – нет выделения «локальных признаков», построения «описаний» и сравнения этих описаний с «эталоном», хранящимся в долговременной памяти. Уже в первые 300 мс анализируется вся сцена и выбирается наиболее важный объект (аттрактор). Никакая последовательная «сборка» найденных с помощью нейронных детекторов типа рецептивных полей Хьюбеля и Визеля не может быть реализована за столь малое время при наличии на сцене многих объектов и текстур.

По данным многих исследователей, и прежде всего Е. П. Кок (одной из наиболее талантливых учениц А. Р. Лурии), левое полушарие при зрительном узнавании работает последовательно, анализируя части целого, а правое работает параллельно, в широком поле внимания. Кроме этого правое полушарие владеет механизмами детального описания конкретных изображений, а левое способно, наоборот, к упрощенному, графическому, схематическому описанию, что отражается в рисунках больных с соответствующими поражениями мозга [4].

Эти и другие данные позволили автору в свое время предложить модель взаимодействия полушарий коры мозга в процессе узнавания [7]. Основными ее пунктами были следующие.

1. При показе нового изображения вычисляется некая функция информативности, позволяющая выбрать несколько наиболее привлекательных для зрения участков неизвестного изображения. Однако как построить такую функцию для сложных цветных сцен было совершенно непонятно, по крайней мере, автору это не удалось.

2. В выделенном участке с помощью детекторов типа рецептивных полей Хьюбеля и Визеля анализируется данный фрагмент изображения, составляется его «образ» и сравнивается с образами частей реальных предметов.

3. Постулировано, что в процессе обучения каждый опознанный фрагмент получает ассоциативную связь с набором конкретных следов, сформированных в правом полушарии при распознавании реальных предметов, видимых в разных ракурсах. Поэтому опознанный фрагмент активирует похожий след, и он визуализируется на «сенсорном экране» в правом полушарии.

4. Этот образ бессознательно воспринимается как реальный, после чего начинается его проверка за счет движений глаз в определенные информативные точки сцены.

Новые данные, полученные в настоящей работе, позволяют пересмотреть описанную выше модель. Прежде всего, они позволяют избавиться от загадочного понятия функции

информативности, наводят на мысль о быстрой параллельной обработке пространственной сцены уже на стадии бессознательного (preattentive) анализа, а также указывают на то, что стратегии движения внимания по изображениям могут варьировать в зависимости от сенсорного типа испытуемых.

Необходимо также учитывать полученные в 90-е гг. дополнительные физиологические данные о наличии в зрительной системе двух каналов передачи и обработки информации, и, соответственно, двух нейронных систем – магноцеллюлярной системы (М-системы) и парвоцеллюлярной системы (Р-системы), которые морфологически начинаются в сетчатке и образуют так называемый «дорзальный» и, соответственно, «вентральный» пути. Первый путь соединен с теменной зоной коры, второй – с височной (см. обзоры: [2; 6]).

М-система отвечает за стимулы в широком поле зрения, включая далекую периферию, содержит на уровне сетчатки клетки, не кодирующие цвет, но срабатывающие на движение и на включение-выключение, т.е. работающие как «он-офф»-детекторы. Ранее, в 80-х гг., такие клетки обозначали термином transient (или У-клетки). Часто М-систему называют системой «ГДЕ», поскольку она связана с функциями пространственной ориентации, определения расстояния и тайминга. Принято считать, что система «ГДЕ» обслуживает функции пространственного зрения и манипулирования предметами в пространстве.

Р-система связана с цветокодирующими нейронами сетчатки (прежнее название – sustained или Х-клетки) и обслуживает узкий (по сравнению со всем полем зрения) фовеальный участок поля зрения. Р-систему часто обозначают как систему «ЧТО», поскольку она отвечает за узнавание объектов.

Важно отметить, что система «ГДЕ» срабатывает раньше системы «ЧТО» и передача в ней сигналов происходит значительно быстрее, чем в системе «ЧТО».

Новая модель анализа зрительных сцен

Приходится допустить, что в зрительной системе уже с первых миллисекунд показа нового изображения (или, после скачка глаз на новое место в реальной обстановке) происходит одновременный анализ всей сцены (или, по крайней мере, значительной ее части по центру фиксации) с помощью большого количества параллельных «процессоров», выделяющих в исходном изображении «экологически» (или «поведенчески») значимые формы. Среди этих процессоров (можно условно называть их «нейронными эталонами») существуют определенные иерархические отношения, а именно, при обучении в процесс зрительного опыта все они приобретают разный «вес», в зависимости от поведенческой значимости выделяемого объекта. Каждый такой объект на изображении сцены приобретает статус **зрительного аттрактора** внимания с определенным весом. В результате сравнения всех активированных эталонов по правилу «побеждает сильнейший» программируется первый скачок в соответствующее место.

Можно предположить, что на самом раннем этапе узнавания М-система первая передает в кору предварительно обработанный низкочастотный, черно-белый «образ» («профиль активности») широкого участка сцены с центром в текущей точке фиксации. Этот «профиль активности» из теменной коры передается в височную, в систему «ЧТО», и может служить своего рода пусковым для активизации целого набора ранее сформированных при обучении нейронных эталонов конкретных предметов. Приходящий затем по Р-системе в височную кору детализированный «профиль активности» способствует дальнейшей сортировке и выбору наиболее близкого нейронного эталона по схеме «побеждает сильнейший». Именно этот эталон актуализуется в сознании (визуализируется), позволяя индивиду последовательно проверять данную «гипотезу» с помощью целенаправленных движений глаз.

Возникает естественный вопрос – каким образом такая схема параллельной обработки может быть реализована в виде нейронных структур?

Для ответа на этот вопрос можно сослаться на современные нейронные вычислительные технологии типа структур «глубокого обучения» (deep learning). Эти структуры являются, по сути, модернизацией нейронных сетей «хопфилдовского типа» [7]. Эта технология позволяет за счет специальных алгоритмов коррекции весов связей между нейронными слоями при обучении (так называемый алгоритм «обратного распространения

ошибки») за разумное время сформировать ограниченный набор «нейронных эталонов», способных работать параллельно и выделять «свои» ключевые образы. В качестве гипотезы можно даже провести аналогию между ритмической активностью коры, проявляющейся как ритмы ЭЭГ, и принципом обратного распространения ошибки, поскольку для хорошего формирования эталонов подобная обучающаяся нейронная система должна циклически «прогонять» стимулы, раз за разом, сверяя активность на выходе с эталонной активностью.

Нейрофизиологические данные о наличии в высших отделах зрительной коры множества зон с ретинотопической проекцией также говорят в пользу возможности такой параллельной обработки.

В описываемой здесь модели именно система «ЧТО» (височная кора) включает весь комплекс нейронных структур с «глубоким обучением».

Заключение. Как данная модель объясняет полученные экспериментальные данные

1. Наши данные показывают, что наиболее «привлекательными» участками сцены являются «странные объекты» и «искаженные формы». В самом деле, с экологической точки зрения необычные объекты требуют быстрого анализа и немедленной реакции.

2. Полученные данные о различиях между двумя группами испытуемых – «аналитиками» и «визуалами» – могут быть объяснены наличием более мощной системы «глубокого обучения» в височной коре правого полушария у «визуалов». В то же время более мощная и быстрая М- система в теменной коре левого полушария у «аналитиков» может хорошо обслуживать узнавание букв, цифр и слов при чтении, а также несложные математические расчеты.

Библиографический список:

1. Белопольский В.И. Взор человека. М.: ИПРАН, 2007.
2. Величковский Б.М. Когнитивная наука. М.: Смысл, 2006.
3. Завалишин Н.В., Мучник И.Б. Модели зрительного восприятия и алгоритмы анализа изображений. М.: Наука, 1974.
4. Кок Е.П. Зрительные агнозии. М.: Медицина, 1967.
5. Левашов О.В. Способ определения траектории смещения произвольного зрительного внимания на изображении. Патент на изобретение № 2545428 от 24.02.2015.
6. Левашов О.В. Функциональная асимметрия магно- и парвоцеллюлярной (М и П систем) (фазической и тонической) при локальных поражениях мозга и при дислексии: нейробиологический подход // Асимметрия. 2009. Т. 3. № 2.
7. Левашов О.В. Вычислительные модели сенсорных систем. М.: ВИНТИ, 1989.
8. Ноттон Д., Старк Л. Движения глаз и зрительное восприятие // Восприятие. Механизмы и модели. М., 1974. С. 226-240.
9. Ярбус А.Л. Роль движений глаз в процессе зрения. М.: Наука, 1965.
10. Levashov O. et.al. Role of the right and the left hemispheres in object recognition: A model // Perception. 1996. Vol. 25. P. 95.
11. Levashov O., Roumyantseva R. Visual attractors for eye movements in early vision // Perception. 2006. Vol. 35. P. 86.

Levashov O.V. Attractors of attention at an early (preattentive) stage of visual scene analysis and a new model of learning

Describes the results of experimental verification of the hypothesis of Notton and stark on “hard” trajectories of inspection for more complex images of spatial scenes (paintings reproductions, advertising, photo real scenes). The hypothesis of Notton and stark was not confirmed. However, we found that visual attention of the person there is a hierarchy of visual “attractors”, where the attention and, consequently, eye movements are directed in the first place. Most “strong” attractors turned out to be areas of images that include complex shapes are unusual, for example, reflections in water, distorted real objects (as in the paintings of the Surrealists). The trajectory of the inspection image

pass through the attractors in sequence, depending on the ratio of their perceptual "weights" and can therefore be different.

Keywords: models of vision, early vision (preattentive vision), analysis of visual scenes, trajectories of image viewing, M - and P-systems in vision, attractors of attention