

**Распознавание типовых действий пользователя при взаимодействии
с цифровым продуктом на основе анализа движения взгляда**

**Recognition of typical user actions when interacting with
a digital product based on gaze movement analysis**

Т. Ю. Посинковский, А. С. Голунова

Омский государственный технический университет, Омск, Российская Федерация

T. Y. Posinkovskiy, A. S. Golunova

Omsk State Technical University, Omsk, Russian Federation

Аннотация. В статье описывается методика определения действий пользователя в графическом пользовательском интерфейсе. Рассматривается эффективность применения различных вариантов интерполяции и фильтрации полученных данных для решения задачи распознавания типовых действий пользователя. Проводится анализ принципов определения фиксаций и саккад с учетом переходных процессов, определяются подходы к выбору базовых признаков наличия паттернов в фиксации. По результатам серии экспериментов делается расчет основных параметров и паттерна типового действия «Чтение текста», анализируются эффективность и полученные результаты.

Ключевые слова: анализ движения глаз, графический пользовательский интерфейс, интерполяция, фильтрация данных

Abstract. The article describes a methodology for determining user actions in a graphical user interface. It examines the effectiveness of applying various interpolation and filtering techniques to the collected data for recognizing typical user actions. The analysis covers the principles of identifying fixations and saccades, taking into account transient processes, and defines approaches for selecting basic features that indicate the presence of patterns in fixations. Based on a series of experiments, the key parameters and patterns of the typical 'Text Reading' action are calculated, and the effectiveness and obtained results are analyzed.

Keywords: eye movement analysis, graphical user interface, interpolation, data filtering

Введение

В настоящее время графические пользовательские интерфейсы цифровых продуктов являются основным и самым популярным методом человеко-машинного взаимодействия. Дизайн-системы унификации таких интерфейсов (такие как, например, Material Design от Google или Apple Human Interface) делают взаимодействие пользователей с цифровым продуктом привычным и удобным, состоящим из относительно небольшого количества повторяющихся типовых действий. Определение и последующая классификация таких действий позволяют структурировать процесс взаимодействия, выявить и изучить возникающие при этом сложности, анализировать пользовательский опыт и проводить дальнейшую оптимизацию интерфейсов.

Одним из наиболее эффективных методов изучения такого взаимодействия является айтрекинг. Этот метод позволяет регистрировать движения глаз пользователя, фиксируя точки внимания, последовательность просмотра элементов интерфейса и длительность фиксаций. Преимущество айтрекинга заключается в его высокой точности и объективности: он позволяет получить данные о том, как пользователь визуально воспринимает интерфейс, не полагаясь на субъективные отчеты. Фиксации (периоды относительной неподвижности взгляда) и саккады (быстрые скачкообразные движения), являются ключевыми индикаторами визуального внимания и когнитивных процессов. Однако исходные данные, получаемые от устройств этого типа, часто содержат шум, выбросы и неравномерный шаг квантования, что затрудняет их непосредственное использование для анализа [1].

Целью статьи является разработка и подтверждение на практике основных элементов методики обработки и анализа данных движения глаз для возможности определения типовых действий пользователей в графическом интерфейсе с помощью математического и статистического анализа характеристик фиксаций и саккад.

Теория

Разделим типовые действия, выполняемые пользователями в графических интерфейсах, на составные части следующим образом:

– Поиск элемента интерфейса. Пользователи перемещаются между элементами интерфейса (меню, вкладки, кнопки) для поиска нужной информации или функций.

– Фокусировка на элементе. Пользователи выбирают объекты (чекбоксы, радиокнопки, элементы списка) для выполнения действий или настройки параметров.

– Ввод данных в элемент. Пользователи вводят информацию в текстовые поля, формы или другие элементы интерфейса.

– Принятие решения. Пользователь выбирает действие (например, нажать кнопку или выбрать пункт меню).

– Ожидание реакции системы. Пользователь ожидает ответа от системы (например, загрузки страницы или выполнения действия).

– Просмотр контента. Пользователи просматривают контент (текст, изображения, видео).

– Реакция на ошибку или уведомление. Пользователь сталкивается с ошибкой или получает уведомление.

Разрабатываемая методика должна иметь возможность анализируя данные движения глаз, достоверно классифицировать такие действия и определять их временные характеристики. Исходя из вышесказанного, выделим в ней следующие этапы:

1. Получение дискретных данных движения глаз пользователя.
2. Устранение неравномерности квантования.
3. Фильтрация исходного сигнала, удаление шумов.
4. Разделение исходных данных на периоды фиксаций и саккад.
5. Вычисление основных статистических характеристик саккад и фиксаций.
6. Достоверное определение по рассчитанным значениям принадлежности каждой фиксации к «паттернам» типовых действий пользователя.
7. Группировка близлежащих фиксаций, относящихся к одному «паттерну», в конечную, определение временных характеристик его начала и окончания.

Рассмотрим этапы более подробно.

Исходные («сырые») данные, полученные непосредственно из айтрекинговых систем, представляют собой временные ряды координат взгляда, последовательность измерений (t_i, x_i, y_i) , где t_i – временные отметки, а x_i и y_i – координаты взгляда в пространстве. Такие ряды обязательно содержат пропуски или неравномерные интервалы между измерениями, что связано с ограничениями аппаратуры, задержками в передаче данных или физиологическими особенностями, такими как моргание или временная потеря сигнала. Если шаг времени $\Delta t_i = t_{i+1} - t_i$ непостоянен, это приводит к появлению неприемлемых ошибок при вычислении скорости и ускорения движения глаз, искажению статистических показателей. Интерполяция данных временных рядов является *критически важным* этапом предварительной обработки «сырых» данных [2]. В рамках проведения экспериментов для целей выбора лучшей был произведен расчет метрик точности трех основных видов интерполяции: линейной, кубической (РСНП) и сплайн-интерполяции.

Шум в исходных данных может быть вызван как микродвижениями глаз, так и аппаратными помехами. Фиксации характеризуются низкочастотными компонентами сигнала (обычно ниже 2 Гц). Саккады в своем составе имеют частотные компоненты до 20–30 Гц. Тремор глаз и аппаратные помехи – высокочастотные компоненты (выше 100 Гц и выше 30 Гц соответственно). Отсутствие фильтрации шума временных рядов координат взгляда может приводить к ложному обнаружению саккад вследствие формирования высокочастотным шумом ложных пиков скорости, которые будут ошибочно интерпретированы как саккады. Так же возможно и искажение фиксаций: шум вызывает колебания координат взгляда, что значительно понижает точность их идентификации, что особенно критично, так как фиксации являются ключевым элементом анализа определения «паттернов» типовых действий пользователя. В рамках проведения экспериментов по полученным данным был произведен расчет метрик точности трех методов фильтрации: скользящее среднее, медианный фильтр, низкочастотный фильтр.

Определение фиксаций и саккад в подготовленном дискретном сигнале возможно по анализу скорости движения глаз в каждой точке временного ряда [3]. Общую скорость движения глаз можно вычислить по формуле

$$\vartheta = \sqrt{\vartheta_x^2 + \vartheta_y^2}, \quad (1)$$

где для численного вычисления скоростей ϑ_x и ϑ_y по координатам x и y соответственно в дискретных временных рядах может быть использована разностная аппроксимация

$$\vartheta_x \approx \frac{x_{i+1} - x_i}{t_{i+1} - t_i},$$

$$\vartheta_y \approx \frac{y_{i+1} - y_i}{t_{i+1} - t_i},$$
(2)

где $x_{i+1} - x_i$, $y_{i+1} - y_i$ – разности между текущей и последующей координатой x и y соответственно, а $t_{i+1} - t_i$ – шаг времени, после проведения аппроксимации одинаковый для всех временных отсчетов.

Далее по данным вектора общей скорости, рассчитанного по формуле (1) для каждого временного отсчета становится возможно определить его принадлежность к саккаде:

$$\text{saccade_indices}_i = \begin{cases} 1, & \text{если } \vartheta_i > \text{threshold} \\ 0, & \text{в противном случае} \end{cases},$$
(3)

где threshold – порог отсечения по скорости, saccade_indices_i – элемент вектора временных отсчетов, принадлежащих саккадам. Если скорость взгляда превышает некий заранее установленный пороговый уровень threshold , то в этой точке определяется саккада. Так как саккады формируются и затухают постепенно [4], при группировке их необходимо расширить, добавив слева и справа саккады прилежащие временные отсчеты, скорость в которых меньше, чем threshold_saccade , но больше чем некое значение $\text{threshold_fixation}$. Таким образом саккады будут продлены на краях. Все временные отсчеты, после выполнения процедуры расширения не отнесенные к саккадам, *считаются фиксациями*. Последовательно идущие временные интервалы саккад и фиксаций группируются совместно и формируется их итоговый перечень.

Далее для каждой полученной фиксации вычисляются *базовые признаки паттернов (БПП)* – набор признаков и статистических характеристик фиксации, отражающих вероятности ее принадлежности к различным «паттернам» типовых действий пользователя. Эти признаки должны быть достаточно чувствительными, позволять достоверно характеризовать то или иное типовое действие пользователя и могут выражаться, например, через:

- Коэффициенты изменения, средние значения, максимальное/минимальное значение движения взгляда по осям в саккаде и фиксации.

- Величины скорости и направления движения взгляда по осям.

- Площади фиксаций, соотношение длины фиксации к ее высоте, статистические показатели.

- Взаимное расположение предыдущей и (или) последующей фиксации относительно текущей.

БПП выбираются в диапазоне от 0 до 1, где 0 – полное отсутствие вероятности вхождения фиксации в «паттерн» типовых действий пользователя по данному признаку, а 1 – 100% вероятность. По полученным значениям соответствующих для каждого «паттерна» БПП может быть рассчитана *итоговая функция вероятности (ИФВ)* вхождения фиксации в «паттерн»:

$$\text{ИФВ}_{ij} = \frac{\sum_{i=1}^n (\text{БПП}_i)}{i},$$
(4)

где ИФВ_{ij} – итоговая функция вероятности наличия i -го «паттерна» в j -й фиксации. Далее для всех ИФВ_{ij} становится возможным определить принадлежность фиксации к «паттерну» i по формуле:

$$\text{pattern_indexes}_{ij} = \begin{cases} 1, & \text{если } \text{ИФВ}_{ij} > \text{pattern_threshold}_i \\ 0, & \text{в противном случае} \end{cases},$$
(5)

где $\text{pattern_indexes}_{ij}$ – элемент вектора pattern_indexes_i , определяющего принадлежность j -й фиксации к i -му «паттерну», а $\text{pattern_threshold}_i$ – некая пороговая вероятность, при которой j -я фиксация считается принадлежащей i -му «паттерну». Так как ИФВ_i может нарастать и спадать постепенно, при группировке их так же необходимо расширить, добавив слева и справа фиксации, ИФВ_i которых меньше, чем $\text{pattern_threshold}_i$, но больше чем некое значение $\text{pattern_threshold_min}_i$. Таким образом «паттерн» будет расширен на краях.

Результаты экспериментов

Для практического подтверждения разрабатываемой методики и опытного определения оптимальных значений характеристик и параметров была проведена серия экспериментов.

Экспериментальная группа состояла из 10 испытуемых. Социальная группа – студенты, средний возраст 21 год ($SD = 0,63$). В экспериментальном окружении группа выполняла тестовое задание: поиск списка научных работ определенного автора на сайте <https://elibrary.ru/>, нахождение среди них заранее определенной работы, изучение ее аннотации, рассмотрение и понимание иллюстраций. В тестовом задании выполнялись следующие типовые действия пользователя в интерфейсе: поиск элементов интерфейса, ввод данных, ожидание реакции системы, просмотр текста и графических изображений.

Для получения данных положения взгляда использовался айтрекер Tobii Eye Tracker 5, подключенный к USB порту персонального компьютера. «Сырые» данные через проприетарную библиотеку Tobii Stream Engine последовательно записывались в текстовый файл в следующем порядке: метка времени (в мкс), координаты x и y взгляда (число от 0 до 1 с точностью шесть знаков после запятой). Частота дискретизации сигнала составила 67 Гц. Дальнейшая обработка полученных данных и математическое моделирование производилось в среде MatLab версии R2023a.

Для определения лучшего метода интерполяции по полученным данным был проведен расчет метрик точности интерполяции: среднеквадратичная ошибка (RMSE), средняя абсолютная ошибка (MAE), коэффициент детерминации (R-squared), а также метрик сохранности структуры сигнала: коэффициент корреляции между исходными и интерполированными данными (PCC) и коэффициент сохранения энергии сигнала (Energy Preservation). Расчет производился для трех видов интерполяции: линейной, кубической и сплайн-интерполяции. Полученные результаты для координат x и y приведены в табл. 1

Таблица 1

Полученные значение метрик точности интерполяции

Метрики	Тип интерполяции координаты X			Тип интерполяции координаты Y		
	линейная	кубическая	Сплайн	линейная	кубическая	Сплайн
RMSE	3,836794	1,042373	3,498840	2,409435	0,575160	2,200225
MAE	0,584267	0,287248	0,699112	0,308854	0,117615	0,386759
R-squared	0,999919	0,999994	0,999934	0,999895	0,999994	0,999914
PCC	0,999960	0,999997	0,999967	0,999948	0,999997	0,999957
Energy Preservation	0,999632	0,999940	1,000010	0,999659	0,999948	0,999942

Из анализа табл. 1 следует, что все методы интерполяции могут быть использованы в дальнейших расчетах без значительного снижения достоверности, но кубическая интерполяция имеет минимальное значение среднеквадратичной, средней абсолютной ошибок для всех координат и коэффициент корреляции, и, следовательно, более предпочтительна для использования.

В результате анализа выбран метод кубической интерполяции, как обеспечивающий оптимальный баланс между точностью и сохранением плавности сигнала, что особенно важно для данных движения глаз, где необходимо сохранить монотонность и избежать осцилляций, которые могут быть вызваны сплайн-аппроксимацией.

Для определения лучшего метода фильтрации исследованы метрики качества трех методов фильтрации: скользящее среднее, медианный фильтр, фильтр низких частот (фильтр Баттерворта второго порядка с частотой среза 10 Гц). Размер окна фильтрации определен тремя временными отсчетами. Для дополнительной оценки сигнала к вышеуказанным метрикам использовалась дополнительная: отношение сигнал-шум (SNR). Полученные результаты расчета приведены в табл. 2.

Из анализа табл. 2 следует, что наименьшее значение среднеквадратичной и средней абсолютной ошибок обеспечивает использование медианного фильтра. Он так же имеет максимальный уровень сигнал-шум и максимально достоверно сохраняет энергию и форму полезного сигнала.

Таблица 2

Полученные значение метрик точности фильтрации

Метрики	Тип фильтрации координаты X			Тип фильтрации координаты X		
	Скользящее среднее	Медианный фильтр	Низкочастотный фильтр	Скользящее среднее	Медианный фильтр	Низкочастотный фильтр
RMSE	9,661737	3,920617	41,51543	6,73747	3,960514	24,77930
MAE	2,994565	0,405612	20,25602	1,656784	0,323237	9,866508
SNR	39,77054	48,43265	27,10611	39,09215	43,97890	27,76727
PCC	0,999689	0,999937	0,994302	0,999529	0,999839	0,993585
Energy Preservation	0,999112	0,999884	0,995096	0,999254	0,999982	0,996310

Исходя из вышесказанного, исходный вектор времени заменялся равномерными отсчетами времени, а вектора значений координат x и y интерполировались по новым отсчетам методом кубической интерполяции. Далее, к интерполированным данным применялся медианный фильтр с размером окна равным трем единицам, после чего по формуле (2) вычислялись вектора скоростей ϑ_x и ϑ_y , а по формуле (1) итоговый вектор ϑ – общая скорость перемещения взгляда в каждой точке временного ряда.

При расчете принадлежности временных отсчетов к саккадам по формуле (3) использовались следующие значения порогов отсека по скорости: $\text{threshold_saccade} = 3.2$, $\text{threshold_fixation} = 0.5$. При понижении значения threshold_saccade ниже 3.2 фиксировались частые ложные определения саккад, а при увеличении параметра $\text{threshold_fixation}$ выше значения 0.5 в фиксации попадали крайние временные отсчеты саккад, что приводило к значительному увеличению величин дисперсии и стандартного отклонения фиксаций. На рис. 1 приведен пример распределение саккад и фиксаций на графике скоростей одного из экспериментов.

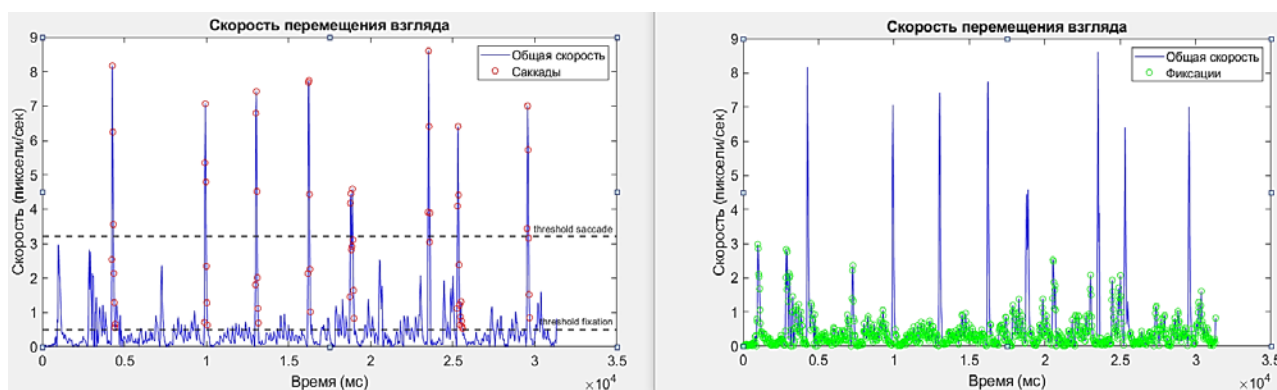


Рис. 1. Распределение саккад и фиксаций на графике скорости перемещения взгляда

После определения перечня фиксаций для каждой из них по формуле (4) вычислялись базовые признаки паттернов (БПП). Для примера рассмотрим определение и расчет БПП паттерна «Чтение текста работы». Из анализа графического изображения паттерна, представленного на рис. 2 следует, что он состоит из определенного количества фиксаций и саккад, идущих друг за другом и обладающих следующими свойствами:

1. Общий характер движения глаз: по оси x – плавное равномерное движения слева направо, по оси y – нахождения взгляда на одной координате с минимальным разбросом во время фиксации, общий вид графика – плавное, без сильных ускорений снижение координат снизу-вверх.
2. Саккады: резкие, с большим ускорением по оси x , практически без ускорения по оси y , идущие последовательно справа налево, похожих продолжительностей и с похожими координатами начала и окончания.
3. Фиксации: по оси x длинные, с небольшим ускорением по оси x и еще меньшим по оси y .

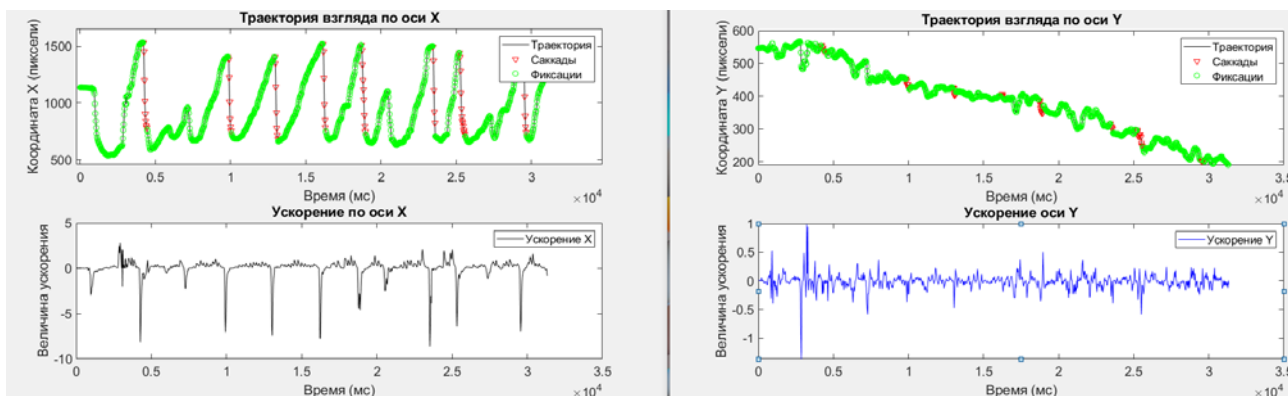


Рис. 2. Графическое представление траектории взгляда и ускорения по оси X и Y в паттерне «Чтение текста работы»

Исходя из вышесказанного были выбраны шесть инвариантных базовых признаков паттерна «Чтение текста работы», устойчивых как к пропорциональному, так и к непропорциональному масштабированию и определен алгоритм формирования их значений (табл. 3).

Таблица 3

Способы расчета БПП паттерна «чтения текста»

Наименование БПП	Диапазон для присвоения значения 1	Диапазон для присвоения значения 0.5
Угол наклона фиксации к оси X в градусах	От –8 до 8	От –12 до –8 или от 8 до 12
Соотношение стандартных отклонений фиксации по осям X и Y (относительная «длина» фиксации)	Более 10	От 4 до 10
Угол наклона саккады, ведущей от окончания текущей фиксации к началу следующей	От –7 до 7	От –12 до –7 или от 7 до 12
Разница координат Y начала текущей и следующей фиксации	От –100 до 100	От –300 до 100 или от 100 до 300
Разница координат X начала текущей и следующей фиксации	От –150 до 150	От –350 до 150 или от 150 до 350
Продолжительность фиксации, сек.	Более 3	От 1.5 до 3

После расчета согласно формуле (4) для каждой саккады перечня БПП по формуле (5) был определен факт принадлежности каждой саккады к паттерну (рис. 3). При этом использовались пороговое значение `pattern_threshold` равное 0.85, и `pattern_threshold_min` равное 0.5.

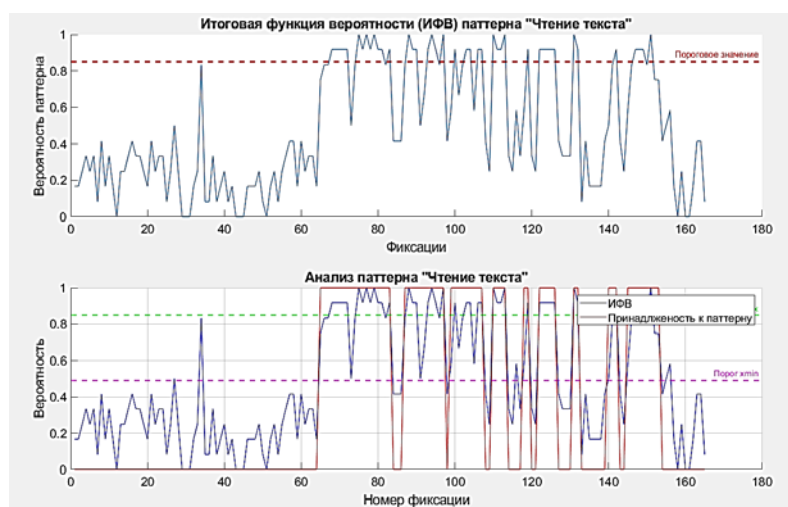


Рис. 3. График определения ИФВ паттерна «Чтение текста» для одного из экспериментов

В результате в серии из 10 экспериментов при использовании методики с описанными выше параметрами удалось определить 92 паттерна «Чтения текста» из 98, выявленных вручную при анализе видеозаписей, что соответствует уровню доверия 93,8% и входит в уровень значимости $\alpha = 0.1$, что является приемлемым результатом.

Выводы и заключение

Разработанная методика позволяет эффективно разделять дискретные данные движения взгляда на фиксации и саккады, рассчитывать для каждой из них статистические, временные и спектральные характеристики, оценивать состав и продолжительность типовых действий, выполняемых пользователем в интерфейсе цифрового продукта. В то же время после проведения серии экспериментов становится очевидно, что для полного понимания причин выбора пользователем того или иного действия, их мотивации при взаимодействии с элементами интерфейса данных только движения глаза становится недостаточно. Процесс анализа необходимо дополнять дополнительными данными, такими, например, как положения мыши, клавиши, нажимаемые пользователем в процессе взаимодействия, информацией о возможных сценариях использования цифрового продукта. Типовые действия необходимо группировать по неким признакам в укрупненные группы. Выход может быть найден в использовании механизмов кластеризации, таких, например, как DBSCAN (Density-based spatial clustering of applications with noise, плотностной алгоритм пространственной кластеризации с присутствием шума) [5], адаптированных к использованию временных рядов, применяемых или к фиксациям, или непосредственно к данным движения взгляда.

Особое внимание необходимо уделить добавлению в методику информации с датчиков биосигналов человека, таких как ВСР и ЭЭГ, анализ которых позволит получить данные об уровнях стресса, увеличении или уменьшении когнитивной нагрузки, что позволит перейти непосредственно к процессу анализа удобства интерфейса, качества его воспринимаемости пользователями.

Разработанная методика в дальнейшем может быть использована в различных прикладных задачах, таких как исследование пользовательского опыта, особенности взаимодействия с интерфейсом пользователей с различными типами расстройств нервной системы и изучение когнитивных процессов.

Список источников

1. Duchowski A. Eye Tracking Methodology: Theory and Practice. Springer, 2007. 218 p.
2. Chatfield C. The Analysis of Time Series: An Introduction. New York : Chapman and Hall/CRC, 2003. 352 p.
3. Eye Tracking: A Comprehensive Guide to Methods and Measures / K. Holmqvist, M. Nyström, R. Andersson [et al.]. Oxford University Press, 2011. 543 p.
4. Nyström M., Holmqvist K. An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data // Behavior Research Methods. 2010. Vol. 42, no. 1. P. 188–204.
5. Митин Г. В., Панов А. В. Модификация алгоритма dbscan с использованием гибридных подходов к определению границ кластеров для обработки потоковых данных // Электронный научный журнал «ИТ-Стандарт». 2023. № 4. P. 36–57.