

Techniques and tools
for human-machine interface
security and usability assessment

Training
A. Gordeev
Edited by V.S. Kharchenko

Preparation and calibration of eye-tracker

Research scenario and selection of respondents

Eye tracking and verbal protocol

Visualization of results



Techniques and tools for HMI security and usability assessment. Training



TRAINING

TECHNIQUES AND TOOLS FOR
**HUMAN-MACHINE
INTERFACE**
SECURITY AND USABILITY
ASSESSMENT

2017



Co-funded by the
Tempus Programme
of the European Union

**Министерство образования и науки Украины
Национальный аэрокосмический университет
им. Н.Е. Жуковского «ХАИ»**

А.А. Гордеев, Д.В. Гордеева

**Методы и инструментальные средства
оценивания безопасности и удобства
использования человеко-машинных
интерфейсов**

**Techniques and tools for human-machine
interface security and usability assessment**

Практикум

Под редакцией В.С. Харченко

**Проект
SEREIN 543968-TEMPUS-1-2013-1-EE-TEMPUS-JPCR
*Modernization of Postgraduate Studies on Security and Resilience
for Human and Industry Related Domains***

2017

Викладені матеріали практичної частини начального курсу «Технології, техніки і інструменти оцінювання інформаційної безпеки і зручності використання» («Techniques and Tools for HMI security and usability assessment»), який був підготовлений в рамках проекту Modernization of Postgraduate Studies on Security and Resilience for Human and Industry Related Domains (reference number 543968-TEMPUS-1-2013-1-EE-TEMPUS-JPCR). Курс присвячений опису та застосування технології айтрекінга, яка застосовується при дослідженні найбільш складних елементів людино-машинного взаємодії. Курс орієнтований на студентів спеціальностей «Кібербезпека», «Комп'ютерна інженерія», «Комп'ютерні науки» та «Програмна інженерія» при вивченні технологій проектування та оцінювання призначених для користувача інтерфейсів програмного забезпечення. Курс також може бути корисний для UX / UI дослідників і проектувальників.

Рецензенти:

- Мохор Владимир Владимирович, доктор технических наук, профессор, директор института проблем моделирования в энергетике им. Г.Е. Пухова, НАН Украины;
- Заславский Владимир Анатольевич, доктор технических наук, профессор, профессор кафедры математической информатики Киевского национального университета им. Тараса Шевченко.

Гордеев А.А., Гордеева Д.В.

Технологии, техники и инструменты оценивания информационной безопасности и удобства использования. Практикум / под ред. В.С. Харченко – Министерство образования и науки Украины, Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ». 2017. – 96 с.

Изложенные материалы практической части учебного курса «Методы и инструментальные средства оценивания информационной безопасности и удобства использования» («Techniques and Tools for HMI security and usability assessment»), который был подготовлен в рамках проекта Modernization of Postgraduate Studies on Security and Resilience for Human and Industry Related Domains (reference number 543968-TEMPUS-1-2013-1-EE-TEMPUS-JPCR).

Курс посвящен описанию и применению технологии айтрекинга, которая применяется при исследовании наиболее сложных элементов человеко-машинного взаимодействия.

Курс ориентирован на студентов специальностей «Кибербезопасность», «Компьютерная инженерия», «Компьютерные науки» и «Программная инженерия» при изучении технологий проектирования и оценки пользовательских интерфейсов программного обеспечения. Курс также может быть полезен для UX / UI исследователей и проектировщиков.

Библиогр.: 18 наименований, рисунков – 35. таблиц – 5.

Утверждено на заседании ученого совета Национального аэрокосмического университета имени Н.Е. Жуковского «ХАИ» (протокол № 6 от 22 февраля 2017 г.).

УДК 004.054

© Гордеев А.А., Гордеева Д.В., 2017

This work is subject to copyright. All rights are reserved by the authors, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms, or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

СПИСОК СОКРАЩЕНИЙ

ПВП	–	Параллельный вербальный протокол
РВП	–	Ретроспективный вербальный протокол
ПО	–	Программное обеспечение
	–	
UX	–	User experience
AOI	–	Area of interest

Этот проект финансируется при поддержке Европейской комиссии. Эта публикация (сообщение) отражает мнения только авторов, и Комиссия не может нести ответственность за любое использование содержащейся в нем информации.

This project has been funded with support from the European Commission. This publication (communication) reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

1. ПОДГОТОВКА К ИССЛЕДОВАНИЯМ И КАЛИБРОВКА АЙТРЕКЕРА

Форма занятия: лабораторная работа

Цель — получить навыки подготовки к исследованиям человеко-машинного взаимодействия с использованием айтрекинга.

Образовательные задачи: изучить процесс подготовки к исследованиям человеко-машинного взаимодействия с использованием айтрекинга.

Практические задачи: проведение процедуры калибровки Tobii Eyetracker X2 60.

Исследовательские задачи: оценить качество калибровки Tobii Eyetracker X2 60.

Теоретический материал

Создание нового проекта

В качестве примера проведения и настройки процедуры калибровки рассмотрим калибровку в программном обеспечении Tobii Studio [1]. Для этого установим программу Tobii Studio на персональный компьютер и подключим Eye-tracker X2-60. Процедуры установки Tobii Studio и подключения Eye-tracker X2-60 являются достаточно простыми и интуитивно-понятными, поэтому мы не будем на них останавливаться.

Запустим Tobii Studio (рис. 1.1) и создадим новый проект. Для создания нового проекта нажмем на кнопку «Create New Project» в Tobii Studio (рис. 1.1). Далее открывается новое окно (рис. 1.2) «Create new project» с полями для ввода исходной информации: названия проекта, его описания, автора и пути для сохранения проекта. После того как были заполнены все обязательные поля и нажата кнопка ввода информации, открывается новое окно для создания первого теста «Create first test» (рис. 1.3).

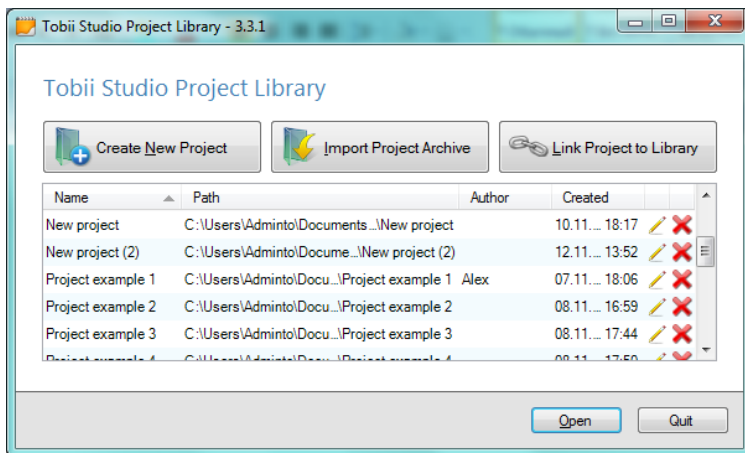


Рисунок 1.1 – Выбор или создание нового проекта в Tobii Studio

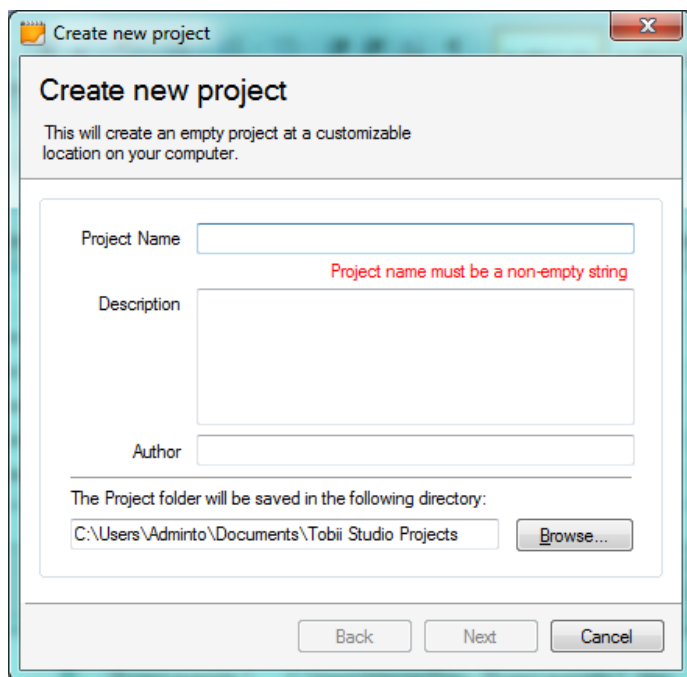
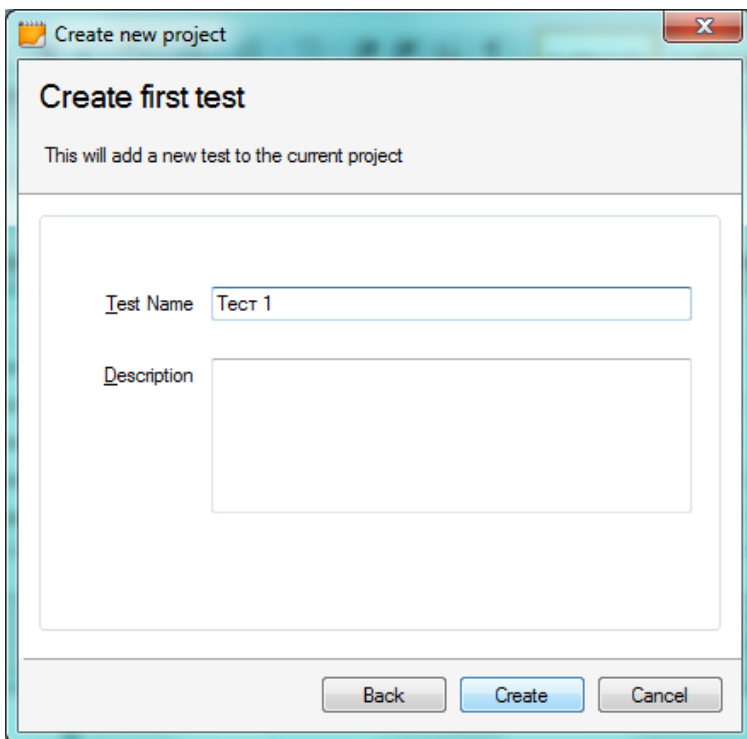


Рисунок 1.2 – Поля для заполнения при создании нового проекта в Tobii Studio

Далее открывается окно для выбора типа объекта исследования. Мы выбираем объект типа «Web». Такой выбор реализуется перетаскиванием типа объекта на светло-серую область, которая располагается в центре окна под серой стрелкой (рис. 1.4). Далее открывается новое окно для ввода информации об объекте исследования, а именно добавления ссылки веб-сайта, размера исследуемой области с возможностью настройки фиксированного времени показа стимула (т.е. веб-страницы) (рис. 1.5). На рис. 1.6 представлена заполненная и настроенная форма Tobii Studio, которая дает возможность начать процедуру исследования. Процедура исследования начинается после нажатия на кнопку «Start recording».



The image shows a software window titled "Create new project". Inside, there's a section titled "Create first test" with a subtitle "This will add a new test to the current project". Below this, there are two input fields: "Test Name" containing the text "Тест 1" and "Description" which is empty. At the bottom of the window, there are three buttons: "Back", "Create", and "Cancel".

Рисунок 1.3 – Форма для создания первого теста

Процедура калибровки

Необходимым условием проведения исследования является корректная калибровка инфракрасных камер айтрекера. Прежде всего, они должны улавливать зрачок и отражение роговицы (рис. 1.7). Если камера не обнаружит один из этих двух элементов вследствие, например, того, что глаза респондента будут закрыты или его голова будет вне головной коробки, айтрекер не будет фиксировать фокус внимания респондента и данные будут потеряны.

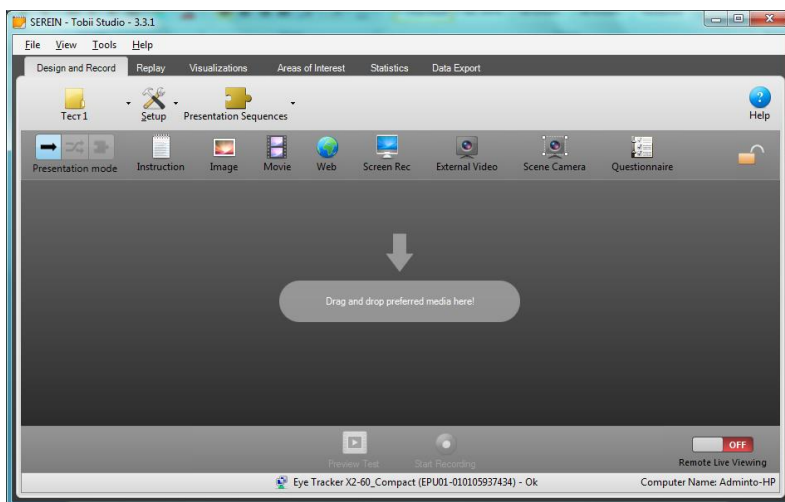


Рисунок 1.4 – Форма выбора типа объекта исследования

Более худшей ситуацией, чем потеря данных, является их искажение или шум в данных. Например, если, зрачок респондента будет перекрыт темным объектом, таким как ресницы или же роговице будут мешать сторонние отражения, записанные данные о движении взгляда респондента будут искажены. Такая ситуация называется смещением. Отметим, что потерянные данные можно легко обнаружить и восстановить, а искаженные же данные идентифицировать намного сложнее. Недостоверные данные могут быть обнаружены совершенно случайно, но исследователи, как правило, целенаправленно не занимаются поиском

искаженных данных. Поэтому такие данные остаются незамеченными, создавая шум.

Web Element Setup

This element opens a web browser and lets the participant navigate from a predefined web page. You need to specify a start URL.

Element Name
Web Element

Browser Settings

Start URL

☒ Fixed browser size Left 20 Top 20
Width 1326 Height 688

☐ Fullscreen

Properties

End recording with F10 or Escape

☐ End recording after: 20 seconds (0,3 minutes)

OK Cancel

Рисунок 1.5 – Форма для настройки данных о веб-странице

Механическое регулирование камеры

Для некоторых айтрекеров настройки камер устанавливаются автоматически и закрыты для регулирования исследователями. Однако существуют айтрекеры, которые дают исследователям возможность увидеть «захват» камерами айтрекера зрачков и роговиц респондента и при необходимости самим настроить (повернуть) камеру таким образом, что бы зрачки респондентов были точно идентифицированы (рис. 1.8). Как только зрачки и роговицы респондента идентифицированы, респондентов

1. Подготовка к исследованиям и калибровка айтрекера

необходимо попросить посмотреть на края раздражителя несколько раз, чтобы убедиться, что айтрекер «не теряет» зрение респондентов.

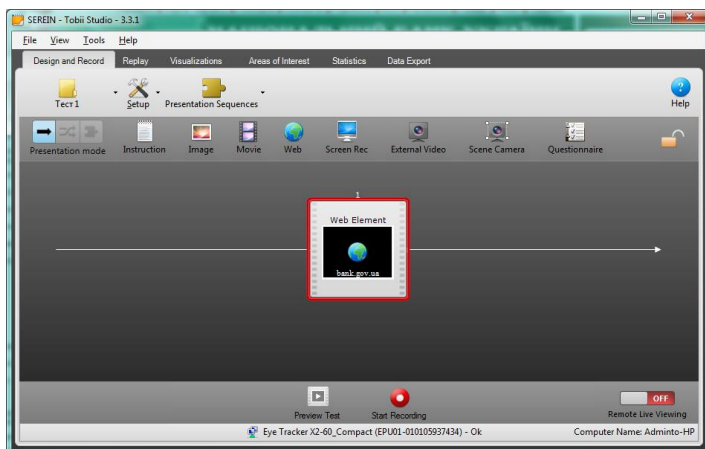


Рисунок 1.6 – Форма для запуска исследований

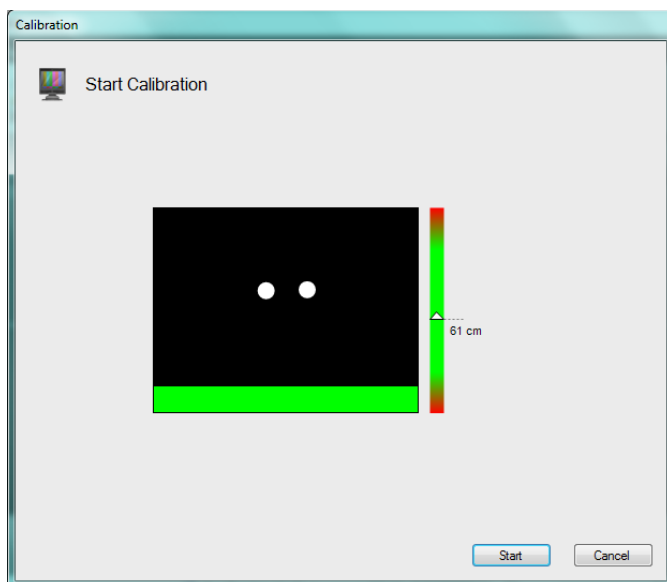


Рисунок 1.7 – Идентификация зрачков респондента



Рисунок 1.8 – Корректное нахождение зрачков и роговиц респондента

Глаза респондента также могут не смотреть в сторону относительно экрана монитора. Исследователь может выявить такую проблему визуально (рис. 1.9). Проблема смещения глаз респондента решается за счет изменения позиции респондента перед экраном монитора.

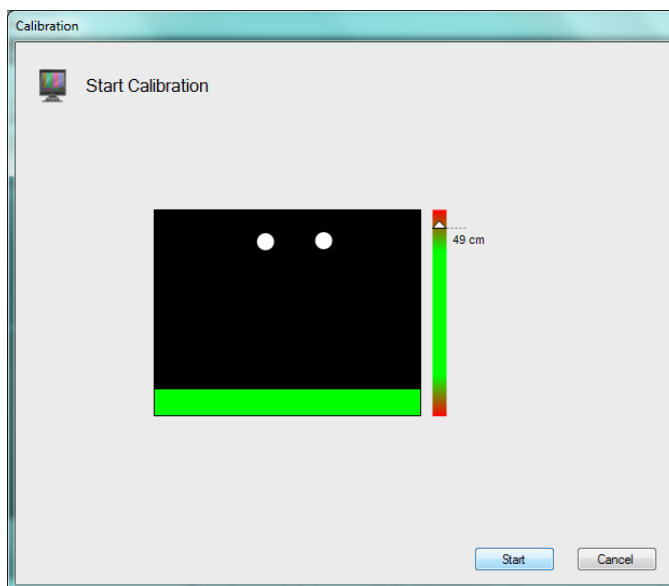


Рисунок 1.9 – Идентификация зрачков респондента

Айтрекеры отдельных производителей могут иметь свои особенности. Детальная информация о них обычно включена в руководство использования айтрекера. Отметим, что в [2] авторы предлагают ряд советов по настройке айтрекера.

Возможные причины недостаточного качества калибровки

Представим причины проблем, которые могут повлиять на распознавание отражения зрачка и роговицы, что в свою очередь ведет к смещениям в данных или их потере (Таблица 1.1) [3].

Процедура калибровки

Процедура калибровки проводится до начала каждой сессии, следовательно, до начала записи данных. Для этого требуется, чтобы каждый респондент перед проведением исследований смотрел на ряд предлагаемых точек. Точки могут быть представлены на бумаге (рис. 1.10), на экране монитора (рис. 1.11) или нанесены специальным инфракрасным маркером. Респондент должен смотреть в центр каждой представляемой точки пока она не исчезнет или его не попросят смотреть на следующую точку.

Таблица 1.1 - Возможные причины проблем при настройке айтрекера

Причина проблемы	Описание проблемы	Решение проблемы
Очки	Очки могут вызвать проблему, которая заключается в создании дополнительных отражений тени оправы, вызывая помехи в обнаружении зрачков	Поверните немного камеру, чтобы сдвинуть отражение от отражения зрачка и роговицы
Контактные линзы	Пузырьки воздуха под линзами могут преломлять свет и формировать несколько отражений	Можно незначительно уменьшить фокус камеры для респондентов, которые носят линзы
Макияж	Данные айтрекера может	

1. Подготовка к исследованиям и калибровка айтрекера

	исказить макияж на ресницах	У исследователя должен быть комплект для снятия макияжа, который подготовлен заранее специально для респондентов
Обвисшие веки или длинные ресницы	Обвисшие веки и длинные ресницы могут частично или полностью закрыть часть области видимости	Необходимо разместить айтрекер немного ниже его обычного положения, что бы его камеры могли бы записывать движения глаз

Число калибровочных точек не является величиной постоянной и может меняться в зависимости от особенностей айтрекера и выбранных настроек. Обычно количество точек для калибровки больше, чем две и меньше двадцати. Например, в Tobii Studio можно настраивать количество точек и выбирать 2, 5 или 9 точек (рис. 1.12). Так же существуют айтрекеры, которые предлагают только одну точку для калибровки айтрекера (рис. 1.13). Состояние текущих исследований будущих технологий айтрекинга предполагает, что коммерческие айтрекеры в будущем не будут нуждаться в калибровке.

Во время калибровки айтрекер анализирует направление зрения респондента, при этом отображая точки, на которые просят посмотреть респондента. Точки при этом расставляются в нескольких различных частях экрана. Такая информация потом интерполируется и экстраполируется во все другие места поля зрения респондента.



Рисунок 1.10 – Пример пятиточечной калибровки

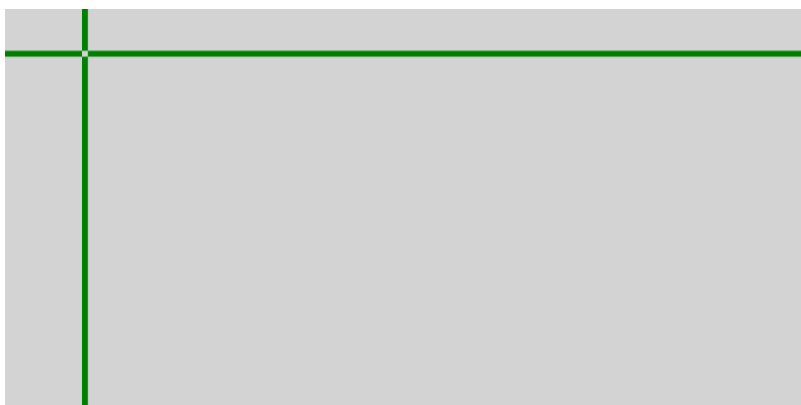


Рисунок 1.11 – Обозначение точек для калибровки на экране монитора

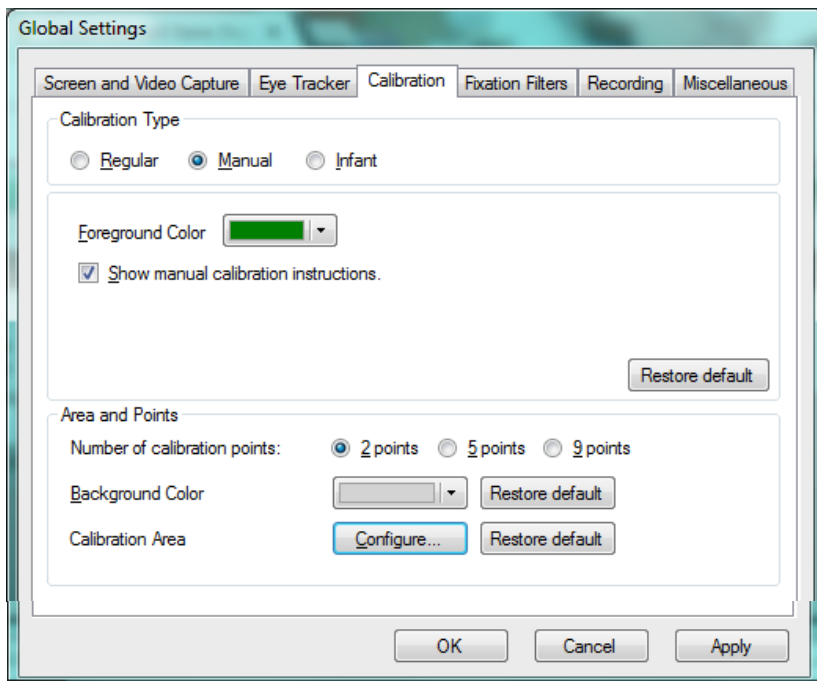


Рисунок 1.12 – Настройка параметров калибровки



Рисунок 1.13 – Пример одноточечной калибровки

Каждый респондент должен пройти калибровку перед началом сессии, поскольку каждый из них имеет свои особенности зрения.

Условия калибровки

Ключевым условием калибровки является соблюдение таких же условий, при которых будет проходить и последующее исследование, иначе данные могут быть неточными. Также исследователю необходимо:

- во-первых, убедиться, что уровень света не изменился между двумя этапами (пилотом и реальном исследовании);
- во-вторых, айтрекер должен находиться на одинаковом расстоянии относительно глаз респондентов на этапе проведения калибровки и при реальном исследовании.

Если исследователь применяет дистанционный айтрекер, необходимо постараться обеспечить респонденту максимально удобную позицию относительно айтрекера (в пределах оптимального расстояния от айтрекера) и также попросить его оставаться относительно неподвижно. Если исследователь использует наголовной айтрекер, его необходимо аккуратно разместить на голове у респондента до начала проведения процедуры калибровки таким образом, чтобы он позднее не сместился. Если после калибровки респондент неосознанно сместил свой стул с исходной позиции (ближе или дальше относительно айтрекера), или он существенно изменил свою позицию (позу) на стуле во время проведения сессии, то процедуру следует повторить.

Также отметим, что, если респондент планирует принимать участие в исследовании с надетыми очками, то и процедуру калибровки он должен проходить также в надетых очках. Идеальным вариантом считается тот, при котором исследователь до начала исследований знает, кто из респондентов будет участвовать в исследовании в надетых очках. Обычно такие нюансы выясняются перед началом исследований, т.е. на этапе анкетирования респондентов.

Проверка калибровки и перекалибровка

Перед проведением исследований рекомендуют оценить качество калибровки. Некоторые программные продукты

поддерживают такую процедуру (рис. 1.14). Если возможности для оценивания нет, то можно попросить респондентов посмотреть на заранее определенные точки (например, точки на картинке, которые были сделаны заранее или определенные буквы на упаковке продукта) после проведения калибровки. Это необходимо для сравнения реального и необходимого фокуса внимания респондента. Если были установлены факты значительного смещения в точках реального и необходимого фокуса, то необходимо отрегулировать камеру при необходимости и повторить процедуру калибровки.

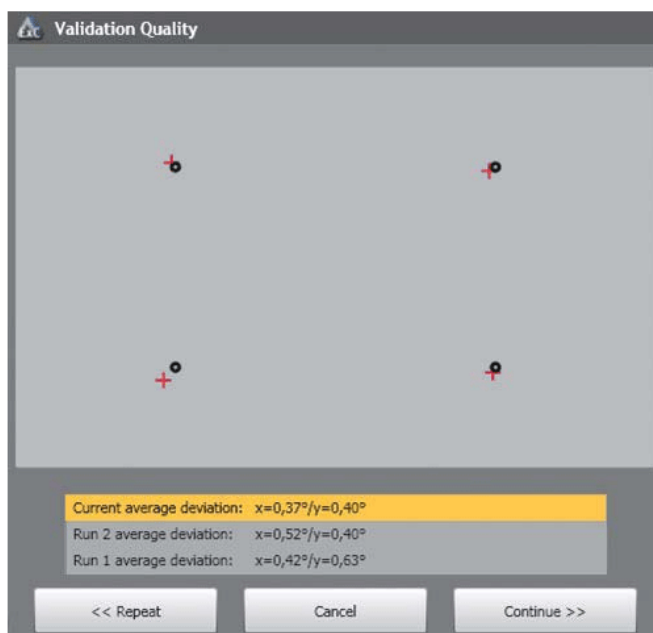


Рисунок 1.14 – Результаты четырехточечной калибровки с отклонениями по оси X и Y

Если необходимо поправить айтрекер или провести процедуру перекалибровки для респондента, то нужно без лишних переживаний сообщить об этом респонденту, что это обычная процедура, которая займет всего несколько дополнительных минут.

При планировании общего времени для сессии нужно обязательно учесть время на процедуру калибровки и возможную перекалибровку. Количество времени, которое следует добавить для сессии, зависит от модели айтрекера (некоторые айтрекеры более эффективны), возраста группы респондентов (обычно у молодых людей качество калибровки выше, у более старших) и опыта исследователя (опытным исследователям требуется меньше времени).

Неудачная калибровка

Если после нескольких процедур перекалибровки не удалось добиться приемлемого уровня калибровки айтрекера, исследования необходимо продолжать, поскольку респондент готов к дальнейшей работе. Для проведения исследований необходимо приглашать большее количество респондентов, чем планировалось привлечь ранее. Например, вместо 10 респондентов можно приглашать 12. Сессии, в которых возникли проблемы с калибровкой необходимо зафиксировать и в дальнейшем, скорее всего, исключить полученные данные такой сессии из общего исследования.

Для того, чтобы включить часть полученных данных сессии, которые имели проблемы с калибровкой, необходимо зафиксировать на экране проблемные области. Если они находятся вдалеке от области интересов исследований, то можно продолжить анализировать эти данные, при этом, не ухудшая достоверности полученных результатов (рис. 1.15).

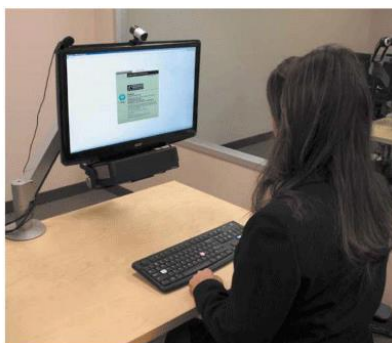
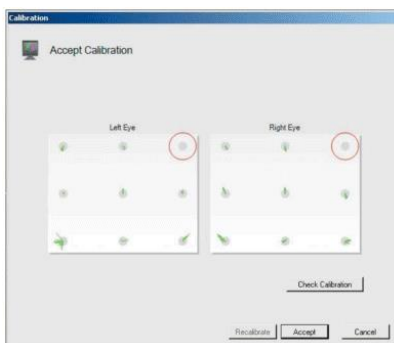


Рисунок 1.15 – Результаты калибровки: калибровка прошла неудачно для правого верхнего угла экрана монитора

Ход работы

1. Установить программное обеспечение Tobii Studio;
2. Подключить айтрекер Tobii Eyetracker X2 60;
3. Провести процедуру калибровки для одного респондента;
4. Оценить качество калибровки и, если есть необходимость, провести перекалибровку айтрекера.

Требования к содержанию отчета

Отчет должен содержать следующее:

- краткое описание процедуры подключения айтрекера;
- результаты калибровки айтрекера;
- результаты оценивания качества калибровки.

Контрольные вопросы

1. Что такое айтрекер и для чего он применяется?
2. Из каких составных частей состоит айтрекер?
3. Для чего необходима процедура калибровки?
4. Опишите наиболее частые причины проблем при настройке айтрекера.
5. Опишите решения наиболее частых проблем при настройке айтрекера.
6. Опишите основные условия проведения калибровки.
7. Как оценить качество калибровки?
8. Как можно компенсировать недостаточное качество калибровки за счет респондентов?

2. РАЗРАБОТКА СЦЕНАРИЯ ИССЛЕДОВАНИЯ И ПОДБОР РЕСПОНДЕНТОВ

Форма занятия: лабораторная работа

Цель – получить навыки подготовки сценария исследования и подборе респондентов.

Образовательные задачи: изучить методику подготовки сценария исследований и подбора группы респондентов.

Практические задачи: подготовить сценарий для проведения исследования и составить анкету для подбора группы респондентов.

Исследовательские задачи: верифицировать сценарий для исследований и анкету для подбора респондентов.

Теоретический материал

Основа для составления сценария

Планирование тестирования начинается не с составления текста заданий сценария, а с детальной проработки целей и вопросов исследования совместно с проектной командой. Рассмотрим процесс формирования основы сценария.

Важные сценарии

«Важными сценариями», как правило, являются пользовательские сценарии, которые влияют на бизнес цели или связаны с тестированием. К ним можно отнести:

- наиболее частотные (например, отправка сообщения в мессенджере);
- влияющие на бизнес-цели (например, работа с платежной формой);
- связанные с обновлением (те, которые затронули редизайн или внедрение новой функциональности).

Известные проблемы

Исследование должно давать ответ на причины бизнес-проблем сервиса. Иногда проблемы в интерфейсах уже известны

разработчикам или пользователям и исследователю необходимо собрать подробности и конкретику. Например, можно также использовать в исследовании обращения пользователей в службу поддержки с вопросами, например, по форме оплаты веб-сервиса и др.

Вопросы

У команды разработчиков программного продукта также могут быть потенциальные вопросы для исследования: например, обращают ли пользователи внимание на баннер, рекламирующий дополнительные услуги; понятны ли названия определенных разделов и др.

Гипотезы

Гипотезы – это то, во что преобразуются известные проблемы и вопросы команды разработчиков и исследователей. Идеальным считается случай, при котором заказчик приходит к исследователю с готовыми гипотезами. Например, «наши клиенты оплачивают сервис только с использованием мобильного телефона, при этом оплачивают также и комиссию. Возможно, пользователи не видят возможности выбора для более выгодного способа оплаты». Если у заказчика гипотез нет, а есть только желание исследовать удобство использования пользовательских интерфейсов, тогда задача формулировки таких гипотез возлагается на исследователя.

Исследователь может сформулировать такие гипотезы самостоятельно либо же во взаимодействии с командой разработчиков. Прежде всего, необходимо выяснить, существуют ли элементы дизайна, над которыми много спорили разработчики и которые могут потенциально иметь проблемы. Исследователю рекомендуется самостоятельно поработать с продуктом, чтобы найти потенциальные сложности для пользователей, которые важно и можно протестировать. Все вышеперечисленное поможет исследователю составить список тех элементов (заданий, вопросов, проверок), которые должны войти в итоговый сценарий.

Метод сбора данных

Исследователю важно продумать и заранее подготовиться к сбору необходимых данных. Традиционно применяются следующие варианты:

– **наблюдение.** Во время выполнения заданий респондент остается наедине с продуктом и ведет себя так, как считает нужным. Комментарии респондента собираются посредством опросников и общения с модератором после теста. Это наиболее простой и понятный для респондента метод. Он обеспечивает более естественное поведение респондента. Очевидным недостатком метода является отсутствие возможности фиксирования полного объема информации;

– **мысли вслух.** Долгое время этот метод чаще остальных использовался в тестированиях удобства использования пользовательских интерфейсов. Якоб Нильсен в свое время называл [4] его главным инструментом оценки удобства использования. Его суть заключается в том, что исследователь просит респондента озвучивать все мысли, возникающие у него при работе с интерфейсом, и комментировать все свои действия, например, таким образом: «Сейчас я собираюсь добавить этот товар в корзину. А где же кнопка? А, вот она. Ой, я забыл посмотреть, какой там был цвет».

Метод помогает объяснить поведение респондента и какие эмоции вызывает у него текущее взаимодействие. Однако у него есть определенные недостатки. Перечислим некоторые из них:

– во-первых, для респондентов не естественно все время «думать вслух». Они будут часто замолкать, и исследователям придется постоянно напоминать о том, что им необходимо продолжать озвучивать свои мысли вслух;

– во-вторых, задания при таком методе выполняются несколько дольше, чем в реальной жизни. Кроме того, часть респондентов начинает пользоваться продуктом более вдумчиво. Проговаривая причины своих действий, они стараются действовать более рационально, и исследователи могут не уловить какие-либо интуитивные моменты их поведения;

– **активное вмешательство модератора (исследователя).** Такой метод лучше всего подходит для тестирования концепций и прототипов. Во время выполнения заданий исследователь активно взаимодействует с респондентом: выясняет в нужные моменты причины его поведения и задает уточняющие вопросы. В некоторых случаях исследователь даже может выдавать

незапланированные задания, вытекающие из диалога с респондентом. Такой метод позволяет собрать максимальное количество качественных данных, однако его можно применять, только если исследователь обладает достаточно высоким опытом. Некорректно сформулированные или не вовремя заданные вопросы могут сильно повлиять на поведение и впечатления респондента и даже сделать результаты тестирования недостоверными. Также при использовании этого метода практически нельзя посчитать соответствующие метрики;

– **ретроспективный метод.** Это комбинация первых двух методов. Респондент сначала выполняет все задания без вмешательства исследователя, а потом перед ним проигрывается видеозапись его работы, и он комментирует свое поведение и отвечает на вопросы исследователя. Главный недостаток такого метода заключается в увеличении времени тестирования.

Основная задача исследователя при выборе метода сбора данных заключается в получении достаточного объема данных, сохранив максимальную естественность поведения респондента.

Очень важным элементом сбора данных является применение технологии айтрекинга. Видя фокус текущего внимания респондента, исследователь может лучше понять его поведение, не задавая лишних вопросов.

Метрики успешности проведения исследований

Метрики - это количественные показатели результатов исследования пользовательских интерфейсов. В результате тестирования исследователь всегда получает перечень недостатков, которые были найдены в интерфейсе. Метрики же позволяют понять, насколько все хорошо или плохо, а также сравнить с другим проектом или предыдущими версиями дизайна.

Приведем в качестве примера следующие метрики:

– **успешность выполнения заданий.** При расчете такой метрики исследователи обычно выделяют три группы оценивания успешности:

- респондент справился с заданием без проблем - 100%;
- респондент столкнулся с проблемами, но выполнил задание самостоятельно - 50%;
- не справился с заданием - 0%.

Если из 12 респондентов 4 справились с заданием легко, 6 - с проблемами, а 2 респондента на справились, то средняя успешность при выполнении задания составит 58%;

– **время выполнения заданий.** Такая метрика применяется только для задач сравнения. Рассмотрим, например, задачу подбора товара в интернет-магазине. Пользователи должны быстро находить фильтры и другие элементы интерфейса, связанные с поиском товаров, но сам процесс выбора будет занимать у них разное время. Женщины при выборе туфель обычно готовы просмотреть 20 страниц выдачи. И это необязательно значит, что на первых страницах не было подходящих товаров или что они не видят фильтры. Часто им просто хочется увидеть все варианты;

– **частотность проблем.** Любой отчет о юзабилити-тестировании содержит список проблем, с которыми столкнулись респонденты. Количество респондентов, которые столкнулись с проблемой является показателем ее частотности в рамках теста. Такую метрику можно применять только в том случае, если респонденты выполняли абсолютно одинаковые задания. Если же в тесте были вариации или задания были не четко сформулированы, а составлены на основании интервью, то посчитать частотность будет сложно. Необходимо будет не только считать количество респондентов, которые столкнулись с проблемой, но и оценивать, сколько респондентов могли бы столкнуться с проблемой (например, выполняли схожую задачу или заходили в тот же раздел). Тем не менее, значение такой метрики позволяет исследователю понять, какие проблемы стоит исправлять в первую очередь;

– **субъективная удовлетворенность.** Это субъективная оценка респондентом удобства работы с продуктом. Выявляется она при помощи опросников, которые респонденты заполняют в процессе или после тестирования. Существуют стандартные опросники. Например, System Usability Scale, Post-Study Usability Questionnaire или Game Experience Questionnaire для игр. Исследователь также может разработать свой собственный опросник.

Респонденты должны действовать максимально естественно и в одинаковых условиях. Поэтому желательно обеспечить следующее:

- единые точки старта. Одни и те же задания для разных респондентов должны начинаться с одной и той же точки человеко-машинного взаимодействия. Можно также после каждого задания просить респондентов вернуться на главную страницу;

- отсутствие вмешательств. Любое общение с исследователем может повлиять на метрики эффективности, например, если исследователь невольно подскажет что-то респонденту. Это также увеличивает время выполнения задания;

- порядок заданий. Чтобы компенсировать эффект обучения в сравнительном тестировании, необходимо менять порядок появления исследуемых программных продуктов для разных респондентов. Например, половина респондентов может начать с вашего проекта, а половина - с конкурентного;

- критерии успешности. Исследователю следует заранее определить, какое именно поведение он считает успешным для выполнения задания. Например, допустимо ли, чтобы при подборе товара в интернет магазине респондент не пользовался фильтрами.

Трактовка метрик

Необходимо помнить, что классическое юзабилити-тестирование - это качественное исследование и полученные значения метрик в первую очередь иллюстративны. Они дают общее представление о разных сценариях исследований, позволяя увидеть проблемные места. Например, настройки аккаунта вызывают больше сложностей, чем регистрация в системе. Они могут показать динамику изменений, если исследования проходят регулярно. То есть метрики дают возможность определить, какие изменения в новом дизайне позволяют выполнять задачу быстрее.

Необходимость в метриках

Не каждый отчет о юзабилити-тестировании содержит значения метрик. Их сбор и анализ требует времени и накладывает ограничения на методы проведения теста. Обычно результаты расчета метрик нужны в следующих случаях:

- доказать то, что в продукт нужно внести изменения. Для лиц, принимающих решения, цифры наглядны, понятны и привычны. Когда исследователь обращает внимание на то, что 10 из 12 респондентов не смогли оплатить товар или что регистрация в системе занимает в среднем в два раза больше времени, чем у конкурентов. Такие данные придают результатам исследования больший вес;

- сравнить. Если исследователь сравнивает продукт с конкурентными, ему также не обойтись без метрик. Иначе он сможет увидеть достоинства и недостатки разных проектов, но не сможет оценить, какое место исследуемый продукт занимает среди них;

- увидеть изменения. Метрики хороши для регулярных тестов одного и того же продукта после внесения изменений. Они позволяют увидеть прогресс после внесенных изменений в интерфейсы, обратить внимание на те места, которые остались без улучшений. Использовать значения метрик исследователь может как доказательную базу, которая покажет руководству преимущества инвестиций в изменения пользовательских интерфейсов. Или хотя бы для понимания того, что исследователь достиг результатов и движется в нужном направлении;

- иллюстрировать, сделать акцент. Цифры хорошо помогают подчеркнуть важные проблемы. Иногда исследователи рассчитывают метрики для наиболее проблемных и (или) важных элементов теста.

Способ фиксации данных исследований

В современном мире разработки программных продуктов UX-исследователи должны стараться как можно быстрее выдавать команде разработчиков результаты своих наблюдений.

Чтобы оптимизировать время на анализ, следует заранее заготовить шаблон для ввода заметок в процессе теста. Для этого можно применять специализированное программное обеспечение (например, Noldus Observer или Morae Manager), но на практике наиболее гибкими и универсальными являются обычные таблицы. Для применения таблиц (рис. 2.1) в исследованиях необходимо заранее разметить и включить в них вопросы для респондентов, места для ввода информации о проблемах, обнаруженных в

заданий, а также гипотезы (для каждого респондента следует делать отметку о подтверждении или опровержении гипотезы).

Также можно воспользоваться более автоматизированными следующими решениями:

- Usability Test Data Logger от Userfocus [5]. Это шаблон Excel для ввода информации о наблюдениях для каждого респондента. В него встроен таймер, который измеряет время выполнения заданий, также автоматически генерируются графики времени и успешности;

- Rainbow Spreadsheet [6]. Наглядная таблица для совместной работы исследователя и команды разработчиков. Большинство записей можно делать прямо во время теста. Если исследователь не успел вовремя сделать записи, то ему лучше записать все, что он помнит, сразу после теста.

	Респондент 1	Респондент 2	Респондент 3	Респондент 4
Задание 1				
Заметил ли функцию А?				
В каком месте искал возможность Б?				
Проблемы и наблюдения по заданию				
...				

Рисунок 2.1 – Пример таблицы для фиксации данных

Подготовка к тестированию

Помимо метода, метрик и протокола проведения исследования, необходимо обратить внимание на следующее:

- **формат общения с исследователем.** Исследователь может находиться в одной комнате с респондентом. В этом случае ему будет легко вовремя задавать вопросы, однако присутствие исследователя может влиять на респондента. Он может начать задавать вопросы исследователю, провоцируя того явно или неявно ему подсказывать.

Лучше всего хотя бы в части процесса исследований оставлять респондента наедине с программным продуктом. В этом случае поведение респондента становится более раскованным и естественным. При этом для возможной связи с исследователем можно настроить микрофон;

– **способ постановки заданий.** Задания может озвучивать исследователь. Но в этом случае, несмотря на единый протокол тестирования, текст задания им может каждый раз произноситься по-разному. Особенно это актуально, если тестирование проводят несколько исследователей. Иногда даже небольшие различия в формулировках могут поставить респондентов в разные начальные условия.

Для того, чтобы избежать таких проблем, лучше выдавать респондентам задания в распечатанном виде на листах бумаги или экране монитора. Также необходимо обеспечить:

– **создание естественных условий.** Даже в лабораторных условиях сложно создать условия, которые были бы естественными для респондентов при тестировании. Здесь стоит подумать про обстановку, в которой будет использоваться программный продукт: отвлекает ли что-то респондента, шумно ли ему, хороший ли интернет, жарко или холодно и т.д.;

– **план тестирования для заказчика.** Это также важный этап подготовки, поскольку он вовлекает проектную команду разработчиков. Исследователь может не посвящать заказчика во все методологические особенности теста, но обязательно следует ему показать, какие будут задания и что планируется исследовать. Встречаются случаи, когда исследователь не может учесть особенности проекта, а может быть, у команды проекта появятся дополнительные идеи и гипотезы. Пример такого плана приведен на рис. 2.2;

– **план отчета.** Обычно отчет пишется после проведения исследования, но рекомендуется составлять план отчета еще до проведения тестов, также основываясь на целях и задачах исследования. С таким планом исследователь может проверить свой сценарий на полноту, а также подготовить наиболее удобные формы, чтобы зафиксировать данные для последующего анализа;

Текст задания	Что проверяем
«Вспомните, какую бытовую технику вы недавно покупали. Какие были критерии выбора? Давайте попробуем при помощи этого сайта подобрать вам что-то по этим же критериям»	• Находят ли правильную категорию • Заметность фильтров • Есть ли сложности с фильтром по цене • Достаточность фильтров • И т.д.

Рисунок 2.2 – План тестирования

– **протокол юзабилити-тестирования.** Как правило, он состоит из следующих частей: инструктаж или брифинг (приветствие, описание мероприятия, подписание документов); вводное интервью (небольшое интервью об использовании продукта, контексте и сценариях); работа с продуктом (задания тестирования); сбор итоговых впечатлений от продукта на основании опыта тестирования.

Подбор респондентов осуществляется путем предварительного анкетирования. В анкете необходимо указывать все важные критерии респондентов, которые также важны для исследований. Респонденты могут подбираться самими исследователями путем размещения объявлений в социальных сетях. Также подбором респондентов могут заниматься агентства по подбору персонала. Т.е. такая работа передается на аутсорсинг сторонней профильной компании.

В качестве примера ниже приводится план пользовательского тестирования (сценарий) удобства использования, который предлагается респондентам для ознакомления при исследовании сайта Киберфорума dessertcon.com, логотипов Киберфорума и флаера.

Пример плана пользовательского тестирования



План пользовательского
тестирования удобства использования (юзабилити)
(UX user experience scenario)

Продукты	Сайт конференции http://dessertcon.com , логотипы конференции и флаер киберфорума,
Подготовил	Игорь Короткий
Утвердил	Александр Гордеев

Содержание

1. Резюме;
2. Объект исследования;
3. Целевая аудитория;
4. Сценарии использования;
5. Метод;
6. Приложения.

1. Резюме

DeSSerT 2016 (Dependable Systems, Services and Technologies) – ежегодная международная научно-практическая конференция. Она посвящена вопросам гарантоспособности, надежности, информационной безопасности, кибербезопасности, зеленым технологиям и др. для критических и бизнес-критических доменов.

В этом году параллельно с конференцией будет проходить Cyber Forum.

Цель исследований – определить уровень удобства использования динамического интернет сайта конференции, выбрать наиболее привлекательный логотип конференции и проанализировать восприятие флаера киберфорума.

Результатом исследования будет краткий отчет.

К исследованию привлекается 11 респондентов.

2. Объект исследования

Публично-доступная версия веб-сайта конференции, включая функциональность для регистрации, варианты логотипов и флаер киберфорума.

Основная гипотеза следующая:

– веб-сайт эффективным (т.е. 80% пользователей ознакомятся с информацией о конференции и пройдут регистрацию менее чем за 10 мин.);

– при ознакомлении с информацией о конференции на сайте фиксации взгляда будут короче, а сакады длиннее, а при регистрации, наоборот, фиксации длиннее, а сакады короче.

3. Целевая аудитория

Планируется, что аудитория респондентов была следующая:

- возраст участников от 20 лет;
- уровень технической грамотности средний или высокий;
- респонденты должны иметь опыт участия в конференциях и проходить процедуру регистрации.

4. Сценарии использования

Во время данного исследования респондентам будет предложены следующие действия:

- ознакомиться с информацией о конференции на сайте;
- пройти процедуру регистрации на сайте конференции;
- выбрать один из вариантов предложенных логотипов конференции;
- ознакомиться с флаером киберфорума.

5. Метод исследования

5.1 Тип исследования

Выполнение пользователем заранее заданных сценариев использования приложения в присутствии исследователя с записью направления взгляда пользователя, без озвучивания действий и мыслей во время прохождения тестирования.

5.2 Оборудование

Исследование будет проводиться в отдельной комнате, где одновременно будут находиться исследователь, респондент и наблюдатель.

Респонденту будет предоставлен ноутбук с доступом в интернет с открытым в окне браузера веб-сайтом, перечень логотипов, флаер киберфорума в электронном виде и будут озвучены инструкции. Также для исследований будет применяться Eye-tracker X2 60.

После выполнения сценариев респондент ответит на заготовленные вопросы по сложности выполнения задач и озвучит свои предложения по улучшению сайта, флаера киберфорума и логотипа в произвольной для него форме.

5.3 Измеряемые и контролируемые параметры

Время выполнения заданий с разбивкой по этапам: ознакомление с информацией о конференции, регистрация на сайте, логотипа и ознакомление с флаером киберфорума.

Средняя продолжительность фиксаций и частота переходов будет измерена для каждого этапа отдельно.

5.4 Отчет

Отчет будет предоставлен в виде твердой копии с указанием полученных качественных и количественных параметров и предложений по улучшению интерфейса веб-сайта киберфорума, выборе логотипа и информацией о флаере киберфорума.

5.5 Гипотеза

Будем считать, что все респонденты справятся с заданием успешно. На выполнение заданий для веб-сайта у каждого респондента уйдет не более 10 минут.

6. Приложения

ПРИЛОЖЕНИЕ 1. АНКЕТА РЕСПОНДЕНТА

Респондент № _____

Возраст _____

Пол _____

Профессия _____

Как часто участвуете в конференциях:

☐ 1 раз в год; ☐ 3 раза в год; ☐ больше 5 раз в
☐ 2 раза в год; ☐ 5 раз в год; год.

Всегда ли регистрируетесь на сайте конференции:

☐ всегда; ☐ иногда;
☐ всегда, когда есть такая ☐ никогда.
возможность;

Можете ли Вы назвать сайты конференций с хорошей организацией процедуры регистрации:

☐ перечень сайтов: _____; ☐ не помню;

__ сайты такие есть, но назвать не __ вспомню и отвечу
могу; позже.
__ таких сайтов нет;

ПРИЛОЖЕНИЕ 2. ФОРМА РЕГИСТРАЦИИ ВЫПОЛНЕНИЯ ЗАДАНИЙ

Задание №1

Ознакомление с информацией на сайте конференции и регистрация

Успешное выполнение: да / нет

Время на ознакомление ____, время на регистрацию ____.

Задание №2

Выбор логотипа

Успешное выполнение: да / нет

Время на выбор ____.

Задание №3

Ознакомление с флаером киберфорума.

Успешное выполнение: да / нет

Время на ознакомление ____.

ПРИЛОЖЕНИЕ 3. ДОПОЛНИТЕЛЬНЫЙ АНАЛИЗ ЭМОЦИЙ РЕСПОНДЕНТА

Необходимо обвести картинку, которая соответствует эмоциональному состоянию (одна картинка на одно задание).

Задание №1. Ознакомление с информацией на сайте конференции и регистрация

Чтобы выполнить задание № 1 обходимо обвести соответствующее эмоциональное состояние при:

- ознакомлении с информацией на сайте конференции (рис. 2.3);



Рисунок 2.3. – Варианты эмоционального состояния при ознакомлении с информацией на сайте конференции

– ознакомлении с информацией на сайте конференции при регистрации (рис. 2.4);



Рисунок 2.4. – Варианты эмоционального состояния при ознакомлении с информацией на сайте конференции

Задание №2. Выбор логотипа

Чтобы выполнить задание № 2 обходимо обвести соответствующее эмоциональное состояние:



Рисунок 2.5. – Варианты эмоционального состояния при ознакомлении с логотипом конференции

Задание №3. Ознакомление с флаером киберфорума

Чтобы выполнить задание № 3 обходимо обвести соответствующее эмоциональное состояние:



Рисунок 2.6. – Варианты эмоционального состояния при ознакомлении с флаером конференции

Ход работы

1. Выбрать объект исследования и предварительно его изучить;
2. Разработать анкету для респондентов, которые будут участвовать в исследовании пользовательских интерфейсов;
3. Разработать сценарий исследования.

Требования к содержанию отчета

Отчет должен содержать следующее:

- анкету для респондентов;
- сценарий исследований.

Контрольные вопросы

1. С какой целью разрабатывается анкета для респондентов?
2. Каким образом разрабатывается сценарий исследований?
3. Можно ли проводить исследования без заранее подготовленного сценария?
4. С какой целью формируется гипотеза исследований?
5. Какие основные методы проведения исследований пользовательских интерфейсов Вам известны?
6. В чем особенность применения ретроспективного метода исследований?
7. Какие метрики успешности проведения исследований Вам известны?
8. Какие Вам известны инструментальные средства фиксации данных исследований?

3. МЕТОДИКА ПРОВЕДЕНИЯ АЙТРЕКИНГА И ВЕРБАЛЬНЫЕ ПРОТОКОЛЫ

Форма занятия: лабораторная работа

Цель – получить навыки проведения айтрекинга с использованием вербальных протоколов.

Образовательные задачи: изучить методику айтрекинга для исследования человеко-машинного взаимодействия.

Практические задачи: провести айтрекинг с использованием инструментального средства Tobii Studio.

Исследовательские задачи: исследовать влияние вербальных протоколов на качество полученных результатов исследований.

Теоретический материал

Методика проведения айтрекинга

Рассмотрим методику проведения айтрекинга для исследования удобства использования человеко-машинного взаимодействия. Она включает в себя следующие последовательные этапы (рис. 3.1):

- подготовительный этап (1) предназначен для проведения предварительной работы перед началом исследований. Включает в себя следующие работы: предварительный анализ объекта исследований, выбор типа айтрекера, разработку сценария исследования, формирование группы респондентов, подключение и настройка оборудования;

- исследовательский этап (2) предназначен непосредственно для проведения необходимых исследований. Включает в себя следующие работы: пилотное тестирование, ознакомление респондентов со сценарием, запись результатов исследований, визуализацию полученных результатов, выбор областей интересов, выбор и расчет метрик;

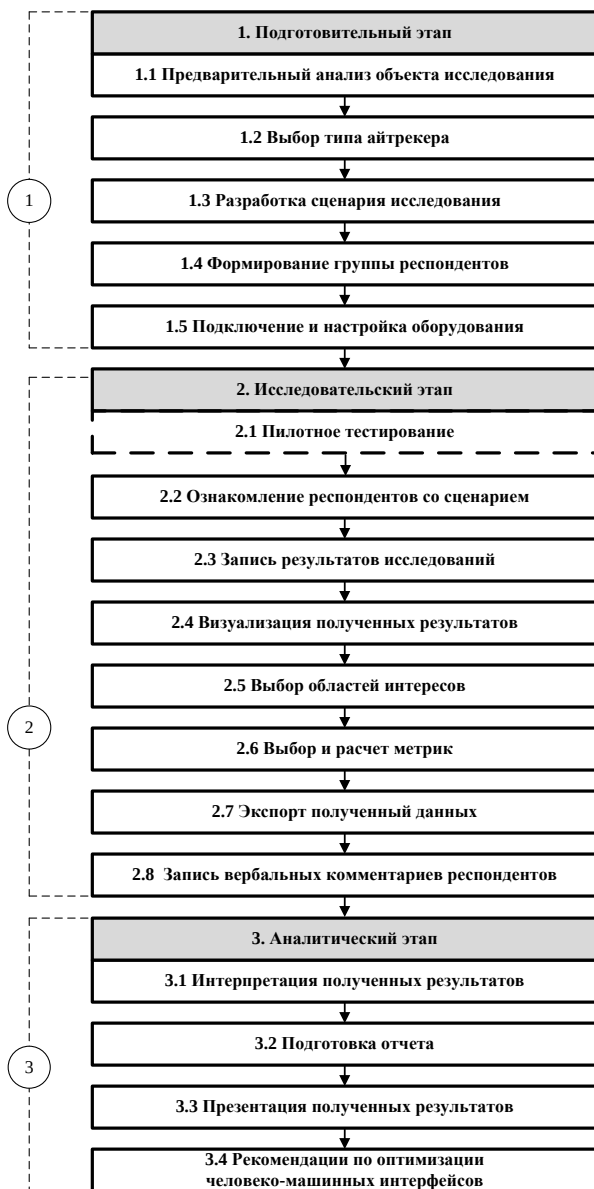


Рисунок 3.1 – Последовательность этапов методики оценивания usability человеко-машинного взаимодействия

– аналитический этап (3) предназначен непосредственно для анализа и представления полученных результатов. Включает в себя следующие работы: интерпретацию полученных результатов, подготовку отчета, презентацию полученных результатов, рекомендации по оптимизации человеко-машинных интерфейсов.

Рассмотрим более детально каждую работу.

1. Подготовительный этап

1.1 Предварительный анализ объекта исследований. Целью работы является анализ свойств и выявление особенностей объекта исследований. При выполнении такой работы необходимо проанализировать все множество свойств и особенности объекта. Допущенные неточности могут повлиять на результаты исследований в последующих этапах.

1.2 Выбор типа айтрекера. Целью такой работы является аргументированный выбор типа айтрекера. Тип айтрекера должен быть подобран с учетом особенностей исследований. Выбор необходимо сделать прежде всего между дистанционным и наголовным типами айтрекеров.

1.3 Разработка сценария исследования. Целью данной работы является разработка последовательности действий, которые должен выполнить каждый респондент во время проведения исследований. Здесь следует учесть следующее: сценарий должен быть описан вербально и заранее, т.е. до проведения реального исследования. Перед проведением исследований необходимо провести его верификацию. Как правило, ее проводят несколько независимых экспертов, имеющих опыт проведения таких исследований. Если в результате верификации было установлено, что сценарий является объективно сложным для выполнения и (или) чрезмерно длительным по времени, то, как правило, его разбивают на несколько составных частей.

1.4 Формирование группы респондентов. Целью такой работы является подбор участников для исследований, которые непосредственно вовлекаются в процесс исследования человеко-машинных интерфейсов программного обеспечения. В данной работе следует учесть, что критерии выбора респондентов должны коррелироваться с общей целью исследований и потенциальными пользователями исследуемых пользовательских интерфейсов.

Например, при проведении исследований человеко-машинных интерфейсов для программного обеспечения систем мониторинга информационной безопасности, необходимо формировать группу респондентов из специалистов, которые имеют опыт работы с подобными системами.

1.5 Подключение и настройка оборудования. Целью такой работы является подготовка оборудования (портативного компьютера, айтрекера и веб-камеры), его запуск и необходимая для проведения исследований настройка. Отметим, что в процессе исследования должны быть записаны данные полученные от айтрекера и веб-камеры. Также отметим, что на портативном компьютере должно быть установлено специализированное программное обеспечение, например, Tobii Studio, которое обеспечивает и поддерживает работу айтрекера.

2. Исследовательский этап

2.1 Пилотное тестирование. Перед проведением реальных исследований настоятельно рекомендуется провести пилотное исследование с айтрекером. Даже если все необходимые требования для проведения исследования соблюдены и исследователь полностью уверен в высоком качестве подготовки к исследованиям, пилотное тестирование часто помогает обнаружить неожиданные проблемы.

При проведении пилотного тестирования, как правило, преследуются 4 цели:

- первая цель заключается в устранении всех технических проблем. Исследователь должен быть уверен, что функции аппаратного и программного обеспечения айтрекера выполняются надежно и в полном объеме. Также успешно формируются все выходные файлы и данные. Если при работе программного и (или) аппаратного обеспечения обнаружены проблемы, то для их решения следует обратиться к службе технической поддержки компании-производителя;

- вторая цель заключается в проведении пилотного исследования для того, чтобы убедиться, что сценарий не имеет сложных для его выполнения участков и все инструкции поняты респондентам;

– третья цель такого тестирования заключается в том, чтобы исследователь получил опыт подготовки проведения конкретного исследования с учетом всех его особенностей. Достижение такой цели также очень важно, поскольку опытный модератор допускает меньше ошибок, получает более качественные данные и может контролировать продолжительность и качество исследований. Даже самым опытным модераторам рекомендуется проводить пилотное тестирование, особенно для сложных исследований;

– четвертая и последняя цель пилотного тестирования заключается в устранении недостатков в описательной части сценария. Даже если сценарий расписан достаточно подробно, существуют особенности, которые влияют на однозначность и точностью восприятия инструкций сценария. Для анализа описательной части сценария приглашаются также независимые эксперты, предпочтительнее из числа опытных исследователей.

Пилотное тестирование не является обязательной работой и необходимость ее проведения определяет сам исследователь.

2.2 Ознакомление респондентов со сценарием. Цель такой работы заключается в изучении респондентом описания последовательности действий, который он должен выполнить в рамках исследований. Следует учесть, что во время ознакомления со сценарием, респондент может задавать исследователю уточняющие вопросы. Если вопросы касаются, исследований и (или) сценария, респондент должен получить полные ответы, которые удовлетворяют его любопытство. Если сценарий в полном объеме не будет понятен респонденту или его некоторые вопросы останутся без ответа, это может негативно сказаться на результатах исследований.

2.3 Запись результатов исследований. Эта работа заключается в накоплении и сохранении данных о зрительных маршрутах респондентов, которые фиксирует айтрекер. Достаточно часто исследователи с использованием веб-камеры во время проведения исследований фиксируют мимику лиц респондентов.

2.4 Визуализация полученных результатов. Целью такой работы является предварительный просмотр и анализ полученных

визуальных результатов (зрительные маршруты, тепловые карты и др.).

2.5 Выбор областей интересов. Работа заключается в определении областей на объекте исследований, которые представляют собой, обычно, проблемные области и связи с этим, представляют наибольший интерес для исследователей. Фактически на этом этапе исследований происходит локализация площади объекта исследования.

2.6 Выбор и расчет метрик. В этой работе исследователь должен определяет профиль метрик, значения которых будут рассчитаны для каждой отдельной области интереса.

2.7 Экспорт полученных результатов. Такая работа заключается в экспорте необходимых результатов исследований. Например, к таким результатам могут быть отнесены статические или динамические иллюстрации формирования тепловых карт и зрительных маршрутов, расчетные значения метрик или же сформированные графики.

2.8 Запись вербальных комментариев респондентов. Исследовательский этап заканчивается важной работой, результаты которой применяются для уточнения и трактовки полученных результатов. В своих комментариях пользователи могут отразить свое мнение и указать на конкретные недостатки пользовательских интерфейсов.

3. Аналитический этап

3.1 Интерпретация полученных результатов. Цель работы заключается в максимально точной и полной трактовке полученных результатов исследований. Такая работа требует от исследователя определенной концентрации при анализе полученных результатов исследований, внимательности к деталям и опыта проведения таких исследований. Ее лучше выполнять в группе от 2 и более исследователей.

3.2 Подготовка отчета. Целью работы является наполнение и оформление документа результатами полученных исследований, их интерпретацией и предварительными выводами о выявленных недостатках в человеко-машинных интерфейсах.

3.3 Презентация полученных результатов. Такая работа заключается в представлении полученных результатов заказчику или представителям заказчика в понятных для них терминах.

3.4 Рекомендации по оптимизации человеко-машинных интерфейсов. Целью работы является набор конструктивных предложений, направленных на улучшение человеко-машинных интерфейсов.

Вербальные протоколы

Вербальные комментарии респондентов к задачам и отзывы о протестированных дизайнах, являются неотъемлемой частью многих UX-исследований, как с применением айтрекера, так и без него. Они предоставляют полезные качественные данные, которые углубляют понимание поведения и движения глаз респондентов в ходе исследования.

Параллельные и ретроспективные вербальные протоколы

Существует два основных типа вербальных протоколов: параллельные и ретроспективные. В параллельном вербальном протоколе (ПВП, также известном как «Протокол размышления вслух» или «Фиксация «мыслей вслух» (Thinking Aloud Protocol)), респонденты формулируют свои мысли в режиме реального времени во время выполнения задачи. При ретроспективном вербальном протоколе (РВП), как следует из названия, респонденты делятся своими воспоминаниями после завершения задания.

Оба вербальных протокола, как параллельный, так и ретроспективный, имеют свои преимущества и недостатки (табл. 3.1). В табл. 3.2 выполнено сравнение двух типов протоколов в зависимости от их влияния при сборе данных.

Выбор протокола для айтрекинга

При проведении айтрекинга, как правило, рекомендуется применять ретроспективный вербальный протокол, потому что комментарии респондентов, полученные после выполнения задачи, не могут повлиять на выполнение самой задачи настолько, на сколько при использовании параллельного вербального протокола. Если необходимы точные измерения производительности (например, время выполнения задачи или количество ошибок) и реалистичное отражение поведения

респондентов, то для айтрекинга необходимо использовать ретроспективный вербальный протокол.

Таблица 3.1 - Преимущества и недостатки параллельного и ретроспективного вербальных протоколов [7]

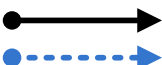
Вербальный протокол	Преимущества	Недостатки
ПВП (параллельный вербальный протокол или «протокол мыслей вслух») 	Респонденты произносят свои мысли вслух сразу, когда выполняют набор указанных задач	Вербализации могут негативно влиять на процесс выполнения задачи
РВП (ретроспективный вербальный протокол) 	Задача выполняется непрерывно	Респонденты по окончании выполнения задачи могут забывать причины, их выполнения
 Выполнение задачи Вербализация выполнения задачи		

Таблица 3.2 - Сравнение параллельного и ретроспективного вербальных протоколов

Критерий сравнения	Параллельный вербальный протокол (ПВП)	Ретроспективный вербальный протокол (РВП)
Влияние на продолжительность сессии	ПВП является более эффективным способом сбора обратной связи, чем РВП; его использование приводит к	РВП удваивает, а иногда даже утраивает время, затраченное на выполнение задач, поскольку задачи и вербальные отчеты

Критерий сравнения	Параллельный вербальный протокол (ПВП)	Ретроспективный вербальный протокол (РВП)
	сокращению продолжительности сессий тестирования, так как вербальный отчет и выполнения задачи происходят одновременно	являются последовательными
Влияние на исследователя и респондентов	ПВП является более привлекательным как для исследователей, так и респондентов, чем РВП, поскольку все действия комментируются, пока задача выполняется	Исследователям может надоесть просматривать выполнение каждой задачи по два раза, а респондентам ее выполнять, а затем просматривать процесс выполнения (это как смотреть фильм, а затем просмотреть его снова с комментариями режиссера)
Влияние на вербализацию	Оба (ПВП и РВП) могут создавать неполные словесные отчеты, потому что некоторые когнитивные процессы (более автоматизированные) не являются доступными для вербализации. Попытки объяснить эти действия могут привести к неправильному описанию. Качество вербализаций является участникомзависимым. Некоторые люди могут легче выражать свои мысли, чем другие. Проблемы самопрезентации	

Критерий сравнения	Параллельный вербальный протокол (ПВП)	Ретроспективный вербальный протокол (РВП)
	(например, желание казаться умным или страх смущения) могут также повлиять на вербальные отчеты	
	Говоря о своих мыслях в режиме реального времени, респонденты не успевают забыть то, о чем они думали в то время, когда выполняли задачу	Поскольку респонденты говорят о своих мыслях после того, как они выполнили задачу, они могут быть не в состоянии вспомнить все. Это может привести к уменьшению числа вербализаций или к (непреднамеренно) сфабрикованным объяснениям. Чтобы нивелировать ухудшение памяти, РВП следует вводить сразу после каждой задачи, задачи должны быть короткими, и следует использовать стимуляторы памяти (например, видеовоспроизведения задания)
Влияние на выполнение	ПВП ухудшает производительность и влияет на показатели производительности	Поскольку респонденты опрашиваются после того, как задача выполнена, РВП не

Критерий сравнения	Параллельный вербальный протокол (ПВП)	Ретроспективный вербальный протокол (РВП)
	. Если время выполнения задачи измеряется, размышление вслух часто увеличивает это время. ПВП также влияет на количество успешно выполненных задач. Для некоторых задач ПВП может улучшить производительность, помогая структурировать мысли респондентов. Для других задач ПВП может увеличить познавательную нагрузку на респондентов и ухудшить выполнение задачи.	влияет на поведение респондентов и на показатели производительности, особенно во время выполнения задачи. Тем не менее, респонденты, которые знают, что их будут опрашивать после выполнения задачи, могут стараться выполнить задачу более вдумчиво и запоминать, что они делали
Влияние на движение глаз	ПВП влияет на движение глаз, потому что респонденты, вербализируя свой мыслительный процесс, выполняют больше фиксаций и, как правило, исследуют больше	РВП, как правило, не оказывает никакого влияния на движение глаз, но если респонденты знают, что после выполнения задачи их будут опрашивать, они могут подойти к выполнению задачи

Критерий сравнения	Параллельный вербальный протокол (ПВП)	Ретроспективный вербальный протокол (РВП)
	областей, чем обычно	по-разному, что может повлиять на их результаты айтрекинга

При использовании ретроспективного вербального протокола необходимо учитывать следующие аспекты:

- для каждой сессии необходимо дополнительное время;
- для того, чтобы помочь респондентам вспомнить свои действия и мысли, им нужны подсказки, например, видеоповторы задачи;
- респонденты будут забывать меньше, если задачи будут короткими и если, словесный опрос будет проводиться сразу же после каждой задачи, а не по завершению всех задач.

Необходимо учитывать тот факт, что респонденты могут начать комментировать человеко-машинное взаимодействие или свои действия, даже если их об этом не просят. Чтобы этого не произошло, можно предложить респондентам выполнить задачу так, как будто бы они находятся дома одни, или вообще выйти из лаборатории.

Параллельный вербальный протокол также используется в айтрекинге. Чаще всего он используется в исследованиях, которые направлены на выявление проблем удобства использования. В этих исследованиях айтрекинг часто используется в качестве инструмента для наблюдения, а не с целью анализа и обобщения данных о движении глаз впоследствии. Таким образом, даже если респонденты смотрят на вещи, о которых они говорят больше, чем обычно, это не является огромной проблемой. Преимуществами параллельного вербального протокола в данном случае является сокращение тестовых сессий, более простая работа с респондентами, отсутствие недостатков, связанных с необходимостью запоминания информации.

Подсказки, используемые в ретроспективном вербальном протоколе

При использовании ретроспективного вербального протокола после выполнения каждой задачи исследователь просит респондентов описать свои мысли во время выполнения задачи. Проблема, связанная с таким подходом основана на долгосрочной памяти, которая может быть ошибочной. Респондент через определенное время может не помнить, например, какие действия он выполнял в начале задания, поэтому ему необходимо давать соответствующие подсказки.

Есть три типа подсказок, которые часто используются при ретроспективном вербальном протоколе:

- **сам стимул**: респондента просят повторить то, как он выполнял задачу, но на этот раз, озвучивая мысли, которые были у него, когда он в первый раз выполнял задачу;

- **видеоповтор действий оператора**: респонденту показывают видеовоспроизведение процесса выполнения им задачи, и в это время он описывает свои мысли на каждом этапе процесса выполнения задачи;

- **видеоповтор айтрекинга**: респонденту показывают видеовоспроизведение процесса выполнения им задачи с маршрутом движения его глаз.

В табл. 3.3 представлено сравнение трех видов подсказок на основании их влияния на память респондента и их влияния на устный отчет.

Примеры инструкций респондентам для проведения исследований айтрекинга с использованием ретроспективного вербального протокола

Перед исследователем UX-тестирования стоят две важные задачи:

1. Обеспечить респонденту возможность выполнять задачи в условиях, наиболее приближенных к естественным;

2. Получить максимум информации о причинах действий респондента и о его эмоциях, вызванных работой с программным продуктом.

Часто модераторы зацикливаются именно на второй задаче, постоянно задавая респонденту вопросы «почему?», «зачем?»,

«как?» и т. д. Но на самом деле главное для исследователя – «научиться правильно молчать», поскольку в реальной ситуации респондент будет пользоваться продуктом самостоятельно. Задавая слишком много вопросов, фокус внимания респондента может смещаться с самого программного продукта на его обсуждение, а для исследования наиболее ценными являются наблюдения за деятельностью респондента. Вопрос может сбить его с мысли и отвлечь от задачи. Возможно, вернувшись к задаче, респондент начнёт действовать по-другому. Поэтому во время проведения исследования рекомендуется использовать ретроспективный вербальный протокол, т.е. задавать вопросы после проведения тестирования. Или, в зависимости от задачи, необходимо дождаться, когда респондент, выполняя задание, дойдёт до логической завершающей точки, и потом задавать вопросы.

Таблица 3.3 - Сравнение подсказок, используемых при ретроспективном вербальном протоколе

Подсказка	Влияние подсказки на память	Влияние подсказки на вербальный отчет
Сам стимул	Имея возможность, лучше обратиться к стимулу (например, при тестировании веб-сайта), чем попросить прокомментировать действия при выполнении задачи устно без каких-либо подсказок. Респонденты могут забывать свои действия, поэтому исследователь должен быть готов им о них напоминать	Сам стимул не оказывает деструктивного влияния на устное описание процесса выполнения задачи, но может стать отвлекающим моментом, и респонденты начнут забывать свои действия
Видеоповтор	Видеовоспроизведение помогает респондентам	Видеовоспроизведение редко нарушает устный

	вспомнить свои действия во время выполнения этой задачи	отчет, особенно если респонденты имеют право приостановить воспроизведение, когда это необходимо
Повтор айтрекинга	Движения глаз обеспечивают дополнительные подсказки. Респонденты могут видеть не только то, что они сделали, но и то, куда они смотрели, что должно помочь им правильно восстановить свои мысли. Такие подсказки особенно полезны при длительных паузах без каких-либо действий (например, когда респонденты проводят много времени на веб-странице, прежде чем кликнуть на что-то). Несмотря на повтор маршрута движения глаз, респонденты могут по-прежнему забывать, почему они куда-то смотрели, куда они смотрели и о чем они думали	Повтор маршрута движения глаз может помешать вербальному отчету из-за первоначального удивления респондентов и увлечению просмотром именно движения глаз. Кроме того, респонденты, как правило, не готовы обрабатывать большое количество информации, которая поступает при таких повторях. Эти эффекты могут быть уменьшены путем напоминания респондентам о задачах, которые они выполняли, чтобы они привыкли видеть маршрут движения своих глаз. Также можно замедлять воспроизведение и дать возможность респондентам приостанавливать и перематывать видео. Просмотр маршрута движения собственных

		глаз также может помочь респондентам более осознанно описать свое поведение взгляда в последующих задачах
--	--	---

Ниже приведены примеры инструкций, которые можно использовать при проведении айтрекинга с использованием ретроспективного вербального протокола:

Исследователь дает респонденту указания для выполнения задачи.

Пожалуйста, выполните задачу таким образом, как если бы Вы находились дома. Если у вас есть какие-либо отзывы о поставленной задаче или тестируемом веб-сайте по ходу выполнения задачи, Вы будете иметь возможность поделиться ими со мной чуть позже. А сейчас, пожалуйста, выполняйте задачу так, как будто меня здесь нет.

Модератор напоминает респондентам инструкции по выполнению задачи. Респондент выполняет задачу.

Давайте поговорим о том, что Вы только что сделали. Иногда бывает трудно вспомнить все, что происходило во время выполнения задачи, поэтому для того, чтобы все вспомнить, мы будем просматривать видеозапись процесса выполнения Вами задачи. Пока мы смотрим видео, я хотел бы, чтобы Вы рассказали мне обо всем, о чем Вы думали в этот момент.

Модератор включает видеоповтор с маршрутом движения глаз.

Движущаяся точка указывает на место, в котором Вы осуществляли поиск. Мы воспроизводим видео на половинной скорости, поэтому у Вас есть время объяснить мне, все, о чем Вы думали в эти моменты. Вы можете сделать паузу или перемотать видео в любое время с помощью этих кнопок.

Модератор демонстрирует работу кнопок управления для остановки видеовоспроизведения, а затем перематывает видео к началу.

Пожалуйста, расскажите мне, о чем Вы думали, когда выполняли это задание.

Если респондент перестает проговаривать информацию, тогда исследователь задает наводящие вопросы.

О чем Вы думали в этот момент?

Не спрашивайте респондентов, почему они смотрели на те элементы, которые попали в маршрут движения их глаз. Респонденты должны использовать информацию о движении взгляда как подсказку, чтобы вспомнить задание, а не для того, чтобы оправдать свои действия (в случае трудностей выполнения задачи).

Ретроспективный вербальный протокол с использованием видеоповтора маршрута движения глаз

Мы выяснили, какие виды подсказок могут быть использованы при ретроспективном вербальном протоколе и провели их сравнение в зависимости от их влияния на память и на вербальный отчет. Теперь необходимо выяснить, какие виды подсказок рекомендуется использовать при проведении айтрекинга с использованием ретроспективного вербального протокола. Мы проведем сравнительный анализ использования РВП с использованием видеоповтора маршрута движения глаз с другими протоколами.

Исследования по оценке эффективности использования различных вербальных протоколов были проведены с использованием различных метрик, среди которых в качестве метрик использовали количество слов, полученных от респондентов, количество сделанных замечаний, количество обнаруженных проблем с удобством использования, рейтинги удовлетворенности пользователей с использованием вербального протокола.

Более детальное описание особенностей использования ретроспективного вербального протокола можно найти в источниках, которые представлены в табл. 3.4.

Таблица 3.4 - Сравнение РВП с использованием видеоповтора маршрута движения глаз с другими протоколами в различных исследованиях

Автор	РВП с использованием видеоповтора	Параллельный вербальный протокол	РВП без использования подсказок
-------	-----------------------------------	----------------------------------	---------------------------------

3. Методика проведения айтрекинга и вербальные протоколы

	процесса выполнения задания		
Elling S. [7]	+	-	-
Hansen J. P. [8]	+	-	-
Hyrskykari A [9]	-	+	-
Russo J.E. [10]	+	+	-
Eger N. [11]	+	+	-
Tobii Technology [12]	-	-	+

Даже в исследованиях, в которых использовались аналогичные метрики, сравнение результатов оказалось сложной задачей по двум причинам. Во-первых, ретроспективный вербальный протокол с использованием видеоповтора айтрекинга не всегда сравним с другими протоколами с использованием таких же метрик. Во-вторых, необходимо учитывать особенности использования видеоповтора маршрута движения глаз. В некоторых исследованиях видеоповтор показывали респондентам на полной скорости [7], в других – на половинной скорости [9, 12]. В одном случае респондентам предоставлялась возможность контролировать видеовоспроизведение, например, делая паузу или замедляя его [10, 12], а в другом это было запрещено [7].

Сравнение РВП с использованием видеоповтора маршрута движения глаз с другими протоколами и ПВП

Несмотря на все различия между исследованиями, использование РВП с видеоповтором маршрута движения глаз дает «лучшие» результаты, чем параллельный вербальный протокол, за счет использования респондентами большего количества слов в вербальном отчете [9, 10], предоставления большего количества комментариев [9], большего количества обнаруженных проблем в удобстве использования [9, 11], и большей удовлетворенности респондентов использованием в исследовании вербального протокола [11].

В то же время, иногда использование параллельного вербального протокола по сравнению с РВП с видеоповтором маршрута движения глаз оказывается более предпочтительным, что связано с получением большего количества слов от респондентов в РВП протоколе. Например, если проводимое исследование имеет жесткие временные рамки, ПВП обеспечит более короткие сессии.

Если сравнивать видеоповторы айтрекинга РВП и РВП, то нельзя четко определить, какой из них является более предпочтительным. Использование РВП с видеоповтором айтрекинга дает большее количество слов в вербальном отчете по сравнению с РВП [8, 10], но в результатах исследований компании Tobii [12] такое преимущество РВП не было обнаружено. Использование видеоповтора действий респондента с подсказками и айтрекинга привели практически одинаковому количеству комментариев [8, 12] и количеству обнаруженных проблем с удобством использования [7, 11, 12]. В тоже время в исследовании [7] отмечается, что респонденты предпочитают просматривать именно айтрекинг, а не видеоповтор действий респондента с подсказками.

Результаты этих исследований указывают на то, что улучшения в части воспоминаний респондентов, которые возникают вследствие просмотра видеоповтора айтрекинга, могут быть нивелированы самим характером подсказок, то есть респонденты могут отвлекаться на сам видеоповтор айтрекинга. Поэтому респонденты, которые знакомы с процессом айтрекинга, меньше отвлекаются при его воспроизведении, чем респонденты, которые впервые видят видеоповтор айтрекинга.

Целевые методы исследования

РВП с видеоповтором айтрекинга основан на том, что респонденты, видя движения своих глаз, лучше будут вербализировать свои действия и мысли. Другой способ с видеоповтором айтрекинга – это целевое исследование. При его использовании исследователь постоянно наблюдает за поведением и движением глаз респондента во время выполнения задачи и, основываясь на этой информации, просит респондента подробно остановиться на интересующих исследователя деталях.

Существует два варианта целевых исследований: триггерный параллельный вербальный протокол и выборочные напоминания.

Триггерный параллельный вербальный протокол

Триггерный ПВП является альтернативой ПВП. Он заключается в том, что респонденту не требуется озвучивать свои мысли, параллельно выполняя задачи. Вместо этого, исследователь опрашивает респондента, о чем он думает, когда его поведение (в том числе поведение движения глаз) представляет интерес для исследования. Такой подход требует возможности просмотра маршрута движений глаз респондента в режиме реального времени (рис. 3.2).

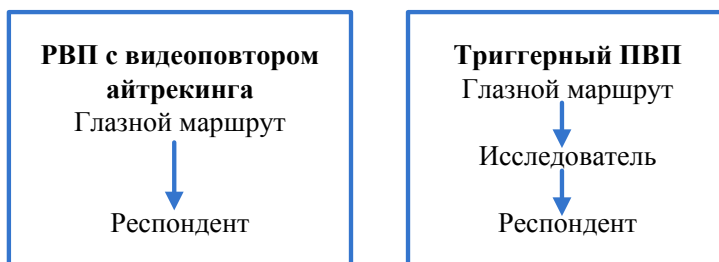


Рисунок 3.2 - Сравнение вербальных протоколов

Выборочные напоминания

Выборочные напоминания являются альтернативой ретроспективному вербальному протоколу. Во время выполнения задачи, исследователь делает заметки о любом интересном поведении респондента (в том числе и его движении глаз) и просит респондента прокомментировать интересующие исследователя моменты выполнения задачи уже после завершения исследования. Респондентам не показывают видеоповтор процесса выполнения всей задачи для того, чтобы они черпали информацию из своей памяти о том, что и почему они делали в определенный момент проведения исследований. Респондентам показывают только часть стимула (например, веб-страницу, на которой возникла проблема) и просят, описать мысли, которые возникали в момент выполнения задачи. Для активации памяти респондентов могут применяться в

качестве подсказок отрывки видеоповтора действий респондента или видеоповтора айтрекинга выполнения задачи.

Преимущества целевых методов исследования

Существует несколько преимуществ использования целевых методов исследования по сравнению с традиционными вербальными протоколами. Целенаправленное исследование гарантирует то, что респонденты будут говорить о том, что имеет отношение к целям исследования. Это связано с тем, что при использовании традиционных вербальных протоколов респонденты могут много говорить об аспектах выполнения задачи, не связанных с фокусом исследования.

Целевое исследование укорачивает тестовые сессии, особенно когда выборочные напоминания используются вместо ретроспективного вербального протокола. Но при использовании видеоповтора действий респондента или видеоповтора айтрекинга как сигналов для активации памяти, может потребоваться дополнительное время для нахождения интересующих отрывков в видеоповторе. Чтобы сделать этот процесс более эффективным, исследователь должен делать пометки в видеозаписи или, по крайней мере, записывать интересующие его временные интервалы.

Использование триггерного «протокола мыслей вслух» вместо традиционного параллельного вербального протокола уменьшает смысловую нагрузку на респондентов, связанную с непрерывной вербализацией. Однако это происходит за счет увеличения нагрузки на исследователя. Исследователь должен обращать внимание на действия респондента и на движение его глаз, и быстро распознавать события, которые могут быть интересны для исследования и требуют более детального внимания к ним. Таким образом, проводить исследования с использованием триггерного «протокола мыслей вслух» и выборочных напоминаний могут только более опытные исследователи.

Ход работы

1. Выбрать соответствующий вербальный протокол для проведения исследований;
2. На основе разработанного сценария провести исследование человеко-машинного взаимодействия с использованием айтрекинга в соответствии с последовательностью этапов, которая представлена на рис. 3.1. и соответствующим вербальным протоколом;
3. Результаты исследований отразить в отчете.

Требования к содержанию отчета

Отчет должен содержать следующее:

- результаты каждого этапа исследований;
- обоснование выбора типа вербального протокола.

Контрольные вопросы

1. Перечислите основные этапы проведения айтрекинга?
2. В чем заключается подготовительный этап исследований и какие работы в него входят?
3. В чем заключается исследовательский этап исследований и какие работы в него входят?
4. В чем заключается аналитический этап исследований и какие работы в него входят?
5. Что такое пилотное тестирование? Когда и для чего его проводят?
6. Что такое вербальный протокол?
7. Перечислите основные виды вербальных протоколов и укажите их особенности.

4. ВИЗУАЛИЗАЦИЯ ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ, ВЫБОР ОБЛАСТЕЙ ИНТЕРЕСОВ И МЕТРИЧЕСКАЯ ОЦЕНКА

Форма занятия: лабораторная работа

Цель – получить навыки анализа визуальной информации, выбора областей интересов, профилирования метрик и интерпретации полученных расчетных результатов.

Образовательные задачи: изучить технику профилирования метрик и выбора зон интересов.

Практические задачи: интерпретировать полученные результаты расчета метрик и визуальной информации.

Исследовательские задачи: в соответствии с аналитическим этапом методики провести интерпретацию полученных расчетных значений метрик визуальной информации.

Теоретический материал

Визуализация полученных результатов

Рассмотрим возможные варианты визуализации полученных результатов:

Зрительные маршруты (зрительные графы)

Зрительные маршруты – это тип визуальной интерпретации полученных результатов исследований. Зрительные маршруты включает последовательность фиксаций (окружностей) и саккад (линий) (рис 4.1) [13]. Размер точки пропорционален длительности фиксации, чем фиксация длиннее, тем больше размер точки. Нумерация и размер точек фиксации зависят от настроек, которые определяются перед началом исследований. Поскольку фиксации пронумерованы, поэтому они передают не только пространственную, но и временную информацию.

Зрительные маршруты для видео

Зрительные маршруты для видео – это динамическая визуализация, представленная как движущиеся точки или

перекрестия. Они накладываются сверху на стимул, например, веб-сайт, картинки или видео записи (рис. 4.2) [14].

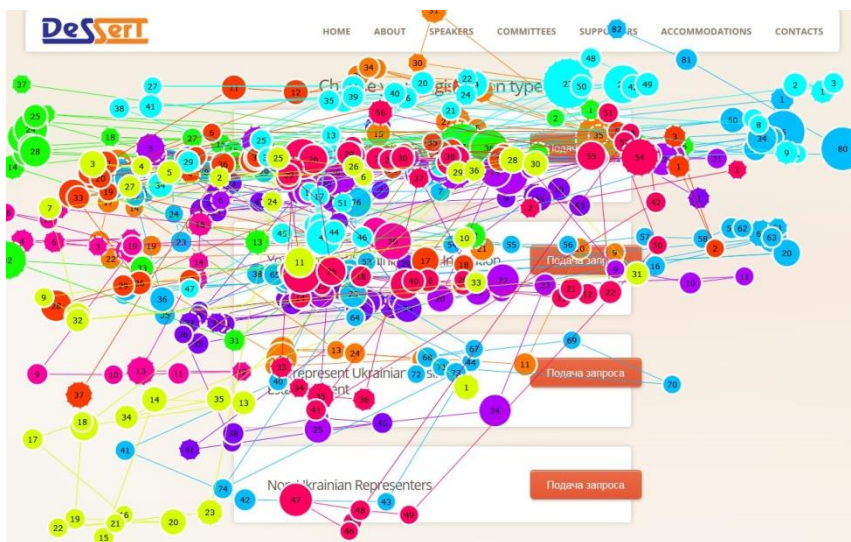


Рисунок 4.1 – Зрительные маршруты для веб-страницы страницы регистрации сайта киберфорума [13]

Повторы зрительных маршрутов для видео можно смотреть в режиме реального времени посредством интерактивной программы просмотра. В UX исследованиях такая возможность позволяет исследователю придумать вопросы для респондента во время сессии, не только на основе того, что респондент делает, но и на основе того, на что он смотрит. Кроме того, в режиме реального времени зрительные маршруты для видео также делают исследования более привлекательным для заинтересованных сторон, особенно когда респондент испытывает сложности в вербальном изложении мыслей или, когда задачи должны быть выполнены в полной тишине.

Тепловая карта

Тепловая карта – это двумерное представление данных, в котором значения айтрекинга показаны в виде цветов. Тепловые

4. Визуализация полученных результатов, выбор областей интересов и метрическая оценка

карты имеют интуитивный характер, поскольку их цветовая гамма соответствует общепринятому представлению температуры: желтый теплее, чем зеленый или синий, оранжевый теплее, чем желтый, и красный – горячий (рис. 4.3) [13]. Т.е. количество «тепла» пропорционально уровню представляемой меры.



Рисунок 4.2 – Зрительные маршруты для видео наложенные на видео «Earth Day: Give Earth a Hand» организации «Greenpeace»

(сверху экранная форма видео без зрительных маршрутов для видео, снизу видео со зрительными маршрутами для видео)

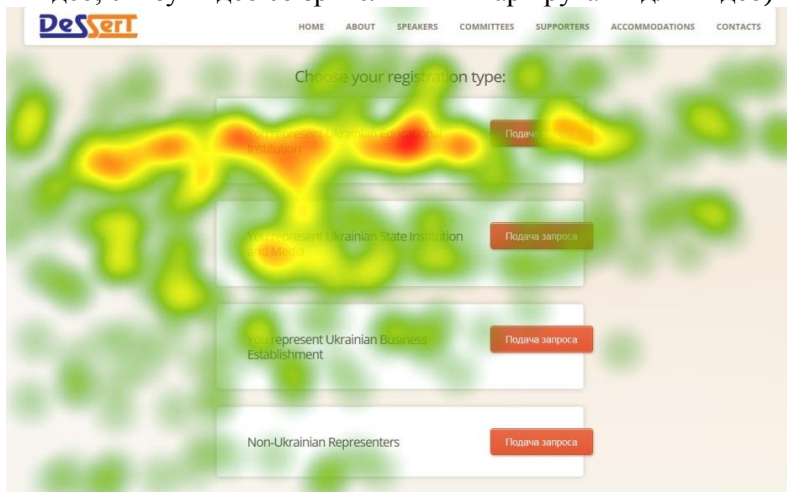


Рисунок 4.3 – Тепловая карта, наложенная на регистрационную форму сайта киберфорума [13]

Как правило, пособие для ПО айттрекинга подробно объясняют алгоритмы, используемые в программах для создания тепловых карт. Основная идея заключается в том, что все фиксации проецируются на фоновое статическое изображение. Цвету отведена небольшая зона непосредственно вокруг центра каждой фиксации. Чем ближе к центру фиксации, тем значения цвета будут выше. Значения перекрывающихся областей складываются, что обеспечивает плавные цветовые карты. Таким образом, чем выше плотность фиксаций в одной локации небольшой площади, тем теплее цвет тепловой карты.

Размер оптимальной выборки респондентов для получения тепловой карты до конца не установлен. Существуют несколько мнений по этому поводу. Рассмотрим наиболее известные из них. Можно говорить, что минимальная выборка респондентов это один респондент, однако существует мнение, что лучшей выборкой может являться 30 респондентов.

Идея о 30 респондентах в качестве оптимального размера выборки для айттрекинга возникла в результате появления онлайн-

документ, который подготовили Кара Пернис и Якоб Нильсен [15]. По мнению авторов, 30 (или иногда 32) респондентов могут производить «стабильную» достоверную тепловую карту. Такое число респондентов было получено авторами упомянутого документа в результате следующих манипуляций с данными исследований. Они взяли данные айтрекинга для 60 респондентов и создали шесть тепловых карт, каждая из которых содержала данные 10 различных респондентов. Когда тепловые карты были сопоставлены друг с другом, было установлено, что они слишком разнятся между собой. Когда всех респондентов разделили на три группы по 20 человек, аналогичным образом проанализировали тепловые карты, которые также не были похожими. Только тогда, когда были созданы две тепловые карты, каждая из которых включала данные от 30 респондентов, авторы пришли к выводу, что такие карты являются схожими. Поэтому Пернис и Нильсен пришли к выводу, что 30 респондентов является оптимальной выборкой респондентов.

Другого мнения придерживается Ага Бойко [3]. В своих работах, посвященных исследованиям с использованием айтрекинга, она не согласна с тем, что минимальным количеством респондентов является число 30. Для опровержения такой теории она приводит следующие доводы:

- во-первых, сходство тепловых карт оценивали визуально, осматривая невооруженным взглядом. Статистические или другие научные методы использованы не были;

- во-вторых, особенности исследовательских данных при формировании тепловой карты не были учтены. Изменяя, например, верхний порог цветовой гаммы, можно достаточно хорошо корректировать идентичность тепловых карт для визуальной идентификации;

- в-третьих, подгруппы из 10, 20 и 30 респондентов, казалось, были сформированы только один раз из большей группы 60 респондентов. Более предпочтительным подходом можно считать формирование таких групп случайным образом.

Таким образом, тепловые карты могут быть получены на основе данных от одного респондента, однако они являются

гораздо более точными в качестве заполнителя для визуализации данных, полученных от нескольких респондентов.

Карта непрозрачности

Карта непрозрачности (рис. 4.4) представляет собой тип результата, который отображает области, в которых имели место визуальные фиксации респондента. Его часто называют «картой реверсного тепла», потому что, хотя результат подобен тепловой карте, районы, на которые респондент фокусировал свое зрение просматриваются тем отчетливее, чем больше на них фокусировали свое внимание респонденты. Те части экрана, на которые респонденты меньше всего обращали свое внимание, отображаются более темными [16].

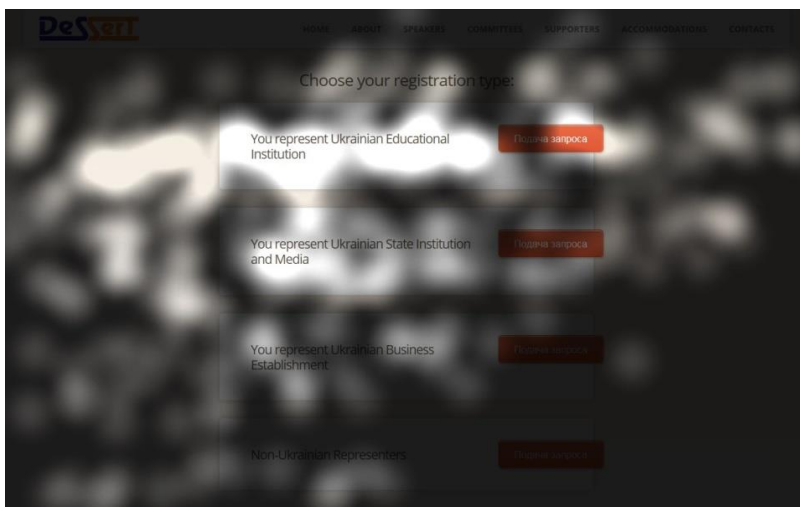


Рисунок 4.4 – Карта непрозрачности, наложенная на регистрационную форму сайта киберфорума [13]

Пчелиный рой

Чтобы проверить визуальные шаблоны, которые были агрегированы для всех пользователей, можно использовать инструмент анализа результатов айтирекинга – пчелиный рой. Этот инструмент отображает точки созерцания во времени для всех

респондентов в виде маленьких точек. Двигаясь по временной шкале, эти точки напоминают пчелиный рой, где, несмотря на явно хаотичные движения отдельных членов роя, общее внимание респондентов, по-видимому, следует некоторыми более строгими правилами. Как правило, в первые секунды респонденты смотрят на самые интересные элементы исследуемого объекта. Для иллюстрации возможностей пчелиного роя мы рассмотрим, например, исследование веб-сайта киберфорума [13]. Для этого респондентов попросили посмотреть на одну веб-страницу киберфорума [13]. После первой секунды исследования внимание респондентов сосредоточено только на одном элементе веб-страницы киберфорума (рис. 4.5). Аналогичная процедура была проведена через 2 секунды (рис. 4.6). Далее после 5 секунд исследований внимание респондентов концентрируется на других элементах объекта (рис. 4.7).

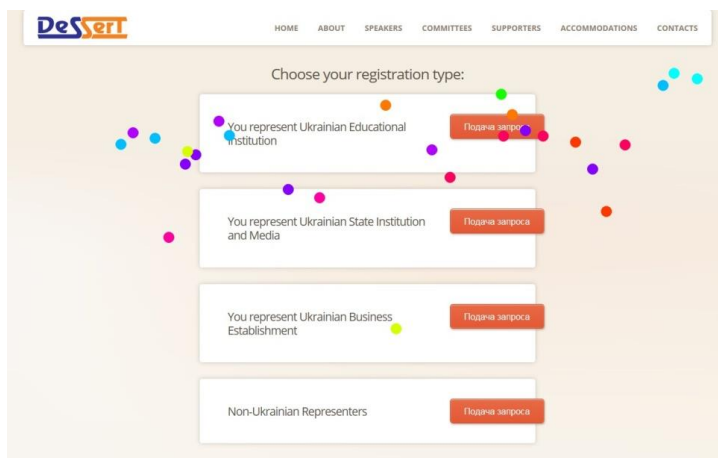


Рисунок 4.5 – Пчелиный рой [13] после первой секунды просмотра веб-страницы киберфорума

Кластеры

Инструмент кластеров (рис. 4.8) визуализирует области визуального внимания несколько иначе, чем инструмент тепловые карты. Для определения областей кластера устанавливается пространственный порог, который ограничивает расстояние между

4. Визуализация полученных результатов, выбор областей интересов и метрическая оценка

двумя фиксациями перед разделением его на разные кластеры. Для каждого кластера анализируется информация о том, какой процент респондентов, имеющих фиксации, включены в кластер. С помощью этого инструмента исследователь может также увеличивать масштаб по временным сегментам.

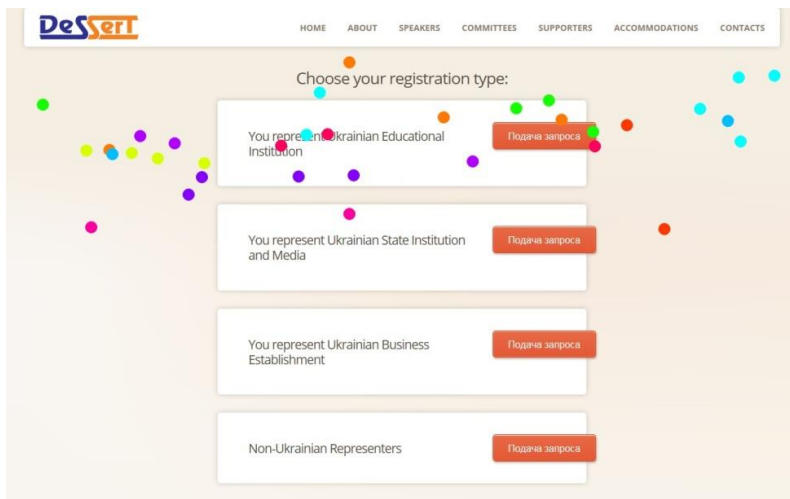


Рисунок 4.6 – Пчелиный рой [1] после второй секунды просмотра веб-страницы киберфорума

4. Визуализация полученных результатов, выбор областей интересов и метрическая оценка

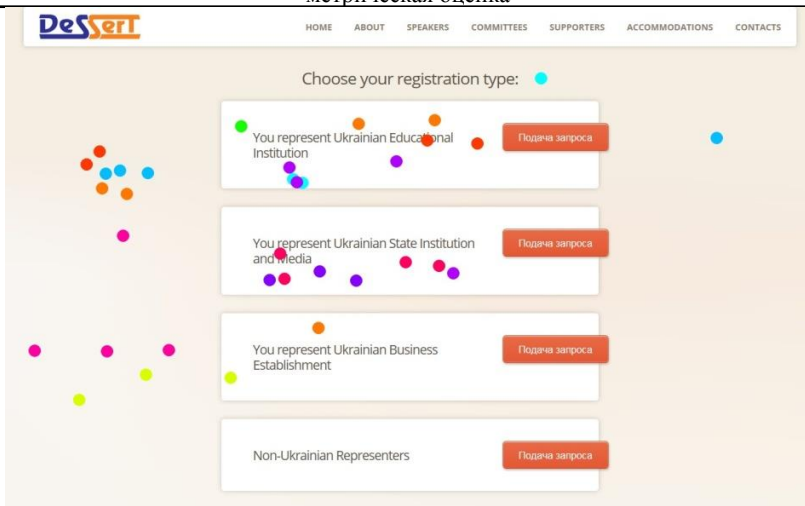


Рисунок 4.7 – Пчелиный рой [13] после пятой секунды просмотра веб-страницы киберфорума



Рисунок 4.8 – Кластеры страницы киберфорума [13]

Области интересов

Айтрекинг, как правило, применяется для исследования, например, не целого веб-сайта и даже не веб-страницы целиком. Как правило, объектом исследования являются отдельные части веб-страниц, которые требуют более детального дополнительного исследования с использованием айтрекера. Такие проблемы не могут быть объяснены и устранены при помощи применения наименее технологичных методов. Например, к таким проблемам на веб-страницах могут быть отнесены следующие:

- почему пользователи интернет-магазина не нажимают кнопку купить, после того, как наполнили виртуальную корзину покупками?
- почему пользователи интернет-сайтов не замечают или не нажимают на баннеры сайта;
- какой из вариантов дизайна веб-сайта лучше?
- новый дизайн сайта не хуже ли существующего?
- какие элементы веб-сайта попадают в фокус внимания пользователя?

Программное обеспечение Tobii Studio поддерживает возможность определения как единичных (рис. 4.9) так и множественных (рис. 4.10) областей интересов.

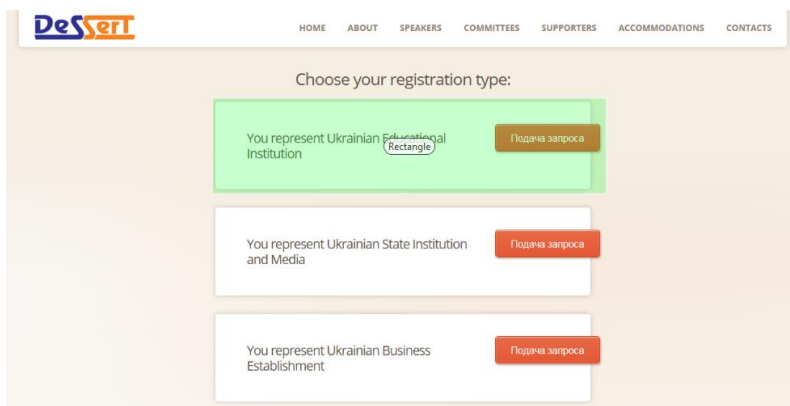


Рисунок 4.9 – Единичная область интересов

Метрики оценивания

4. Визуализация полученных результатов, выбор областей интересов и метрическая оценка

Метрики дают исследователям возможность рассчитать их величину для одного объекта исследования и сравнить ее с величиной такой же метрики, рассчитанной для другого объекта исследования. Метрики айтрекинга, как правило, помогают провести более детальное измерение, т.е. не только на целевой основе, но и для отдельных компонентов задач.

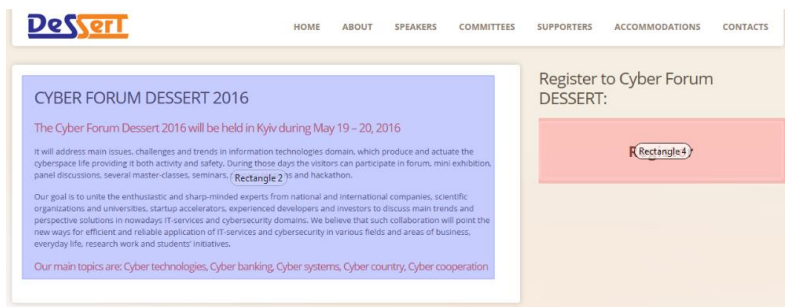


Рисунок 4.10 – Множественные области интересов

Отметим, что важным аспектом является выбор необходимой номенклатуры метрик для исследования с использованием айтрекинга с учетом особенностей объекта исследования.

Каждый исследователь перед началом исследования должен принять решение о том, какие метрики он будет применять при исследовании, и на какие вопросы он должен получить ответы от респондентов, чтобы получить необходимую информацию от них для расчета метрик. Т.е. исследователь должен сформировать соответствующий профиль метрик.

Существует свыше ста метрик айтрекинга. Некоторые из них достаточно просто рассчитать и соответственно, интерпретировать полученные таким образом результаты. К ним можно отнести следующие: суммарное время, которое потратил респондент, рассматривая какой-либо объект или какое количество респондентов заметили и рассматривали объект исследования.

Отметим, что исходные данные для расчета таких метрик достаточно просто получить на основе результатов исследований.

Одни и те же метрики могут представлять собой различные когнитивные явления в контексте разных стимулов и целей.

Например, о чем исследователям может говорить ситуация, когда респондент смотрит на объект исследования достаточно долго. С одной стороны, это может означать, что объект заинтересовал респондента и тем самым привлек его внимание, а с другой стороны, может иметь место ситуация, когда респондент испытывает сложности в понимании или решении какой-либо задачи [17].

Типы метрик

В работе [2] были выделены четыре основных типа метрик айтрекинга: (1) метрики движения, (2) метрики положений, (3) метрики численности и (4) метрики латентности и расстояния. Поскольку этот курс рассчитан, прежде всего, на будущих исследователей-практиков, будем группировать метрики айтрекинга в соответствии с типами объектов исследований, на которые они направлены. Несмотря на то, что такая классификация метрик может выглядеть не достаточно полной, тем не менее, для исследователей она будет полезной, потому что поможет им быстро идентифицировать необходимые метрики для исследований. Рассмотрим такие группы метрик. Они привязаны к двум основным вопросам исследования пользовательского опыта: (1) вопросы, связанные с привлечением внимания и (2) вопросы, связанные с производительностью. Соответственно метрики привлечения внимания могут ответить на следующие вопросы:

- какая позиция размещения на веб-странице делает рекламу наиболее запоминающейся?
- какой дизайн для рекламы наиболее желателен?
- какой дизайн упаковки дольше сохраняет интерес у клиентов?
- привлекает ли новая реклама больше внимание к ключевому продукту по сравнению с текущей версией рекламы?

Метрики производительности отвечают на следующие вопросы:

- упрощает ли обновленный дизайн макета веб-страницы поиск и сравнение информации?
- легко ли отличить новую этикетку медикаментов от существующей?

– проще ли воспринять информацию от нового буклета, чем от существующего?

– упрощает ли новый GPS навигатор работу водителя на дороге по сравнению с существующей системой?

Метрики привлекательности

Существует три подгруппы метрик привлекательности: (1) метрики области заметности, (2) метрики области интереса и (3) метрики эмоционального возбуждения. Рассмотрим такие группы метрик более детально:

– метрики области заметности ориентированы на распознавание объектов, т.е. насколько просто что-либо заметить;

– метрики области интереса ориентированы на оценивание интереса к объекту, с того момента, как только он был замечен;

– метрики эмоционального возбуждения ориентированы на оценивание представления о необходимости в объекте. Такие метрики часто используются для оценивания эффективности рекламы на веб-странице, упаковки продукции и на каком-либо другом объекте, который должен привлекать внимание без дополнительных для этого усилий.

Метрики области заметности

Метрики зоны заметности применяются для оценивания видимости участка или объекта. С их помощью рассчитывается информация о количестве респондентов, которые заметили необходимый объект, а также о том, насколько быстро они замечают объект. Такие метрики зависят от визуальной заметности области, от размера этой области, ее расположения или графического представления.

Другой фактор, который влияет на заметность области внимания, помимо визуального внимания, это осознания важности области. Менее выделяющиеся элементы могут привлечь внимание пользовательским опытом, который может напоминать пользователю, что в тех или иных областях внимания содержится важная информация. Например, если пользователь знает, что навигационная система находится сверху, то он, прежде всего туда и будет смотреть, даже, если она незаметна.

К метрикам области заметности относятся следующие:

– метрика процента респондентов, которые обращают внимание на область интереса. Это информативная метрика, которая выделяет заинтересовавшихся респондентов. Чтобы рассчитать такую метрику, достаточно определить критерии фиксации (т.е. область интереса), а далее определить респондентов, которые как минимум один раз обратили внимание на область интересов;

– метрика количества фиксаций перед первой фиксацией в области интереса. Для оценивания заметности так же можно измерить количество фиксаций на стимуле (веб-странице), до самой первой фиксации в области интереса (например, рекламный баннер). Например, значение этой метрики позволит установить следующее:

- какое количество первых фиксаций в разрезе респондентов приходится на объект исследования;

- какое количество первых фиксаций в разрезе респондентов приходится на сторонние объекты;

– метрика времени фиксации до первой фиксации в области интереса. Такую метрику следует использовать, когда требуется определить, как быстро была замечена интересующая область интереса. Может показаться, что данная метрика повторяет предыдущую метрику количества фиксаций перед первой фиксацией в области интереса, однако это не так. Например, для заинтересованной стороны (например, для заказчика исследований) значение времени фиксации является более понятной метрикой, в отличие от количества фиксаций. Заинтересованной стороне проще понять, когда исследователь, например, объявляет, что «до момента обращения внимания на рекламу нового продукта респондент затратил 5 секунд с момента открытия веб-браузером соответствующей веб-страницы», чем, например, если исследователь, объявит, что респондент сделал 19 фиксаций. Когда речь идет о времени, исследователю не нужно объяснять, что представляет собой фиксации и каким образом информация о них применяется в исследованиях.

Метрики области интереса

Целью разработчиков рекламы и потребительской упаковки товара является не только создание ее заметной, но и надолго

4. Визуализация полученных результатов, выбор областей интересов и метрическая оценка

запоминающейся в памяти респондентов. Такое сочетание невозможно в том случае, когда реклама и упаковка товара на полке быстро привлекают внимание (как метрики заметности), но не могут удержать его из-за нехватки информативности и актуальности. Пример на рис. 4.11 иллюстрирует разницу между понятиями заметности и интереса (область интереса AOI1 – Area of interest – только привлекает внимание респондента - заметность, а область интереса AOI2 является областью реального интереса).

Группа метрик области интересов включает в себя следующие:

- метрика количества фиксации в области интереса. Получить значения таких метрик не сложно, поскольку необходимо только подсчитать количество фиксаций в области интереса. Например, рис. 4.11. иллюстрирует, что респонденты проявили больший интерес к области интереса AOI 2 (7 фиксаций) нежели к области интереса AOI 1 (1 фиксация);

- метрика общего времени задержки в области интереса. Общее время задержки в области интереса, это время от начала работы с сайтом до окончания работы с сайтом и подсчетом всех посетителей.



Рисунок 4.11. – Иллюстрация различий между заметностью и интересом

Важно помнить, что время задержки зависит не только от количеств фиксаций, но также от их длительности. Длительность фиксации, также может быть связана с трудностями в обработки информации. Тем не менее, когда исследователь использует время задержки как метрику интереса, ему следует, прежде всего, убедиться в том, что области интересов сравниваемых между собой продуктов, не отличаются в рамках средней длительности производимых фиксаций;

– метрика процента времени в области интереса. Чтобы рассчитать эту метрику, исследователю необходимо знать общее время внимания в области интереса (область домашней страницы) и общее время внимания на стимуле (домашней странице) в течении всей работы с веб-страницей.

Значения такой метрики, например, может дать возможность исследователю сделать следующие выводы «Из общего количества респондентов, смотрящих на домашнюю страницу, 11% обратили внимание на необходимую рекламу». Такая информация имеет больше смысла, чем простая фраза «Среднее количество времени, потраченное участниками на просмотр главной рекламы на домашней странице - 0,9 секунд». Для более полной интерпретации полученной информации лучше всего, конечно, представить информацию следующим образом «Из общего количества времени, которые респонденты потратили на просмотр домашней страницы 8,2 секунды и 11% из него (0,9 секунд) было потрачено на просмотр необходимой рекламы».

При расчете метрик интереса, прежде всего, следует определиться с множеством исходных данных. Расчет метрик может осуществляться для всех респондентов, которые участвовали в исследовании в независимости от того, смотрели ли они в область интереса. Или же для расчетов могут быть использованы данные только тех респондентов, которые обращали свое внимание на область интереса.

Лучше всего определить количество или процент респондентов, которые все-таки смотрели в область интереса и

потом определить, насколько большой был их интерес. Например, «большинство респондентов (80%), которые прошли мимо полки с детскими продуктами, как минимум один раз посмотрели на детский монитор Sony. Из общего количества времени респонденты потратили на просмотр полки - 3 секунды, 50% из которых они потратили на просмотр детского монитора Sony - 1,5 секунды».

Метрики эмоционального возбуждения

Зрачок – это отверстие в цветной части глаза (радужной оболочке), которая контролирует количество света, попадающего в глаза. Если свет яркий, то зрачок уменьшается, чтобы уменьшить количество света, который попадает на глаз. В темноте - все наоборот, зрачок увеличивается, чтобы позволить получить большее количество света. Известно, что зрачок также расширяется, когда человек встречает кого-то, кто привлекает его внимание.

Так как зрачки реагируют на эмоциональное состояние, размер зрачка может быть использован как показатель возбуждения, причиной которого может стать веб-сайт, продукт или реклама. Ниже приведены несколько примеров, которые объясняют особенности применения такой метрики.

Особенность 1. Размер зрачка зависит от нескольких факторов

Зрачок большого размера не всегда является результатом сильного возбуждения. Расширенные зрачки во время выполнения задания так же могут являться результатом высокой умственной нагрузки. Суженные зрачки, с одной стороны, могут быть скорее симптомом сонливости и усталости, чем низкой возбужденности стимула. Поэтому, перед тем, как измерять зрачок по эмоциональному возбуждению, необходимо определить, являются ли стимулы равными с точки зрения умственной нагрузки, которую они вызывают (например, с помощью индекса общей нагрузки, такого, например, как Индекс Загрузки Задач НАСА [18]). Необходимо планировать исследование таким образом, чтобы минимизировать усталость респондента (возможно, уменьшить количество заданий или продолжительность сессии),

например, за счет использования противовеса для компенсации любых эффектов усталости.

Размер зрачка очень чувствителен к изменениям света: он уменьшается при ярком свете и увеличивается в темноте. Известно, что магнитуда изменения зрачка от светового блика на много больше, чем магнитуда изменений, причиненных эмоциональными возбуждениями и когнитивными процессами. Для того, чтобы контролировать эти условия, исследования должны проводиться в комнате без окон (или с черными оконными занавесками) и уровни искусственного освещения должны быть неизменными на протяжении всего исследования. Если исследование проходит на улице или в другом месте, где натуральное освещение не может быть устранено, диаметр зрачка будет измерен неправильно.

Особенность 2. Расширение зрачка не указывает на валентность эмоций

Даже если исследователь считает, что эмоциональные причины раздражения зрачков отсутствуют во время выполнения задания, зрачки могут изменять свой диаметр также вследствие получения входных данных. В то время, как известно, значительное количество результатов исследований в поддержку утверждения, что положительные эмоции увеличивают размер зрачка, то нет определенной точки зрения такого поведения в части негативных эмоций. Известно, что результатами негативных эмоций является как расширение, так и сужения зрачка. Например, если зрачки респондентов были расширены больше нормы, когда им был представлен дизайн А нежели дизайн Б. В таком случае для установления истины необходимо использовать другие методы сбора данных, такие, например, как интервью, чтобы убедиться, что дизайн А лучше или хуже дизайна Б.

Особенность 3. Применение метрики диаметра зрачка

Метрику изменения диаметра зрачка можно применить для всего стимула, но нельзя применить ее для каждой индивидуальной области стимула. Существуют две причины, почему диаметр зрачка не может использоваться для отдельной области интереса:

– во-первых, чтобы значок начал расширяться в ответ на раздражитель необходимо дополнительное время ($\sim 0.2-0.8$ секунды). Для достижения пика может уйти даже больше времени;

– во-вторых, размер значка постоянно изменяется, а это создает дополнительный шум в данных. Лучший способ избежать шума, это аккумулировать значения всех метрик в начале работы со стимулом и подсчитать среднее значение в конце. Это также связано с особенностью применения айтрекинга, поскольку айтрекеры считывают изменения размера значка каждый раз, когда взгляд респондента падает на стимул (60 раз в секунду на 60 Hz айтрекера).

Метрики производительности

В то время, когда метрики привлекательности определяют влияние интерфейсов и дизайна на осведомленность, интерес и желание респондента, помогая выявлять проблемы в интерфейсах и дизайне, метрики производительности определяют, насколько хорошо респонденты достигают своих собственных целей. Используя значения метрик производительности, исследователь имеет возможность оценивать удобство использования интерфейса и привлекательность дизайна, включая, насколько стимул способствует умственной нагрузке, требованию к обработке информации, дизайну, уровню эффективности поиска и т.д.

Метрики умственной нагрузки

Умственная нагрузка отражает взаимосвязь между когнитивными требованиями, которые предъявляются респондентом и ограниченными познавательными ресурсами респондента. Чем выше уровень умственной нагрузки респондента, тем выше вероятность того, что производительность будет хуже, пример, когда водитель автомобиля при интенсивном движении в незнакомой местности во время неблагоприятных погодных условий может привести к высокой умственной нагрузке и снижению качества вождения. Очень часто при проведении исследований трудно установить изменения производительности в результате изменений умственной нагрузки. Хотя, именно метрики умственной нагрузки служат хорошим способом определения реальной производительности.

Наиболее часто используемой метрикой умственной нагрузки является диаметр зрачка, который так же может отображать эмоциональное возбуждение. Как уже отмечалось, отображение определенных метрик айтрекера зависит от целей сценария и стимулов, которые используются в исследовании. Изменения размера зрачка респондента, который медленно просматривает фотографии, будет указывать на его эмоциональное возбуждение (если другие факторы под контролем). Но когда того же респондента попросят решить сложную математическую задачу, диаметр его зрачка станет метрикой умственной нагрузки.

Умственная нагрузка, как правило, представляет интерес при исследовании интерфейсов и дизайна, который используются в ситуациях с высокими когнитивными потребностями, и когда низкая эффективность может иметь критические последствия, например, в транспортной или медицинской сферах. Измерение умственной нагрузки будет более подходящим при исследовании навигационной системы автомобиля, кабины самолета, инфузионного насоса, или даже заявки в колл-центре, чем, например, при исследовании веб-сайта или онлайн-магазина.

Метрики когнитивной обработки

Одной из метрик когнитивной обработки является средняя продолжительность фиксации. Продолжительность фиксации обычно колеблется в пределах от 100 мс до половины секунды, в среднем 200-250 мс. для чтения и 280-330 мс. для просмотра сцены. Визуально, более долгие фиксации, как правило, обозначаются большими окружностями и, как правило, означают более глубокую обработку и более трудоемкое извлечение информации.

Трудность такой обработки может исходить из различных источников, в том числе необычной терминологии, неясных образов, сложных понятий, и высокой плотности информации.

Средняя продолжительность фиксации часто используется в качестве зависимой переменной при сравнении стимулов с большим количеством информации, например, медицинская карта пациента, инструкции различных типов, информационные сайты или приложения центра обработки вызовов. Среднюю меру продолжительности фиксации иногда путают со временем

выдержки, но время выдержки относится ко всему времени, потраченному респондентом на исследование области интересов, которая включает в себя также все индивидуальные длительности фиксации.

Пример. Сложности когнитивной обработки и умственная нагрузка

Как уже было замечено, сложность когнитивной обработки и умственной нагрузки в некоторой степени связанные между собой метрики, поскольку они обе учитывают требования, предъявляемые к респонденту. Когда сложность стимула увеличивается, увеличивается и средняя продолжительность фиксации (т.е. размер). Основное различие между этими двумя метриками заключается в их индивидуальной диагностике. Средняя продолжительность фиксации может помочь исследовать конкретные области стимула, в то время как метрика диаметра зрачка является более глобальной. Диаметр зрачка может быть использован для оценки стимула, как в целом, так и в контексте дополнительных требований к пользователям, не относящимися к раздражителю (например, шум и временное ограничение), которые могут появляться во время выполнения задачи.

Метрики нахождения цели

Чтобы найти правильную ссылку на веб-странице, кнопку на устройстве, или лекарства на полке, респондент должен успешно пройти два этапа:

- во-первых, он должен определить местонахождение цели, проявляя внимание к ней;
- во-вторых, он должен понимать цель, с которой он ищет нужный объект, понимать суть объекта и его отношение к своей цели.

Метрики нахождения цели относятся к первому этапу этого процесса, а метрики узнаваемости цели относятся ко второму этапу. Аккумулирование метрик нахождения и узнаваемости цели имеет смысл только для решения такой задачи, в которой необходимо выбрать определенный элемент (или элементы).

Группа метрик нахождения цели включает следующие метрики:

Метрика респондентов, которые были сфокусированы на цели

Такая метрика является метрикой эффективности поиска. Можно предположить, что цели, которые выбрали респонденты (например, нажатие на правую ссылку), не означают, что такие активности могут относиться к метрике интереса, которая используется для определения уровня популярности чего-либо. Однако, если большому проценту респондентов не удалось выбрать цель (например, нажали на неправильную ссылку), то в таком случае очень важным является установить причины, которые привели к такому результату. Первым шагом на пути определения причины является определение количества респондентов, которые были сфокусированы на цели, по крайней мере, один раз.

Если часть респондентов смотрела на цель несколько раз, поиск цели является проблемой. Такая ситуация может быть вызвана плохой общей компоновкой экрана или же быть результатом неоптимального расположения или визуального представления цели, в частности, или других элементов (например, другие элементы могут действовать в качестве отвлекающих). С другой стороны, если большинство респондентов смотрели на цель, но никто ее не выбрал, то проблема, скорее всего, связана с пониманием цели.

Метрика количества фиксаций до первой фиксации

Это метрика эффективности поиска, т.е. чем меньше количество фиксаций до первой фиксации на цели, тем более эффективным является поиск. Поскольку внимание респондентов фиксировалось на цели более чем один раз, то часть названия метрики «до первой фиксации» является важной смысловой частью общего определения метрики. На рис. 4.12 целью является ссылка «Моя корзина», расположенная на левой стороне страницы, ниже логотипа сайта. Респондент рассматривал область интереса 3 раза, но первой его фиксацией стала фиксация №10, т.е. он сделал 9 фиксаций до первой фиксации. Эта метрика должна применяться совместно с процентным соотношением респондентов, которые вообще заметили цель.

4. Визуализация полученных результатов, выбор областей интересов и метрическая оценка

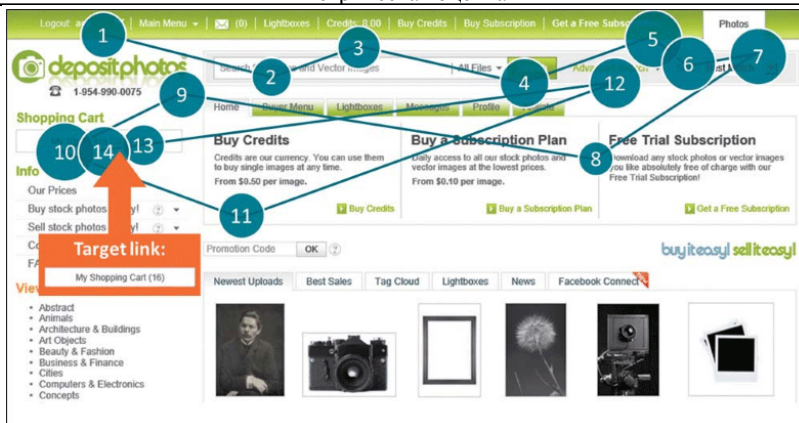


Рисунок 4.12. – Поиск необходимой цели на сайте

Метрика времени до первой фиксации

Эта метрика связана с предыдущей метрикой, и часто используется вместо метрики количества фиксаций из-за более интуитивно понятного характера концепции времени, чем понятия фиксаций. Как и предыдущая метрика, данная метрика должна применяться совместно с процентным соотношением респондентов, которые заметили цель.

Метрики узнаваемости цели

После того, как объект был найден и зафиксирован, он должен быть воспринят и соответствовать цели респондента.

Успешность узнаваемости цели зависит от того, насколько лёгок для понимания объект, который связан с содержанием цели (скажем, ссылка на этикетке), разборчив и представлен (возможно, появится ссылка на выбор). Иногда проблемы с распознаванием целей также связаны с другими ссылками, кнопками, или объектами, которые конкурируют с объектом за внимание респондента. Например, контрольный дисплей, который, кажется, имеет аналогичную функцию управления, которую ищет респондент, или же лекарственные средства в аптеке, расположенные вблизи нужного препарата на аптечной полке, могут помешать идентифицировать объект.

В рассматриваемую группу вошли следующие метрики:

– **метрика количества зрительных визитов на цели до ее выбора на объекте.** Обычно, простая в распознании цель должна была быть выбрана при первоначальном ее опознавании, а не после нескольких зрительных визитов. После того, как респондент обращает внимание на цель, он должен быть в состоянии идентифицировать объект в качестве цели. Например, на рис. 4.11 респондент заметил цель дважды, прежде чем он нажал на нее. Первый визит состоял из одной фиксации (# 10), а второй визит состоял уже из двух фиксаций (# 13 и # 14);

– **метрика времени от первой фиксации на выбранном объекте.** Время, которое требуется с момента начала фиксации цели, до того момента, когда респондент выбирает его (например, нажав по нему мышкой) может варьироваться, но более короткая продолжительность указывает на более значимые цели. Программное обеспечение, которое поддерживает анализ результатов айтрекинга, позволяет получить значение такой метрики.

Стоит отметить, что расчет значений метрики к количеству фиксаций для первой фиксации на цели и времени до первой фиксации на цели может быть излишним, поскольку оба показателя предоставляют одну и ту же информацию. Но, несмотря на это, для заинтересованных лиц более привычным будет являться значение показателя времени до первой фиксации на цели.

Ход работы

1. Провести предварительный анализ визуальной информации;
2. Определить области интереса;
3. Сформировать профиль метрик;
4. Провести расчеты метрик;
5. Проанализировать полученные результаты исследования и сделать соответствующие выводы.

Требования к содержанию отчета

Отчет должен содержать следующее:

- визуальную информацию в виде зрительных маршрутов, тепловых карт, карт непрозрачности, пчелиного роя и кластеров;
- результаты предварительного анализа визуальной информации;
- обоснование выбора типа вербального протокола;
- обоснование выбора зон интересов;
- профиль метрик и их расчетные данные.

Контрольные вопросы

1. Какие виды формирования визуальной информации поддерживает айтрекинг (Tobii Studio)?
2. В чем особенность визуализации информации в виде зрительных маршрутов? Для чего они применяются?
3. В чем особенность визуализации информации в виде тепловых карт? Для чего они применяются?
4. В чем особенность визуализации информации в виде карт непрозрачности? Для чего они применяются?
5. В чем особенность визуализации информации в виде пчелиного роя? Для чего он применяется?
6. В чем особенность визуализации информации в виде кластеров? Для чего они применяются?
7. В чем заключаются сложности при анализе визуальной информации?
8. Что такое метрики и для чего они применяются?
9. Что такое профиль метрик?
10. Для чего необходимы области интересов?

ЛИТЕРАТУРА

1. Веб-сайт компании Tobii (<http://tobii.com>);
2. Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Jarodzka Halszka, Joost van de Weijer. Eye tracking: A comprehensive guide to methods and measures. Oxford University Press. 2011, 560 p;
3. Aga Bojko. Eye Tracking the User Experience: A Practical Guide to Research. Rosenfeld Media, 2013, 320 p.;
4. Jakob Nielsen. Thinking Aloud: The #1 Usability Tool. (<https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/>);
5. Information about software «Usability Test Data Logger» (<http://www.userfocus.co.uk/resources/datalogger.html>);
6. The Rainbow Spreadsheet: A Collaborative Lean UX Research Tool (<https://www.smashingmagazine.com/2013/04/rainbow-spreadsheet-collaborative-ux-research-tool/>);
7. Elling S. Retrospective Think-Aloud Method: Using Eye Movements as an Extra Cue for Participants' Verbalizations / S. Elling, L. Lentz, M. de Jong // Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Vancouver, Canada, 2011). – p. 1161– 1170;
8. Hansen J. P. The Use of Eye Mark Recordings to Support Verbal Retrospection in Software Testing / J. P. Hansen // Acta Psychologica, 1991. – Vol. 76 (1). – p. 31– 49;
9. Hyrskykari A. Gaze Path Stimulation in Retrospective Think-Aloud / A. Hyrskykari, S. Ovaska, P. Majaranta, Kari-Jouko Räihä, M. Lehtinen // Journal of Eye Movement Research, 2008. – Vol. 2 (4). – p. 1– 18;
10. Russo J. E. A Software System for the Collection of Retrospective Protocols Prompted by Eye Fixations / J. E. Russo // Behavior Research Methods and Instrumentation , 1979. – Vol. 11 (2). – p. 177– 179;
11. Eger N. Cueing Retrospective Verbal Reports in Usability Testing Through Eye-Movement Replay / N. Eger, L. J. Ball, R. Stevens, J. Dodd // Proceedings of the 21st British HCI Group Annual

Conference on People and Computers (Lancaster, UK, 2007). – p. 129–137;

12. Tobii Technology. Retrospective Think Aloud and Eye Tracking: Comparing the Value of Different Cues When Using the Retrospective Think Aloud Method in Web Usability Testing [Electronic resource] / Tobii Technology, 2009. – Access point: <http://acuity-ets.com/downloads/>

RTA%20and%20eyetracking%20whitepaper.pdf;

13. Веб-сайт киберфорума (<http://dessertcon.com>);

14. Видео ролик организации Гринпис «Earth Day: Give Earth a Hand» (<https://www.youtube.com/watch?v=Ep9MFiWXR8M&t=30s>);

15. Kara Pernice and Jakob Nielsen. How to Conduct Eyetracking Studies(https://media.nngroup.com/media/reports/free/How_to_Conduct_Eyetracking_Studies.pdf);

16. Kavoura Androniki, Sakas Damianos P., Tomaras Petros. Strategic Innovative Marketing: 4th IC-SIM, Mykonos, Greece 2017, – 764 p.;

17. W. Albert, T. Tullis. Measuring the User Experience / // Second Edition: Collecting, Analyzing, and Presenting Usability Metrics (Interactive Technologies). Morgan Kaufmann; 2 edition, July 17, 2013. – 320 p.;

18. Sandra G. Hart. NASA-Task Load Index (Nasa-Tlx); 20 Years Later / NASA, 2006 (https://humanfactors.arc.nasa.gov/groups/TLX/downloads/HFES_2006_Paper.pdf).

АНОТАЦІЯ

Гордєєв О.О., Гордєєва Д.В. **Технології, техніки та інструменти оцінювання інформаційної безпеки і зручності використання. Практикум** / під ред. В.С. Харченко - Міністерство освіти і науки України, Національний аерокосмічний університет ім. Н.Е. Жуковського «ХАІ». 2017. - 96 с.

УДК 004.054

Викладені матеріали практичної частини начального курсу «Технології, техніки і інструменти оцінювання інформаційної безпеки і зручності використання» («Techniques and Tools for HMI security and usability assessment»), який був підготовлений в рамках проекту Modernization of Postgraduate Studies on Security and Resilience for Human and Industry Related Domains (reference number 543968-TEMPUS-1-2013-1-EE-TEMPUS-JPCR).

Курс присвячений опису та застосування технології айтрекінга, яка застосовується при дослідженні найбільш складних елементів людино-машинного взаємодії.

Курс орієнтований на студентів спеціальностей «Кібербезпека», «Комп'ютерна інженерія», «Комп'ютерні науки» та «Програмна інженерія» при вивченні технологій проектування та оцінювання призначених для користувача інтерфейсів програмного забезпечення. Курс також може бути корисний для UX / UI дослідників і проектувальників.

Бібл. – 18 найменувань, рисунків – 35.

ЗМІСТ

СПИСОК СКОРОЧЕНЬ	3
1.ПІДГОТОВКА ДО ДОСЛІДЖЕНЬ І КАЛІБРУВАННЯ АЙТРЕКЕРА	4
2. РОЗРОБКА СЦЕНАРІЮ ДОСЛІДЖЕННЯ І ВИБІР РЕСПОНДЕНТІВ	19
3. МЕТОДИКА ПРОВЕДЕННЯ АЙТРЕКІНГА І ВЕРБАЛЬНІ ПРОТОКОЛИ.....	35
4. ВІЗУАЛІЗАЦІЯ ОТРИМАНИХ РЕЗУЛЬТАТІВ, ВИБІР ОБЛАСТЕЙ ІНТЕРЕСІВ І МЕТРИЧНА ОЦІНКА.....	57
ЛИТЕРАТУРА.....	81
АНОТАЦІЯ.....	83
ЗМІСТ.....	84
ABSTRACT.....	85
CONTENT.....	86
ДОДАТОК А. НАВЧАЛЬНА ПРОГРАМА КУРСУ	87
ЗМІСТ.....	94

SUMMARY

UDC 004.054

Gordieiev O., Gordieieva D. **Techniques and tools human-machine interface security and usability assessment** / Edited by Kharchenko V. S. – Department of Education and Science of Ukraine, National Aerospace University named after N. E. Zhukovsky “KhAI”, 2017. – 96 p.

The materials of the practical part of the practical course "Techniques and tools human-machine interface security and usability assessment", developed in the framework of the TEMPUS SEREIN project "Modernization of Postgraduate Studies on Security Resilience for Human and Industry Related Domains" (543968-TEMPUS-1-2013-1-EE-TEMPUS-JPCR) are described.

The course is devoted to the description and application of eye-tracking technology, which is used in the study of the most complex elements of human-machine interaction.

The course is aimed at students of the specialties "Cyber Security", "Computer Engineering", "Computer Science" and "Software Engineering" in studying the technologies of designing and evaluating user interface software. The course may also be useful for UX / UI researchers and designers.

Bible - 18 items, drawings - 35.

CONTENT

ACRONYMS	3
1. PREPARATION FOR RESEARCH AND CALIBRATION OF THE EYE-TRACKER.....	4
2. DEVELOPMENT OF RESEARCH SCENARIO AND SELECTION OF RESPONDENTS	19
3. TECHNIQUE OF EYE-TRACKING AND VERBAL PROTOCOLS.....	35
4. RESULTS VISUALIZATION, SECTION OF AREA OF INTERESTS AND METRICALLY ASSESSMENT.....	57
REFERENCES.....	81
ABSTRACT.....	85
CONTENTS	86
APPENDIX A. TEACHING PROGRAM	87

ПРИЛОЖЕНИЕ А. УЧЕБНАЯ ПРОГРАММА КУРСА

DESCRIPTION OF THE MODULE

TITLE OF THE MODULE	Code
Techniques and tools human-machine interface security and usability assessment	

Teacher(s)	Department
Coordinating: Oleksandr Gordieiev Others: Dariya Gordieieva	Cybersecurity and information systems department

Study cycle	Level of the module	Type of the module
Engineer	A	Full-time tuition. Compulsory

Form of delivery	Duration	Langage(s)
Full-time tuition	One semester	English (Ukrainian)

Prerequisites	
Prerequisites: Psychology; Computer graphics; Software development.	Co-requisites (if necessary): Software QA;

Credits of the module	Total student workload	Contact hours	Individual work hours
4	120	46	74

Aim of the module (course unit): competences foreseen by the study programme		
The aim of module is to study the use high-technology technique for human-machine interfaces security assessment. To acquire skills and abilities of conducting for user-interfaces security assessment and looking and exceptions user errors.		
Learning outcomes of module (course unit)	Teaching/learning methods	Assessment methods
At the end of course, the successful student will be able to: 1. Forming of focus groups	Interactive lectures, Learning in laboratories, Just-in-Time	Module Evaluation Questionnaire

Приложение А. Учебная программа

for research.	Teaching	
2. Select of eye-tracker type depending from features of research.	Interactive lectures, Learning in laboratories, Just-in-Time Teaching	Module Evaluation Questionnaire
3. Tuning of eye-tracking process including calibration of eye-tracker.	Interactive lectures, Learning in laboratories, Just-in-Time Teaching	Module Evaluation Questionnaire
4. Tune-up of eye-tracking results visualisation and select of types visualisation.	Interactive lectures, Learning in laboratories, Just-in-Time Teaching	Module Evaluation Questionnaire
5. Profiling of metrics for eye-tracking results assessment.	Interactive lectures, Learning in laboratories, Just-in-Time Teaching	Module Evaluation Questionnaire
6. Interpretation of received eye-tracking results and looking for problems user interfaces.	Interactive lectures, Learning in laboratories, Just-in-Time Teaching	Module Evaluation Questionnaire
7. Implementation of eye-tracking technologies for looking for problems in security domain.	Interactive lectures, Learning in laboratories, Just-in-Time Teaching	Module Evaluation Questionnaire

Themes	Contact work hours	Time and tasks for individual work
--------	--------------------	------------------------------------

Приложение А. Учебная программа

	Lectures	Consultations	Seminars	Practical work	Laboratory work	Placements	Total contact work	Individual work	Tasks
1. Software quality models and standards: relationship between security, usability and greenness 1.1 ISO 25010 software quality model review 1.2 Evolution and correlation of software quality models characteristics security, usability and greenness 1.3 Preparation for research and calibration of the eye-tracker	6				6			18	1.4 Analysis of software usability models and software security models with use SSA-technique
2. Introduction in eye-tracking technologies 2.1 Eye-tracking conception 2.2 Eye-tracking application necessity 2.3 Eye-tracker types and technical characteristics 2.4 Technique of eye-tracking and verbal protocols	6				6			18	2.5 Differences between eye-trackers for HMI security assessment
3. Eye-tracking metrics and visualization types 3.1 Metrics taxonomy 3.2 Metrics attractiveness 3.3 Performance metrics 3.4 Visualization types of eye-tracking research results 3.5 Technique of eye-tracking and verbal protocols	6				6			18	3.6 Additional eye-tracking metrics for security assessment
4. Eye-tracking technique assessment and eye-tracking application in security domain 4.1 Eye-tracking assessment technique 4.2 Possible additional applications eye-tracking for security, usability and control 4.3 Standard ANSI/ISA-101.01 2015 «Human machine interfaces for process automation systems» particularities	6				6			18	4.5 Additional biometric technologies for HMI security assessment

Приложение А. Учебная программа

4.4 Results visualization, section of area of interests and metrically assessment								
Total	24			24			72	

Assessment strategy	Weight in %	Deadlines	Assessment criteria
Lecture activity, including fulfilling special self-tasks	30	7,14	85% – 100% Outstanding work, showing a full grasp of all the questions answered. 70% – 84% Perfect or near perfect answers to a high proportion of the questions answered. There should be a thorough understanding and appreciation of the material. 60% – 69% A very good knowledge of much of the important material, possibly excellent in places, but with a limited account of some significant topics. 50% – 59% There should be a good grasp of several important topics, but with only a limited understanding or ability in places. There may be significant omissions. 45% – 49% Students will show some relevant knowledge of some of the issues involved, but with a good grasp of only a minority of the material. Some topics may be answered well, but others will be either omitted or incorrect. 40% – 44% There should be some work of some merit. There may be a few topics answered partly or there may be scattered or perfunctory knowledge across a larger range. 20% – 39% There should be substantial deficiencies, or no answers, across large parts of the topics set, but with a little relevant and correct material in places. 0% – 19% Very little or nothing that is correct and relevant.
Learning in	50	7,14	85% – 100% An outstanding piece of

Приложение А. Учебная программа

laboratories		work, superbly organised and presented, excellent achievement of the objectives, evidence of original thought. 70% – 84% Students will show a thorough understanding and appreciation of the material, producing work without significant error or omission. Objectives achieved well. Excellent organisation and presentation. 60% – 69% Students will show a clear understanding of the issues involved and the work should be well written and well organised. Good work towards the objectives. The exercise should show evidence that the student has thought about the topic and has not simply reproduced standard solutions or arguments. 50% – 59% The work should show evidence that the student has a reasonable understanding of the basic material. There may be some signs of weakness, but overall the grasp of the topic should be sound. The presentation and organisation should be reasonably clear, and the objectives should at least be partially achieved. 45% – 49% Students will show some appreciation of the issues involved. The exercise will indicate a basic understanding of the topic, but will not have gone beyond this, and there may well be signs of confusion about more complex material. There should be fair work towards the laboratory work objectives. 40% – 44% There should be some work towards the laboratory work objectives, but significant issues are likely to be neglected, and there will be little or no appreciation of the complexity of the problem. 20% – 39% The work may contain some correct and relevant material, but most
--------------	--	--

Приложение А. Учебная программа

			issues are neglected or are covered incorrectly. There should be some signs of appreciation of the laboratory work requirements. 0% – 19% Very little or nothing that is correct and relevant and no real appreciation of the laboratory work requirements.
Module Evaluation Quest	20	8,16	The score corresponds to the percentage of correct answers to the test questions

Author	Year of issue	Title	No of periodical or volume	Place of printing. Printing house or internet link
Compulsory literature				
Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Jarodzka Halszka, Joost van de Weijer	2011	Eye tracking: A comprehensive guide to methods and measures		Oxford University Press
Aga Bojko	2013	Eye Tracking the User Experience: A Practical Guide to Research.		Rosenfeld Media
W. Albert, T. Tullis	2013	Measuring the User Experience		Morgan Kaufmann; 2 edition

СОДЕРЖАНИЕ

СПИСОК СОКРАЩЕНИЙ.....	3
1. ПОДГОТОВКА К ИССЛЕДОВАНИЯМ И КАЛИБРОВКА АЙТРЕКЕРА.....	4
2. РАЗРАБОТКА СЦЕНАРИЯ ИССЛЕДОВАНИЯ И ПОДБОР РЕСПОНДЕНТОВ.....	19
3. МЕТОДИКА ПРОВЕДЕНИЯ АЙТРЕКИНГА И ВЕРБАЛЬНЫЕ ПРОТОКОЛЫ.....	35
4. ВИЗУАЛИЗАЦИЯ ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ, ВЫБОР ОБЛАСТЕЙ ИНТЕРЕСОВ И МЕТРИЧЕСКАЯ ОЦЕНКА.....	57
ЛИТЕРАТУРА.....	81
АНОТАЦІЯ.....	83
ЗМІСТ.....	84
ABSTRACT.....	85
CONTENT.....	86
ПРИЛОЖЕНИЕ А. УЧЕБНАЯ ПРОГРАММА КУРСА...	87
СОДЕРЖАНИЕ.....	94

Гордєєв Олександр Олександрович
Гордєєва Дар'я Валеріївна

**ТЕХНОЛОГІЇ, ТЕХНІКИ ТА ІНСТРУМЕНТИ
ОЦІНЮВАННЯ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ І УДОБСТВА
ВИКОРИСТАННЯ**

Практикум
(російською мовою)

Редактор Харченко В.С.

Комп'ютерна верстка
О.О. Гордєєв

Зв. план, 2017

Підписаний до друку 03.09.2016

Формат 60x84 1/16. Папір офс. №2. Офс. друк.

Умов. друк. арк. 71,76. Уч.-вид. л. 73,52. Наклад 100 прим.

Замовлення 2. Ціна вільна

Національний аерокосмічний університет ім. М. Є.
Жуковського
"Харківський авіаційний інститут"
61070, Харків-70, вул. Чкалова, 17
<http://www.khai.edu>

Віддруковано ФОП Лисенко І.Б.
61070, Харків-70, вул. Чкалова, 17, моторний корпус, к. 147
Свідectво про внесення суб'єкта видавничої справи в державний
реєстр видавців, виготовлювачів і розповсюджувачів видавничої
продукції ДК №2607 от 11.09.06 р.