

Київський національний університет імені Тараса Шевченка
Міністерство освіти і науки України
Київський національний університет імені Тараса Шевченка
Міністерство освіти і науки України

Кваліфікаційна наукова
праця на правах рукопису

ЗЕЛІНСЬКИЙ СЕРГІЙ ВЯЧЕСЛАВОВИЧ

УДК 004.5

ДИСЕРТАЦІЯ

**МУЛЬТИМОДАЛЬНА ВЗАЄМОДІЯ ТА АНАЛІЗ ПОВЕДІНКИ
КОРИСТУВАЧІВ З ВИКОРИСТАННЯМ ВІДСТЕЖЕННЯ ПОГЛЯДУ,
ЖЕСТІВ ТА ВИРАЗІВ ОБЛИЧЧЯ**

Спеціальність 123 Комп'ютерна інженерія
Галузь знань 12 Інформаційні технології

Подається на здобуття наукового ступеня доктора філософії

Дисертація містить матеріали власних досліджень. Використання ідей,
результатів і текстів інших авторів мають посилання на відповідне джерело

_____ Зелінський Сергій Вячеславович

Науковий керівник:
кандидат фізико-математичних наук, доцент
Бойко Юрій Володимирович

Київ — 2025

АНОТАЦІЯ

Зелінський С.В. Мультимодальна взаємодія та аналіз поведінки користувачів з використанням відстеження погляду, жестів та виразів обличчя. — Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 123 «Комп'ютерна інженерія». — Київський національний університет імені Тараса Шевченка, Київ, 2025.

Робота присвячена дослідженню та розробці нових підходів до мультимодальної безконтактної взаємодії людини з комп'ютером у цифровому середовищі з використанням відстеження погляду, розпізнавання жестів рук та виразів обличчя. Окрему увагу приділено удосконаленню підходів до аналізу поведінки користувачів у веб-середовищі, зокрема дослідженню візуальної уваги, яка традиційно залишається поза межами типових аналітичних інструментів, що ґрунтуються на кліках, журналах подій та записах сесій.

У роботі обґрунтовано потребу у створенні доступних засобів як для аналізу поведінки користувачів у веб-середовищі з урахуванням зорової уваги, так і для реалізації безконтактної взаємодії з цифровими об'єктами на основі погляду, жестів і виразів обличчя без використання спеціалізованого обладнання. На основі аналізу наукових джерел сформульовано дослідницькі прогалини, які охоплюють відсутність доступних інструментів для аналізу поведінки користувачів у веб-середовищі із врахуванням зорової уваги, обмежене використання поєднання погляду та жестів у браузерній взаємодії, а також нерозкритий потенціал виразів обличчя як засобу управління.

У дисертації реалізовано три взаємопов'язані напрями дослідження:

1. Інтеграція відстеження погляду із записом веб-сесій — реалізовано кросплатформне розширення для веб-браузера, що забезпечує запис змін інтерфейсу та дій користувача із синхронізованим

відображенням зорової уваги, без потреби у попередній зміні коду веб-сторінок.

2. Безконтактна взаємодія на основі поєднання погляду та жестів рук — реалізовано систему для маніпулювання об'єктами (перетягування, масштабування, обертання), яка поєднує вибір об'єкта за допомогою погляду та трансформацію за допомогою жестів, що імітують природні дії користувача.
3. Використання виразів обличчя для безконтактної взаємодії — запропоновано та реалізовано підхід, у якому дії задаються через налаштовувані комбінації мимічних форм змішування. Це дозволяє створювати гнучкий інтерфейс для взаємодії без потреби у фізичних пристроях введення.

Ефективність запропонованих рішень оцінено експериментально із залученням користувачів, які виконували типові завдання в цифровому середовищі. Результати дослідження засвідчили позитивну динаміку навчання, високу зацікавленість у використанні безконтактної взаємодії та практичну придатність запропонованих методів у реальних умовах.

Наукова новизна дисертаційної роботи полягає у комплексному підході до розробки мультимодальних засобів безконтактної взаємодії з цифровими інтерфейсами та аналізу поведінки користувачів на основі використання погляду, жестів і виразів обличчя із застосуванням стандартного обладнання. Вперше реалізовано кросплатформний інструмент аналізу поведінки користувачів у веб-середовищі, який поєднує запис сесій взаємодії з відображенням зорової уваги користувача без необхідності модифікації коду веб-сторінок чи використання спеціалізованого обладнання. Вперше запропоновано та реалізовано систему безконтактної взаємодії у веб-браузері, яка поєднує вибір об'єктів за допомогою погляду та їх трансформацію за допомогою жестів рук, що імітують природні дії

користувача. Крім того, вперше у веб-середовищі реалізовано підхід до безконтактної взаємодії на основі виразів обличчя, де дії задаються шляхом налаштовуваних комбінацій мимічних форм змішування.

Усі запропоновані рішення функціонують у веб-браузері із використанням лише стандартного обладнання та орієнтовані на широке практичне використання без потреби у спеціалізованому програмному чи апаратному забезпеченні.

Практична цінність роботи полягає у розробці доступних і гнучких засобів для безконтактної взаємодії та аналізу поведінки користувачів, які працюють у сучасних веб-браузерах без потреби у спеціалізованому обладнанні. Запропоновані рішення мають потенціал для впровадження у сферах допоміжних технологій, освіти, промисловості, медицини, а також у дослідженнях користувацького досвіду та інших прикладних контекстах.

Перспективи подальших досліджень передбачають підвищення точності розпізнавання погляду, жестів і виразів обличчя, адаптацію систем до індивідуальних особливостей користувачів, впровадження методів машинного навчання для автоматичного аналізу поведінки, а також інтеграцію розроблених підходів у практичні середовища, де безконтактна взаємодія є критично важливою.

Ключові слова: мультимодальна взаємодія, безконтактна взаємодія, відстеження погляду, розпізнавання жестів, вирази обличчя, запис веб-сесій, людино-комп'ютерна взаємодія, аналіз поведінки користувачів, комп'ютерний зір, машинне навчання, інтерфейси користувача, допоміжні технології, веб-застосунки.

SUMMARY

Zelinskyi S.V. Multimodal interaction and user behavior analysis using eye tracking, gestures, and facial expressions. — A qualification scientific work in the form of a manuscript.

Dissertation for the degree of Doctor of Philosophy in specialty 123 «Computer Engineering». — Taras Shevchenko National University of Kyiv, Kyiv, 2025.

This dissertation focuses on the investigation and development of new approaches to multimodal hands-free human–computer interaction in digital environments, based on eye tracking, hand gesture recognition, and facial expressions. Particular attention is given to improving approaches for analyzing user behavior in web environments, specifically by studying visual attention, which traditionally remains beyond the scope of standard analytical tools based on clicks, event logs, and session recordings.

The dissertation substantiates the need for accessible solutions for both the analysis of user behavior in web environments with consideration of visual attention and for hands-free interaction with digital objects based on gaze, gestures, and facial expressions, without the use of specialized hardware. Based on the analysis of scientific sources, key research gaps have been identified, including the lack of accessible tools for user behavior analysis in web environments that account for visual attention, the limited use of gaze–gesture interaction in browser-based systems, and the underexplored potential of facial expressions as a means of control.

The dissertation comprises three interrelated research directions:

1. Integration of eye tracking with web session recording — a cross-platform browser extension was developed to record interface changes and user actions with synchronized visualization of gaze data, without requiring modifications to the web page source code.
2. Hands-free interaction using combined gaze and hand gestures — a system was implemented for manipulating objects (dragging, scaling, rotating) by

combining gaze-based selection with gesture-based transformation, imitating natural user behavior.

3. Facial expression–based hands-free interaction — a novel approach was proposed and implemented in which actions are triggered through configurable combinations of facial blendshape activations, allowing the construction of flexible interaction interfaces without the need for physical input devices.

The effectiveness of the proposed solutions was evaluated experimentally by involving users in typical digital tasks. The results demonstrate a positive learning curve, strong interest in hands-free interaction, and the practical applicability of the developed methods in real-world scenarios.

The scientific novelty of this dissertation lies in a comprehensive approach to the development of multimodal hands-free interaction tools for digital interfaces and the analysis of user behavior, based on the use of gaze, hand gestures, and facial expressions with standard consumer-grade hardware. For the first time, a cross-platform user behavior analysis tool has been implemented, combining session recording with synchronized visualization of gaze data in web environments, without requiring modifications to web page code or the use of specialized equipment. Additionally, a browser-based hands-free interaction system has been proposed and implemented, enabling object selection through gaze and manipulation through hand gestures that imitate natural user behavior. Furthermore, for the first time in web environments, a facial expression–based interaction approach has been implemented, in which actions are triggered through configurable combinations of facial blendshape activations.

All the proposed solutions operate within web browsers using only standard consumer-grade hardware and are designed for broad practical application without the need for specialized software or hardware.

The practical value of the work lies in the development of accessible and flexible tools for hands-free interaction and user behavior analysis that operate in modern web browsers without the need for specialized equipment. The proposed solutions have potential applications in assistive technologies, education, industrial and clinical settings, and user experience research, among other applied contexts.

Future research directions include improving the accuracy of gaze, gesture, and facial expression recognition; adapting systems to individual user characteristics; applying machine learning methods for automated behavior analysis; and integrating the proposed approaches into practical environments where hands-free interaction is critically important.

Keywords: multimodal interaction, hands-free interaction, eye tracking, gesture recognition, facial expressions, web session recording, human–computer interaction, user behavior analysis, computer vision, machine learning, user interfaces, assistive technologies, web applications.

СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА

Статті у наукових фахових видання України:

1. Zelinskyi, S., & Boyko, Y. (2024). Integrating session recording and eye-tracking: development and evaluation of a Chrome extension for user behavior analysis. *Radioelectronic and Computer Systems*, 2024(3), 38–54. <https://doi.org/10.32620/reks.2024.3.03>
2. Zelinskyi, S., & Boyko, Y. (2024). Exploring Gaze-Gesture Interaction on the Web: A Comparison with Mouse Input for Object Manipulation. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 3, 33–42. <https://doi.org/10.32782/it/2024-3-4>
3. Zelinskyi, S., & Boyko, Y. (2024). Using Facial Expressions for Custom Actions: Development and Evaluation of a Hands-Free Interaction Method. *Computer Systems and Information Technologies*, 4, 116–125. <https://doi.org/10.31891/csit-2024-4-14>

Праці, які засвідчують апробацію матеріалів дисертації:

1. Zelinskyi, S., Boyko, Y., & Didmanidze, I. (2024). Evaluating the Usability of Gaze-Gesture Interaction for Object Manipulation on the Web. *International Scientific-Practical Conference Modern Challenges and Achievements in Information and Communication Technologies – 2024*
2. Zelinskyi, S., Boyko, Y. (2024). Hands as Pointers: A Touchless Gesture-Based Approach for Web Objects Manipulation. *XI Всеукраїнська науково-практична конференція студентів, аспірантів та молодих вчених*, 308–310.
3. Zelinskyi, S., & Boyko, Y. (2024). Facial Blendshapes for Custom Action: An Experimental Evaluation Using Facial Expressions for Human-Computer Interaction. *Modeling, Control and Information Technologies*, 7, 174–176. <https://doi.org/10.31713/mcit.2024.050>

4. Zelinskyi, S., & Boyko, Y. (2024). Leveraging Large Language Models for Free-Form Voice Command Manipulation of 3D Objects: A Prototype Study. *Proceedings of the 50th International Scientific and Practical Conference*, 247–250. <https://doi.org/10.52058/50>

ЗМІСТ

Перелік умовних позначень.....	14
ВСТУП.....	16
Розділ 1. Теоретичні основи аналізу поведінки користувачів та мультимодальної взаємодії у цифровому середовищі.....	22
1.1 Методи аналізу поведінки користувачів у веб-середовищі.....	22
1.1.1 Лог-файли та аналітичні системи.....	22
1.1.2 Карти кліків та теплові карти уваги.....	24
1.2 Технології запису веб-сесій.....	26
1.2.1 Основні підходи до запису веб-сесій.....	27
1.2.2 Відмінності між підходами до запису веб-сесій.....	30
1.3 Технології відстеження погляду.....	30
1.3.1 Стаціонарні пристрої відстеження погляду.....	32
1.3.2 Мобільні (носимі) пристрої відстеження погляду.....	34
1.3.3 Програмні рішення на основі веб-камери.....	36
1.3.4 Порівняльний аналіз технологій відстеження погляду.....	38
1.3.5 Використання технології у веб-аналізі поведінки.....	39
1.4 Аналіз та вибір технології відстеження погляду.....	39
1.5 Інтеграція технології відстеження погляду із записом сесій.....	41
1.6 Використання жестів у НСІ.....	42
1.7 Взаємодія на основі погляду та жестів.....	43
1.8 Вирази обличчя як засіб взаємодії у НСІ.....	45
1.9 Виявлені дослідницькі прогалини.....	50
1.10 Формулювання загальної задачі дослідження.....	51

Розділ 2. Інтеграція інструментів запису сесій та відстеження погляду для аналізу поведінки користувачів у веб-середовищі	53
2.1 Формалізація задачі запису веб-сесій з інтегрованими даними відстеження погляду.....	53
2.2 Архітектура системи та реалізація.....	55
2.2.1 Технічний огляд.....	55
2.2.2 Основні компоненти.....	56
2.2.3 Процес роботи системи.....	58
2.2.4 Процес калібрування	59
2.2.5 Обробка даних погляду.....	62
2.2.6 Захист персональних даних і конфіденційність	64
2.3 Користувацький інтерфейс	65
2.4 Оцінювання системи на основі користувацького дослідження	69
2.5 Аналіз результатів користувацького дослідження	71
2.6 Виклики та обмеження.....	75
2.7 Перспективи подальших досліджень.....	75
2.8 Висновки.....	77
Розділ 3. Безконтактна взаємодія з використанням технології відстеження погляду та жестів у веб-середовищі	78
3.1 Архітектура системи та реалізація.....	79
3.1.1 Загальний технічний огляд	79
3.1.2 Основні компоненти.....	80
3.1.3 Процес калібрування	82
3.1.4 Процес взаємодії.....	83
3.2 Інтерфейс користувача	86

	12
3.3 Оцінювання ефективності запропонованого підходу до взаємодії	89
3.4 Аналіз результатів	91
3.5 Обговорення результатів.....	92
3.6 Виклики та обмеження.....	93
3.7 Сфери потенційного застосування.....	94
3.8 Перспективи подальших досліджень.....	95
3.9 Висновки.....	96
Розділ 4. Безконтактне керування цифровими системами з використанням виразів обличчя	98
4.1 Формалізація використання мімічних форм змішування в поточному дослідженні	98
4.2 Архітектура системи	100
4.3 Методи та критерії оцінювання ефективності системи	101
4.4 Порівняння взаємодії на основі виразів обличчя та комп'ютерної миші у задачі редагування тексту	103
4.4.1 Налаштування експерименту	103
4.4.2 Аналіз отриманих результатів.....	106
4.4.3 Обговорення.....	108
4.4.4 Висновки.....	109
4.5 Оцінка навчального ефекту безконтактної взаємодії на основі виразів обличчя	110
4.5.1 Вибір мімічних форм змішування для цього дослідження	111
4.5.2 Огляд системи та налаштування експерименту	112
4.5.3 Визначення користувацьких дій	114
4.5.4 Процес виконання користувацьких дій	116

4.5.5 Оцінювання ефективності взаємодії та аналіз результатів	117
4.5.6 Обговорення.....	121
4.5.7 Висновки.....	124
Висновки.....	126
Список використаних джерел.....	129
Додатки.....	138

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

AoI	Області інтересу (Areas of Interest)
API	Інтерфейс прикладного програмування (Application Programming Interface)
AR	Доповнена реальність (Augmented Reality)
CSS	Каскадні таблиці стилів (Cascade Style Sheets)
DOM	Об'єктна модель документа (Document Object Model)
EEG	Електроенцефалографія (Electroencephalography)
HCI	Взаємодія людини з комп'ютером (Human-Computer Interaction)
HTML	Мова гіпертекстової розмітки (Hyper Text Markup Language)
IDE	Інтегроване середовище розробки (Integrated Development Environment)
IMU	Інерціальний вимірювальний модуль (Inertial Measurement Unit)
IR	Інфрачервоне випромінювання (Infrared)
JSON	Нотація об'єктів JavaScript (JavaScript Object Notation)
LLM	Велика мовна модель (Large Language Model)
LSTM	Довготривала короткочасна пам'ять (Long Short-Term Memory)
SEO	Оптимізація для пошукових систем (Search Engine Optimization)
SPA	Односторінковий застосунок (Single Page Application)
UI	Користувацький інтерфейс (User Interface)

URL	Уніфікований локатор ресурсу (Uniform Resource Locator)
UX	Користувацький досвід (User Experience)
VR	Віртуальна реальність (Virtual Reality)

ВСТУП

Обґрунтування вибору теми дослідження. Зі зростанням складності цифрових інтерфейсів та підвищенням інтересу до аналізу поведінки користувачів традиційні методи, що ґрунтуються на кліках, журналах подій або записах сесій, демонструють суттєві обмеження. Вони фіксують дії користувача, але не враховують його зорову увагу, що ускладнює глибше розуміння поведінки у веб-середовищі. Технології відстеження погляду відкривають нові можливості для аналізу того, як користувач сприймає інтерфейс, однак їх використання без спеціалізованого обладнання залишається недостатньо дослідженим.

Паралельно з цим зростає зацікавленість у розробці альтернативних підходів до взаємодії з цифровими об'єктами, які не потребують фізичних пристроїв введення. Мультимодальні підходи, що поєднують відстеження погляду, розпізнавання жестів рук та виразів обличчя, мають потенціал забезпечити інтуїтивну, гнучку та доступну взаємодію. Попри активний розвиток таких методів у сфері віртуальної та доповненої реальності, їх впровадження у веб-середовище ускладнене технічними обмеженнями — зокрема необхідністю встановлення додаткового програмного забезпечення або використання спеціалізованих апаратних засобів.

Таким чином, дослідження нових засобів аналізу поведінки користувачів, а також розробка доступних підходів до безконтактної взаємодії на основі стандартного обладнання — є актуальним напрямом для створення більш інклюзивних, природних і гнучких цифрових інтерфейсів.

Мета і завдання дослідження. Метою дисертаційної роботи є розробка підходів для аналізу поведінки користувачів у веб-середовищі з урахуванням зорової уваги шляхом інтеграції запису веб-сесій із відстеженням погляду, а також створення та оцінка підходів до безконтактної мультимодальної взаємодії з

цифровими об'єктами на основі відстеження погляду, жестів рук та виразів обличчя.

Для досягнення поставленої мети в роботі передбачено вирішення таких дослідницьких завдань:

1. дослідити методи аналізу поведінки користувачів у веб-середовищі та виявити їхні ключові обмеження;
2. проаналізувати існуючі підходи до безконтактної мультимодальної взаємодії на основі погляду, жестів та виразів обличчя, а також визначити їх технічні та концептуальні обмеження;
3. розробити підхід до інтеграції запису веб-сесій із відстеженням погляду для забезпечення поглибленого аналізу взаємодії користувача з інтерфейсом;
4. реалізувати прототипи систем мультимодальної взаємодії, що поєднують відстеження погляду, жести рук та вирази обличчя для керування цифровими об'єктами;
5. провести експериментальну оцінку ефективності запропонованих рішень на основі користувацьких досліджень;
6. визначити технічні та функціональні обмеження розроблених підходів і окреслити напрями їх подальшого вдосконалення.

Об'єктом дослідження є процеси взаємодії людини з комп'ютерними системами із застосуванням методів введення даних на основі відстеження погляду, розпізнавання жестів рук і виразів обличчя.

Предметом дослідження є методи, засоби та технології розробки, реалізації та експериментальної оцінки інтерактивних і аналітичних систем, що забезпечують мультимодальну взаємодію людина–комп'ютер через відстеження погляду,

розпізнавання жестів рук і виразів обличчя, а також інтеграцію аналізу поведінки користувачів шляхом поєднання запису веб-сесій та даних про візуальну увагу.

Методи дослідження. У дисертаційній роботі застосовано методи комп'ютерного зору та рішення машинного навчання для реалізації відстеження погляду, визначення ключових точок рук та обличчя, а також зчитування інтенсивності мимічних форм змішування, що відповідають виразам обличчя. Для отримання координат ключових точок та інтенсивності мимічних форм змішування використовувалися попередньо натреновані моделі глибокого навчання.

Для збору даних про поведінку користувачів у веб-середовищі використовувалися технології запису веб-сесій, що дозволило фіксувати як візуальні зміни інтерфейсу, так і дії користувача. Зібрані дані додатково підлягали базовій попередній обробці для зменшення шуму, зокрема в координатах погляду.

Архітектура створених систем моделювалася з використанням уніфікованої мови моделювання UML. Програмна реалізація виконувалася мовою JavaScript із застосуванням середовища Node.js та сучасних інструментів веб-розробки.

Для оцінки ефективності запропонованих підходів було проведено серію експериментальних досліджень за участю користувачів. У ході експериментів аналізувались такі аспекти, як зручність використання, інтуїтивність інтерфейсів та практична придатність реалізованих рішень у типових сценаріях взаємодії.

Структура дисертаційної роботи. Дисертаційна робота складається з чотирьох розділів, які охоплюють теоретичне підґрунтя дослідження, практичну реалізацію запропонованих рішень, а також результати їх експериментальної перевірки.

У першому розділі проаналізовано сучасні підходи до аналізу поведінки користувачів у веб-середовищі, а також до мультимодальної безконтактної взаємодії на основі погляду, жестів та виразів обличчя. Визначено наукові

прогалини у кожному з напрямів і сформульовано загальну дослідницьку задачу, що стала основою для подальших етапів роботи.

Другий розділ присвячено розробці підходу до аналізу поведінки користувачів у веб-середовищі з урахуванням зорової уваги. Описано реалізацію кросплатформного розширення для веб-браузера, яке забезпечує запис сесій із синхронізованим відображенням погляду користувача без необхідності внесення змін до коду веб-сайтів, на яких здійснюється спостереження. Це дозволяє використовувати інструмент на будь-якому веб-ресурсі без попередньої інтеграції, на відміну від більшості аналітичних систем.

У третьому розділі представлено метод безконтактної взаємодії з цифровими об'єктами у веб-браузері, що базується на поєднанні погляду для вибору об'єктів та жестів рук для їх трансформації. Наведено архітектуру рішення, приклади використання та результати його експериментальної оцінки.

Четвертий розділ присвячено використанню виразів обличчя як альтернативного засобу взаємодії у цифрових середовищах. Запропоновано підхід, у якому дії задаються шляхом налаштовуваних комбінацій мімічних форм змішування. Оцінено зручність використання запропонованого прототипу та потенціал його практичного застосування.

Наукова новизна отриманих результатів. Наукова новизна дисертаційної роботи полягає в обґрунтуванні, розробці та експериментальній перевірці нових підходів до аналізу поведінки користувачів у веб-середовищі та до мультимодальної безконтактної взаємодії з цифровими об'єктами на основі погляду, жестів і виразів обличчя із використанням стандартного обладнання.

У межах першого напрямку, а саме аналізу поведінки користувачів, вперше запропоновано та реалізовано кросплатформний підхід, що поєднує запис веб-сесій із синхронізованим відображенням зорової уваги. Рішення реалізовано у вигляді розширення для веб-браузера та не потребує модифікації коду веб-застосунків, що

забезпечує його використання на будь-яких веб-ресурсах без попередньої інтеграції.

Удосконалено підходи до збору та репрезентації поведінкових даних, що полягають в об'єднанні інформації про дії користувачів, зміни структури веб-сторінки та напрямок зорової уваги в межах єдиного запису сесії. Це створює передумови для подальшого автоматизованого аналізу взаємодії та виявлення семантично важливих елементів інтерфейсів.

Такий підхід дозволяє здійснювати комплексний аналіз взаємодії користувачів з інтерфейсом, поєднуючи контекст дій із напрямком візуальної уваги.

У межах другого напрямку, а саме мультимодальної взаємодії на основі погляду та жестів, вперше запропоновано та реалізовано підхід до безконтактного маніпулювання цифровими об'єктами у веб-середовищі із використанням лише стандартної веб-камери. Користувачі можуть обирати об'єкти за допомогою погляду, а їх трансформація (перетягування, масштабування, обертання) здійснюється за допомогою жестів рук. Рішення не потребує спеціалізованого обладнання, що розширює сфери його потенційного застосування.

У межах третього напрямку, а саме взаємодії на основі виразів обличчя, вперше у веб-середовищі реалізовано підхід, у якому дії задаються через налаштовувані комбінації мимічних форм змішування. На відміну від підходів, що використовують фіксований набір мимічних команд, запропоноване рішення забезпечує можливість створення індивідуальних конфігурацій виразів і зв'язування їх з певними діями.

Усі розроблені рішення є кросплатформними, функціонують у веб-браузері, не потребують спеціалізованого програмного чи апаратного забезпечення, що робить їх доступними для широкого кола користувачів і придатними до впровадження в різних сферах — від досліджень користувацького досвіду до допоміжних технологій.

Структура та обсяг дисертації. Дисертаційна робота складається зі вступу, чотирьох розділів, висновків та списку використаних джерел. Загальний обсяг роботи складає 139 сторінку з яких основний зміст викладено на 136 сторінках. Дисертація містить 29 рисунків, 10 таблиць та список використаних джерел, що включає 65 найменувань.

РОЗДІЛ 1. ТЕОРЕТИЧНІ ОСНОВИ АНАЛІЗУ ПОВЕДІНКИ КОРИСТУВАЧІВ ТА МУЛЬТИМОДАЛЬНОЇ ВЗАЄМОДІЇ У ЦИФРОВОМУ СЕРЕДОВИЩІ

Цей розділ дисертаційного дослідження присвячено теоретичним засадам аналізу поведінки користувачів та безконтактної мультимодальної взаємодії у цифровому середовищі. У ньому проаналізовано сучасні наукові підходи до вивчення поведінки користувачів, зокрема із врахуванням зорової уваги, а також підходи до взаємодії, що базуються на використанні погляду, жестів і виразів обличчя. Окрему увагу приділено існуючим технічним рішенням, їх перевагам і обмеженням. Узагальнення результатів аналізу дозволило виявити наукові прогалини, які стали підґрунтям для формулювання загальної дослідницької задачі, що розв'язується у подальших розділах роботи.

1.1 Методи аналізу поведінки користувачів у веб-середовищі

Аналіз поведінки користувачів є ключовим аспектом досліджень у сфері зручності використання, цифрового маркетингу (digital marketing) та веб-розробки. Розуміння того, як користувачі взаємодіють із веб-сайтами та застосунками, допомагає покращити інтерфейс, підвищити ефективність навігації та оптимізувати користувацький досвід (user experience). У дослідженні [1] наводиться детальний опис методів і принципів HCI та підкреслюється важливість аспектів зручності використання при проектуванні інтерактивних систем. Існує кілька підходів до аналізу поведінки користувачів у веб-середовищі, кожен із яких має свої переваги та обмеження.

1.1.1 Лог-файли та аналітичні системи

Одним із найстаріших методів аналізу поведінки користувачів є збір та обробка лог-файлів (файлів журналу) серверів. Лог-файли містять детальну інформацію про всі запити, що надходять на веб-сервер, включаючи IP-адресу користувача, час запиту, URL-адресу відвіданої сторінки, відомості про веб-

браузер і пристрій користувача, а також джерело трафіку (реферальні сайти, пошукові системи, прямі переходи за посиланням). Ці дані дозволяють відстежувати загальні тренди веб-трафіку та аналізувати навантаження на сервер.

Однією з переваг таких систем є можливість автоматичного збору даних без впливу на продуктивність веб-сайту. Можливість аналізу великих обсягів інформації протягом тривалого часу робить цей метод корисним для таких задач, як пошукова оптимізація (SEO), безпековий моніторинг (виявлення підозрілої активності), а також аналіз помилок (наприклад, 404-сторінки).

Одним з ключових недоліків цього підходу є обмежена деталізація поведінки користувача. Оскільки в файлах журналу фіксується інформація лише про відвідування сторінок, а не реальні взаємодії користувача зі сторінкою такі як, рух миші, кліки, гортання сторінки. Іншим недоліком підходу із використанням лог-файлів є відсутність візуалізацій — аналіз проводиться із використанням текстових файлів, що ускладнює інтерпретацію даних. Також цей метод не дозволяє отримувати аналітику в реальному часі, оскільки аналіз файлів журналу зазвичай виконується із затримкою.

Щоб подолати ці обмеження, були розроблені більш гнучкі та інтуїтивно зрозумілі веб-аналітичні платформи, такі як Google Analytics, Facebook Pixel, тощо. Сучасні веб-аналітичні платформи надають такі можливості, як:

- *Візуалізація даних* — графіки, діаграми, звіти, що спрощують аналіз.
- *Моніторинг у реальному часі* — можливість відстеження активності користувачів у момент її здійснення.
- *Глибший аналіз поведінки* — відстеження подій (кліки, гортання веб-сторінки, завантаження файлів), сегментація аудиторії, аналіз шляхів користувачів.

Таким чином, лог-файли залишаються важливим джерелом даних, особливо для технічного аналізу веб-сайту. Але для глибшого розуміння поведінки користувачів застосовуються сучасні аналітичні інструменти, які надають

інтерактивну, зручну та детальну інформацію про взаємодію користувачів із веб-інтерфейсом.

1.1.2 Карти кліків та теплові карти уваги

Для аналізу взаємодії користувачів із веб-сторінками широко використовуються теплові карти (heatmaps), які дозволяють візуально відобразити найбільш активні ділянки сторінки. Ці карти допомагають визначити, які елементи інтерфейсу привертають увагу, а які залишаються непоміченими, що є важливим для покращення користувацького досвіду, оптимізації навігації, тощо.

Серед основних типів теплових карт можна виділити три основні, а саме теплові карти кліків (click heatmaps), теплові карти руху миші (mouse movement heatmaps) та теплові карти гортання сторінки (scroll heatmaps).

Теплові карти кліків. Теплові карти цієї категорії відображають області з найбільшою кількістю натискань. Вони дозволяють оцінити ефективність розташування кнопок, посилань та інтерактивних елементів. Також цей вид аналізу допомагає виявити проблеми пов'язані зі зручністю використання, наприклад, коли користувачі натискають на неінтерактивні елементи, очікуючи певної реакції системи.

Теплові карти руху миші. Використовуються для аналізу траєкторії переміщення курсору та області з найбільшою концентрацією рухів. Вважається, що рух миші може корелювати з візуальною увагою користувача. Проте цей метод має певні обмеження, оскільки користувач може читати текст у центрі екрану, залишаючи курсор на боковій панелі. На Рис. 1.1 наведено приклад теплової карти руху миші.

Теплові карти гортання сторінки. Дана категорія теплових карт використовується для визначення, до якого рівня сторінки користувачі прокручують контент. Це дозволяє оцінити, які секції отримують найбільше переглядів, а які залишаються поза увагою. Такі теплові карти є особливо корисними для аналізу довгих сторінок, блогів та інтернет-магазинів, оскільки

допомагають зрозуміти, чи досягають користувачі ключових елементів (кнопок, форм реєстрації, цінових блоків, тощо).

Теплові карти мають низку переваг, що пояснюють їхню популярність серед маркетологів та UX-дизайнерів. Однією з таких переваг є візуалізація даних. Вона дозволяє швидко оцінити поведінкові шаблони користувачів та дає змогу оперативно визначити неефективні зони користувацького інтерфейсу. Як результат, це дозволяє збільшити показник конверсії (виконання користувачем цільової дії, наприклад, оформлення підписки або покупки) шляхом оптимізації розташування кнопок, форм, та іншого важливого контенту. Іншою перевагою є автоматичний збір даних, який не потребує безпосередньої участі користувачів — дані кліків, рухів курсору миші, прокрутки сторінок збираються у фоновому режимі.

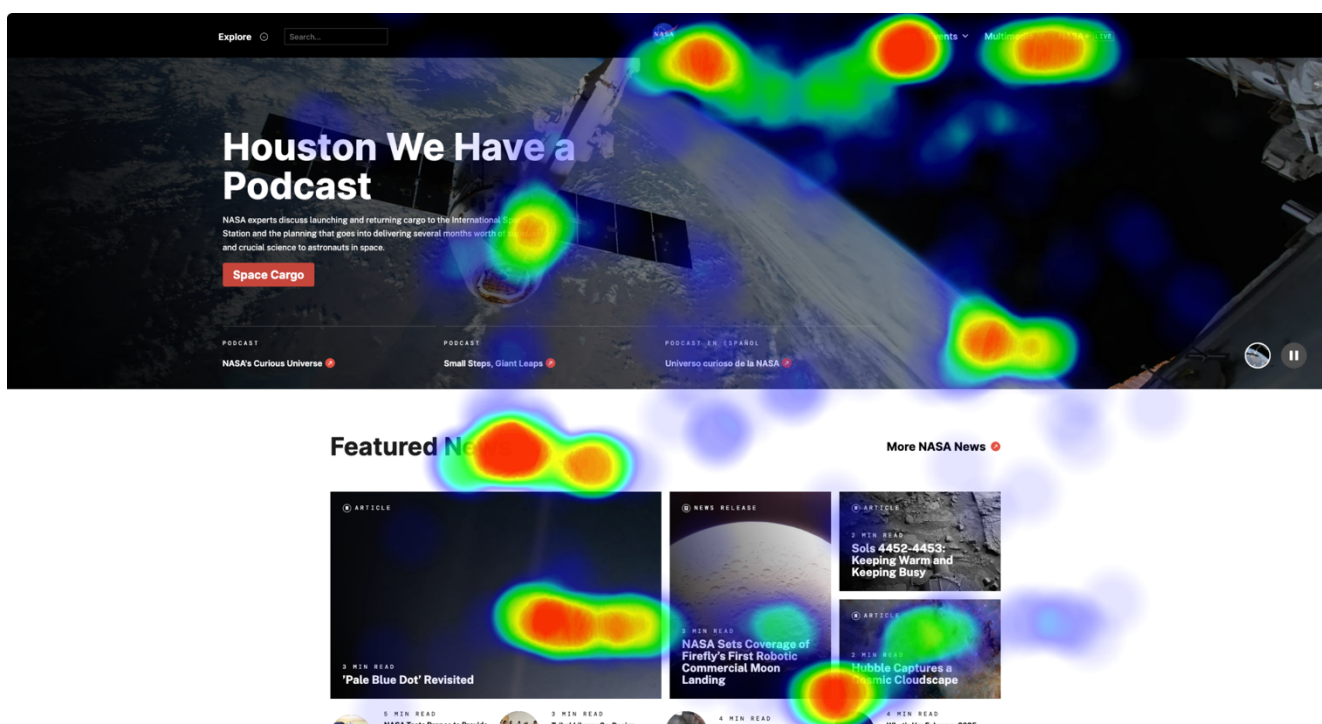


Рис. 1.1. Приклад теплової карти руху миші

Попри численні переваги теплових карт, вони мають і певні обмеження. Одним із таких обмежень є те, що теплові карти не відображають реальний рівень уваги користувачів, оскільки аналізується лише взаємодія з мишею, а не рухи очей. Також, використовуючи лише теплові карти, складно пояснити мотиви поведінки користувачів, оскільки теплові карти відображають лише ключові точки взаємодії, що не пояснюють причини, які спонукали користувачів до певних дій. Слід зазначити, що аналіз, що базується лише на використанні теплових карт, може містити похибки, адже деякі зони інтерфейсу можуть виглядати «популярними» через випадкові натискання.

Описані вище обмеження підтверджуються проведеними дослідженнями. Так у дослідженнях [2], [3], [4] було показано, що традиційні методи аналізу поведінки користувачів, такі як данні кліків, опитування користувачів, і навіть сучасні інструменти веб-аналітики, зачасти не можуть надати детальне розуміння того, як користувачі взаємодіють з інтерфейсами. Ці методи обмежені у можливостях співставляти дії користувачів з їх візуальною увагою, що залишає прогалини під час аналізу поведінки користувачів.

Щоб подолати ці обмеження, часто поєднують теплові карти із записом сесій користувачів або більш точними методами аналізу уваги, зокрема відстеженням погляду (eye tracking). Так, наприклад, поєднання інструментів запису сесій дозволяє не лише аналізувати рухи миші, але й безпосередньо визначати, на які елементи користувач дивиться, як довго і в якій послідовності. Такий підхід забезпечує значно глибше розуміння поведінки користувачів, що буде розглянуто в наступних секціях.

1.2 Технології запису веб-сесій

Запис веб-сесій (web session recording) є одним з ключових інструментів для аналізу поведінки користувачів у веб-середовищі. Використання цього підходу дозволяє записувати та відтворювати взаємодію користувачів із веб-інтерфейсом,

що є цінним для оцінки зручності використання, виявлення недоліків та обмежень інтерфейсу та оптимізації дизайну.

Сучасні рішення для запису веб-сесій застосовують різні підходи до збору даних, що впливають на точність відтворення, продуктивність системи та вимоги до зберігання інформації. У цьому розділі розглянуто основні методи запису веб-сесій, їхні переваги та обмеження, а також наведено аналіз актуальних тенденцій у використанні цих технологій.

1.2.1 Основні підходи до запису веб-сесій

На сьогоднішній день існують три основні підходи до запису веб-сесій. Цими підходами є: запис змін DOM (DOM mutation recording), запис подій (event logging) та гібридний підхід (hybrid approach). Кожен підхід має свої переваги і обмеження, а також контекст застосування.

Запис змін DOM. Цей підхід базується на реєстрації змін у DOM-структурі веб-сторінки. Використовуючи MutationObserver API, система фіксує додавання, видалення та модифікацію HTML-елементів, атрибутів і текстового контенту у вигляді покрокової історії змін.

Однією з ключових переваг цього підходу є висока точність відтворення сесій. Оскільки при записі сесії реєструються структурні та візуальні зміни веб-сторінки, при відтворенні є можливість забезпечити повну реконструкцію вигляду сторінки такою, яка вона була у момент взаємодії з нею користувача. Цей підхід також є ефективним у поєднанні з SPA-застосунками (Single Page Application), розробленими з використання таких бібліотек як Angular, React, Vue, оскільки фіксує зміни без перезавантаження сторінки. Також, оскільки під час відтворення сесій відображаються анімації, динамічні оновлення контенту та CSS-зміни, використання цього підходу надає можливість аналізувати складні взаємодії користувача із веб-сторінкою.

Попри переваги у точності відтворення записаних веб-сесій, підхід із записом змін DOM має свої обмеження. Одним з таких обмежень є високе

навантаження на систему, оскільки постійне відслідковування змін DOM може впливати на продуктивність веб-сторінки. Також, у випадках, коли веб-сайт містить багато інтерактивного контенту, збереження записаних сесій може потребувати значного обсягу даних. Іншим обмеженням може бути питання конфіденційності. Оскільки під час запису веб-сесій інтерфейс може відображати «чутливі дані» (sensitive data) користувачів (наприклад, адресу електронної пошти, номери банківських карток, тощо) постає потреба у фільтрації таких даних.

Прикладом систем, що використовують підхід із записом змін DOM, є бібліотека з відкритим вихідним кодом rrweb [5].

Запис подій. Цей підхід базується на реєстрації подій користувача, таких як кліки на елементах інтерфейсу, переміщення курсору, заповнення форм, гортання сторінки, навігація між сторінками. На основі записаних подій система спрощено реконструює взаємодію користувача.

Однією з ключових переваг цього методу є мінімальний вплив на продуктивність веб-застосунку, оскільки система реєструє лише події, а не зміни у структурі сторінки. В результаті це знижує вимоги до обсягу збереження даних, оскільки записані події займають менше пам'яті порівняно із збереженням змін DOM. Іншою перевагою підходу із записом подій є швидкий доступ до аналітичних даних, що є корисним для базового аналізу поведінки користувачів.

Проте цьому підходу притаманні такі обмеження, як низька точність. Записані події не дозволяють точно зрозуміти, як змінювався інтерфейс. Також, при використанні цього підходу не фіксуються зміни, що відбуваються без явної дії користувачів (спливаючі підказки, анімації, тощо). Все це призводить до втраченого контексту взаємодії.

З огляду на обмежену точність та неможливість повного відтворення взаємодії, використання лише запису подій на практиці є малопоширеним. Натомість сучасні інструменти застосовують гібридний підхід, що дозволяє

поєднати ефективність підходу із записом подій з точністю підходу із записом змін DOM.

Гібридний підхід. Сучасні системи запису веб-сесій поєднують підхід із записом подій та періодичне фіксування змін DOM. Такі системи записують ключові події користувача, такі як кліки, гортання сторінки, введення тексту, натомість фіксація змін DOM відбувається при значних змінах, наприклад, відкриття спливаючих вікон, навігація між екранами SPA-застосунку.

Даний підхід забезпечує баланс між продуктивністю та точністю, оскільки відбувається запис лише ключових подій користувача та суттєвих змін DOM, що дозволяє знизити навантаження на браузер та сервер. Як результат, у порівнянні з повним записом змін DOM, зменшується і обсяг даних, необхідний для збереження записаних сесій. Гібридний підхід підтримує всі типи веб-додатків — підходить як для традиційних веб-сайтів, так і для динамічних SPA-рішень.

Серед обмежень можна виділити складність реалізації, оскільки необхідно точно визначати, коли записувати зміни DOM, щоб при цьому не втрачати контекст взаємодії.

Більшість провідних рішень використовують гібридний підхід, оскільки він забезпечує оптимальний баланс між точністю та продуктивністю. Прикладами таких рішень є: Hotjar¹, Microsoft Clarity², LogRocket³. З огляду на те, що перелічені рішення є комерційними продуктами із закритим вихідним кодом, це може обмежувати їхнє використання в академічних дослідженнях та малих стартапах із обмеженим бюджетом.

¹ <https://www.hotjar.com/>

² <https://clarity.microsoft.com/>

³ <https://logrocket.com/>

1.2.2 Відмінності між підходами до запису веб-сесій

Підхід із повним записом змін DOM передбачає реєстрацію всіх змін у структурі сторінки в реальному часі, що забезпечує максимально точно відтворення користувацької взаємодії та змін веб-сторінки. Цей підхід дозволяє фіксувати навіть найменші візуальні зміни, такі як анімації, оновлення контенту та зміни стилів CSS. Завдяки цьому він є особливо ефективним для аналізу складних інтерфейсів, зокрема односторінкових веб-застосунків (SPA).

Гібридний підхід поєднує запис ключових подій користувача та періодичні знімки DOM (DOM snapshots). Це дозволяє зменшити обсяг збережених даних та знизити навантаження на систему, що робить його придатним для аналізу поведінки великої кількості користувачів. Однак через періодичну фіксацію стану сторінки можуть втрачатися окремі проміжні зміни, що може вплинути на точність відтворення взаємодії.

Підхід із повним записом змін DOM забезпечує покадрове відтворення всіх змін на сторінці, що є критично важливим для глибокого аналізу зручності використання та поведінки користувачів. Водночас, гібридний підхід зосереджується на спрощеному аналізі поведінкових шаблонів у великих вибірках користувачів, оскільки оптимізує зберігання даних.

Оскільки наше дослідження спрямоване на детальний аналіз поведінки користувачів та точне відтворення їхньої взаємодії з веб-інтерфейсом, для реалізації було обрано підхід із повним записом змін DOM. Він забезпечує найвищу точність у дослідженні зручності використання, дозволяючи ретельно аналізувати динамічні зміни інтерфейсу та особливості користувацької взаємодії.

1.3 Технології відстеження погляду

Відстеження погляду (eye tracking) — це технологія, яка дозволяє визначати напрямок погляду користувача та аналізувати його візуальну увагу. Вона відіграє ключову роль у дослідженнях зручності використання, нейронауках, маркетингу та інших сферах.

Відстеження погляду забезпечує розуміння когнітивних процесів і поведінки користувача, які неможливо спостерігати базуючись на традиційних даних про взаємодію користувачів [6]. Ця технологія надає можливість виявлення шаблонів візуальної уваги, які є вирішальними для розуміння намірів користувача та процесів прийняття рішень [7], проблем зручності використання [8], [9], і навіть підвищення рівня кібербезпеки комп'ютерних систем [10].

Технологія відстеження погляду використовується в різних сферах, включаючи психологію, взаємодію людини з комп'ютером, маркетинг, а також медичні дослідження [11] для вивчення на чому та яким чином користувачі фокусують свою зорову увагу. Слід зазначити, що технологія відстеження погляду, окрім використання у контексті досліджень взаємодії користувачів з цифровими системами, також може використовуватися як метод введення та взаємодії [12].

У дослідженні [10] було запропоновано використання технології відстеження погляду, як метод ідентифікації користувачів, таким чином забезпечуючи додатковий рівень захисту з використанням унікальних візуальних шаблонів. Іншим прикладом застосування цієї технології є визначення емоцій користувачів, що асоціюються або з взаємодією користувача із цифровою системою — [13], чи емоцій, що виникають під час перегляду медіа-контенту [14].

Нещодавнє дослідження розглянуло використання відстеження погляду для оцінки ефективності реклами шляхом аналізу 112 статей, опублікованих у період з 1979 по 2019 роки. Було показано, що підходи із використанням відстеження погляду надають детальне миттєве уявлення про те, на що люди звертають увагу під час перегляду реклами, тим самим зменшуючи кількість помилок, які виникають при покладанні лише на пам'ять та думки респондентів. Це допомагає розуміти, як різні елементи та їх характеристики, такі як колір, розмір, а також напрямок погляду впливають на ефективність оголошень [15].

В залежності від принципу роботи, існують три основні типи систем відстеження погляду (трекерів погляду): стаціонарні, мобільні та програмні рішення на основі веб-камери.

1.3.1 Стаціонарні пристрої відстеження погляду

Стаціонарні трекари погляду функціонують на основі інфрачервоної (ІЧ) технології, яка забезпечує високу точність вимірювання. Їхня робота ґрунтується на використанні ІЧ-діодів, що випромінюють інфрачервоне світло, а також високошвидкісних камер, які фіксують відбите від рогівки світло (corneal reflection method). Отримані дані обробляються за допомогою спеціалізованих алгоритмів, що обчислюють напрямок погляду на основі зміщення світлових відбитків.

Характеристики. Основні характеристики стаціонарних трекерів погляду наведені у Табл. 1.1.

Параметр	Значення
Розмір	Компактні пристрої ($\approx 20\text{--}40$ см) або вбудовані в монітори
Частота запису	250–2000 Гц (залежно від моделі)
Точність	$\sim 0.1^\circ$ (висока)
Затримка	< 5 мс
Сценарії застосування	Лабораторні дослідження, UX-тестування, когнітивні науки

Табл. 1.1. Основні характеристики стаціонарних трекерів погляду

На Рис. 1.2 наведено приклади стаціонарних систем відстеження погляду, зокрема Tobii Pro Spectrum, що працює з частотою до 2 000 Гц і характеризується дуже високою точністю, а також EyeLink 1000 Plus, який підтримує частоту до 2 000 Гц і широко використовується у когнітивних дослідженнях.

Переваги та обмеження. Стаціонарні трекери погляду забезпечують найвищу точність серед усіх технологій відстеження погляду завдяки використанню інфрачервоних камер та високочастотного запису (250–2 000 Гц). Вони ідеально підходять для наукових досліджень, таких як аналіз когнітивних процесів, UX-тестування, психофізіологічні експерименти та нейромаркетинг. Їхня низька затримка (<5 мс) дозволяє детально реєструвати навіть швидкі мікрорухи очей (мікросаккади), що важливо для вивчення візуального сприйняття та когнітивного навантаження. Також ці системи забезпечують високу стабільність вимірювань, оскільки вони фіксуються у статичному положенні, що зменшує вплив рухів голови.

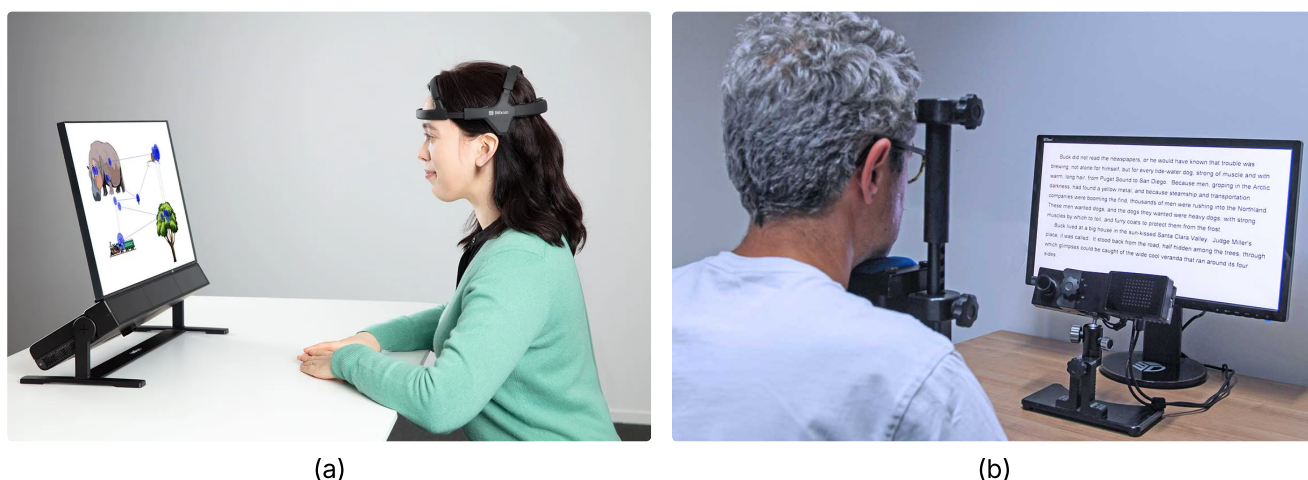


Рис. 1.2. Приклади стаціонарних систем відстеження погляду: (a)⁴ — Tobii Pro Spectrum, (b)⁵ — EyeLink 1000 Plus

Однак основним обмеженням стаціонарних трекерів є те, що вони можуть використовуватися лише в контрольованому середовищі. Це обмежує їхнє застосування у ситуаціях, де важливо досліджувати поведінку користувачів у

⁴ <https://www.tobii.com/products/eye-trackers/screen-based/psychology-neuroscience-use-cases>

⁵ <https://www.sr-research.com/eyelink-1000-plus/>

реальному житті, наприклад, під час руху або взаємодії з фізичними об'єктами. Ще одним суттєвим обмеженням є висока вартість, яка може досягати від 10 000 до 50 000 дол. США, що робить такі системи доступними лише для великих лабораторій або комерційних досліджень. Крім того, освітлення в приміщенні може впливати на точність розпізнавання, хоча більшість сучасних систем використовують ІЧ-фільтрацію для мінімізації цього ефекту.

Попри ці обмеження, стаціонарні трекери залишаються золотим стандартом у дослідженнях, що потребують високою точності та де важлива максимальна деталізація даних. Це робить цей тип систем відстеження погляду незамінними для наукових експериментів та професійного UX-аналізу.

1.3.2 Мобільні (носимі) пристрої відстеження погляду

Пристрої відстеження погляду цього типу зазвичай реалізуються у формі розумних окулярів (smart glasses) із вбудованими датчиками та камерами. Вони оснащені мініатюрними камерами, розташованими поблизу очей, які записують рух очей із частотою 50–200 Гц, а також зовнішньою камерою, що фіксує навколишнє середовище. Зібрані дані обробляються з використанням алгоритмів просторового зіставлення, які поєднують інформацію з декількох камер для точного визначення напрямку погляду в реальному середовищі.

Основні характеристики. Основні характеристики мобільних систем відстеження погляду подані у Табл. 1.2.

На Рис. 1.3 наведено приклади мобільних систем відстеження погляду, зокрема Tobii Pro Glasses 3, які працюють з частотою 100 Гц, мають легку конструкцію та підтримують бездротову передачу даних через Wi-Fi, а також Pupit Labs Core, що характеризуються широкими можливостями налаштування для проведення наукових досліджень.

Параметр	Значення
Розмір	Окуляри вагою ~50–100 г

Частота запису	50–200 Гц
Точність	~0.2° (середня)
Затримка	10–30 мс
Сценарії застосування	Аналіз поведінки у реальному середовищі, медицина, нейромаркетинг

Табл. 1.2. Основні характеристики мобільних (носимих) систем відстеження погляду



(a)



(b)

Рис. 1.3. Приклади мобільних (носимих) систем відстеження погляду: (a)⁶ — Tobii Pro Glasses 3, (b)⁷ — Pupil Labs Core

Переваги і обмеження. Мобільні (носимі) трекери погляду поєднують високу точність із можливістю використання в реальних умовах, що робить їх незамінними в дослідженнях поведінки користувачів за межами лабораторії. Основною перевагою є гнучкість, оскільки пристрій можна носити як окуляри, що дозволяє відстежувати погляд під час руху, взаємодії з фізичними об'єктами або

⁶ <https://www.tobii.com/products/eye-trackers/wearables/tobii-pro-glasses-3/consumer-ux-use-cases>

⁷ <https://pupil-labs.com/products/core>

виконання реальних завдань. Вони широко використовуються у когнітивних дослідженнях, аналізі поведінки покупців, віртуальній реальності та медицині.

Однак мобільні трекери мають і певні обмеження. Через меншу частоту запису (50–200 Гц), порівняно зі стаціонарними системами (до 2 000 Гц), вони менш придатні для досліджень мікрорухів очей. Крім того, фізичний комфорт залишається проблемою: окуляри можуть спричиняти дискомфорт при тривалому носінні, а їхній дизайн може не підходити для всіх форм обличчя. Ще один виклик — калібрування, оскільки рух голови може впливати на точність розпізнавання. Також мобільні трекери залишаються дорогими, із вартістю від 5 000 до 15 000 дол. США, що обмежує їхнє використання у масових дослідженнях.

Попри ці обмеження, мобільні трекери залишаються універсальним компромісом між точністю та можливістю роботи в реальних середовищах, що робить їх цінним інструментом у багатьох сферах досліджень.

1.3.3 Програмні рішення на основі веб-камери

Такі рішення ґрунтуються на використанні звичайних веб-камер у поєднанні з алгоритмами комп'ютерного зору (computer vision) для оцінки напряму погляду. Вони застосовують методи машинного навчання для аналізу положення зіниць і відстеження рухів голови. Камери зазвичай працюють із частотою 30 – 60 Гц, що робить подібні системи придатними для масштабного використання завдяки їх доступності та простоті впровадження.

Прикладом програмного рішення для відстеження погляду є WebGazer.js — бібліотека з відкритим вихідним кодом, що надає можливості відстеження погляду у веб-браузерах [16].

Основні характеристики. У Табл. 1.3 наведено основні характеристики програмних рішень відстеження погляду, які використовують веб-камери.

Параметр	Значення
Розмір	Будь-який пристрій з веб-камерою

Частота запису	30–60 Гц
Точність	~1° (низька)
Затримка	50–100 мс
Сценарії застосування	Веб-аналітика, онлайн-дослідження UX, освітні платформи

Табл. 1.3. Основні характеристики програмних систем відстеження погляду на основі веб-камери

Переваги і обмеження. Програмні рішення для відстеження погляду є найдоступнішим варіантом, оскільки вони не потребують спеціалізованого обладнання і можуть працювати, використовуючи звичайні веб-камери. Це робить їх ідеальними для масштабних досліджень, таких як веб-аналітика, UX-дослідження, онлайн-освіта та навіть адаптивні інтерфейси. Використання алгоритмів комп'ютерного зору та машинного навчання дозволяє компенсувати деякі обмеження у точності, а глибокі нейромережі допомагають адаптувати систему до різних користувачів. Такі рішення легко інтегруються у веб-застосунки та можуть використовуватися без додаткового калібрування, що робить їх зручними для неконтрольованих умов використання.

Проте основним обмеженням програмних рішень є нижча точність порівняно з апаратними трекерами (~1° проти ~0.1° у стаціонарних системах). Це обумовлено тим, що стандартні веб-камери працюють із частотою лише 30–60 Гц, тоді як професійні пристрої можуть фіксувати погляд із частотою понад 1 000 Гц. Також зовнішні фактори, такі як освітлення, якість камери та положення голови, можуть значно впливати на точність розпізнавання. Ще одним обмеженням є затримка в реальному часі (50–100 мс), яка може бути критичною для застосунків, що потребують миттєвого реагування.

Попри ці недоліки, програмні трекери є ідеальним компромісом між доступністю та функціональністю, що дозволяє впроваджувати технології

відстеження погляду у масштабовані цифрові рішення, не потребуючи високоякісного обладнання. Завдяки розвитку глибокого навчання та оптимізації алгоритмів, ці системи можуть стати ще більш точними у майбутньому.

1.3.4 Порівняльний аналіз технологій відстеження погляду

У Табл. 1.4 наведено основні характеристики кожного з описаних типів систем відстеження погляду.

Тип трекера	Точність	Частота запису	Вартість	Основні застосування
Стационарні	~0.1°	250–2 000 Гц	Висока (від 10 000 дол. США)	Лабораторні дослідження, UX, когнітивні науки
Мобільні	~0.2°	50–200 Гц	Висока (від 5 000 дол. США)	Аналіз поведінки, нейромаркетинг, медицина
Програмні	~1°	30–60 Гц	Низька (до 500 дол. США)	Веб-аналітика, UX, онлайн-дослідження

Табл. 1.4. Порівняльна таблиця технологій відстеження погляду

Комерційні апаратні рішення, такі як Tobii та EyeLink мають перевагу у високій точності та надійності даних відстеження погляду. Ці системи широко використовуються у дослідженнях і індустрії загалом для вивчення візуальної уваги користувачів. Проте, висока вартість та необхідність використання спеціалізованого обладнання робить ці рішення менш доступними для звичайних користувачів та академічних дослідників.

Запропонована в цій роботі система використовує програмний підхід, що дозволяє отримати масштабоване рішення для аналізу поведінки користувачів,

зберігаючи доступність без необхідності використання високовартісного обладнання.

1.3.5 Використання технології у веб-аналізі поведінки

Завдяки доступності веб-камер та розвитку алгоритмів комп'ютерного зору програмні рішення стали особливо актуальними у сфері аналізу поведінки користувачів у веб-середовищі. Такі системи дозволяють аналізувати зони візуальної уваги під час взаємодії з веб-сторінками, визначати елементи інтерфейсу, які залишаються поза увагою або, навпаки, становлять найбільше зацікавлення. Крім того, відстеження погляду можна поєднувати з записом сесій, що дає змогу дослідникам отримувати глибше розуміння шаблонів поведінки користувачів.

Запропоноване в цій роботі рішення поєднує програмне відстеження погляду з записом веб-сесій користувачів, що дозволяє аналізувати їхню поведінку без використання високовартісного обладнання. Це робить систему доступною для масштабного впровадження та використання у академічних та практичних контекстах.

1.4 Аналіз та вибір технології відстеження погляду

Для вибору системи відстеження погляду у поточному дослідженні було розглянуто декілька програмних рішень. Під час аналізу акцент робився на рішення з відкритим вихідним кодом (що дозволить зменшити витрати і полегшити підтримку такого рішення у майбутньому) та можливостями кросплатформності (що підвищило б доступність такого рішення для користувачів різних операційних систем).

Одним з таких рішень є OGAMA (Open Gaze and Mouse Analyzer) [17]. Це програмне рішення із відкритим вихідним кодом для аналізу рухів очей та курсора миші в експериментах на основі слайд-шоу. OGAMA оптимізовано для найкращої роботи з фіксованими зображеннями або слайд-шоу, що обмежує його

використання під час роботи з динамічним та інтерактивним контентом. Крім того, OGAMA не є кросплатформним рішенням, що також обмежує використання цього інструменту при роботі на інших операційних системах.

Іншим інструментом з відкритим вихідним кодом, призначеним для точного відстеження погляду у контрольованих експериментальних середовищах є GazeParser [18]. Дане рішення використовує мову програмування Python для розширеного аналізу даних та інтегрується з такими бібліотеками як PsychoPy. Проте, незважаючи на переваги, стільникова природа (desktop-based) цього рішення ускладнює його інтеграцію з інструментами запису веб-сесій, щоб забезпечити повне та просте у використанні рішення, яке поєднує запис сесій із відстеженням погляду.

WebGazer.js — бібліотека з відкритим вихідним кодом, яка надає можливості відстеження погляду у веб-браузерах використовуючи стандартну веб-камеру. Бібліотека використовує методи комп'ютерного зору та машинного навчання для визначення точки, в яку дивиться користувач на екрані (gaze point). WebGazer.js робить технологію відстеження погляду більш доступною для звичайних користувачів через виключення необхідності у використанні спеціалізованого обладнання. Проте, це рішення має обмеження, пов'язані з точністю та необхідністю калібрування.

Недавнє дослідження перевірило можливість використання WebGazer.js для досліджень у ранньому дитинстві. Було встановлено, що, хоча віддалений веб-метод на основі WebGazer.js призводить до більш шумних даних і вищого рівня вибуття (тобто більшої кількості учасників, які припинили участь у дослідженні до його завершення) порівняно з традиційними лабораторними системами відстеження руху очей, він все ж таки забезпечує значущі висновки щодо передбачення цілеспрямованих дій [19].

Інше дослідження виявило, що WebGazer.js забезпечує точні та достовірні вимірювання напрямку погляду для передбачення думок, не пов'язаних із

завданням (TUT, task-unrelated thought), та розуміння прочитаного, демонструючи його потенціал для масштабованого й доступного в реальному часі виявлення когнітивного стану в освітніх онлайн-середовищах [20].

1.5 Інтеграція технології відстеження погляду із записом сесій

Незважаючи на потенційні переваги інтеграції технології відстеження погляду із записами сесій, існує небагато інструментів, які ефективно поєднують ці дві технології. У роботі було розглянуто декілька прикладів таких інструментів.

GazeRecorder є одним із прикладів такої системи, яка забезпечує можливість відстеження погляду, використовуючи веб-камеру, та можливість запису веб-сесій [21]. Вона надає можливість відстеження погляду у реальному часі та генерації теплових карт, використовуючи агреговані дані від декількох користувачів. Це робить дану систему економічно вигідним та ефективним рішенням для дослідження зручності використання та поведінки користувачів у веб-середовищах. Одним з обмежень використання GazeRecorder є те, що його інтеграція потребує доступу та модифікації кодової бази веб-сайту. Іншим обмеженням є те, що GazeRecorder не є рішенням з відкритим вихідним кодом, що обмежує можливість його додаткового налаштування та розширення іншими розробниками.

iTrace є іншим прикладом інструменту, що інтегрує технологію відстеження погляду, але в інтегроване середовище розробки (IDE), для аналізу поведінки користувача під час виконання завдань з розробки програмного забезпечення [22]. Для забезпечення комплексного розуміння взаємодії розробників та когнітивних процесів система записує та аналізує, як дії користувачів так і детальні дані відстеження погляду, включаючи контекст (програмний код, артефакти, тощо). Однак через те, що iTrace спеціалізується виключно на середовищах IDE, його застосування для аналізу поведінки користувачів у веб-середовищі є обмеженим.

1.6 Використання жестів у НСІ

В останні роки методи взаємодії з цифровими системами на основі жестів набувають дедалі більшої популярності. У дослідженні [23] технології, що використовують жести для взаємодії, пропонується класифікувати на дві основні групи:

1. *Візуально-орієнтовані технології* — ґрунтуються переважно на відстеженні за допомогою камери або сенсорів, коли жести рук користувачів записуються на відстані. До цієї категорії належать звичайні відеокамери, інфрачервоні/глибинні 3D-камери та системи захоплення руху.
2. *Носимі пристрої* — технології, що базуються на пристроях, які користувачі носять на собі, таких як рукавички, акселерометри, кільця або браслети.

Перевагами використання жестів для взаємодії з комп'ютерними системами є їх інтуїтивна зрозумілість. Так, систематичний огляд жестової взаємодії в доповненій реальності (AR) свідчить, що жести в повітрі особливо ефективні в контекстах, де користувачі зосереджені на фізичних завданнях, наприклад, у медичних або промислових умовах, оскільки такі жести дозволяють взаємодіяти з цифровими елементами без необхідності повністю концентруватися на самій системі. У тому ж дослідженні також наголошується на перевагах простих жестів, оскільки вони покращують зручність використання, будучи легкими для запам'ятовування і потребуючи мінімального навчання [24].

Жестова взаємодія використовується у багатьох сферах діяльності людини. Так, у дослідженні [25] розглядаються приклади застосування інтерфейсів на основі комп'ютерного зору і розпізнавання жестів рук у таких сферах як медичні системи, допоміжні технології, розваги, управління кризовими ситуаціями, ліквідація наслідків катастроф та взаємодія людини з роботами.

У дослідженні [26] було розроблено систему керування жестами для медичної візуалізації, яка дозволяє користувачам інтуїтивно маніпулювати 3D-медичними зображеннями за допомогою рухів рук. В іншому дослідженні було реалізовано безконтактний інтерфейс, заснований на жестах, для пошуку медичних зображень [27].

Підхід до взаємодії з використанням жестів для маніпулювання об'єктами у веб-середовищі, який дає змогу користувачам взаємодіяти з 2D-об'єктами за допомогою відстеження рук і жесту «щипок» (pinch), було запропоновано у дослідженні [28]. За результатами попереднього користувацького дослідження, учасники позитивно оцінили інтуїтивність запропонованого методу та плавність взаємодії, яку він забезпечує.

1.7 Взаємодія на основі погляду та жестів

Як вже було зазначено раніше в межах цього розділу дисертаційного дослідження, технологія відстеження погляду знайшла широке застосування у сфері взаємодії людини з комп'ютером, зокрема для аналізу поведінки користувачів.

Інший приклад застосування цієї технології представлений у дослідженні [12], де розглядається використання відстеження погляду як методу введення та взаємодії з цифровими системами. Вказування на графічні елементи є однією з основних задач у сфері взаємодії людини з комп'ютером (HCI) [29]. У цьому контексті технологія відстеження погляду має перевагу порівняно з традиційними методами введення, такими як комп'ютерна миша, завдяки високій швидкості [30]. Це дає змогу користувачам швидко вибирати об'єкти, просто фокусуючи на них погляд.

Сучасні методи взаємодії людини з цифровими системами можуть поєднувати кілька модальностей введення. Так, жести рук можуть доповнювати відстеження погляду, забезпечуючи більш природній спосіб маніпулювання об'єктами, що, в свою чергу, дає змогу виконувати такі дії, як обертання,

масштабування та перетягування. У дослідженні [31] було запропоновано підхід, що поєднує погляд і жести рук для взаємодії у тривимірному віртуальному просторі.

У [32] представлено набір технік взаємодії, заснованих на поєднанні погляду та жестів рук. Результати показали, що така комбінація може перевершувати системи, які використовують лише погляд або лише жести. Інше дослідження також засвідчило, що техніки взаємодії, у яких погляд використовується як допоміжний засіб, перевершують введення, здійснене виключно жестами, і надають уявлення про компроміси при поєднанні погляду з прямими чи непрямими, а також просторовими або семантичними жестами рук у вільному просторі [33].

У дослідженні [34] було запропоновано просторову техніку взаємодії під назвою Gaze-Grasp Pose Interaction (GG Interaction), призначену для маніпулювання об'єктами у тривимірному віртуальному середовищі. Ця техніка поєднує відстеження погляду з жестом захоплення, що дозволяє зручно та інтуїтивно вибирати як малі, так і віддалені об'єкти, зменшуючи фізичну втому користувача та усуваючи невідповідність між видимістю об'єкта і положенням руки. Проведений експеримент показав, що GG Interaction перевершує стандартну техніку ray-casting у завданнях вибору об'єктів, особливо за умов часткового перекриття. Тестування за участю 20 користувачів підтвердило, що запропонований метод забезпечує швидший вибір, нижчу частоту помилок і вищу природність взаємодії порівняно з традиційними підходами.

У дослідженні [35] продемонстровано портативну гарнітуру для відстеження погляду у вигляді окулярів із вбудованими камерами для фіксації руху очей і сцени. Система дозволяла користувачам вибирати об'єкти поглядом і керувати ними за допомогою жестів рук, забезпечуючи мобільну та безконтактну взаємодію. Водночас було відзначено, що паралаксні помилки та обмеження в

точності розпізнавання жестів залишаються викликом для подальшого вдосконалення подібних рішень.

Систему віртуальної миші, яка дає змогу керувати курсором за допомогою рухів очей і рук, було представлено у [36]. Система застосовує методи комп'ютерного зору та обробки зображень, що дозволяє виконувати всі базові дії миші без фізичного контакту з пристроєм. За результатами тестування за участю 58 користувачів встановлено, що жести руками є точнішими, ніж жести очима, особливо під час імітації кліків.

В останні роки поєднання відстеження погляду та жестів рук широко досліджувалося в контексті віртуальної реальності (VR). Так у дослідженні [37] було розроблено прототип VR-системи, в якій користувачі могли дивитися на об'єкти та маніпулювати ними за допомогою рук. Результати експериментального дослідження засвідчили високий рівень занурення у віртуальне середовище, досягнутий завдяки цьому підходу.

Попри численні переваги, метод взаємодії на основі поєднання погляду та жестів має і низку викликів. У дослідженні [38] автор розглядає принципи проектування та проблеми, пов'язані з інтерфейсами, що використовують взаємодію за допомогою погляду та жести стиснення пальців (pinch interaction). Серед визначених принципів — розподіл функцій між очима та руками, підвищення точності погляду, мінімізація хибних спрацьовувань та забезпечення сумісності з іншими методами керування. Крім того, наголошується на проблемах адаптації користувачів, точності виконання дій, часу реакції системи та труднощах, що виникають при взаємодії з об'єктами, розташованими на різних відстанях.

1.8 Вирази обличчя як засіб взаємодії у НСІ

Останніми роками вирази обличчя розглядаються як перспективний засіб взаємодії людини з цифровими системами. Подібно до технологій взаємодії, заснованих на жестах рук, системи, які використовують вирази обличчя умовно поділяються на:

1. *Візуально-орієнтовані* — ґрунтуються на використанні камер або сенсорів для дистанційного фіксування мімічних змін обличчя користувача.
2. *Носимі пристрої* — базуються на застосуванні сенсорів, що прикріплюються до тіла користувача, зокрема до обличчя (наприклад, окуляри з вбудованими сенсорами або електроди, прикріплені до шкіри).

У дослідженні [39] було розглянуто застосування виразів обличчя для «природної взаємодії» з цифровими пристроями, такими як смартфони та планшети. Зокрема, було проаналізовано сценарій використання електронного пристрою для читання книг (eBook reader), де користувачі могли керувати функціоналом за допомогою виразів обличчя. Запропонована система розпізнавала емоційні вирази та мімічні жести, інтерпретуючи їх як команди введення. Незважаючи на ефективність у створенні природної взаємодії, система підтримувала лише обмежений набір базових жестів, не проходила тестування в завданнях, що потребують високої точності, і не була адаптована до різноманітних потреб користувачів. Це зумовило обмежене застосування підходу у складніших сценаріях, зокрема для маніпулювання об'єктами.

У дослідженні [40] вивчалось використання виразів обличчя та жестів рук у сфері охорони здоров'я та розумних середовищ. Автори припускають, що такі методи взаємодії можуть бути корисними для користувачів із фізичними обмеженнями. Хоча у роботі було продемонстровано ефективність застосування мімічних жестів та жестів рук для взаємодії в розумних середовищах, дослідження не включало оцінювання точності системи під час виконання складних завдань. Це може обмежувати можливості застосування наведеного підходу у більш складних сценаріях взаємодії, таких як точне керування пристроями.

У роботі [41] було представлено підхід до аналізу виразів обличчя в контексті людино-комп'ютерної взаємодії, заснований на комп'ютерному зорі. Основна увага приділялася вдосконаленню комунікації між користувачем і машиною, зокрема для осіб з обмеженою рухливістю. Проте дослідження зосереджувалося переважно на аналізі та прогнозуванні активності м'язів, приділяючи недостатньо уваги інтеграції в прикладні системи.

Інше дослідження [42] запропонувало систему, що поєднує локалізацію ключових точок обличчя, розпізнавання виразів та відстеження погляду для реалізації інтерактивних функцій. Водночас система не враховувала складність зіставлення виразів обличчя з конкретними діями, що обмежує її прикладну ефективність.

У дослідженні [43] було реалізовано підхід «передавального навчання» (transfer learning) для розпізнавання емоцій у взаємодії людини з комп'ютером. Автори підкреслюють потенційну користь розпізнавання емоцій для таких галузей, як обслуговування клієнтів, е-навчання, робототехніка тощо. Однак дослідження зосереджувалося на діагностиці афективних станів, а не на активній взаємодії, що обмежує його застосування у безконтактних системах керування.

Подібні висновки наведено у [44], де йдеться про розпізнавання емоцій на основі виразів обличчя під час виконання завдань за комп'ютером. Дослідження вказує, що вирази обличчя не завжди точно відображають внутрішній емоційний стан користувача, особливо в контексті тривалого навантаження.

У дослідженні [45] було розроблено систему цифрового аватара з реалістичною синхронізацією виразів обличчя користувача для забезпечення природної взаємодії людини з машиною. Система ґрунтується на параметризованій 3D-моделі, реалізованій за допомогою технології мімічних форм змішування (blendshapes). Незважаючи на досягнення у реалістичному відображенні міміки, система не була призначена для виконання точних дій з об'єктами.

У дослідження [46] було представлено метод взаємодії, що використовує м'язову активність обличчя для керування аватаром у віртуальному середовищі. Система працює із використанням електродів, що прикріплені до шкіри користувача, які виявляють підшкірну м'язову активність за допомогою електроміографії. Використовуючи прототип системи користувачі мали змогу передавати вирази обличчя у віртуальному середовищі без необхідності використання камери. Система розглядається як прототип для реабілітації м'язів обличчя, зокрема для пацієнтів із гіпомімією.

Робота [47] описує систему OCOSense — смарт-окуляри, оснащені оптоміографічними сенсорами та інерційним вимірювальним блоком (Inertial Measurement Unit, IMU) — пристроєм, що об'єднує акселерометри, гіроскопи та іноді магнітометри для визначення положення, орієнтації та руху в просторі. Така комбінація дає змогу розпізнавати вирази обличчя та мімічні жести. Отримані дані передаються на мобільний пристрій для подальшої обробки. Система орієнтована на контекстно-орієнтовану взаємодію в середовищах AR/XR та має перспективи для використання в афективних обчисленнях і моніторингу психічного здоров'я.

У низці досліджень [48], [49], [50] акцент зроблено на моделюванні поведінки комп'ютерної миші за допомогою виразів обличчя. Наприклад, у [48] реалізовано систему керування веб-браузером через мімічні жести, такі як моргання чи закривання одного ока. Результати тестування засвідчили високу точність розпізнавання та задоволеність користувачів. У [49] за допомогою бібліотеки OpenCV було реалізовано інтерфейс, де користувач міг керувати курсором, імітувати кліки, прокрутку тощо — лише за допомогою рухів обличчя та голови. Автори відзначили потенціал системи як альтернативного засобу введення для користувачів із фізичними обмеженнями.

У низці інших досліджень автори досліджували поєднання виразів обличчя з іншими модальностями. У роботі [51] представлено мультимодальну систему FaceSwitch, яка поєднує відстеження погляду та мімічні жести та надає змогу

взаємодіяти з користувацьким інтерфейсом шляхом наведення погляду для вибору об'єкта та використання мімичних жестів для виконання відповідних команд. Жести асоціюються з конкретними подіями введення (клацання, прокрутка), виконуючи роль «механічних перемикачів». Кількісні показники, отримані в ході користувацького дослідження, свідчать про те, що запропонована парадигма взаємодії покращує ефективність традиційних опцій доступності, таких як управління лише за допомогою погляду або керування поглядом у поєднанні з механічним перемикачем, та водночас наближається за швидкістю та точністю до традиційної взаємодії з використанням миші.

У дослідженні [52] було розроблено систему із використанням методів комп'ютерного зору для дитини з дегенеративним нервово-м'язовим захворюванням, яка надавала змогу взаємодіяти з комп'ютером за допомогою розпізнавання виразів обличчя. Система використовувала стандартну веб-камеру та метод головних компонент (principal component analysis) для класифікації виразів обличчя, розпізнаючи вирази «посмішка» та «поцілунок» для активації функцій миші. Попри ефективність у імітації базових дій мишею, ця система не підтримує складніші сценарії взаємодії.

У [53] було представлено систему, яка поєднує відстеження погляду та мімичні рухи для керування курсором миші, та модуль перетворення мовлення в текст (Speech-to-Text) — для виконання текстового вводу голосом. Розробка орієнтована на користувачів із важкими формами паралічу. При такому стані можливості людини можуть бути обмежені лише рухами очей, головою та використанням голосових команд. Автори відзначили інтуїтивність інтерфейсу та покращення досвіду користувача завдяки мультимодальній взаємодії.

Окрім систем, що базуються на комп'ютерному зорі, перспективним напрямом є використання нейрокомп'ютерних інтерфейсів, які зчитують електричну активність мозку користувача. У роботі [54] було реалізовано браузерну систему, що дозволяє здійснювати управління за допомогою ментальних

команд на основі сигналів електроенцефалографії (Electroencephalography, EEG) — технології, що реєструє електричну активність мозку за допомогою електродів, розміщених на поверхні голови. Автори відзначають потенціал таких підходів для збирання інформації про стан користувача в реальному часі, що відкриває можливості для побудови адаптивної взаємодії.

1.9 Виявлені дослідницькі прогалини

Аналіз літератури та наявних рішень у сфері аналізу поведінки користувачів і підходів до мультимодальної взаємодії дозволив виявити низку актуальних наукових прогалин, які стали підґрунтям для формулювання дослідницьких напрямів цієї роботи.

У галузі аналізу поведінки користувачів у веб-середовищі простежується нестача інструментів, які дозволяють не лише фіксувати дії користувача, а й аналізувати його візуальну увагу. Хоча існують окремі рішення для запису веб-сесій (наприклад, rrweb) та для відстеження погляду (зокрема WebGazer.js), відсутні цілісні інструменти, що поєднують обидві технології в єдиному середовищі. Крім того, наявні підходи або потребують модифікації коду веб-застосунків, або не забезпечують кросплатформність і зручність, необхідні для широкого використання у практичному UX-дослідженні.

У контексті безконтактної взаємодії на основі погляду та жестів, більшість попередніх реалізацій ґрунтується на спеціалізованому апаратному забезпеченні — інфрачервоних сенсорах, промислових трекерах або пристроях на зразок Microsoft Kinect. Така апаратна залежність обмежує застосування відповідних систем у реальних умовах. При цьому останні досягнення в галузі комп'ютерного зору створюють можливості для реалізації подібних підходів до взаємодії на базі лише стандартної веб-камери, зокрема — безпосередньо у веб-браузері. Проте, практичних реалізацій, які б дозволяли оцінити ефективність такого підходу у звичних цифрових середовищах, майже не існує.

Ще одним малодослідженим напрямом є використання виразів обличчя як засобу управління у веб-середовищі. Існуючі рішення здебільшого обмежуються фіксованими наборами жестів (наприклад, усмішка, підняття брів), що унеможлиблює гнучке налаштування взаємодії. Крім того, більшість реалізацій орієнтовані на асистивні технології або спеціалізовані пристрої, тоді як потенціал використання виразів обличчя як способу взаємодії у веб-середовищах залишається практично невивченим.

Таким чином, виявлені дослідницькі прогалини стосуються:

- недостатньої кількості інструментів, що поєднують запис веб-сесій із даними про зорову увагу;
- обмеженої доступності мультимодальної взаємодії на основі погляду і жестів у веб-середовищі;
- відсутності рішень, що дозволяють використовувати вирази обличчя як гнучкий та налаштовуваний засіб взаємодії.

Ці обмеження стали основою для визначення трьох напрямів дослідження, розкритих у наступних розділах дисертації.

1.10 Формулювання загальної задачі дослідження

Ураховуючи виявлені дослідницькі прогалини, сформульовано загальну задачу дослідження, яка полягає у розробці доступних мультимодальних засобів безконтактної взаємодії з цифровими об'єктами та інструментів для комплексного аналізу поведінки користувачів у веб-середовищі. Особливу увагу приділено створенню рішень, які поєднують відстеження погляду, розпізнавання жестів і виразів обличчя на основі стандартного обладнання, зокрема веб-камери, без потреби у спеціалізованих пристроях або модифікації кодової бази веб-застосунків.

Задача дослідження передбачає створення архітектурних і програмних рішень, які забезпечують:

- синхронізований запис та відтворення веб-сесій із візуалізацією зорової уваги;

- можливість природної взаємодії з цифровими об'єктами за допомогою погляду та жестів;
- використання виразів обличчя як засобу взаємодії для виконання довільних дій у цифровому середовищі;
- експериментальну перевірку зручності, ефективності та доступності запропонованих підходів.

Таким чином, дисертаційна робота спрямована на вирішення комплексної задачі, що охоплює області аналізу поведінки користувачів, людино-комп'ютерної взаємодії та комп'ютерного зору.

РОЗДІЛ 2. ІНТЕГРАЦІЯ ІНСТРУМЕНТІВ ЗАПИСУ СЕСІЙ ТА ВІДСТЕЖЕННЯ ПОГЛЯДУ ДЛЯ АНАЛІЗУ ПОВЕДІНКИ КОРИСТУВАЧІВ У ВЕБ-СЕРЕДОВИЩІ

Цей розділ дисертаційного дослідження присвячено розробці інструменту для аналізу поведінки користувачів у веб-середовищі з урахуванням зорової уваги. На основі аналізу сучасних підходів до запису веб-сесій і технологій відстеження погляду (розглянутих у Розділ 1) обґрунтовано доцільність їх інтеграції в межах єдиної системи. Такий підхід дає змогу не лише фіксувати дії користувача на веб-сторінці, а й досліджувати, як візуальна увага впливає на прийняття рішень під час взаємодії з інтерфейсом.

У цьому розділі представлено практичну реалізацію запропонованого підходу у вигляді кросплатформного розширення для веб-браузера. Розроблена система поєднує запис веб-сесій із синхронізованим відображенням зорової уваги користувача. Основні ідеї та технічні рішення, що лягли в основу цієї реалізації, були висвітлені в попередній науковій публікації — [55].

2.1 Формалізація задачі запису веб-сесій з інтегрованими даними відстеження погляду

Необхідно створити систему, яка дозволяє записувати та відтворювати веб-сесії, синхронізуючи дані взаємодії користувача з даними відстеження погляду. Система має забезпечувати такі можливості як:

1. *Запис веб-сесії* — збереження послідовності подій, які включають події взаємодії користувача та зміни в DOM.
2. *Запис даних погляду* — збереження координат погляду у часовій шкалі.
3. *Відтворення веб-сесії* — реконструкцію дій користувача разом із відображенням його погляду у відповідні моменти часу.

Запис подій сесії. Під час запису сесії фіксується єдина послідовність подій, що включає:

1. Події взаємодії користувача — дії, які користувач виконує на сторінці (кліки, введення тексту, прокрутка тощо).
2. Події змін у DOM — структурні та візуальні зміни сторінки (додавання, зміна, видалення елементів, зміна атрибутів).

Послідовність усіх подій визначається як:

$$\mathcal{E} = \{t_i, e_i, d_i\}_{i=1}^{N_e}, \quad (2.1)$$

де:

- t_i — часова мітка події;
- e_i — тип події (click, input, scroll, mutation, тощо);
- d_i — додаткові параметри, що описують подію (наприклад, координати кліку або змінене значення елемента).

Запис даних погляду. Дані погляду записуються і зберігаються окремо у вигляді послідовності:

$$\mathcal{G} = \{t_j, x_g(t_j), y_g(t_j)\}_{j=1}^{N_g}, \quad (2.2)$$

де:

- t_j — часова мітка;
- $(x_g(t_j), y_g(t_j))$ — координати погляду у момент часу t_j .

Відтворення сесії. У процесі відтворення сесії відбувається послідовне виконання записаних подій та відображення координат погляду користувача у відповідні моменти часу. Функція відтворення визначається як:

$$R(t) = (E_i \mid t_i \leq t) \cup (G_j \mid t_j = t \text{ або } t_j = \max(t_k \mid t_k < t)), \quad (2.3)$$

де $R(t)$ — стан сесії у момент часу t , що включає:

- E_i — всі події, що сталися до або в момент t ;
- G_j — координати погляду в момент t , а якщо такої точки немає, то береться остання відома точка погляду G_j з міткою часу $t_k < t$.

2.2 Архітектура системи та реалізація

Запропоноване у цьому модулі дисертаційного дослідження рішення реалізовано у вигляді розширення для веб-браузера Chrome. Система фіксує дані взаємодій користувача із веб-сторінкою (рухи миші, кліки, прокрутку сторінки, навігацію, введення даних у форми, тощо), візуальні та структурні зміни веб-сторінки, а також дані відстеження погляду користувача.

Поєднання даних про взаємодію користувача із веб-сторінкою та даних відстеження погляду дозволяє отримати детальне розуміння не лише того, які дії виконував користувач, а й на які елементи веб-сторінки він звертав увагу в момент цих дій. Такий подвійний підхід дає змогу дослідникам і дизайнерам аналізувати поведінку користувачів у комплексі, що може значно підвищити якість досліджень зручності використання та поведінки користувачів.

Однією з переваг запропонованого рішення є те, що система працює у фоновому режимі, не впливаючи на дії користувача під час запису сесії. Це гарантує, що процес збору даних не впливає на реальний досвід використання веб-сторінки, що в свою чергу забезпечує точніші результати аналізу.

2.2.1 Технічний огляд

Для розробки запропонованої системи було використано сучасні веб-технології та бібліотеки, що забезпечують її стабільну та ефективну реалізацію. Основним фреймворком для створення інтерфейсу користувача було обрано React у поєднанні з мовою програмування TypeScript. React є популярною JavaScript-

бібліотекою, що використовує компонентний підхід, який сприяє повторному використанню коду та підтримці модульної структури додатка. Використання TypeScript забезпечує статичну типізацію, що покращує якість коду, підвищує його підтримуваність і зменшує кількість помилок під час виконання.

Для запису веб-сесій користувачів система використовує `rrweb` — бібліотеку з відкритим вихідним кодом, яка надає можливості запису та відтворення веб-сесій. Бібліотека реєструє зміни у DOM та події взаємодії користувачів. Основними перевагами `rrweb` є можливість створювати детальні записи взаємодії з веб-сторінками за мінімального впливу на продуктивність браузера. Це робить її оптимальним рішенням для запису веб-сесій без негативного впливу на досвід користувачів.

Функціональність відстеження погляду була реалізована за допомогою `WebGazer.js` — бібліотеки з відкритим вихідним кодом, що дозволяє здійснювати відстеження погляду в режимі реального часу, використовуючи звичайну веб-камеру. `WebGazer.js` використовує моделі машинного навчання для прогнозування точки, на яку користувач дивиться на екрані. Ця бібліотека була обрана через простоту інтеграції та здатність працювати безпосередньо у веб-браузері, без необхідності використання спеціалізованого обладнання.

2.2.2 Основні компоненти

Архітектура розширення побудована із використанням модульного підходу, що передбачає розподіл системи на кілька незалежних компонентів. Розроблена система складається з наступних модулів: користувацький інтерфейс (UI), модуль запису сесій (`session recording module`), модуль відстеження погляду (`eye-tracking module`), сценарій вмісту (`content script`), фоновий сценарій (`background script`), а також модуль локального збереження даних (`local storage`). Така структуризація забезпечує чітке розмежування функцій, що спрощує підтримку та масштабування розширення у майбутньому. На Рис. 2.1 наведено компонентну діаграму, що ілюструє взаємодію основних модулів розширення.

Користувацький інтерфейс реалізований з використанням бібліотеки React та мови програмування TypeScript, що дозволяє створити інтерактивне середовище для керування процесом запису сесій. Він надає користувачам можливість запускати й зупиняти процес запису сесій, переглядати та керувати даними записаних сесій, а також відтворювати сесії користувачів з інтегрованими даними відстеження погляду.

Основу модуля запису веб-сесій становить бібліотека rrweb, яка фіксує зміни у DOM та взаємодію користувача. Модуль відповідає за ініціалізацію rrweb, керування початком та завершенням запису, а також передачу зібраних даних у сценарій вмісту.

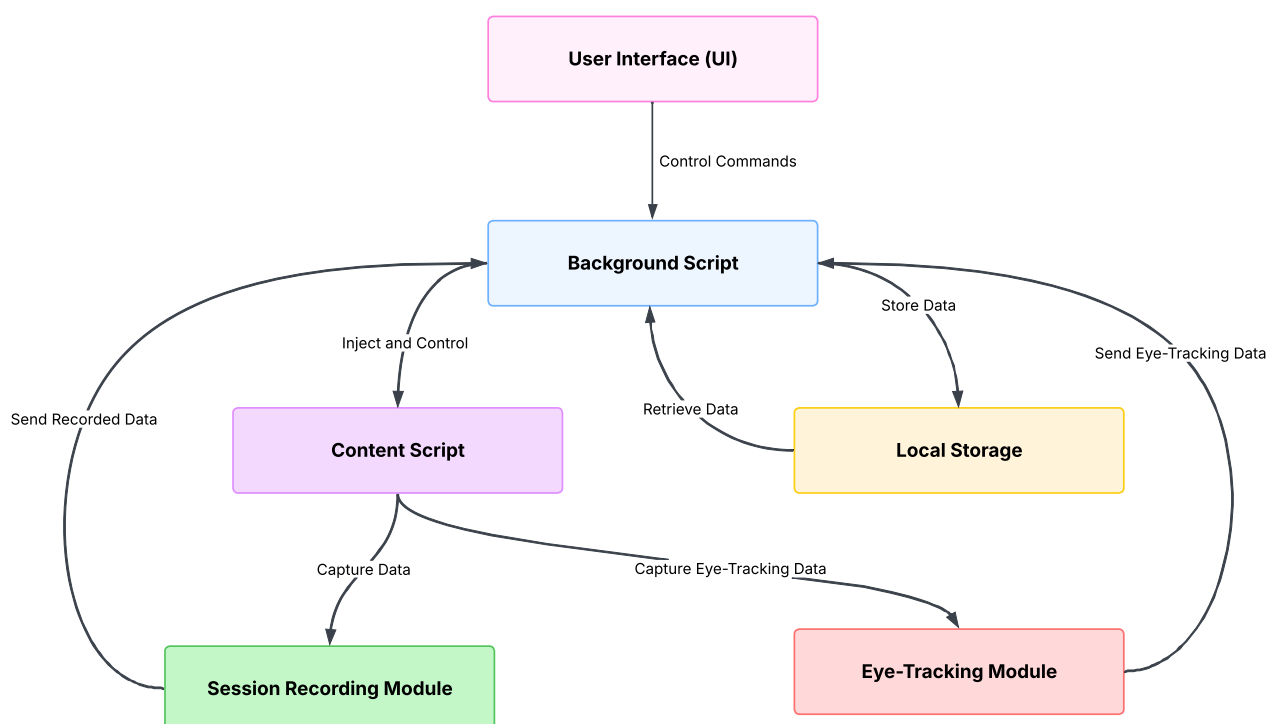


Рис. 2.1. Компонентна діаграма запропонованого розширення

Для збору інформації про положення погляду користувача у модуль відстеження погляду інтегровано бібліотеку WebGazer.js. Функції модуля

включають процес калібрування, запуск і зупинку збору даних та передачу зібраних даних (координат погляду) у сценарій вмісту.

Сценарій вмісту вбудовується у веб-сторінки та відповідає за ініціалізацію та зв'язок із модулями запису сесій та відстеження погляду. Він отримує дані про зміни у DOM та події взаємодії користувача, а також дані відстеження погляду, та передає отримані їх до фонового сценарію, який реалізує основну логіку розширення.

Фоновий сценарій забезпечує обмін даними між користувацьким інтерфейсом та іншими модулями, синхронізує події та зберігає фінальний запис сесій.

Для локального зберігання записаних сесій використовується Web Storage API, що забезпечує збереження, експорт і імпорт даних безпосередньо на пристрої користувача. Це дозволяє працювати із записаними сесіями без необхідності постійного підключення до сервера, що підвищує автономність системи.

Запропонована модульна архітектура розширення забезпечує високу гнучкість системи, дозволяє ефективно поєднувати запис сесій із відстеженням погляду та спрощувати подальше розширення функціональності.

2.2.3 Процес роботи системи

Процес роботи розширення базується на чіткій послідовності кроків, що забезпечують узгоджене збереження та синхронізацію даних про взаємодію та зорову увагу користувачів.

Ініціалізація. Під час активації розширення фоновий сценарій ініціалізує всі необхідні компоненти та вбудовує сценарій вмісту у поточну веб-сторінку.

Початок запису сесії. Коли користувач натискає кнопку запису у користувацькому інтерфейсі розширення, запит передається до фонового сценарію, який надсилає команду сценарію вмісту на запуск модулів запису сесії та відстеження погляду.

Збір даних. Сценарій вмісту отримує дані про зміни у DOM та взаємодію користувача, а також дані відстеження погляду, та передає їх у фоновий сценарій для подальшої обробки.

Завершення запису. Коли користувач зупиняє запис, користувацький інтерфейс передає команду фоновому сценарію, який надсилає команду на завершення запису та збору даних та розпочинає їх обробку.

Обробка та збереження даних. Фоновий сценарій фільтрує отримані дані, застосовуючи до даних відстеження погляду такі алгоритми фільтрування, як ковзне середнє (moving average filtering), пороговий фільтр швидкості (velocity thresholding), фільтрацію на основі відстані (distance-based thinning). Після обробки дані синхронізуються та зберігаються локально, використовуючи модуль локального зберігання для подальшого аналізу та відтворення.

Для наочного представлення цього процесу було розроблено UML-діаграму послідовності з деталізацією взаємодії між основними компонентами системи — Рис. 2.2.

Для забезпечення гнучкості та портативності розширення підтримує функції експорту записаних веб-сесій у файл та імпорту із файлу для відтворення веб-сесій, записаних на інших пристроях.

2.2.4 Процес калібрування

Для забезпечення точної роботи відстеження погляду система використовує процес калібрування, який допомагає WebGazer.js адаптуватися до індивідуальних особливостей користувача. Калібрування є необхідним кроком перед початком взаємодії, оскільки без нього система не може точно визначати точку фіксації погляду.

Хоча WebGazer.js має вбудовану можливість автоматичного калібрування, що здійснюється під час природної взаємодії користувача зі сторінкою (зокрема, під час кліків), на початкових етапах цієї взаємодії модель може бути недостатньо точною через нерівномірний розподіл кліків на екрані. З метою підвищення

точності прогнозу погляду від самого початку сесії, у межах нашої системи передбачено попереднє калібрування з інструкціями для користувача, що дозволяє зібрати навчальні дані з різних частин екрану.

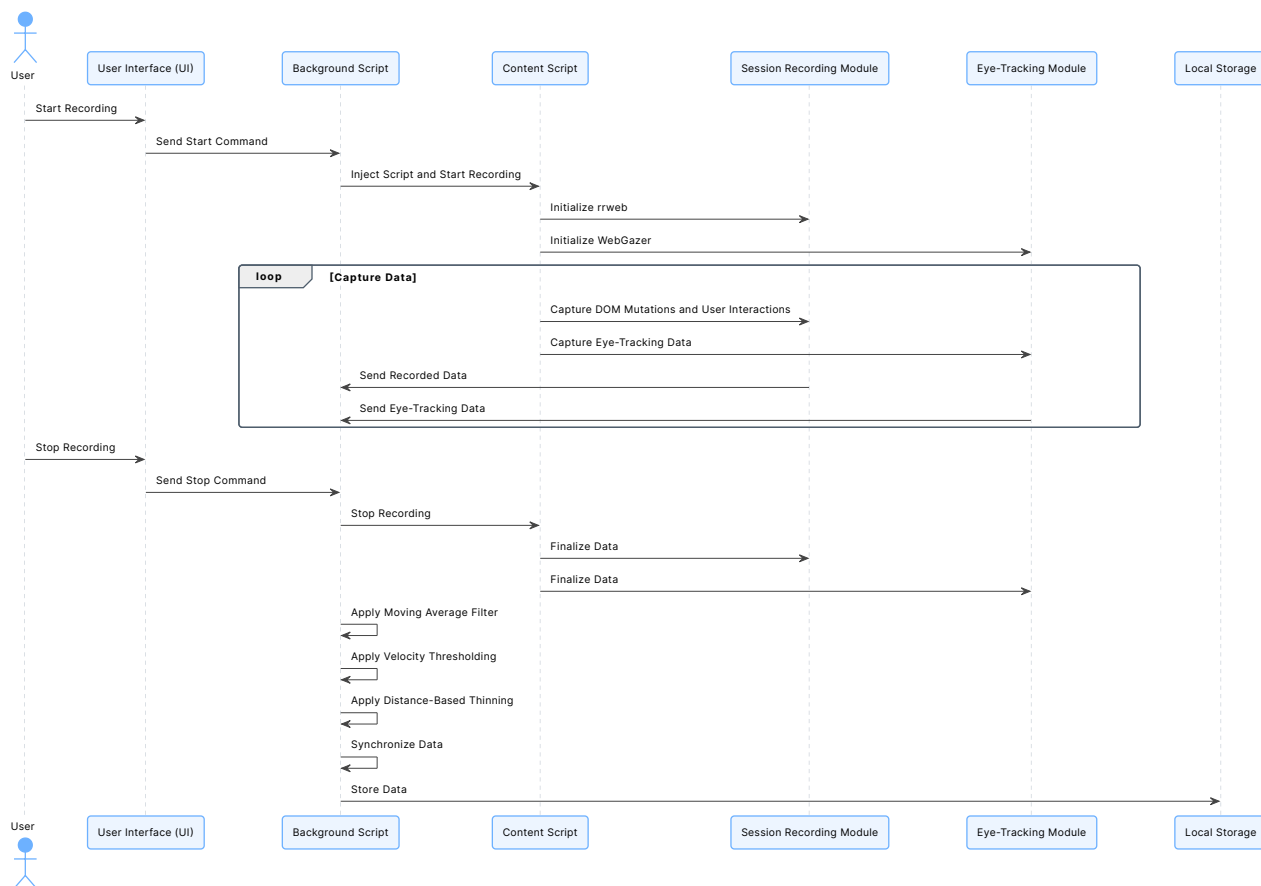


Рис. 2.2. UML-діаграма послідовності процесу запису веб-сесії із інтегрованими даними відстеження погляду

Під час процесу калібрування на екрані користувача відображається дев'ять точок, рівномірно розподілених по екрану. Користувач має сфокусувати увагу на кожній точці та натиснути на неї п'ять разів, щоб додати дані до моделі розпізнавання точки погляду. Після завершення п'яти натискань точка змінює колір на зелений, сигналізуючи про завершення калібрування для цієї області. Процес повторюється для кожної точки, поки всі точки не стануть «зеленими», що

свідчить про завершення процесу калібрування. Після цього система починає використовувати зібрані дані для визначення точки погляду у реальному часі.

На Рис. 2.3 схематично зображено екран користувача, де цифрою 1 позначено приклади невідкаліброваних точок, а цифрою 2 — точки, для яких калібрування завершено.

Точки, додані до моделі під час процесу калібрування, зберігаються локально на пристрої користувача, а отже при повторній ініціалізації модуля відстеження погляду (після перезавантаження веб-сторінки або перезапуску веб-браузера), користувачу не потрібно проходити процес калібрування знову.

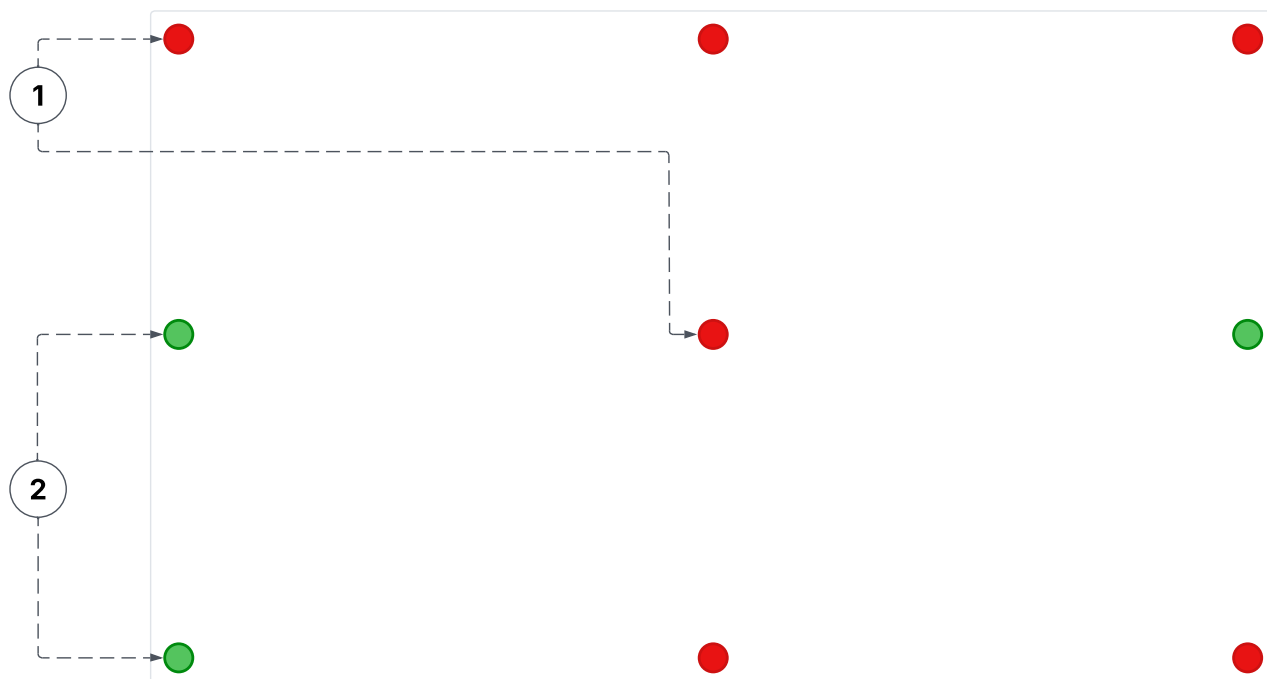


Рис. 2.3. Схематичне зображення екрану користувача з набором точок калібрування, де (1) — невідкалібровані точки, (2) — відкалібровані точки

Проте, оскільки точність відстеження може змінюватися через освітлення, рухи голови або зміну положення тіла користувача, система передбачає можливість повторного калібрування. Користувач може запустити процес повторного

калібрування у будь-який момент, натиснувши відповідну кнопку в меню розширення.

2.2.5 Обробка даних погляду

Щоб зменшити надмірність і зменшити шум у даних відстеженні погляду, під час обробки даних погляду було застосовано алгоритми фільтрації даних.

Дані погляду представлені у вигляді послідовності точок погляду:

$$\mathcal{G} = \{(t_i, x_g(t_i), y_g(t_i))\}_{i=1}^{N_g}, \quad (2.4)$$

де:

- t_i — мітка часу;
- $x_g(t_i), y_g(t_i)$ — координати погляду користувача на екрані в момент часу t_i .

В свою чергу фільтровану послідовність даних погляду можна представити як:

$$\mathcal{G}_{filtered} = F(\mathcal{G}),$$

де $F(\cdot)$ — функція фільтрації.

Фільтр ковзного середнього. Для згладжування короткострокових коливань у даних відстеження погляду використовується фільтр ковзного середнього.

$$G_{smooth}(t_i) = \frac{1}{N} \sum_{j=i-N}^i G(t_j), \quad (2.5)$$

де:

- N — розмір вікна згладжування;
- $G_{smooth}(t_i)$ — згладжена точка погляду.

Під час застосування алгоритму обчислюється середня позиція точки погляду в рухомому вікні розміром 5 кадрів. Це дозволяє отримати стабільніший та плавніший шлях погляду. Розмір вікна було підібрано експериментально, щоб забезпечити баланс між згладжуванням і чутливістю даних.

Порогова фільтрація за швидкістю. Щоб усунути незначні мікрорухи очей, які не відображають реальні зміни уваги користувача, застосовується алгоритм порогової фільтрації за швидкістю. Для цього обчислюється швидкість руху погляду за формулою:

$$v(t_i) = \frac{\|G_{smooth}(t_i) - G_{smooth}(t_{i-1})\|}{t_i - t_{i-1}} \quad (2.6)$$

Вводиться поріг мінімальної швидкості v_{min} і якщо $v(t_i) < v_{min}$, то точка $G_{smooth}(t_i)$ відкидається. Порогове значення $< v_{min}$ було підібрано емпірично, щоб виключити дрібні коливання, зберігаючи суттєві рухи очей, і становить 15 пікселів на кадр.

Прорідження за відстанню. Щоб скоротити кількість точок погляду, що зберігаються, до даних погляду застосовується алгоритм фільтрації «прорідження за відстанню». Це дозволяє зменшити обсяг даних, але при цьому зберегти ключові закономірності руху очей. В результаті до фінальної послідовності точок погляду потрапляють лише ті точки, які знаходяться на значній відстані від попередньої:

$$d(t_i, t_{i-1}) = \|G(t_i) - G(t_{i-1})\| \quad (2.7)$$

Якщо:

$$d(t_i, t_{i-1}) < d_{min},$$

то точка $G(t_i)$ відкидається, де d_{min} — мінімальна відстань між точками погляду. Граничне значення d_{min} було підібрано експериментально, щоб досягти оптимального співвідношення між зменшенням обсягу даних та збереженням важливих шаблонів руху погляду, і становить 20 пікселів.

Таким чином фінальна функція фільтрації даних погляду виглядає наступним чином:

$$F(\mathcal{G}) = \left\{ G_{smooth}(t_i) \mid v(G_{smooth}(t_i)) \geq v_{min}, d(G_{smooth}(t_i), G_{filtered}(t_{i-1})) \geq d_{min} \right\}, \quad (2.8)$$

де:

- $G_{filtered}(t_i)$ — точка погляду у фінальній відфільтрованій послідовності.

На Рис. 2.4 вказано послідовність застосування фільтрів до зібраних даних відстеження погляду.

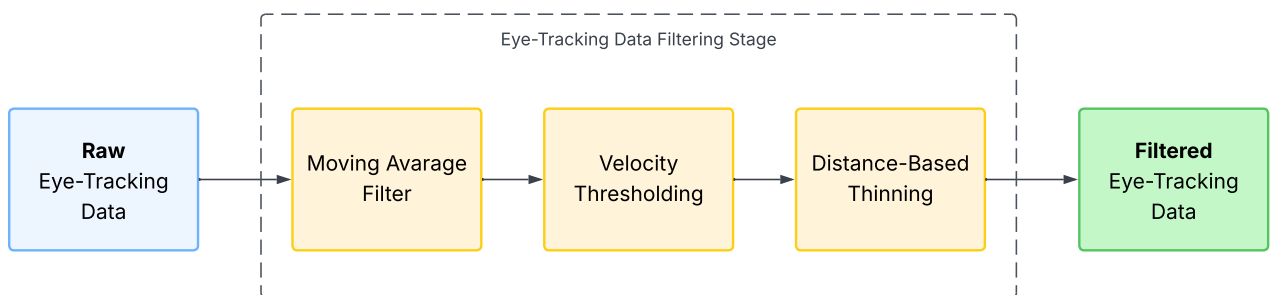


Рис. 2.4. Процес застосування алгоритмів фільтрації до зібраних даних відстеження погляду

2.2.6 Захист персональних даних і конфіденційність

Забезпечення конфіденційності та безпеки даних є одним з ключовим аспектів розробленого розширення, оскільки воно оперує інформацією про веб-сесії взаємодії користувачів та відстеження їхнього погляду. Для захисту особистих даних користувачів реалізовано комплексний підхід, що передбачає мінімізацію

збору інформації, їх анонімізацію та обробку безпосередньо на пристрої користувача.

Розширення спроектоване таким чином, щоб збирати виключно необхідні дані, уникаючи фіксації зайвої персональної інформації. Наприклад, система реєструє лише події взаємодії користувача з веб-сайтом та координати його погляду без збереження будь-яких зображень чи відео. Для додаткового захисту застосовується підхід з анонімізацією даних. Так, наприклад, rrweb автоматично маскує «чутливі» (sensitive) поля вводу, виключаючи їх із записів сесій, а всі значення, що можуть містити особисту інформацію (наприклад, імена чи email-адреси), не зберігаються. Аналогічно, WebGazer.js працює виключно з координатами погляду, не зберігаючи жодного візуального контенту, а всі обчислення відбуваються локально на пристрої користувача.

Завдяки такій архітектурі система гарантує безпеку та анонімність користувачів та мінімізує ризики витоку даних та їхньої подальшої ідентифікації.

2.3 Користувацький інтерфейс

Користувацький інтерфейс (UI) розробленого розширення створено з урахуванням інтуїтивності та зручності використання. Він складається з кількох основних компонентів, що дозволяють користувачам керувати процесом запису, переглядати записані сесії та керувати ними.

Цей розділ містить огляд компонентів користувацького інтерфейсу із відповідними знімками екрана, що ілюструють основні функції розширення.

Контроль процесу запису. Розширення надає інтерфейс для керування записом сесій, що містить кнопки запуску та зупинки, а також індикатор статусу запису.

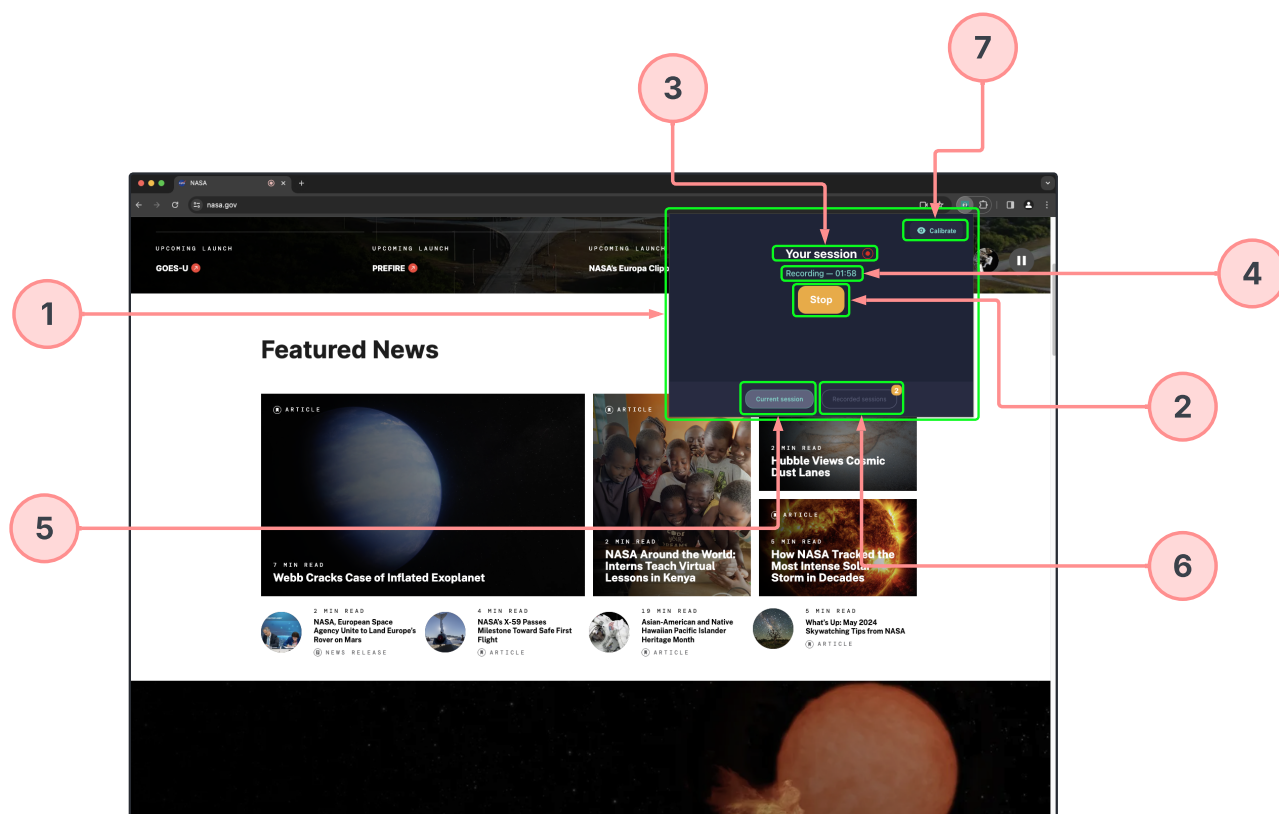


Рис. 2.5. Вікно розробленого розширення — користувацький інтерфейс для запису веб-сесії

На Рис. 2.5 наведено знімок вікна веб-браузера з відкритим спливаючим вікном розширення, де позначені наступні елементи користувацького інтерфейсу:

1. *Спливаюче вікно розширення* — з'являється після натискання на піктограму розширення.
2. *Кнопка «записати / зупинити» («start / stop»)* — запускає або зупиняє процес запису веб-сесії.
3. *Індикатор статусу запису* — пульсуюче червоне коло, що сигналізує про активний запис.
4. *Тривалість запису* — відображає загальний час запису поточної сесії.
5. *Вкладка «Записати сесію»* — перенаправляє користувача на панель керування записом сесії (поточний екран).

6. *Вкладка «Записані сесії»* — перенаправляє користувача на екран, де відображається список локально збережених сесій. Вкладка має круглу помаранчеву позначку, що відображає кількість збережених сесій.
7. *Кнопка «Калібрувати»* — запускає процес калібрування модуля відстеження погляду.

Керування записаними сесіями. Розширення надає користувачам інтерфейс для керування записаними сесіями, який дозволяє експортувати, видаляти та імпортувати сесії. На Рис. 2.6 зображено знімок вікна браузера з відкритим спливаючим вікном розширення, де позначені такі елементи:

1. *Елемент записаної сесії* — містить інформацію про записану сесію, включаючи назву веб-сторінки, URL, дату та час запису та його тривалість.
2. *Кнопка запуску відтворення сесії* — при натисканні відкриває нову вкладку браузера, в якій відображається веб-сторінка із вбудованим інструментом для відтворення обраної сесії.
3. *Кнопки експорту та видалення записів* — кнопка «Експортувати» дозволяє зберегти дані сесії у файл у форматі JSON, а кнопка «Видалити» — видаляє дані сесії з локального сховища даних.
4. *Кнопка «Імпортувати сесію»* — дозволяє завантажити записані сесії з JSON-файлу. Даний функціонал є корисним для відтворення сесій, записаних на інших пристроях.

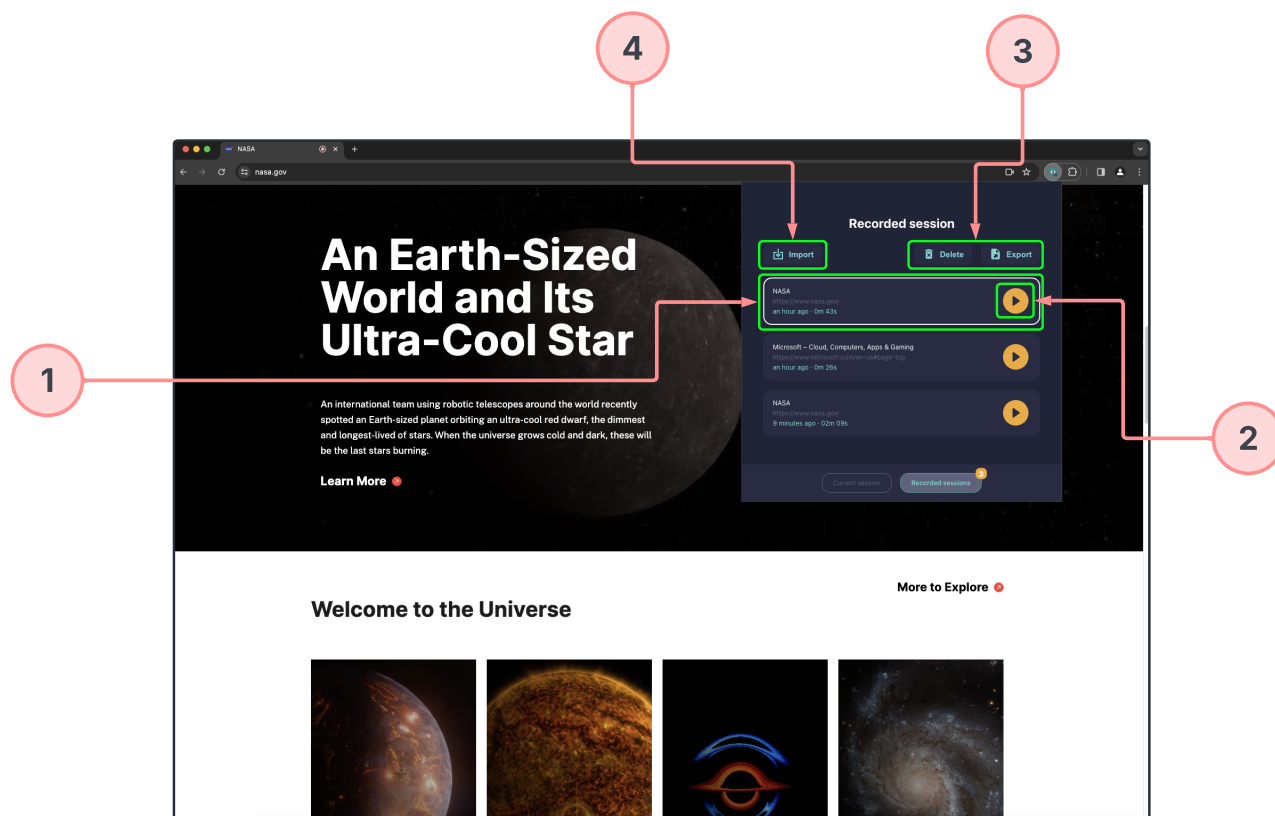


Рис. 2.6. Вікно розробленого розширення — користувацький інтерфейс для керування записаними веб-сесіями

Засіб відтворення сесій. Розширення надає користувацький інтерфейс для відтворення записаних сесій (session replay tool), що відображається у окремій вкладці веб-браузера та дозволяє аналізувати користувацьку взаємодію із інтегрованими даними відстеження погляду.

На Рис. 2.7 зображено знімок вкладки веб-браузера із засобом відтворення сесій, де відмічено наступні елементи:

1. *Індикатор положення курсору миші* — відображає рух курсору миші користувача під час запису сесії.
2. *Індикатор погляду* — відображає точку погляду користувача під час запису сесії.

3. *Повзунок таймлайну* — дозволяє переміщатися між різними моментами запису.
4. *Контрольні елементи відтворення* — панель, що містить такі елементи, як кнопка «відтворити / пауза» для керування відтворенням запису, налаштування швидкості відтворення та пропуск періодів неактивності, що автоматично переміщує користувача до моменту наступної взаємодії у записаній сесії.

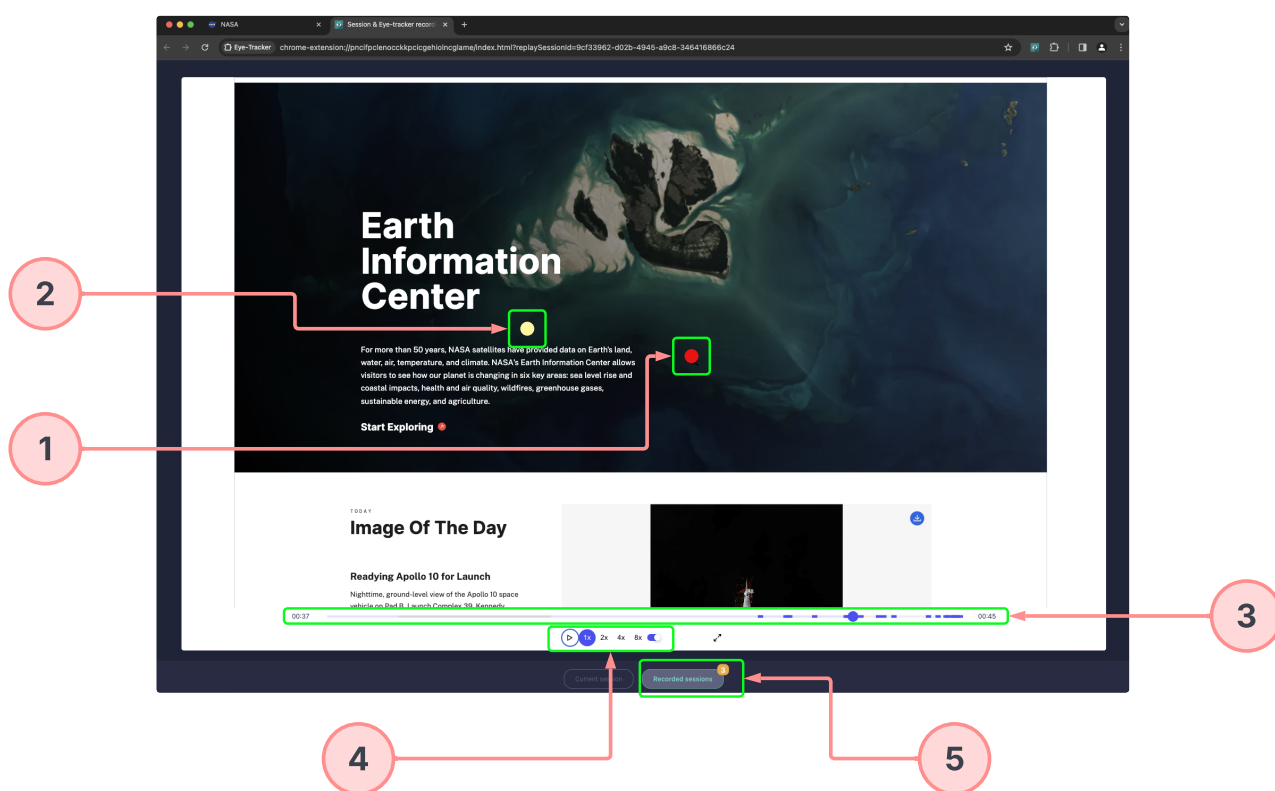


Рис. 2.7. Вікно розробленого розширення — користувацький інтерфейс інструменту відтворення записаних веб-сесій

2.4 Оцінювання системи на основі користувацького дослідження

Для оцінки зручності використання та ефективності розробленого розширення було проведено структуроване користувацьке дослідження. Учасники

взаємодіяли з розширенням у контрольованих умовах, після чого заповнювали детальну анкету для збору кількісних та якісних даних.

Методологія. Дослідження проводилося із залученням користувачів, які могли представляти потенційних кінцевих користувачів розширення, включаючи людей із різним рівнем технічної підготовки.

У дослідженні взяли участь 25 респондентів віком від 18 до 35 років. До вибірки були включені як користувачі з базовим рівнем володіння комп'ютером, так і досвідчені учасники, що дозволило отримати відгуки з урахуванням різного рівня технічної підготовки.

Кожному учаснику було запропоновано виконати набір типових завдань, використовуючи розроблене розширення. Завдання включали:

- *Запис сесії* — користувач проходив процес калібрування модуля відстеження погляду, розпочинав запис, взаємодіяв із веб-сторінкою, а потім зупиняв запис.
- *Відтворення сесії* — користувач переглядав записану сесію, аналізуючи зафіксовану поведінку та дані відстеження погляду.
- *Керування записаними сесіями* — оцінювалась функціональність експорту, імпорту та видалення записаних сесій.

Збір даних. Для кількісної оцінки користувацького досвіду в анкеті було використано шкалу від 1 до 5 із можливістю введення десяткових значень, що дозволило респондентам надати точніший зворотний зв'язок; додатково були включені відкриті запитання для якісного аналізу. Анкета пропонувала користувачам оцінити наступні аспекти зручності використання (usability):

- *інтуїтивність інтерфейсу користувача* — легкість орієнтації та розуміння принципів взаємодії;
- *доступність функцій* — зручність доступу до основних можливостей розширення;

- *продуктивність* — швидкість роботи системи під час виконання типових завдань;
- *навчальна крива* — час, необхідний для опанування основних функцій;
- *загальна задоволеність* — суб'єктивне сприйняття зручності, комфорту та ефективності використання.

Окрім загальних аспектів зручності використання, користувачі також оцінювали окремі функції та характеристики розширення:

- *запис сесій* — ефективність реалізації функціональності запису користувацької взаємодії;
- *відтворення сесій* — якість візуалізації та зручність навігації під час перегляду записів;
- *точність даних відстеження погляду* — відповідність зафіксованих координат фактичному положенню погляду;
- *керування сесіями* — зручність перегляду, видалення, експорту та імпорту записаних даних.

Анкета також містила пункт щодо готовності респондента рекомендувати розширення іншим користувачам, що вважається важливим індикатором загального рівня задоволеності. У розділі відкритих відповідей учасники могли вказати на проблеми, з якими вони стикнулися під час взаємодії із системою, а також запропонувати можливі покращення.

2.5 Аналіз результатів користувацького дослідження

Оцінка зручності використання. Середні бали користувачів щодо характеристик зручності використання представлено в Табл. 2.1. Загальний бал зручності використання становить 4,5 з 5, що розраховано як середнє арифметичне балів окремих характеристик.

Інтуїтивність інтерфейсу (4,4/5) та доступність функцій (4,4/5) отримали високі бали, що свідчить про зрозумілість та легкість навігації. Продуктивність (4,5/5) та навчальна крива (4,5/5) вказують на ефективність роботи розширення та швидку адаптацію користувачів. Загальна задоволеність (4,5/5) підтверджує позитивний досвід взаємодії з розширенням.

Характеристика зручності використання	Середній бал (1–5)
Інтуїтивність інтерфейсу користувача	4,4
Доступність функцій	4,4
Продуктивність	4,5
Навчальна крива	4,5
Загальна задоволеність	4,5

Табл. 2.1. Середні бали користувачів для характеристик зручності використання

Оцінка функцій та характеристик розширення. Результати оцінювання ключових функцій та характеристик розширення наведено в Табл. 2.2. Користувачі надали оцінки як основним функціональним можливостям (запис, відтворення, керування сесіями), так і точності модуля відстеження погляду. Функціональність запису та відтворення сесій отримала високі середні бали (4,1 та 4,3 відповідно), що вказує на їхню ефективність. Керування записаними сесіями (4,2/5) також було позитивно сприйнято, зокрема щодо можливостей імпорту, експорту та видалення записів.

Водночас точність відстеження погляду отримала найнижчий бал (3,8/5), що, згідно з відгуками, пов'язано з періодичними неточностями у фіксації координат погляду. Детальний розподіл балів подано на Рис. 2.8.

Функція / характеристика	Середній бал (1–5)
Запис сесій	4,1
Відтворення сесій	4,3
Точність даних відстеження погляду	3,8
Керування записаними сесіями	4,2

Табл. 2.2. Середні бали користувачів для ключових функцій і характеристик системи

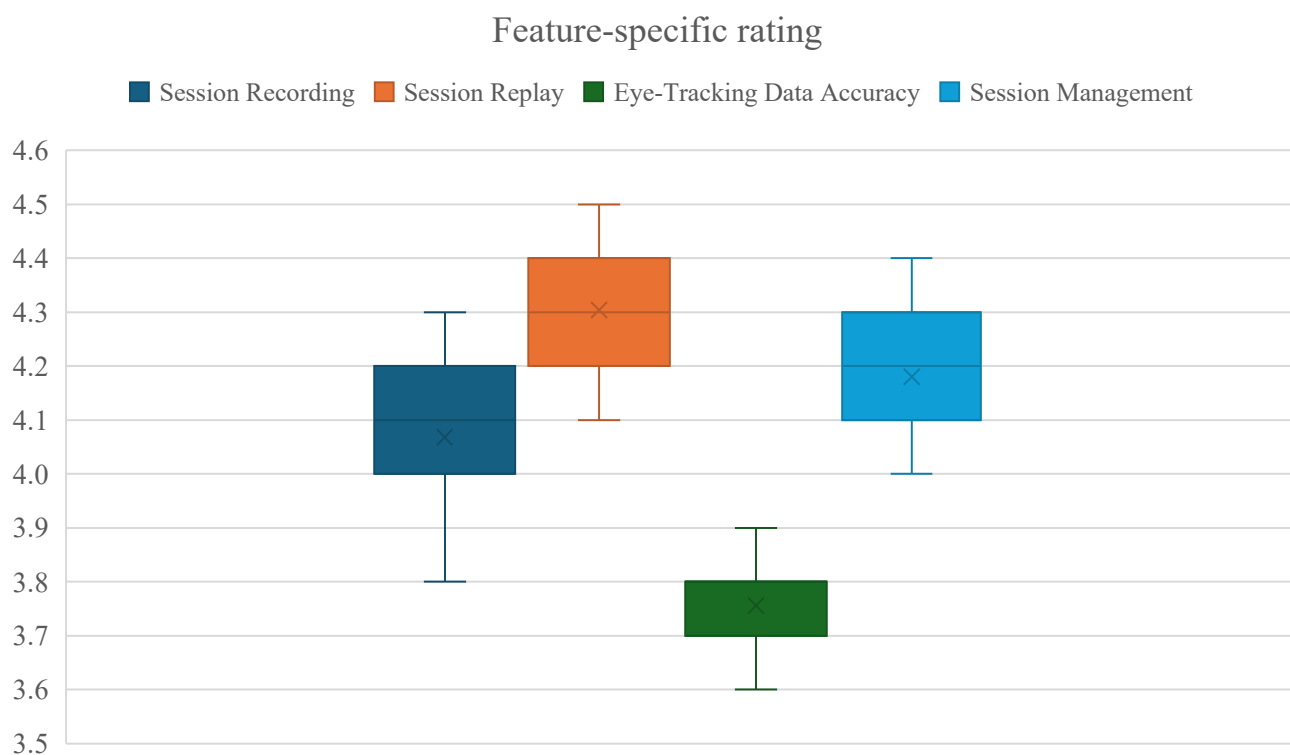


Рис. 2.8. Розподіл балів користувачів для ключових функцій і характеристик системи

Якісний аналіз та зворотний зв'язок. Переважна більшість користувачів відзначила корисність функції відстеження погляду, інтегрованої із записом веб-сесій, однак водночас вказала на необхідність підвищення її точності. Під час перегляду записаних веб-сесій користувачі помічали періодичні неточності в даних відстеження погляду.

У відповідях респонденти часто пропонували шляхи вдосконалення процесу калібрування, зокрема: «додати більше точок калібрування» або замінити на підхід, за якого «користувач відстежує поглядом рухомий об'єкт, що переміщується по екрану комп'ютера». Також у відгуках містилися пропозиції щодо покращення візуальної складової інтерфейсу розширення та інструменту перегляду записаних веб-сесій.

Загальний рівень готовності рекомендувати розширення є високим, що свідчить про позитивний досвід його використання (Рис. 2.9).

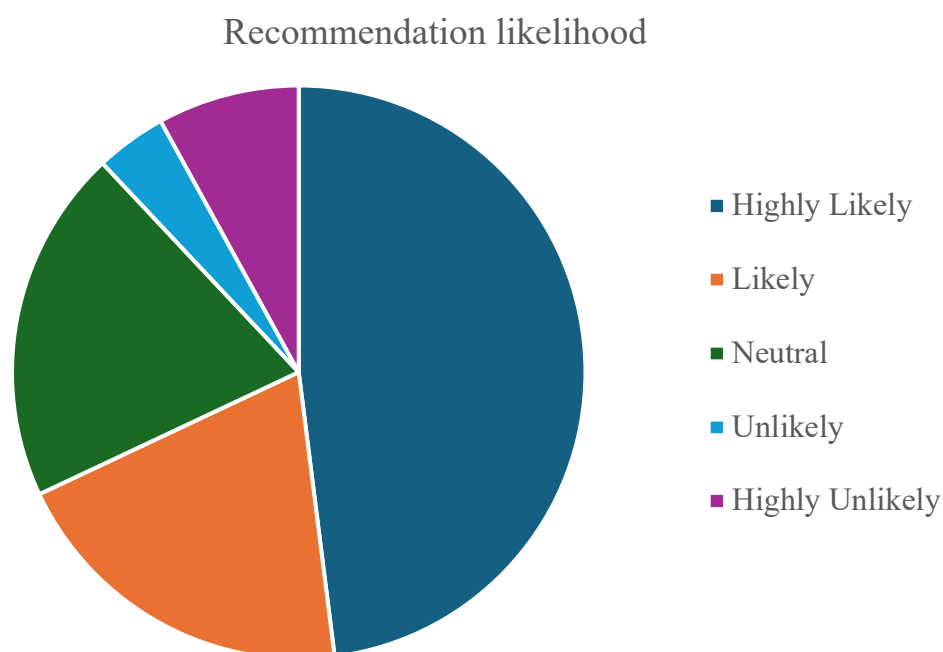


Рис. 2.9. Рівень готовності рекомендувати розширення іншим користувачам

У межах цього розділу не проводилась кількісна оцінка точності алгоритму відстеження погляду, реалізованого за допомогою WebGazer.js, оскільки сам алгоритм не є предметом дослідження. Основна мета полягала у створенні інструменту для поєднання запису веб-сесій із фіксацією зорової уваги користувача, а також у перевірці доцільності його застосування в реальних умовах. Оцінка ефективності системи базувалася на якісному аналізі її функціональності, зручності використання та сприйнятті користувачами під час експерименту.

2.6 Виклики та обмеження

Проведене користувацьке дослідження дало змогу оцінити зручність використання та ефективність розробленого розширення для запису й відтворення веб-сесій із інтегрованими даними відстеження погляду. Загалом учасники відзначили, що розширення є ефективним, а більшість його функцій отримали високі бали. Водночас результати дослідження виявили низку потенційних проблем, які можуть бути усунені або вдосконалені в майбутньому, зокрема — періодичні неточності у даних відстеження погляду.

Розширення отримало високі середні бали за такими характеристиками, як інтуїтивність інтерфейсу користувача, доступність функцій, продуктивність, навчальна крива та загальна задоволеність. Середній бал зручності використання (4,5/5) свідчить про те, що більшість учасників визнали розширення простим у користуванні та ефективним для виконання поставлених завдань.

Також було високо оцінено функціональність запису сесій (4,1), відтворення сесій (4,3) та керування сесіями (4,2), що підтверджує загалом позитивний користувацький досвід.

Попри це, модуль відстеження погляду отримав нижчий середній бал (3,8 з 5). Деякі респонденти повідомили про неточності у визначенні координат погляду, які іноді не відповідали фактичному напрямку зору. Ця проблема є критично важливою, оскільки точність відстеження погляду безпосередньо впливає на можливість детального аналізу поведінки користувачів.

2.7 Перспективи подальших досліджень

На основі аналізу користувацького зворотного зв'язку, поданого в розділі результатів, було визначено ключові напрями вдосконалення. Їхнє впровадження дозволить підвищити зручність використання та ефективність розширення.

Відповідно до результатів проведеного дослідження, одним із пріоритетних напрямів подальшої роботи є покращення точності даних відстеження погляду. Потенційне підвищення точності може бути досягнуте шляхом удосконалення

процесу калібрування. У подальших дослідженнях пропонується розглянути декілька підходів до цього процесу. Один з них передбачає аналіз впливу кількості точок калібрування на точність, використовуючи чинний метод калібрування, за якого користувач фокусується на точці та виконує кілька кліків миші (для кожної точки).

Інший підхід полягає у заміні процесу калібрування на такий, що передбачає слідування поглядом за рухомим об'єктом, який переміщується екраном та зупиняється в заздалегідь визначених точках. Припускається, що такий підхід може бути зручнішим для користувачів, оскільки не потребує багаторазового натискання миші, особливо при великій кількості точок. Крім того, розглядається можливість заміни WebGazer.js на інше програмне рішення з вищою точністю відстеження.

Ще одним перспективним напрямом є інтеграція додаткових засобів візуалізації, зокрема теплових карт (heatmaps) і графіків руху погляду (gaze plots), які дозволили б краще інтерпретувати поведінку користувачів. Окремим напрямом може стати використання моделей машинного навчання для автоматичного виявлення шаблонів взаємодії та зон підвищеної уваги.

Очікується, що впровадження цих покращень дозволить здійснювати глибший аналіз користувацької взаємодії та отримувати більш цінні висновки в межах досліджень зручності використання.

Перспективним напрямом також є застосування формалізованих моделей оцінювання, заснованих на даних відстеження погляду. Зокрема, у роботі [56] запропоновано підхід до кількісної оцінки зручності інтерфейсу на основі фіксацій погляду в межах виділених областей інтересу (Areas of Interest, AoI) — визначених сегментів інтерфейсу, які є важливими з точки зору виконання користувачем певних завдань. Такий підхід дозволяє аналізувати як увагу користувача, так і процес прийняття рішень під час взаємодії з інтерфейсом.

На відміну від подібних досліджень, де області інтересу визначаються вручну, запропонований у цій роботі підхід передбачає збереження повної

структури веб-сторінки під час сеансу взаємодії, включно зі змінами DOM, діями користувача та даними відстеження погляду. Завдяки наявності HTML-структури сторінки з'являється можливість ідентифікувати семантично важливі елементи інтерфейсу (такі як кнопки, поля введення, посилання) та надалі застосовувати формалізовані моделі аналізу, що базуються на фіксаціях погляду. Такий підхід відкриває перспективи для масштабованого автоматизованого оцінювання зручності використання, без необхідності ручного маркування інтерфейсних елементів.

2.8 Висновки

У цьому розділі представлено розробку та експериментальне оцінювання прототипу інструменту для інтеграції запису веб-сесій та відстеження погляду з метою поглибленого аналізу поведінки користувачів у веб-середовищі. Реалізовано архітектуру та створено розширення для браузера Chrome, яке забезпечує синхронізовану реєстрацію змін у DOM, дій користувача та координат його візуальної уваги без потреби у спеціалізованому обладнанні або внесенні змін до коду веб-застосунку, на якому здійснюється запис сесії користувача.

Проведене експериментальне оцінювання підтвердило ефективність основних функцій інструменту: середній бал зручності використання склав 4,5 із 5 для функцій запису, відтворення та керування сесіями. Водночас виявлено обмеження, пов'язані з точністю відстеження погляду, модуль якого отримав нижчий середній бал (3,8 із 5), що обумовлює потребу у подальшому вдосконаленні технологічних компонентів системи.

Основними напрямками подальшого розвитку є підвищення точності відстеження погляду шляхом оптимізації механізмів калібрування, розширення можливостей візуалізації даних (теплові карти, графіки руху погляду) для глибшого аналізу поведінки користувачів, а також впровадження алгоритмів машинного навчання для автоматизованого виявлення закономірностей у поведінкових даних та визначення типових шаблонів взаємодії користувачів.

РОЗДІЛ 3. БЕЗКООНТАКТНА ВЗАЄМОДІЯ З ВИКОРИСТАННЯМ ТЕХНОЛОГІЇ ВІДСТЕЖЕННЯ ПОГЛЯДУ ТА ЖЕСТІВ У ВЕБ- СЕРЕДОВИЩІ

Цей розділ дисертаційного дослідження присвячено розробці підходу до безконтактної взаємодії на основі поєднання погляду та жестів рук у веб-середовищі. Сучасні дослідження у сфері взаємодії людини з комп'ютером (НСІ) дедалі частіше звертаються до природних методів введення, які виходять за межі традиційних пристроїв, таких як миша та клавіатура. Зокрема, відстеження погляду, жести рук та вирази обличчя розглядаються як засоби, здатні зробити цифрову взаємодію більш інтуїтивною та природною.

Вибір об'єктів на екрані є однією з основних операцій в інтерфейсах користувача. Технологія відстеження погляду забезпечує високу швидкість вибору, однак має обмежену функціональність для подальших дій. У цьому контексті жести рук виступають природним доповненням, розширюючи можливості управління за рахунок підтримки операцій перетягування, масштабування та обертання.

Попередні дослідження демонструють ефективність комбінованої взаємодії на основі погляду та жестів, передусім у віртуальних середовищах (VR). Проте широке впровадження подібних рішень поза межами VR, зокрема у веб-просторі, стримується апаратними обмеженнями.

Метою цього розділу є реалізація та експериментальна оцінка підходу до взаємодії на основі погляду та жестів, який працює у веб-середовищі з використанням лише стандартної веб-камери. Такий підхід усуває потребу у спеціалізованому обладнанні та робить технологію доступнішою для кінцевих користувачів.

У межах дослідження було створено веб-орієнтовану систему, яка поєднує вибір об'єктів за допомогою погляду та їх маніпулювання жестами рук. Ефективність запропонованого підходу порівнювалася з традиційним введенням мишею на основі кількісних показників, зокрема часу виконання типових завдань.

Основні результати, що лягли в основу цього модуля, були представлені у вигляді наукових статей і матеріалів конференцій [28], [57], [58].

3.1 Архітектура системи та реалізація

Розроблена система поєднує відстеження погляду та розпізнавання жестів рук для забезпечення природної та інтуїтивної взаємодії з об'єктами у веб-середовищі. Основною метою проєкту є створення доступного та кросплатформного рішення, яке не потребує спеціалізованого обладнання — взаємодія здійснюється за допомогою стандартної веб-камери, яка одночасно фіксує координати погляду користувача та положення його рук.

Однією з ключових функціональних особливостей запропонованого рішення є тісна інтеграція двох каналів введення — погляду та жестів, — що забезпечує безконтактне вибирання та маніпулювання об'єктами у режимі реального часу. Такий підхід не лише підвищує зручність взаємодії, а й розширює можливості застосування системи в різних контекстах: від веб-дизайну та інтерактивних освітніх платформ до віртуальних середовищ і систем дистанційної співпраці.

3.1.1 Загальний технічний огляд

Архітектура системи ґрунтується на використанні сучасних веб-технологій, що дає змогу реалізувати всі функції без встановлення додаткового програмного забезпечення. Всі компоненти функціонують безпосередньо у браузері, забезпечуючи високу сумісність і доступність для кінцевого користувача.

Для реалізації системи були використані такі технології та програмні засоби:

- *WebGazer.js* — бібліотека на основі машинного навчання, що дозволяє визначати точку фіксації погляду користувача у реальному часі за допомогою веб-камери;

- *MediaPipe Hand Landmarker* [59] — компонент з набору інструментів Google MediaPipe, призначений для виявлення та відстеження ключових точок рук та розпізнавання жестів;
- *JavaScript та Web API* — для інтеграції основних модулів та обробки логіки взаємодії;
- *HTML і CSS* — для створення користувацького інтерфейсу;
- *Canvas API* — для графічного відтворення об'єктів та реалізації візуальної частини взаємодії з ними.

3.1.2 Основні компоненти

Запропонована система реалізує поєднання відстеження погляду та розпізнавання жестів рук для забезпечення безконтактної взаємодії з об'єктами у веб-середовищі. Її архітектура включає кілька ключових компонентів, які функціонують у тісній взаємодії, забезпечуючи вибір об'єкта за допомогою погляду та подальше маніпулювання ним за допомогою жестів (Рис. 3.1).

Модуль відстеження погляду. У запропонованій системі відстеження погляду реалізовано за допомогою бібліотеки WebGazer.js, яка використовує веб-камеру для визначення точки фіксації погляду користувача на екрані. Модуль виконує низку ключових функцій: обчислює координати погляду в реальному часі, передає ці дані іншим компонентам системи, а також забезпечує виконання процесу калібрування, що сприяє підвищенню точності визначення точки фіксації.

Зважаючи на те, що WebGazer.js є рішенням, орієнтованим на використання стандартної веб-камери, точність визначення координат може варіюватися залежно від умов освітлення, положення голови користувача та якості виконаного калібрування.

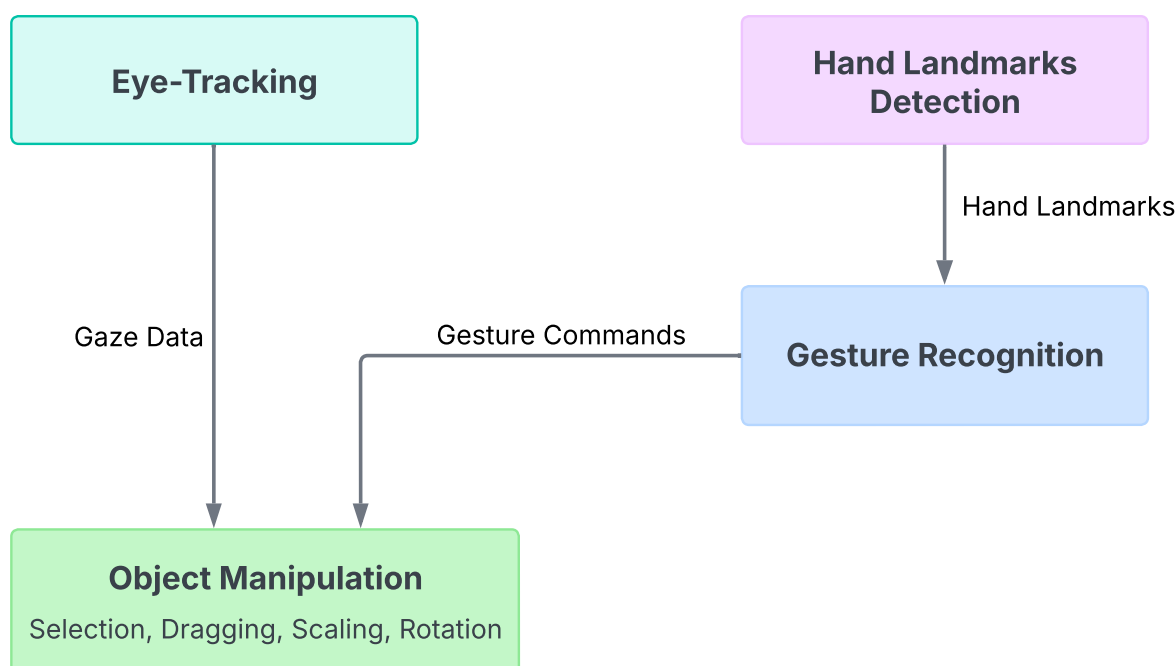


Рис. 3.1. Схема взаємодії компонентів системи

Модуль розпізнавання жестів. Розпізнавання жестів у системі реалізовано із використанням програмного рішення MediaPipe Hand Landmarker, який здійснює аналіз положення руки користувача та визначає просторові координати ключових точок кисті. Основними функціями цього модуля є визначення просторового положення рук, розпізнавання жестів у реальному часі, а також передача відповідних даних до модуля взаємодії.

Завдяки цьому стає можливим інтерпретування жестів, зокрема жесту захоплення (pinch) та рухів рук, що є критично важливими для реалізації маніпулювання об'єктами у віртуальному середовищі.

Для розпізнавання жесту «pinch» у цій роботі було використано простий евристичний підхід, заснований на порівнянні відстані між вказівним і великим пальцем з заданим пороговим значенням. Такий підхід є обчислювально ефективним і достатнім для сценарію з обмеженим набором жестів, де основною метою є оцінка концепції мультимодальної взаємодії.

У цьому дослідженні не проводилася окрема оцінка точності алгоритмів розпізнавання жестів, оскільки фокус був зосереджений на перевірці функціональності інтерактивної системи, що поєднує вибір об'єктів поглядом та їх маніпулювання жестами рук. Жест «pinch» визначався на основі евристики — відстані між вказівним і великим пальцем — що забезпечувало необхідну стабільність у запропонованому сценарії. Експериментальна частина мала на меті оцінити інтуїтивність, зручність та ефективність такого типу безконтактної взаємодії.

Водночас, у низці досліджень демонструється ефективність застосування методів глибинного навчання для класифікації ширшого спектра жестів. Наприклад, у роботі [60] було реалізовано систему розпізнавання жестів української дактильної абетки з використанням MediaPipe та рекурентної LSTM-мережі, досягнуто точності понад 98 % на тестовому наборі. Такий підхід може бути адаптований для задач, де необхідна підтримка більшої кількості жестів або точніше розрізнення схожих жестів.

Модуль взаємодії. Модуль взаємодії відповідає за інтеграцію даних відстеження погляду та розпізнавання жестів, що дає змогу вибирати об'єкти і проводити з ними подальше маніпулювання. Зокрема, реалізовано механізм активації об'єкта на основі фіксації погляду: якщо користувач протягом певного часу зосереджує погляд на об'єкті, система автоматично вважає його вибраним. Подальші дії здійснюються за допомогою відповідних жестів, таких як захоплення, перетягування, масштабування або обертання.

Такий підхід забезпечує інтуїтивну модель взаємодії, в якій користувач ініціює вибір об'єкта за допомогою погляду, а керування ним відбувається із використанням жестів рук.

3.1.3 Процес калібрування

Перед початком взаємодії користувач проходить калібрування системи, що є критично важливим етапом її налаштування. Цей процес дозволяє зменшити

похибки у визначенні координат погляду та підвищити загальну точність взаємодії з об'єктами. Калібрування забезпечує адаптацію системи до індивідуальних особливостей користувача, що, у свою чергу, підвищує надійність і зручність використання розробленого методу.

У межах цього дослідження застосовується процедура калібрування, детально описана в секції 2.2.4 цієї дисертаційної роботи.

3.1.4 Процес взаємодії

Процес взаємодії користувача із запропонованою системою включає вибір об'єкта за допомогою погляду та подальше маніпулювання ним за допомогою жестів. Така комбінація забезпечує інтуїтивне управління цифровими об'єктами у веб-середовищі без потреби у використанні традиційних засобів введення, таких як миша або клавіатура.

Вибір об'єкту здійснюється шляхом фіксації погляду на відповідному елементі інтерфейсу. Система, використовуючи бібліотеку WebGazer.js, визначає координати точки фіксації та, якщо погляд утримується на об'єкті протягом заданого порогового інтервалу, інтерпретує це як намір вибору. Після цього відповідний об'єкт візуально підсвічується, сигналізуючи про його активацію.

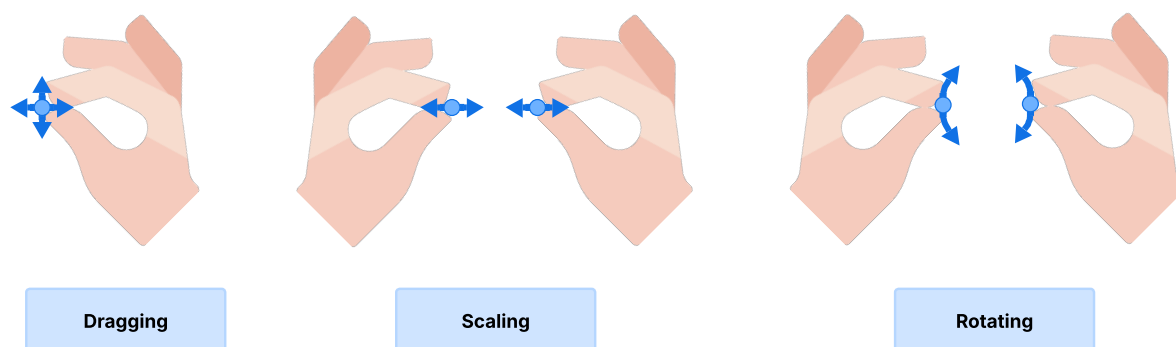


Рис. 3.2. Ілюстрація жестів для перетягування (*dragging*), масштабування (*scaling*) та обертання (*rotating*) об'єктів

Після вибору об'єкта система переходить у режим керування жестами, що дає змогу користувачеві виконувати з ним такі дії, як перетягування (dragging), масштабування (scaling) та обертання (rotating). На Рис. 3.2 проілюстровано відповідні жести для кожної з цих дій.

Перетягування. Якщо користувач виконує жест «захоплення» (pinch), система розпізнає це як команду захоплення об'єкта, після чого він слідує за рухом руки. Нова позиція об'єкта (x', y') визначається за формулою:

$$(x', y') = (x_0 + dx, y_0 + dy), \quad (3.1)$$

де:

$$dx = x_p - x_p^0, \quad dy = y_p - y_p^0$$

- (x_0, y_0) — початкова позиція об'єкту до початку перетягування;
- (x_p, y_p) — поточна позиція точки захоплення (між кінчиками вказівного та великого пальців);
- (x_p^0, y_p^0) — початкова позиція точки захоплення (на початку виконання жесту).

Масштабування. Якщо користувач виконує жест «захоплення» двома руками та змінює відстань між ними, система інтерпретує це як команду на зміну розміру об'єкта. Новий розмір об'єкту S' обчислюється за формулою:

$$S' = S_0 \times s, \quad (3.2)$$

де:

- S_0 — початковий розмір об'єкта;
- s — коефіцієнт масштабу, що розраховується за формулою:

$$s = \frac{d_p}{d_p^0}$$

де:

- d_p — поточна відстань між точками захоплення на обох руках;
- d_p^0 — початкова відстань між точками захоплення (на початку виконання жесту).

Обертання. Система інтерпретує обертальний рух обох рук із жестом «захоплення» як команду на обертання об'єкта. Під час цього жесту користувач переміщує руки по дузі, змінюючи орієнтацію вектора між точками захоплення на правій та лівій руках. Це дозволяє системі обчислити кут повороту θ за допомогою скалярного добутку векторів v_1 та v_2 :

$$\theta = \arccos\left(\frac{v_1 \cdot v_2}{\|v_1\| \|v_2\|}\right), \quad (3.3)$$

де:

$$v_1 = (x_p^r - x_p^l, y_p^r - y_p^l), \quad v_2 = (x_p^{r'} - x_p^{l'}, y_p^{r'} - y_p^{l'}),$$

- v_1 — вектор між точками захоплення правої та лівої руки на початку взаємодії;
- v_2 — відповідний вектор у поточний момент часу.

Напрямок обертання визначається за допомогою псевдоскалярного добутку:

$$\text{cross} = (x_p^r - x_p^l)(y_p^{r'} - y_p^{l'}) - (y_p^r - y_p^l)(x_p^{r'} - x_p^{l'}) \quad (3.4)$$

Таким чином, якщо значення cross:

- позитивне — обертання об'єкта проти годинникової стрілки (counterclockwise);
- негативне — за годинниковою стрілкою (clockwise).

Тоді фінальна формула для кута обертання має вигляд:

$$\theta = \begin{cases} \arccos\left(\frac{v_1 \cdot v_2}{\|v_1\| \|v_2\|}\right), & \text{cross} \geq 0 \\ -\arccos\left(\frac{v_1 \cdot v_2}{\|v_1\| \|v_2\|}\right), & \text{cross} < 0 \end{cases} \quad (3.5)$$

На Рис. 3.3 наведено блок-схему логіки обробки взаємодії користувача із розробленою системою.

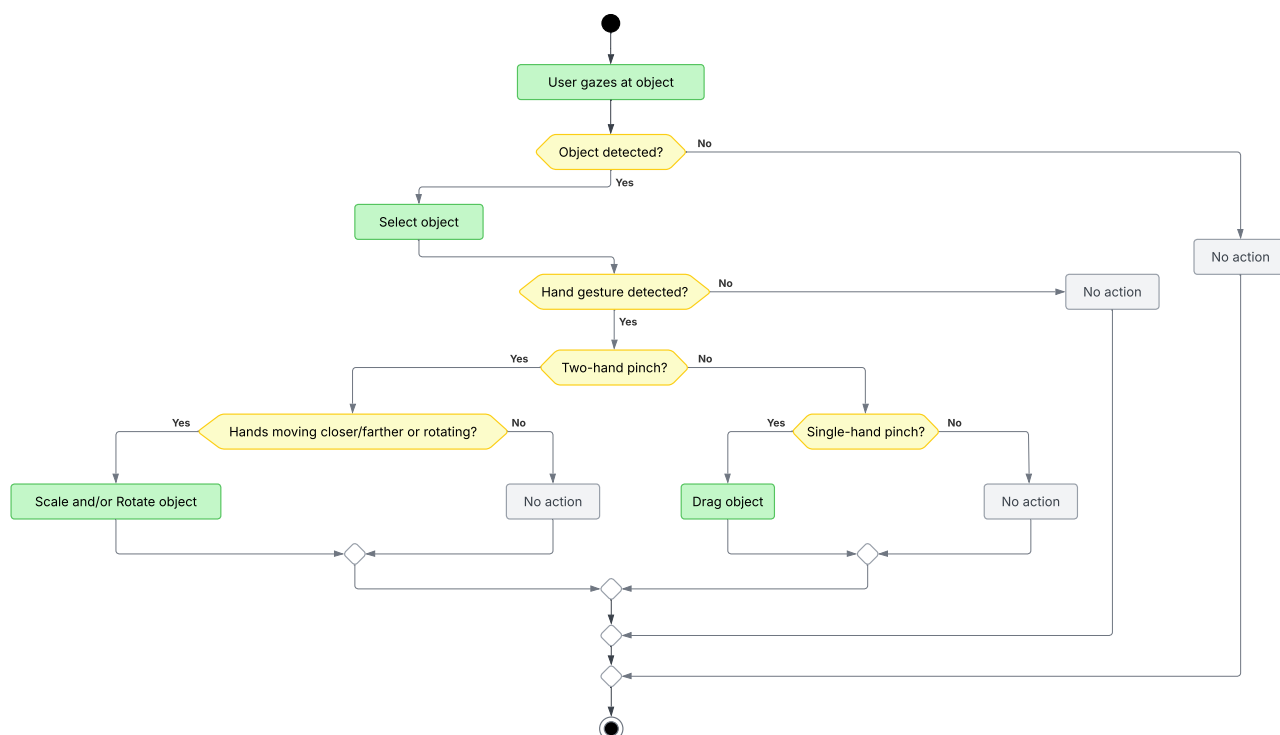


Рис. 3.3. Процес взаємодії користувача із системою на основі погляду та жестів

3.2 Інтерфейс користувача

Інтерфейс користувача (UI) розробленої системи забезпечує інтуїтивно зрозумілий механізм маніпулювання об'єктами за допомогою погляду та жестів рук. Основні елементи інтерфейсу включають відеопотік із веб-камери, інструмент калібрування трекера погляду, вибір рівня завдань та область маніпулювання об'єктами.

Маніпулювання об'єктами. Система дозволяє користувачам вибирати об'єкти за допомогою погляду та застосовувати до них перетворення за допомогою жестів рук. На Рис. 3.4 зображено основний інтерфейс під час маніпулювання обраним об'єктом.

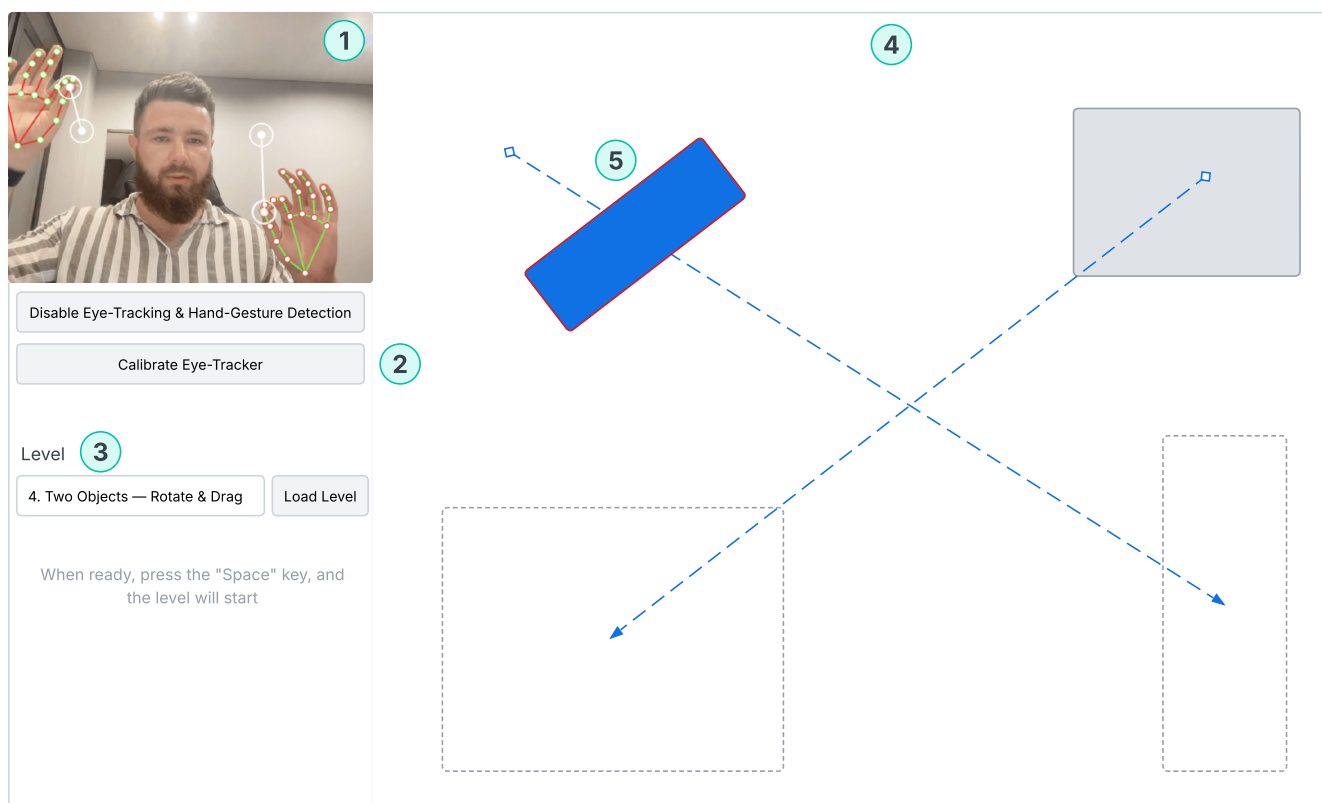


Рис. 3.4. Знімок екрана інтерфейсу користувача розробленої системи — процес маніпулювання захопленням об'єктом

Основні елементи інтерфейсу позначені числовими індикаторами:

1. *Відеопотік* — відображає захоплений потік відео з веб-камери користувача, поверх якого в режимі реального часу відображаються розпізнані ключові точки рук та жести; для розпізнаних жестів відображається початкова точка жесту захоплення (де цей жест був вперше зафіксований), з'єднана лінією з поточною точкою жесту захоплення.

2. *Кнопка «Калібрувати»* — використовується для калібрування модуля відстеження погляду перед початком роботи з системою.
3. *Вибір рівня* — випадаючий список, що дає змогу обрати один з п'яти рівнів експерименту, які відрізняються за складністю.
4. *Область для маніпулювання об'єктами* — містить об'єкти, які необхідно розмістити у відповідних зонах; відповідність між об'єктом та цільовою зоною позначена стрілкою.
5. *Захоплений об'єкт* — об'єкт, до якого наразі застосовується перетворення.

Завершення рівня. Після успішного завершення рівня система відображає час, витрачений на виконання завдання. На Рис. 3.5 представлено знімок екрана інтерфейсу завершеного рівня.

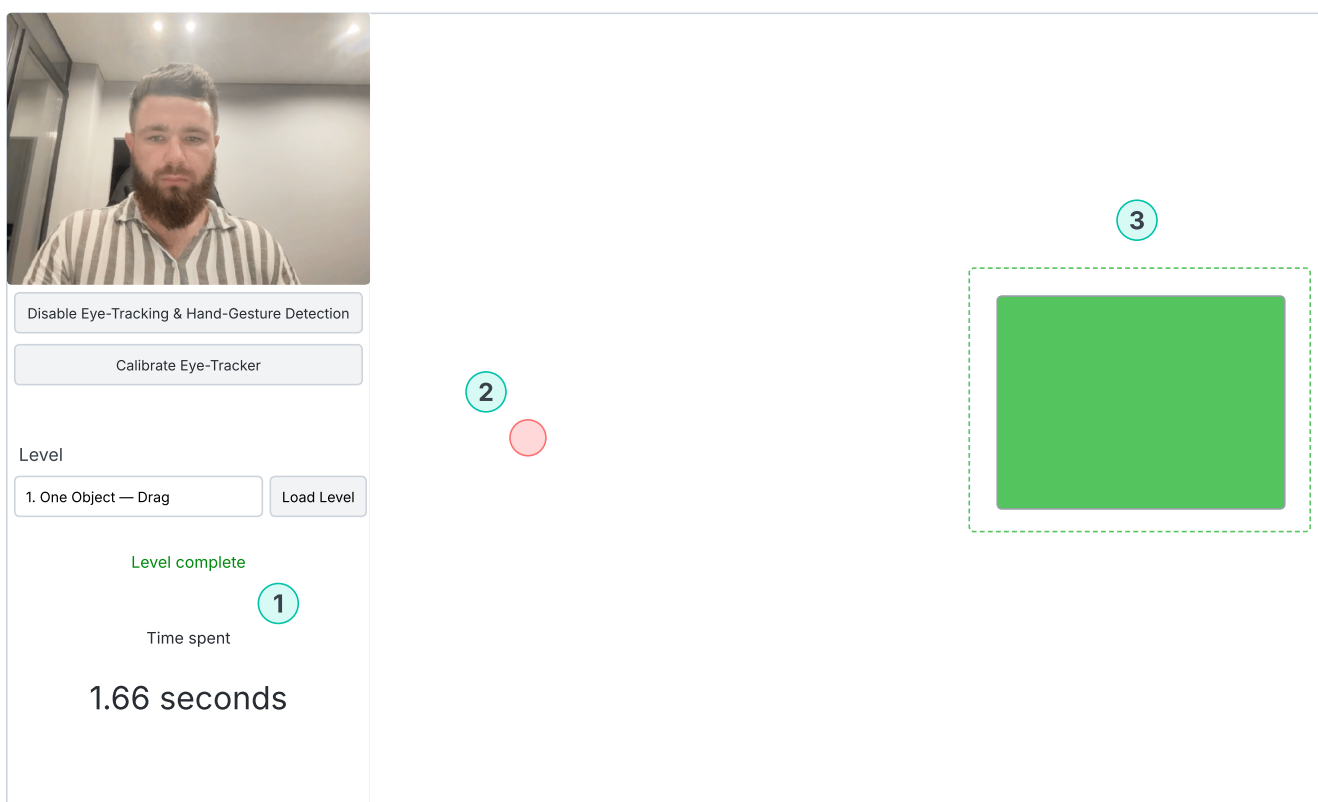


Рис. 3.5. Знімок екрана інтерфейсу користувача розробленої системи — завершений рівень

Окрім уже описаних елементів, на знімку екрана позначено такі складові:

1. *Індикатор часу* — показує, за який час користувач виконав завдання поточного рівня.
2. *Індикатор погляду* — відображає точку фіксації погляду користувача в реальному часі.
3. *Співставлений об'єкт* — об'єкт, що був коректно розміщений у відповідній цільовій області.

3.3 Оцінювання ефективності запропонованого підходу до взаємодії

Для оцінювання ефективності взаємодії на основі погляду та жестів у порівнянні з традиційним введенням за допомогою миші було проведено експериментальне дослідження. Участь у дослідженні взяли 10 респондентів. Кожен учасник виконував завдання на п'яти рівнях, що поступово ускладнювалися, використовуючи обидва методи введення — мишу та запропонований метод на основі погляду та жестів. Кожен рівень проходився п'ять разів для кожного методу.

Рівні були спроектовані з поступовим підвищенням складності за рахунок збільшення кількості об'єктів або типів трансформацій (перетягування, масштабування, обертання), які потрібно застосувати для успішного завершення завдання. Завдання вважалося виконаним, якщо всі об'єкти були розміщені всередині відповідних цільових областей.

На Рис. 3.6 наведено схематичне зображення усіх п'яти рівнів:

1. *Перетягування одного об'єкта* — необхідно перемістити один об'єкт у задану область.
2. *Перетягування кількох об'єктів* — необхідно перемістити два об'єкти у відповідні області.
3. *Масштабування + перетягування* — необхідно зменшити розмір двох об'єктів і помістити їх у відповідні області.

4. *Обертання + перетягування* — необхідно повернути кожен об'єкт і помістити їх у відповідні обласні.
5. *Усі трансформації* — необхідно застосувати всі три типи трансформацій (перетягування, масштабування, обертання), щоб коректно розмістити два об'єкти у відповідних областях.

Основним критерієм оцінювання ефективності був час виконання кожного завдання кожного рівня, зафіксований у секундах. Отримані дані були проаналізовані та порівняні між двома методами введення.

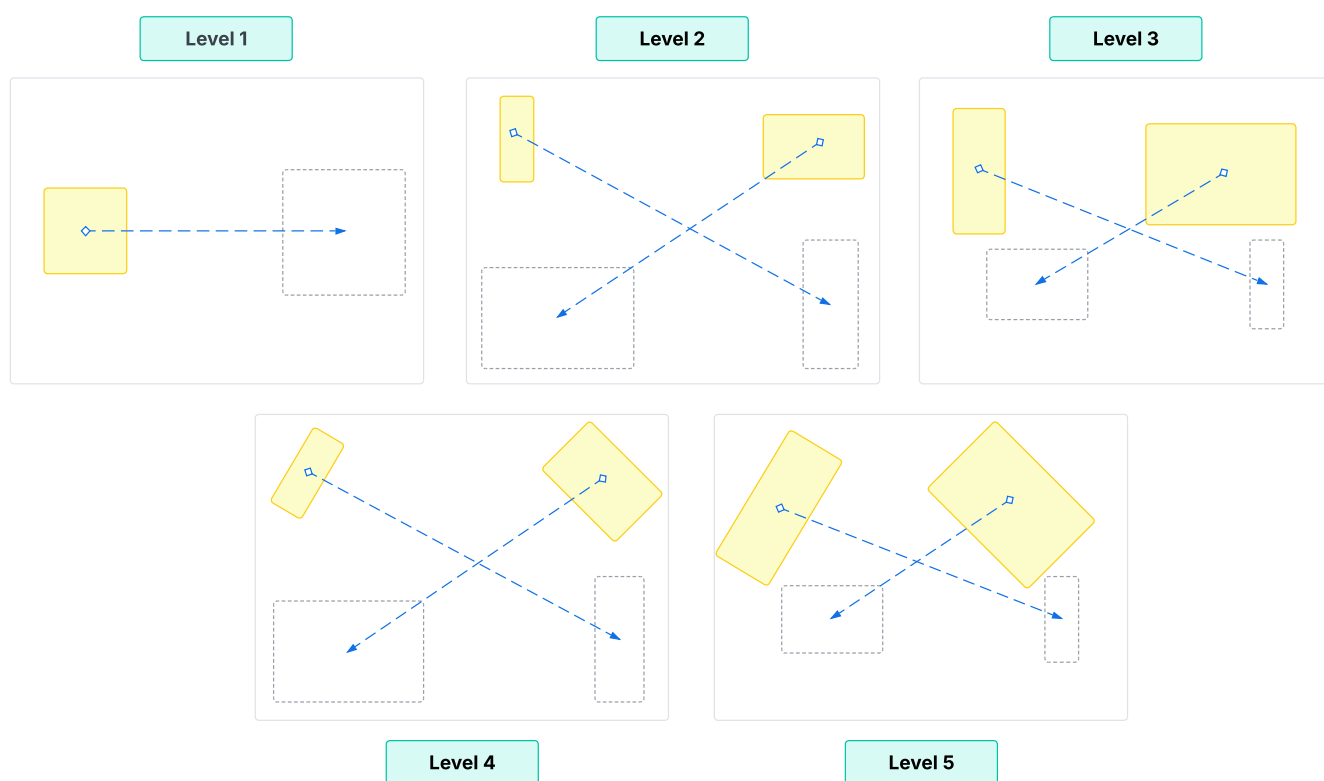


Рис. 3.6. Схематичне зображення усіх п'яти рівнів експериментального дослідження

3.4 Аналіз результатів

Середній час виконання завдань на кожному рівні, агрегований за всіма учасниками, подано на Рис. 3.7. Результати порівняння свідчать, що використання миші забезпечувало вищу швидкість виконання завдань на простіших рівнях. Однак на рівні 5, який вимагав одночасного застосування кількох трансформацій (обертання, масштабування, переміщення), метод взаємодії на основі погляду та жестів продемонстрував вищу ефективність, перевершивши метод введення мишею за середнім часом виконання.

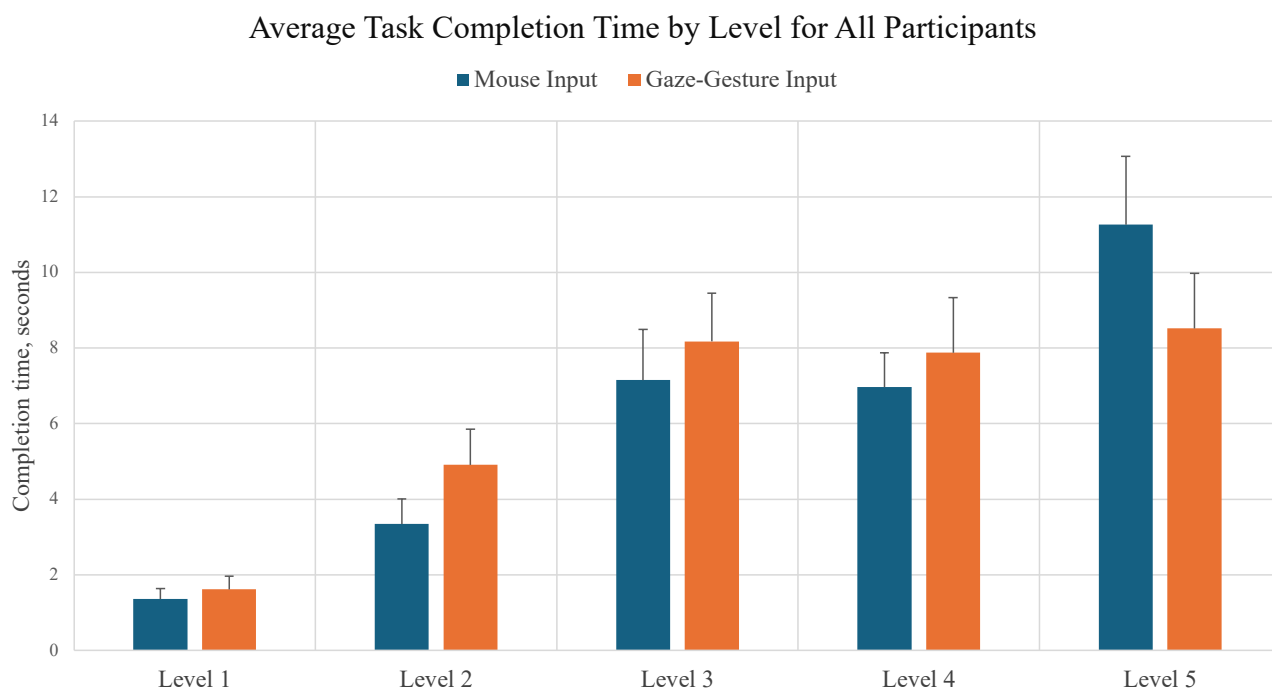


Рис. 3.7. Середній час виконання завдання на кожному з рівнів (усереднений для всіх учасників експерименту)

На Рис. 3.8 представлено діаграми типу "ящик з вусами" (boxplot), які ілюструють розподіл часу виконання завдань для кожного рівня. Аналіз показав, що при використанні миші спостерігався менший розкид значень, особливо на перших рівнях, що свідчить про стабільність цього способу введення. Натомість

при взаємодії на основі погляду та жестів спостерігалася більша варіабельність результатів, що є очікуваним з огляду на новизну методу для учасників.

Примітно, що на рівні 5 метод, який поєднує погляд і жести, не лише досяг рівня ефективності миші, а й перевершив його: медіанна тривалість виконання завдання була нижчою, а розподіл значень — більш зсунутим у бік швидших результатів. Це може свідчити про перевагу одночасного виконання декількох дій за допомогою жестів та погляду порівняно з послідовним виконанням дій мишею.

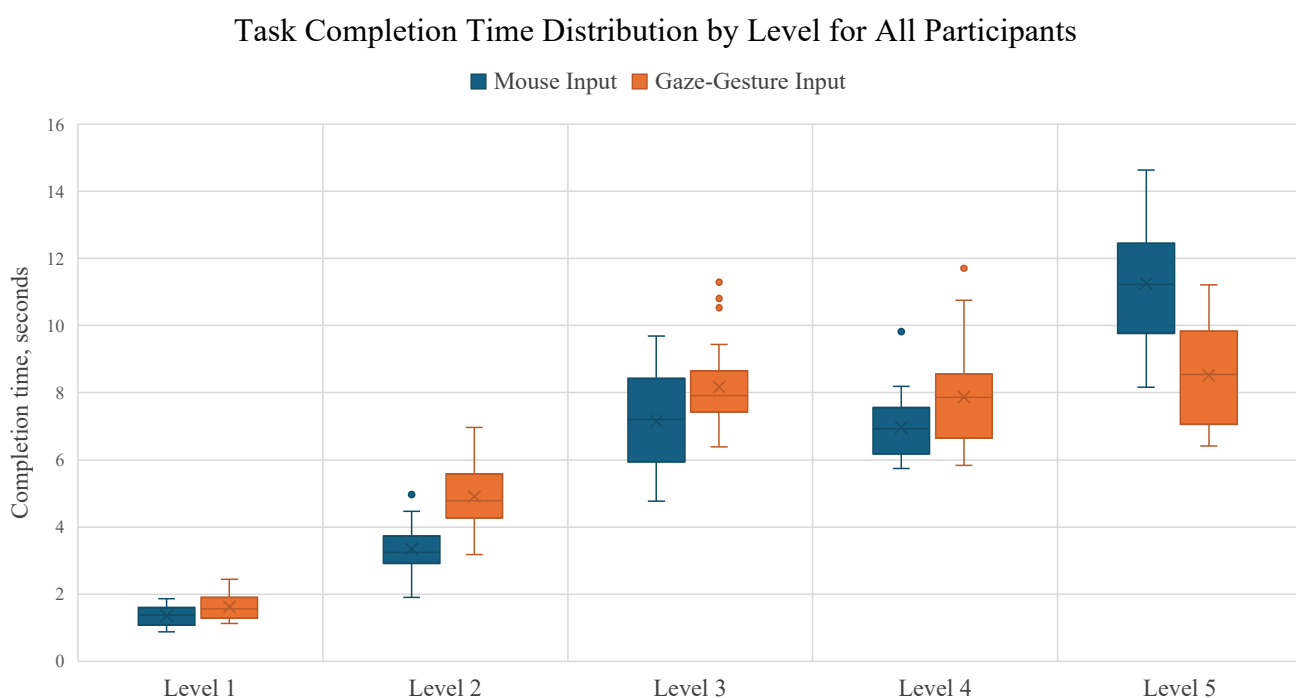


Рис. 3.8. Розподіл часу виконання завдань на кожному рівні (усі учасники експерименту)

3.5 Обговорення результатів

Проведене дослідження ефективності запропонованого підходу до взаємодії на основі погляду та жестів у веб-середовищі дозволило виявити як його переваги, так і поточні обмеження. Загалом учасники позитивно оцінили інтуїтивність та зручність використання запропонованого підходу, особливо в контексті

маніпулювання об'єктами у веб-середовищі. Водночас деякі аспекти потребують подальшого вдосконалення для підвищення стабільності та точності взаємодії.

Порівняльний аналіз із традиційним введенням за допомогою миші засвідчив, що запропонований метод є особливо ефективним у виконанні складних завдань, які передбачають одночасне застосування кількох трансформацій. У таких сценаріях користувачі змогли швидше завершувати завдання завдяки паралельному застосуванню трансформацій масштабування та обертання, у той час як при використанні введення мишею ці дві дії мали виконуватися користувачем послідовно. Проте у випадках простих операцій миша продовжує демонструвати вищу точність і стабільність результатів.

Високий рівень варіативності часу виконання завдань при використанні погляду та жестів можна пояснити новизною підходу для користувачів і потребою у короткому періоді адаптації. Крім того, деякі учасники повідомили про труднощі з точним визначенням положення погляду, що впливало на вибір об'єктів і знижувало загальну точність системи. Цей аспект підкреслює важливість подальшого вдосконалення модуля відстеження погляду.

3.6 Виклики та обмеження

Попри численні переваги запропонованого підходу, система має низку обмежень, які необхідно враховувати під час її впровадження та подальшого вдосконалення. Одним із ключових викликів залишається точність відстеження погляду. Хоча використання бібліотеки WebGazer.js забезпечує високу доступність і просту інтеграцію у веб-середовище, її точність помітно поступається професійним рішенням — особливо за умов змінного освітлення, обмеженої роздільної здатності чи низької якості веб-камери.

Іншим важливим фактором є процес калібрування, від якого безпосередньо залежить якість даних фіксації погляду. Поточна процедура передбачає активну участь користувача, може бути дещо тривалою та впливати на загальну зручність

взаємодії з системою. Цей етап є критично важливим для забезпечення точності, але потребує подальшої оптимізації.

Крім того, апаратні обмеження також істотно впливають на якість роботи системи. Оскільки рішення базується на використанні стандартної веб-камери, воно має нижчу стабільність і точність у порівнянні з професійними трекерами погляду. Такий компроміс між доступністю та точністю є ключовим викликом для розробки подібних інтерфейсів.

Загалом результати експериментального дослідження підтверджують перспективність запропонованого підходу до взаємодії у веб-середовищі, але також вказують на критичні аспекти, які потребують подальшого вдосконалення — насамперед у контексті підвищення точності відстеження погляду та стабільності розпізнавання жестів.

3.7 Сфери потенційного застосування

Розроблена система взаємодії на основі погляду та жестів має широкий спектр можливих сфер застосування. Завдяки веб-орієнтованій реалізації та використанню лише стандартного споживчого обладнання, вона не потребує спеціалізованих апаратних засобів, що робить її доступною для широкого кола користувачів.

Одним із перспективних напрямів є інтеграція подібних рішень у веб-застосунки з інтерактивними об'єктами, зокрема в онлайн-інструменти для дизайну, інтерактивні карти, освітні платформи та віртуальні середовища, де потрібна складна маніпуляція об'єктами, що включає перетягування, масштабування та обертання.

Іншою важливою сферою використання є засоби для віддаленої співпраці, де така система може покращити природність і зручність взаємодії, наприклад, під час спільного перегляду, обговорення або редагування 3D-моделей у реальному часі.

Крім того, запропонована методика має значний потенціал у контексті підвищення доступності цифрових інтерфейсів. Вона може слугувати альтернативним методом введення для користувачів із порушеннями моторики, надаючи змогу інтуїтивно керувати веб-інтерфейсами без застосування традиційних пристроїв введення — миші або клавіатури.

3.8 Перспективи подальших досліджень

Майбутні дослідження можуть бути спрямовані на розширення функціональних можливостей запропонованого підходу, зокрема через інтеграцію взаємодії на основі погляду та жестів із традиційними пристроями введення, такими як миша або сенсорний екран. Це забезпечить гнучкість у виборі інтерфейсного методу відповідно до контексту використання та потреб користувача.

Одним із ключових напрямів удосконалення системи є підвищення точності відстеження погляду. Хоча використання бібліотеки WebGazer.js забезпечує високу доступність та простоту впровадження у веб-середовище, її ефективність залишається чутливою до зовнішніх умов — зокрема, освітлення, положення голови користувача та якості калібрування. У цьому контексті перспективними є дослідження, спрямовані на вдосконалення процедури калібрування — наприклад, шляхом збільшення кількості контрольних точок і подальшого аналізу їхнього впливу на загальну точність системи, або використанням іншого підходу до процесу калібрування.

Перспективним також є розширення набору підтримуваних жестів та реалізація багатокористувацької взаємодії. Це дозволить адаптувати систему до колаборативних платформ, спільного редагування та віддаленої співпраці.

Поєднання технології відстеження погляду з іншими модальностями, зокрема голосовим введенням, відкриває нові можливості. Так, у дослідженні [61] описано прототип, що використовує великі мовні моделі (Large Language Models, LLM) — тип штучного інтелекту, здатний генерувати текст, відповідати на

запитання та інтерпретувати команди, спираючись на попереднє навчання на великих обсягах текстових даних. У цьому випадку LLM використовується для керування об'єктами за допомогою голосових команд. Попередні результати свідчать про інтуїтивність та гнучкість такого підходу. Поєднання голосових команд із вибором об'єктів за допомогою погляду має потенціал для створення нових форм доступної взаємодії, зокрема для осіб із порушеннями моторики, як у веб-середовищі, так і в AR/VR.

3.9 Висновки

У цьому розділі дисертаційного дослідження запропоновано та реалізовано веб-орієнтований підхід до маніпулювання об'єктами, який поєднує використання відстеження погляду для вибору об'єкта та розпізнавання жестів рук для виконання трансформацій, що забезпечує можливість безконтактної взаємодії із застосуванням лише стандартної вебкамери.

Проведене експериментальне оцінювання підтвердило ефективність запропонованого підходу у складних сценаріях взаємодії, що вимагають одночасного виконання кількох трансформацій (перетягування, масштабування, обертання), де час виконання завдань виявився меншим у порівнянні з традиційним введенням за допомогою миші. Водночас встановлено, що для простих завдань, зокрема перетягування об'єктів, традиційне введення залишається ефективнішим.

У процесі дослідження виявлено основні обмеження запропонованого підходу: обмежену точність відстеження погляду, що впливає на точність вибору об'єктів, та необхідність постійного утримання рук у межах поля зору камери для стабільного розпізнавання жестів.

Подальший розвиток системи передбачає удосконалення механізмів калібрування для підвищення точності відстеження погляду, розробку покращених візуальних індикаторів для підтримки користувача під час взаємодії, а також застосування алгоритмів машинного навчання для підвищення точності та стійкості розпізнавання жестів.

Отримані результати підтверджують перспективність комбінованого підходу на основі погляду та жестів рук для створення доступних, інтуїтивно зрозумілих і гнучких підходів у сфері людино-комп'ютерної взаємодії, що сприяє розвитку технологій безконтактної взаємодії у веб-середовищах.

РОЗДІЛ 4. БЕЗКОНТАКТНЕ КЕРУВАННЯ ЦИФРОВИМИ СИСТЕМАМИ З ВИКОРИСТАННЯМ ВИРАЗІВ ОБЛИЧЧЯ

Цей розділ дисертаційного дослідження присвячено дослідженню можливостей безконтактної взаємодії з цифровими інтерфейсами на основі виразів обличчя. Традиційні засоби введення, такі як миша та клавіатура, залишаються основними інструментами управління у цифровому середовищі. Водночас у ситуаціях з обмеженою моторикою або в умовах, де використання рук є ускладненим, виникає потреба у доступних і альтернативних методах взаємодії.

Одним із таких методів є використання виразів обличчя як способу ініціювання дій. Завдяки розвитку технологій комп'ютерного зору та розпізнавання обличчя з'явилася можливість реєструвати мімічні зміни та інтерпретувати їх як команди керування. У межах цього дослідження запропоновано підхід, що ґрунтується на мімічних формах змішування (facial blendshape) [62], які використовуються для створення налаштовуваних виразів обличчя та їх зіставлення з певними діями у цифровому середовищі.

Реалізована система не потребує спеціалізованого обладнання та функціонує із використанням лише стандартної веб-камери, що значно розширює її доступність для користувачів. Метою цього розділу є створення та експериментальне оцінювання прототипу взаємодії на основі виразів обличчя, а також вивчення його потенціалу як альтернативного або додаткового засобу керування.

Результати, отримані в межах цього етапу роботи, були оприлюднені у публікаціях [63], [64].

4.1 Формалізація використання мімічних форм змішування в поточному дослідженні

У цьому дослідженні мімічні форми змішування (blendshapes) є основою для визначення користувацьких дій, реалізованих через вирази обличчя. Вихідні дані модуля MediaPipe Face Landmarker [65] включають набір коефіцієнтів форм

змішування у діапазоні від 0 до 1, де кожен коефіцієнт відповідає певному м'імічному руху або виразу обличчя. У межах запропонованої реалізації форма змішування вважається активованою, якщо її значення дорівнює або перевищує порогове значення 0.5.

Обґрунтування порогового значення. Порогове значення 0.5 було визначено емпірично під час серії попередніх випробувань, що мали на меті знайти оптимальний баланс між чутливістю системи та її точністю. Обране значення забезпечує достатню стійкість до випадкових або незначних м'імічних змін, одночасно зберігаючи природність та зручність користувацької взаємодії.

Формалізація активації виразів обличчя. Для стандартизації процедури обробки м'імічних даних у дослідженні було введено булеву функцію $F(\textit{blendshape})$, яка визначає активність відповідної форми змішування:

$$F(\textit{blendshape}) = \begin{cases} 1, & \textit{blendshape score} \geq 0.5 \\ 0, & \textit{otherwise} \end{cases} \quad (4.1)$$

Для визначення комплексних виразів обличчя, які формуються через комбінацію декількох форм змішування, використовується наступна функція:

$$E_{\textit{expression}} = \bigwedge_{j \in A} F(\textit{blendshape}_j) \wedge \bigwedge_{k \in B} \neg F(\textit{blendshape}_k), \quad (4.2)$$

де:

- $E_{\textit{expression}}$ — логічне значення, що вказує на активність відповідного виразу обличчя $\textit{expression}$;
- A — множина форм змішування, які мають бути активними;
- B — множина форм змішування, які мають бути неактивними.

Наприклад, вираз обличчя «Smile Left» формалізується як:

$$E_{\textit{Smile Left}} = F(\textit{mouthSmileLeft}) \wedge \neg F(\textit{mouthSmileRight}),$$

що відповідає: $A = \{\textit{mouthSmileLeft}\}$, $B = \{\textit{mouthSmileRight}\}$

Відбір мімічних форм змішування. Вихідні дані MediaPipe Face Landmarker включають набір коефіцієнтів для 52-ох форм змішування. Однак, з метою підвищення стабільності й надійності взаємодії, для використання в системі було відібрано лише їхню підмножину. Попереднє тестування проводилося на окремій групі користувачів, з метою визначення легкості активації кожної форми змішування та її стабільності під час повторного відтворення.

До фінального набору увійшли лише ті форми, які демонстрували високу повторюваність, стабільність і зручність відтворення. Натомість було відкинуто форми, які виявилися малопомітними, нестабільними або складними для активації.

4.2 Архітектура системи

Запропонована система реалізує структурований конвеєр (pipeline) для безконтактної взаємодії на основі виразів обличчя. Такий конвеєр трансформує необроблені вхідні дані у відповідні реакції системи шляхом послідовного виконання кількох етапів. Архітектура системи поділяється на дві основні стадії: стадію обробки (processing stage) та стадію виконання дій (action execution stage). Етапи робочого процесу трансформації даних та виконання дій наведено на Рис. 4.1 та детально описано нижче.

Вхідні дані. Система отримує відеопотік у реальному часі за допомогою стандартної веб-камери. Отримані зображення використовуються як джерело даних для аналізу виразів обличчя.

Стадія обробки. Першим кроком на цій стадії є виявлення та оцінка виразів обличчя. Для ідентифікації ключових точок обличчя та обчислення параметрів мімічних форм змішування, що характеризують вирази обличчя та їхню інтенсивність, використовується модуль MediaPipe Face Landmarker. Далі застосовується фільтрація за пороговим значенням: визначається, які форми змішування є активними. Такий підхід запобігає випадковим активаціям і підвищує загальну точність системи.

Стадія виконання дій. На цій стадії здійснюється зіставлення активованих виразів обличчя з наперед визначеними командами (наприклад, навігація, маніпулювання об'єктами тощо). Після цього відповідні команди передаються до модуля управління, який забезпечує їх виконання у цифровому середовищі.

Вихідні дані. Результатом роботи системи є виконання користувацької дії в реальному часі. Для зручності користувача система також надає візуальний зворотний зв'язок у інтерактивному інтерфейсі, що дає змогу негайно оцінити ефективність виконаних дій.

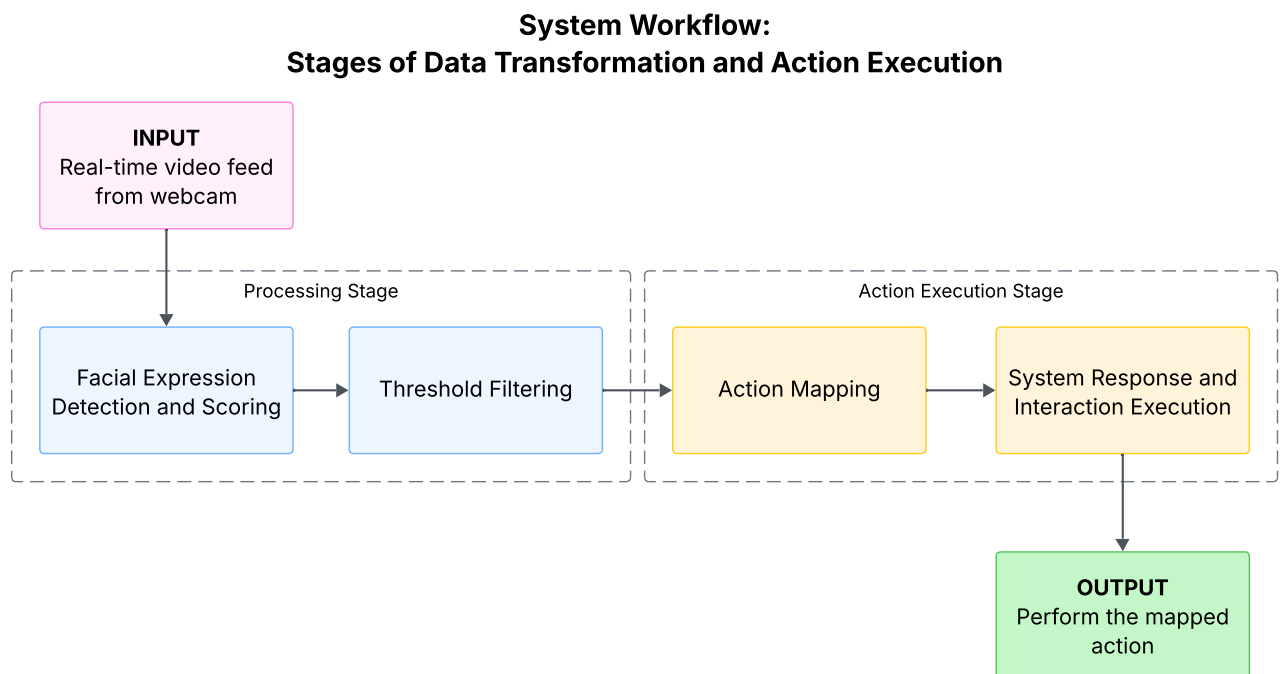


Рис. 4.1. Етапи робочого процесу запропонованої системи

4.3 Методи та критерії оцінювання ефективності системи

Для оцінювання ефективності запропонованої системи використовувалася комбінація кількісних і якісних показників. Вибір метрик ґрунтувався на усталених підходах у галузі взаємодії людини з комп'ютером (HCI), що дало змогу здійснити комплексний аналіз зручності використання та ефективності системи.

Першою кількісною метрикою, спрямованою на оцінювання ефективності взаємодії, виступає час виконання завдань. Цей показник є загальноприйнятим у дослідженнях зручності використання (наприклад, згідно зі стандартом ISO 9241-11) і дозволяє визначити швидкість адаптації користувача до нової системи. Окрім того, аналіз динаміки виконання завдань у межах кількох спроб дає змогу оцінити ефект навчання та зміни у продуктивності.

Для отримання суб'єктивної оцінки зручності, інтуїтивності інтерфейсу та можливих труднощів під час взаємодії було проведено опитування учасників експерименту після завершення виконання завдань. Такі дані дозволяють виявити аспекти, що не завжди виявляються при аналізі кількісних показників. Крім того, респонденти надали пропозиції щодо потенційних сценаріїв застосування розробленого підходу.

Поєднання об'єктивних і суб'єктивних метрик забезпечує комплексну оцінку запропонованої системи, яка охоплює як показники продуктивності (ефективність, стабільність виконання завдань), так і аспекти користувацького досвіду. Такий підхід дозволяє не лише оцінити поточний стан системи, а й окреслити перспективи її застосування в реальних умовах і напрями подальшого вдосконалення.

Запропонована система використовує алгоритми комп'ютерного зору та попередньо навчені моделі для зчитування інтенсивності мімичних форм змішування без потреби додаткового спеціалізованого навчання. У межах цього дослідження не проводилась кількісна оцінка точності розпізнавання виразів обличчя, оскільки основна мета полягала у перевірці доцільності використання виразів як засобу безконтактної взаємодії у типових цифрових сценаріях. Оцінювання зосереджувалося на зручності створення індивідуальних конфігурацій виразів і загальному користувацькому досвіду. Водночас слід зазначити, що питання точності набуває особливої актуальності в контексті побудови адаптивних систем, які передбачають автоматичне налаштування порогових значень

інтенсивності мімічних форм під конкретного користувача. Для забезпечення надійності розпізнавання в умовах індивідуальних фізіологічних особливостей такі системи потребують додаткового аналізу точності.

4.4 Порівняння взаємодії на основі виразів обличчя та комп'ютерної миші у задачі редагування тексту

Метою експерименту було попереднє оцінювання доцільності та зручності використання виразів обличчя як засобу взаємодії під час виконання завдань з редагування тексту. Основне завдання дослідження полягало у порівнянні ефективності та зручності взаємодії на основі виразів обличчя з традиційною взаємодією за допомогою комп'ютерної миші в умовах контрольованого експериментального середовища.

4.4.1 Налаштування експерименту

Дослідження здійснювалося з використанням веб-прототипу, призначеного для виявлення та інтерпретації виразів обличчя. Для експерименту було обрано кілька виразів, які є простими у відтворенні та легко розпізнаються. Кожен вираз визначався на основі відповідних мімічних форм змішування (blendshapes). Формалізовані визначення обраних виразів обличчя наведено у Табл. 4.1.

Вираз обличчя	Формула виразу обличчя на основі мімічних форм змішування ($E_{expression}$)
Left Eye Closed	$F(eyeBlinkLeft) \wedge \neg F(eyeBlinkRight)$
Mouth Pucker	$F(mouthPucker)$
Smile	$F(mouthSmileLeft) \wedge F(mouthSmileRight)$
Jaw Open	$F(jawOpen)$

Табл. 4.1. Вирази обличчя, використані в експерименті, та відповідні формули на основі мімічних форм змішування

На основі зазначених виразів було визначено користувацькі дії для форматування та редагування тексту в імітаційному текстовому редакторі. Відповідність між виразами обличчя та діями системи наведено в Табл. 4.2. Інтерфейс реалізованого прототипу ілюстровано на Рис. 4.2.

Користувацька дія	Формула виразу обличчя
Застосування жирного форматування (bold)	$E_{Left\ Eye\ Closed}$
Застосування курсиву (italic)	$E_{Mouth\ Pucker}$
Застосування підкреслення (underline)	E_{Smile}
Застосування закреслення (strikethrough)	$E_{Jaw\ Open}$
Застосування підсвічування (highlight)	$E_{Left\ Eye\ Closed} \wedge E_{Smile}$
Видалення виділених слів (delete)	$E_{Left\ Eye\ Closed} \wedge E_{Mouth\ Pucker}$
Виділення всіх слів (select all)	$E_{Left\ Eye\ Closed} \wedge E_{Jaw\ Open}$

Табл. 4.2. Визначення користувацьких дій у імітаційному текстовому редакторі на основі обраних виразів обличчя

Механізм активації виразів обличчя. З метою запобігання хибному розпізнаванню та випадковій активації команд, система використовувала механізм підтвердження через клавішу «Shift». Користувачі повинні були утримувати клавішу «Shift» під час демонстрації виразу обличчя, а активація відповідної дії відбувалася лише в момент відпускання клавіші за умови, що у цей момент вираз залишався активним. Інтерфейс надавав зворотний зв'язок у реальному часі, підсвічуючи розпізнаний вираз і відповідну дію до її виконання.

Учасники. У дослідженні взяли участь шість користувачів, усі з яких мали досвід роботи з комп'ютером, але не мали попереднього досвіду використання виразів обличчя для взаємодії з інтерфейсами.

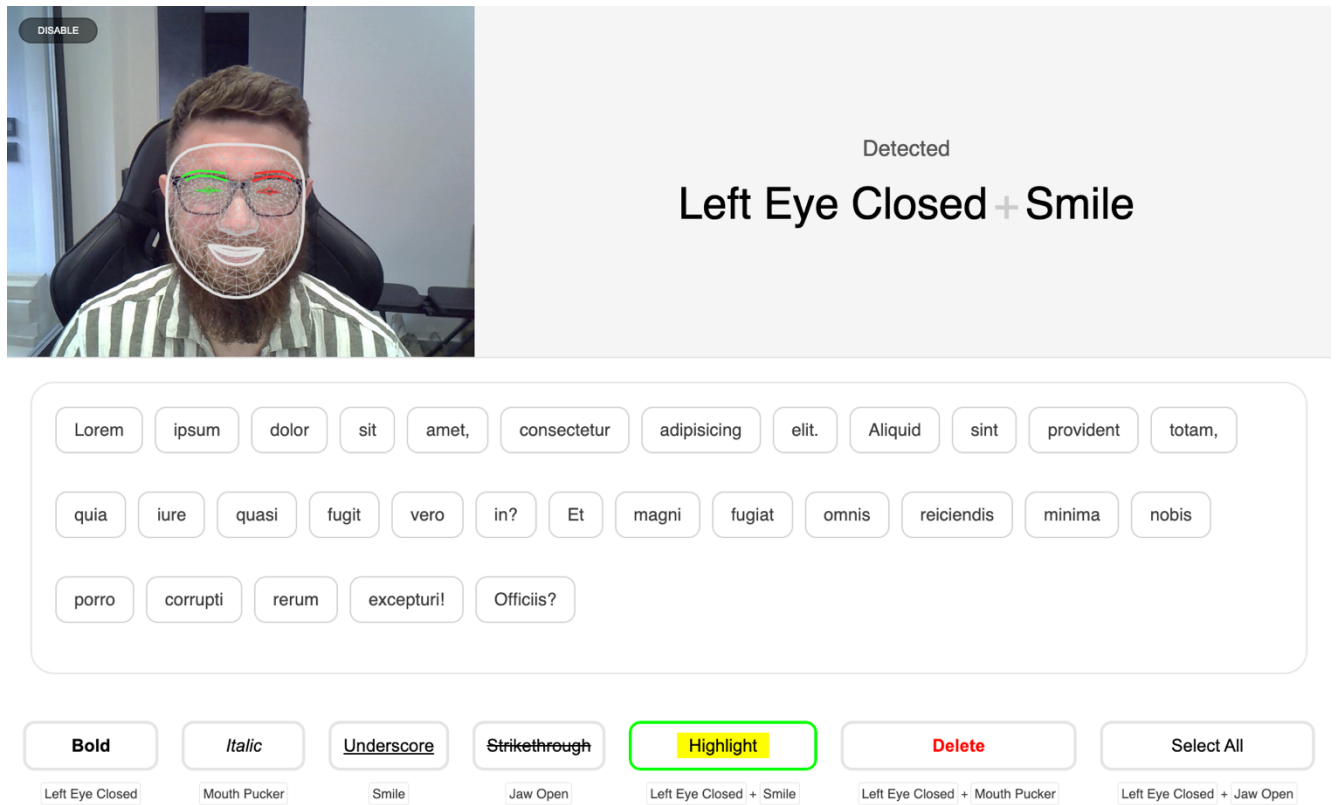


Рис. 4.2. Інтерфейс реалізованого імітаційного текстового редактора, використаного в експерименті

Процедура експерименту. Кожен учасник виконував набір завдань з редагування тексту, використовуючи два різні методи взаємодії:

- *контрольна умова* — традиційна взаємодія з використанням лише комп'ютерної миші;
- *експериментальна умова* — поєднання комп'ютерної миші з виразами обличчя, активація яких здійснювалася за допомогою клавіші «Shift».

У обох експериментальних умовах для взаємодії з інтерфейсом використовувалася комп'ютерна миша. У контрольній умові всі дії виконувалися

виключно за допомогою миші. У експериментальній умові миша використовувалася лише для вибору слів у тексті, тоді як основні команди форматування (жирний шрифт, курсив тощо) активувалися за допомогою виразів обличчя у поєднанні з клавішею «Shift».

У кожному випадку фіксувався час виконання завдання для подальшого аналізу ефективності. Окрім цього, учасники надавали якісний зворотний зв'язок щодо зручності, інтуїтивності та складнощів взаємодії.

Мета експерименту. Основними завданнями експерименту були:

1. Оцінити доцільність застосування виразів обличчя як альтернативного методу введення.
2. Порівняти ефективність виконання завдань за допомогою миші та виразів обличчя.
3. Зібрати відгуки користувачів щодо зручності, легкості запам'ятовування та відтворення виразів обличчя.

Результати цього етапу дослідження стали підґрунтям для подальших удосконалень та реалізації повноцінної системи безконтактної взаємодії у наступному експерименті.

4.4.2 Аналіз отриманих результатів

Час виконання завдань. Порівняльний аналіз тривалості виконання завдань показав, що традиційний метод введення за допомогою комп'ютерної миші був помітно швидшим, ніж запропонований підхід із використанням виразів обличчя. На Рис. 4.3 наведено час виконання завдань для кожного з методів для всіх учасників експерименту.

Середній час виконання завдання за допомогою миші становив 20,34 секунди, тоді як при використанні комп'ютерної миші у поєднанні з виразами обличчя та клавішею «Shift» — 33,45 секунди.

Ці результати свідчать про те, що новий метод потребує більше часу для виконання завдань. Водночас, його застосування було повністю функціональним і дало змогу виконати всі поставлені дії.

Зворотній зв'язок від учасників. У ході експерименту було зібрано якісні відгуки учасників щодо зручності використання системи та складнощів, які виникали під час виконання завдань. Більшість респондентів зазначили, що їм було важко запам'ятати відповідність між виразами обличчя та командами, особливо у випадках комбінованих виразів (наприклад, «Left Eye Closed + Smile» для виділення тексту). Для успішного виконання завдань учасники часто поклалися на візуальні підказки, надані інтерфейсом.

Декілька учасників запропонували удосконалити систему, зокрема — реалізувати більш інтуїтивні відповідності між виразами обличчя та командами або покращити візуальний супровід, що сигналізує про активовану дію.

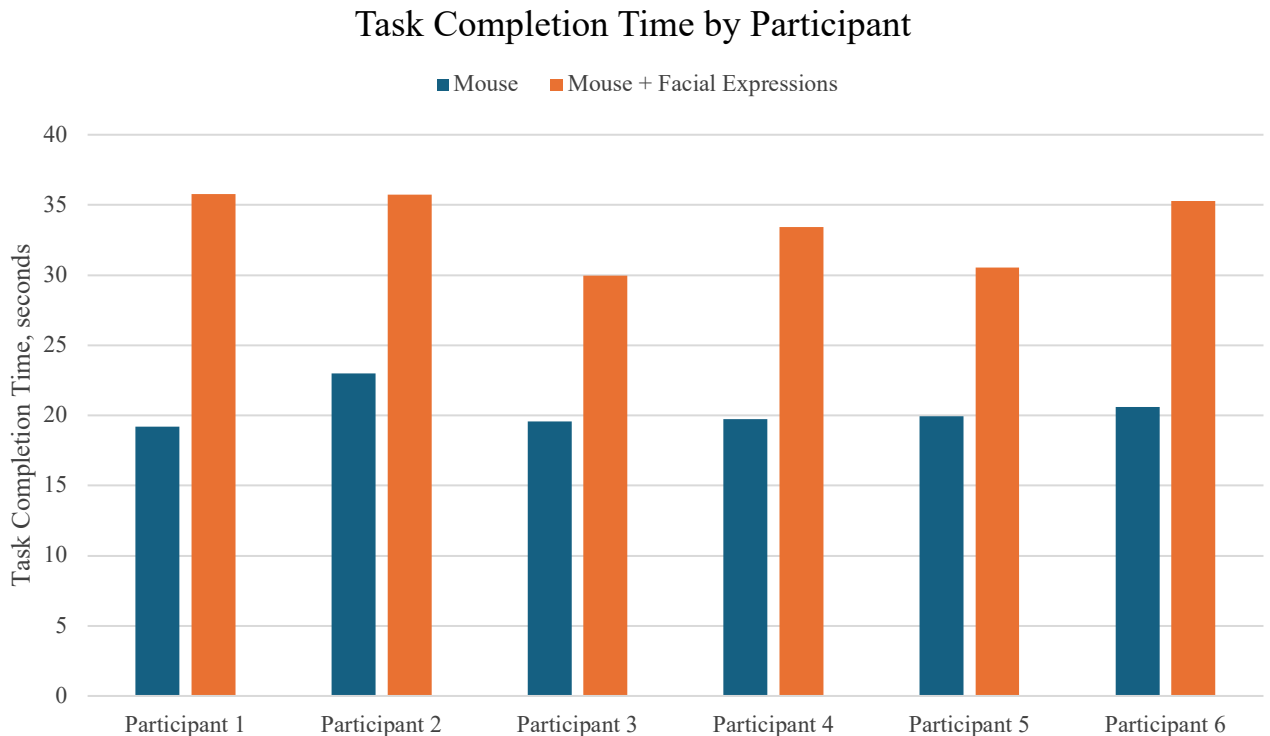


Рис. 4.3. Час виконання завдань із використанням лише комп'ютерної миші та з використанням миші у поєднанні з виразами обличчя

Незважаючи на збільшену тривалість виконання завдань, учасники позитивно оцінили потенціал безконтактного керування. Цей підхід був визнаний перспективним, особливо для сценаріїв багатозадачності або для користувачів з обмеженою рухливістю.

4.4.3 Обговорення

Результати експерименту підтвердили доцільність використання виразів обличчя як засобу взаємодії для виконання дій у текстовому редакторі, проте водночас виявили низку суттєвих викликів, які необхідно врахувати для подальшого вдосконалення запропонованого підходу.

Переваги методу. Однією з ключових переваг запропонованої взаємодії є можливість безконтактного керування. Учасники експерименту позитивно оцінили здатність взаємодіяти з інтерфейсом без використання рук, що особливо актуально для користувачів з обмеженою моторикою або в умовах, коли традиційні засоби введення (миша, клавіатура) є недоступними або незручними.

Другою важливою перевагою є гнучкість у виконанні команд. Комбінування мимічних виразів дозволяє реалізовувати широкий спектр дій, не обмежуючись лише базовими жестами.

До переваг також належить доступність і кросплатформність реалізованої системи. Її веб-орієнтоване середовище та використання стандартної веб-камери дозволяють уникнути потреби у спеціалізованому обладнанні, що забезпечує потенційно широке коло користувачів.

Обмеження та виклики. Незважаючи на переваги, запропонований підхід має низку обмежень. Одним із головних викликів є збільшений час виконання завдань у порівнянні з взаємодією лише комп'ютерною мишею. Це, ймовірно, пов'язано з когнітивними труднощами користувачів, зокрема — складністю запам'ятовування відповідності між виразами обличчя та діями. Особливо це стосувалося комбінованих виразів, які вимагали більшої концентрації уваги.

Додатковим обмеженням стала необхідність утримування клавіші «Shift» як засобу підтвердження виразу. Цей механізм був запроваджений для запобігання випадковим активаціям, однак він ускладнив процес взаємодії та, ймовірно, став додатковим чинником, що вплинув на зростання часу виконання завдань.

Можливості покращення. Одним із напрямів удосконалення є оптимізація відповідності виразів обличчя певним командам. Використання більш інтуїтивних комбінацій або можливість налаштування таких комбінацій користувачем дозволять зменшити когнітивне навантаження.

Іншим перспективним рішенням є модифікація механізму підтвердження команд. Замість використання клавіші «Shift» можливо запровадити альтернативні способи активації, зокрема інші вирази обличчя, голосові команди чи жести голови. Наприклад, для активації режиму введення команд користувач може заплющити ліве око. Після цього він відтворює відповідний вираз обличчя, який відповідає бажаній дії. Для підтвердження виконання команди користувач заплющує праве око. Альтернативно, повторне заплющення лівого ока може слугувати сигналом для скасування команди та виходу з режиму введення.

Також доцільним є впровадження навчального режиму, що включає поступове ознайомлення з командами та удосконалені візуальні підказки — це може покращити досвід першого використання та прискорити навчання.

Значення отриманих результатів. Отримані результати демонструють, що взаємодія на основі виразів обличчя є функціональною, однак потребує додаткових кроків для підвищення ефективності, інтуїтивності та зручності використання. Проведений експеримент став підґрунтям для розробки вдосконаленої версії системи, в якій реалізовано зокрема усунення необхідності натискання клавіші «Shift» та переглянуто відповідність виразів обличчя до користувацьких дій.

4.4.4 Висновки

Проведений експеримент підтвердив доцільність використання виразів обличчя як альтернативного засобу взаємодії для виконання дій у текстовому

редакторі. Результати засвідчили функціональність запропонованого методу безконтактного керування та його потенціал для застосування у цифрових середовищах, зокрема в умовах обмеженого доступу до традиційних засобів введення.

Водночас експеримент виявив низку обмежень, що впливають на швидкість і зручність взаємодії. Зокрема, порівняння з традиційним введенням за допомогою комп'ютерної миші показало, що виконання завдань із використанням виразів обличчя потребує більше часу. Це пояснюється як наявністю додаткового механізму підтвердження через клавішу «Shift», так і когнітивними труднощами, пов'язаними із запам'ятовуванням комбінацій виразів, особливо для складних дій.

Загалом результати вказують на необхідність подальшої оптимізації методу, зокрема спрощення процесу підтвердження команд, удосконалення відповідності виразів обличчя до функціональних дій, а також впровадження адаптивних або навчальних елементів для підтримки користувача.

Отримані висновки стали підґрунтям для вдосконалення системи, реалізованого в межах наступного експерименту. Основні зміни включали усунення необхідності використання клавіші «Shift», оптимізацію відповідностей між виразами обличчя та діями, а також оцінку впливу навчального ефекту на ефективність і зручність використання системи.

4.5 Оцінка навчального ефекту безконтактної взаємодії на основі виразів обличчя

На основі результатів попереднього дослідження було розроблено та реалізовано подальший експеримент [64], описаний у цій секції. Основною метою нового дослідження було забезпечення повністю безконтактної взаємодії, а також усунення виявлених раніше недоліків, зокрема труднощів із запам'ятовуванням виразів обличчя, шляхом впровадження більш інтуїтивно зрозумілих відповідностей між виразами та командами.

4.5.1 Вибір мимічних форм змішування для цього дослідження

Для визначення набору мимічних форм змішування, використаних у межах цього дослідження, було проведено попереднє тестування. Його метою було виокремлення таких виразів обличчя, які є достатньо виразними та можуть стабільно відтворюватися звичайними користувачами. Додатковим критерієм слугував потенціал поєднання окремих форм у комбіновані вирази, що можуть бути використані для активації складніших користувацьких дій (наприклад, поєднання руху брів із відкриванням рота).

У Табл. 4.3 наведено перелік відібраних виразів обличчя та їх формалізоване представлення на основі мимічних форм змішування.

Певні мимічні форми змішування були свідомо виключені з фінального набору через можливі труднощі у практичному використанні. Зокрема, моргання та напрямки погляду не застосовувалися, оскільки мають потенціал викликати ненавмисні активації. Моргання є рефлекторною дією, що ускладнює його використання як надійного тригера, а повне закриття очей знижує візуальний контроль над інтерфейсом, погіршуючи користувацький досвід. Водночас система може бути налаштована на використання одностороннього закриття ока у поєднанні з іншими виразами, як-от посмішка — подібна комбінація застосовувалась у попередньому дослідженні [63].

Напрямок погляду (*gaze direction*) не реалізовувався у межах цього експерименту, проте може бути ефективним у специфічних сценаріях, зокрема для зміни ракурсу під час навігації у тривимірному середовищі — наприклад, коли обертання камери відбувається відповідно до напрямку погляду користувача.

Вираз обличчя	Формула виразу обличчя на основі мимічних форм змішування ($E_{expression}$)
Brows Up	$F(browInnerUp) \wedge F(browOuterUpLeft) \wedge F(browOuterUpRight)$

Brows Down	$F(\text{browDownLeft}) \wedge F(\text{browDownRight})$
Smile	$F(\text{mouthSmileLeft}) \wedge F(\text{mouthSmileRight})$
Smile Left	$F(\text{mouthSmileLeft}) \wedge \neg F(\text{mouthSmileRight})$
Smile Right	$\neg F(\text{mouthSmileLeft}) \wedge F(\text{mouthSmileRight})$
Mouth Pucker	$F(\text{mouthPucker})$
Jaw Open	$F(\text{jawOpen})$

Табл. 4.3. Визначення виразів обличчя на основі мимічних форм змішування

Застосування зазначених форм змішування дозволяє реалізувати стабільну та інтуїтивну систему безконтактної взаємодії, орієнтовану на зниження ймовірності помилкових активацій і забезпечення зручності для кінцевого користувача.

4.5.2 Огляд системи та налаштування експерименту

Розроблена система для оцінювання містить інтерактивну сітку, яка використовується для аналізу здатності користувачів виконувати завдання з підбору кольорів і чисел за допомогою взаємодії на основі виразів обличчя. Налаштування експерименту було спрямоване на перевірку ефективності та зручності застосування запропонованого підходу до взаємодії.

Інтерфейс користувача. Графічний інтерфейс системи представлений у вигляді сітки, що складається з комірок, частина з яких визначена як інтерактивні. Інтерактивні комірки містять колір (один із чотирьох доступних) та число з діапазону [0–9]. Над кожною інтерактивною коміркою розміщено круглу мітку, яка відображає цільові значення — колір та число, яких користувач має досягти в комірці. Користувач може переміщатися між інтерактивними комірками сітки та змінювати їхні властивості за допомогою виразів обличчя, зіставляючи значення комірки з відповідною цільовою міткою.

Ціль завдання. Користувач розпочинає виконання завдання з верхнього лівого кута сітки та послідовно переміщується між інтерактивними комірками. Для кожної комірки необхідно встановити такі значення кольору та числа, які збігаються з цільовими параметрами, вказаними у круглій мітці. Завдання вважається завершеним після того, як усі інтерактивні комірки набудуть відповідних цільових значень.

Ключові елементи інтерфейсу. На Рис. 4.4 зображено знімок екрана інтерфейсу користувача розробленої системи.

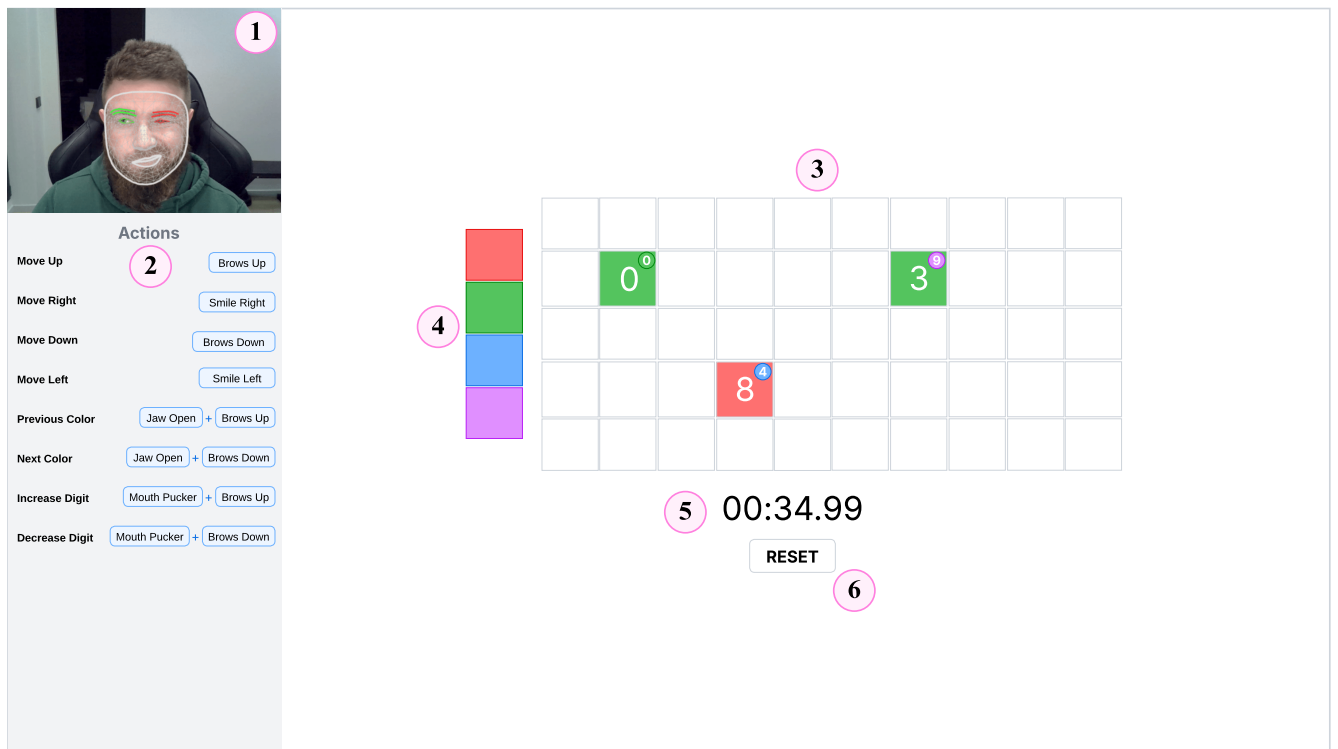


Рис. 4.4. Інтерфейс користувача розробленої системи

На рисунку позначено такі основні елементи:

1. *Відеопотік* — відображає поточне зображення з веб-камери користувача разом із мітками, що позначають виявлені ключові точки обличчя в реальному часі.

2. *Конфігурація дій* — містить перелік доступних користувацьких дій і відповідних їм мімічних виразів.
3. *Інтерактивна сітка* — основне поле для виконання завдання, яке складається з інтерактивних комірок із цільовими параметрами.
4. *Палітра кольорів* — містить набір із чотирьох доступних кольорів, які можна обирати для зміни кольору комірок.
5. *Секундомір* — відображає час виконання завдання, починаючи з моменту першої взаємодії; після завершення завдання фіксується загальний час.
6. *Кнопка скидання* — скидає поточний стан сітки та ініціює новий сеанс взаємодії.

4.5.3 Визначення користувацьких дій

На основі обраних виразів обличчя, описаних у попередній секції, було сформовано набір користувацьких дій, що забезпечують навігацію інтерактивною сіткою, зміну кольору та числового значення в інтерактивних комірках. Відповідність між кожною дією та виразом обличчя подано в Табл. 4.4.

Користувацька дія	Формула виразу обличчя
Переміститися на одну клітинку вгору (Move Up)	$E_{Brows\ Up}$
Переміститися на одну клітинку вниз (Move Down)	$E_{Brows\ Down}$
Переміститися на одну клітинку вліво (Move Left)	$E_{Smile\ Left}$
Переміститися на одну клітинку вправо (Move Right)	$E_{Smile\ Right}$
Вибрати попередній колір (Select Previous Color)	$E_{Jaw\ Open} \wedge E_{Brows\ Up}$

Вибрати наступний колір (Select Next Color)	$E_{Jaw\ Open} \wedge E_{Brows\ Down}$
Зменшити число на одиницю (Decrease Number)	$E_{Mouth\ Pucker} \wedge E_{Brows\ Down}$
Збільшити число на одиницю (Increase Number)	$E_{Mouth\ Pucker} \wedge E_{Brows\ Up}$

Табл. 4.4. Визначення користувацьких дій на основі обраних виразів обличчя

На Рис. 4.5 наведено ілюстрації виразів обличчя, що відповідають кожній з наведених дій у Табл. 4.4.

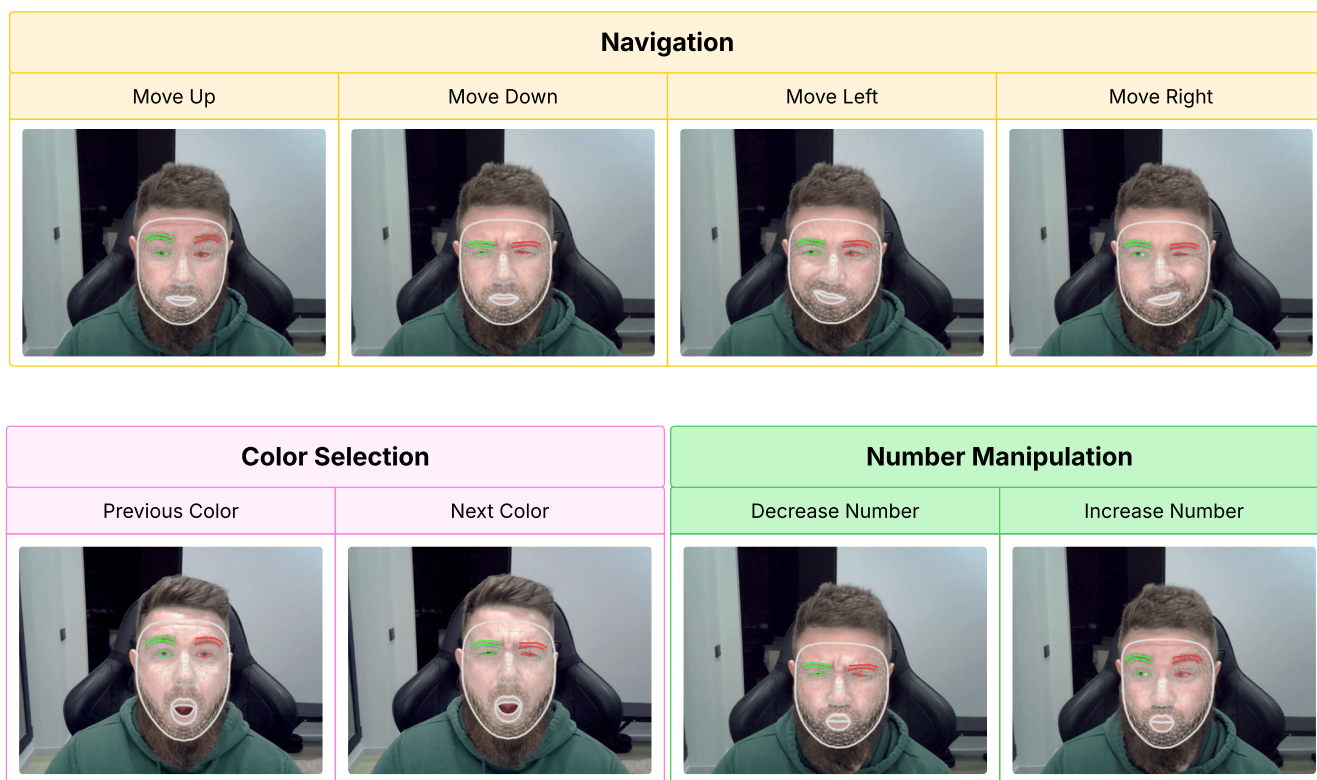


Рис. 4.5. Ілюстрація виразів обличчя, які відповідають кожній з доступних дій користувача

4.5.4 Процес виконання користувацьких дій

Механізм виконання користувацьких дій на основі виразів обличчя побудований таким чином, що кожна дія ініціюється негайно після розпізнавання відповідного виразу. Дія виконується лише один раз, навіть якщо вираз залишається активним. Для застосування наступної дії користувач повинен повернути обличчя у нейтральний стан, що запобігає повторній активації тієї самої команди. Таким чином, після завершення дії система переходить у режим очікування повернення обличчя до нейтрального положення, перш ніж розпочати розпізнавання нової команди. З метою спрощення діаграми, етап повернення обличчя до нейтрального стану не відображено на Рис. 4.6.

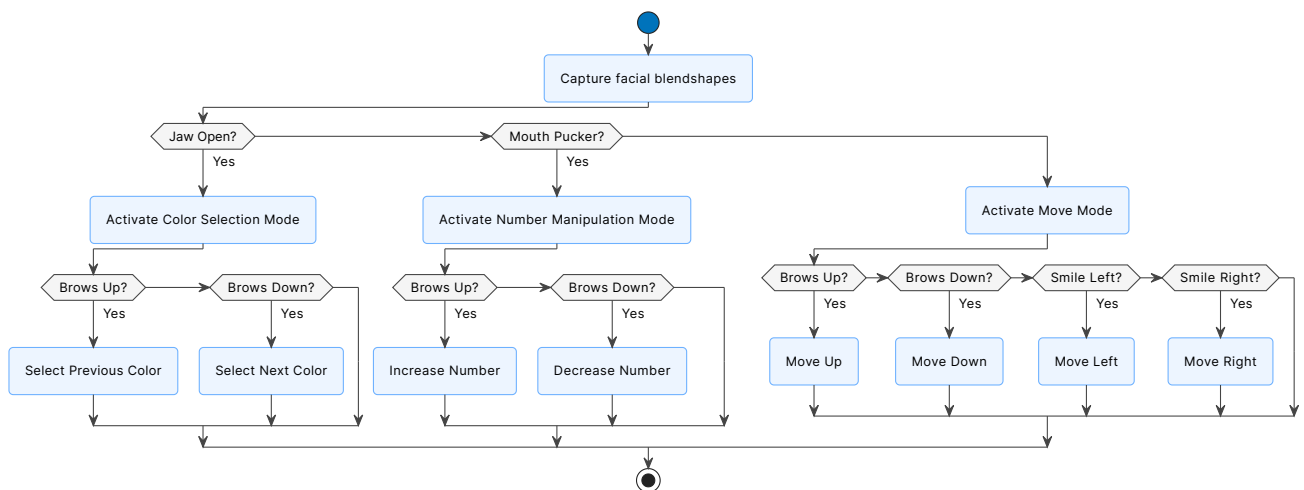


Рис. 4.6. Діаграма процесу розпізнавання дій користувача з використанням виразів обличчя

Дії, що потребують поєднання двох виразів. Окремі дії, зокрема зміна кольору або числового значення, вимагають одночасного відтворення двох мімічних виразів. У таких випадках один із виразів виконує роль активатора режиму взаємодії:

- вираз «Jaw Open» активує режим вибору кольору;
- вираз «Mouth Pucker» — режим зміни числа.

Система перебуває в обраному режимі доти, доки користувач підтримує активуючий вираз. У межах активного режиму користувач може змінювати значення, піднімаючи або опускаючи брови. Для повторного виконання дії (наприклад, збільшення числа на одиницю) необхідно, щоб лише брови повернулися у нейтральне положення — активуючий вираз при цьому може залишатися активним. Така логіка дозволяє здійснювати послідовні дії без необхідності повного відновлення нейтрального стану обличчя.

4.5.5 Оцінювання ефективності взаємодії та аналіз результатів

У дослідженні взяли участь 16 осіб віком від 19 до 34 років. Кожен учасник виконував завдання зі співставлення кольорів і чисел у межах інтерактивної сітки. Завдання повторювалося п'ять разів (п'ять спроб), при цьому фіксувався час виконання для аналізу ефекту навчання.

Аналіз навчальної кривої. Аналіз динаміки часу виконання завдань на різних етапах експерименту виявив характерні зміни продуктивності користувачів. Як показано на Рис. 4.7, спостерігався чіткий ефект навчання — середній час виконання зменшувався з кожною наступною спробою, що свідчить про поступове засвоєння системи.

Однак у третій та четвертій спробах було зафіксовано незначні коливання продуктивності. Згідно з отриманими відгуками, це зумовлено двома чинниками: (1) фізичною втомою м'язів обличчя, яка ускладнювала відтворення виразів, та (2) зниженням концентрації через надмірну впевненість у власних навичках після кількох успішних спроб.

Індивідуальні особливості продуктивності. Помітна була варіативність між учасниками: зокрема, учасники 4 та 10 демонстрували стабільно вищий час виконання, що може свідчити про індивідуальні труднощі адаптації до запропонованого методу взаємодії. Натомість, учасник 16 показав найшвидші результати, що свідчить про здатність окремих користувачів швидко пристосовуватися до нового методу взаємодії.

До четвертої–п'ятої спроб більшість учасників досягли стабільного часу виконання завдання, що свідчить про можливість досягнення передбачуваного рівня продуктивності після короткого періоду навчання.

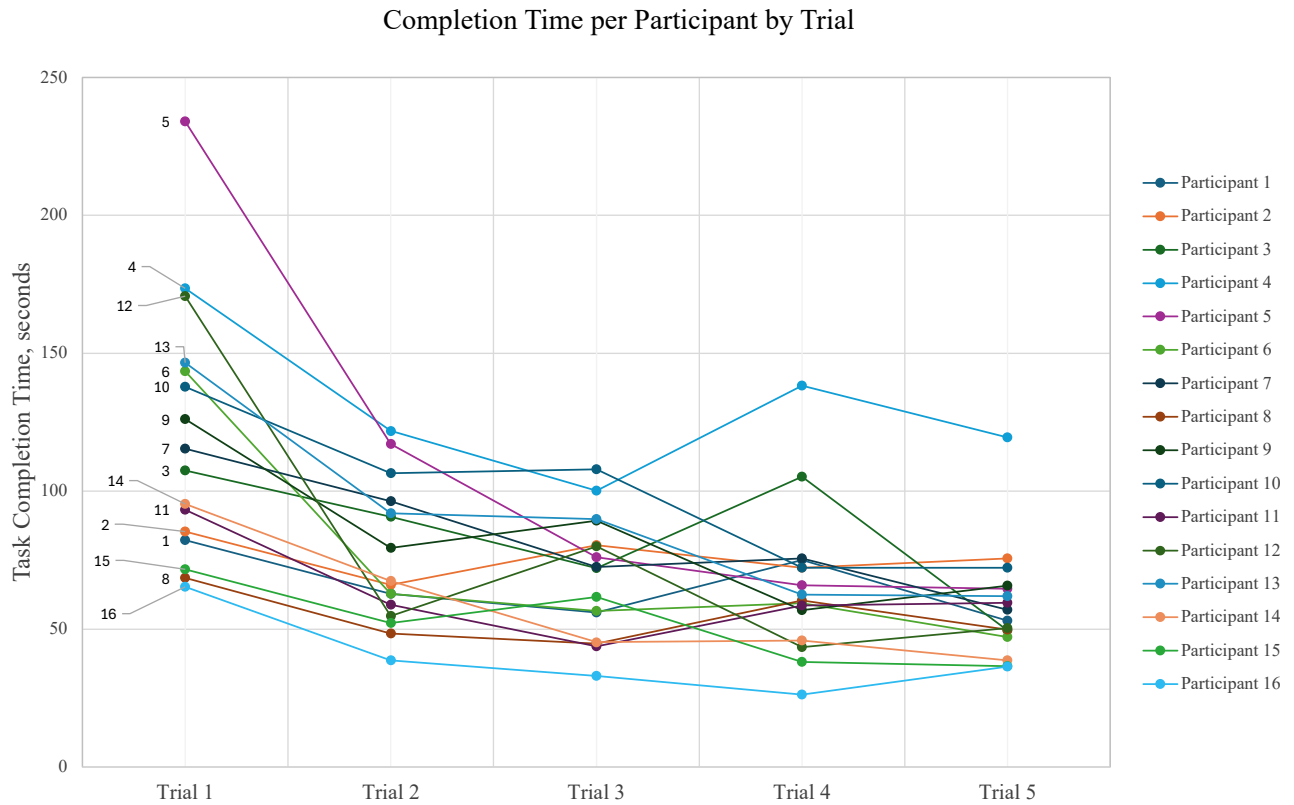


Рис. 4.7. Час виконання завдання для кожної спроби для кожного з учасників

Статистичний аналіз ефекту навчання. Для оцінки процесу навчання було проаналізовано час виконання завдань протягом п'яти послідовних спроб. Як показано на Рис. 4.8(а), середній час виконання завдань демонструє стійку тенденцію до зменшення, а зменшення стандартного відхилення (error bars) свідчить про покращення точності виконання. Загалом середній час виконання зменшився на 51 % — з ≈ 120 секунд у першій спробі до ≈ 59 секунд у п'ятій. Для перевірки статистичної значущості цього результату було проведено парний t-тест

між першою та п'ятою спробами: $t(15) = 5.99, p < 0.0001$, що підтверджує наявність вираженого ефекту навчання.

На Рис. 4.8(b) наведено розподіл часу виконання завдань серед учасників. Отримані дані свідчать про зменшення варіативності продуктивності, що підтверджує зростання стабільності виконання завдань після певного періоду тренування.

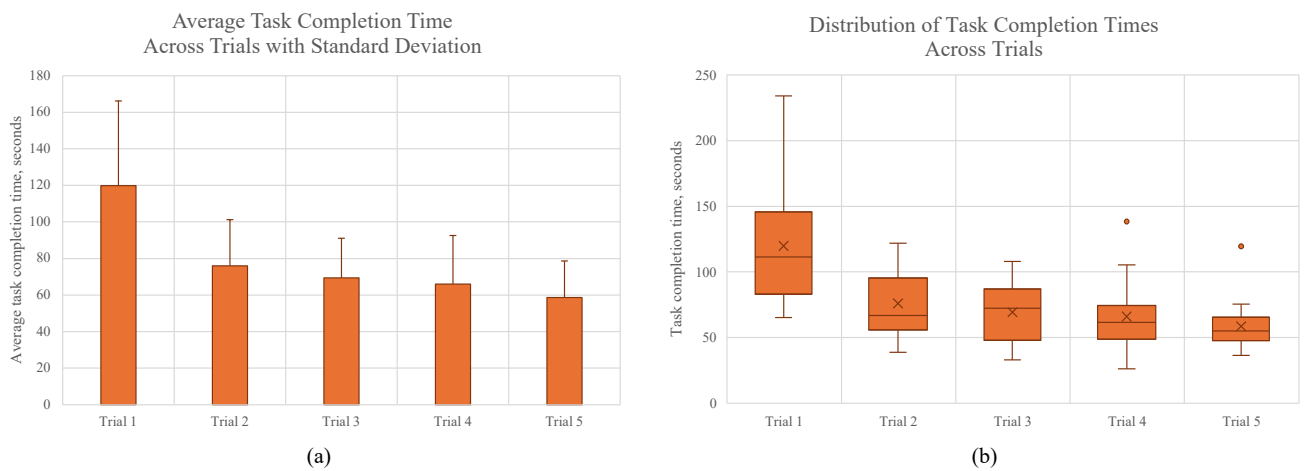


Рис. 4.8. Аналіз кривої навчання на основі часу виконання завдань: (a) — середній час виконання із стандартним відхиленням; (b) — розподіл часу виконання завдань серед учасників

Оцінка складності виразів обличчя. Учасники надали зворотний зв'язок щодо складності відтворення окремих виразів обличчя. Як показано на Рис. 4.9, різні вирази сприймалися з різним рівнем складності. Для аналізу цього аспекту було проведено хі-квадрат тест $\chi^2(7) = 18.0, p = 0.012$, який підтвердив статистично значущу різницю у сприйнятті складності відтворення різних виразів обличчя.

Найскладнішою виявилась дія «Jaw Open + Brows Down», яку зазначили як важку у виконанні 9 учасників. Це вказує на те, що комбіновані вирази, що вимагають одночасної координації декількох м'язів, вимагають більшого зусилля для відтворення.

Найменше складнощів викликали вирази «Brows Up», «Brows Down» та «Smile Left» — кожен із них був відмічений як складний у відтворенні лише одним учасником, що свідчить про їхню відносну простоту у виконанні.

Вираз «Smile Right» виявився складним у відтворенні для трьох учасників, зокрема один з них повідомив про фізіологічні обмеження в контролі правої частини обличчя, пов'язані з неврологічним станом, що ускладнює незалежне заплющення правого ока. Це підкреслює важливість забезпечення гнучкості при проєктуванні системи для врахування індивідуальних фізичних особливостей користувачів.

Отримані результати свідчать, що комбіновані вирази обличчя, такі як «Jaw Open + Brows Down», можуть потребувати оптимізації алгоритмів розпізнавання, щоб полегшити їхнє використання. Також важливо враховувати проблеми асиметричного контролю обличчя, щоб збільшити доступність та універсальність системи.

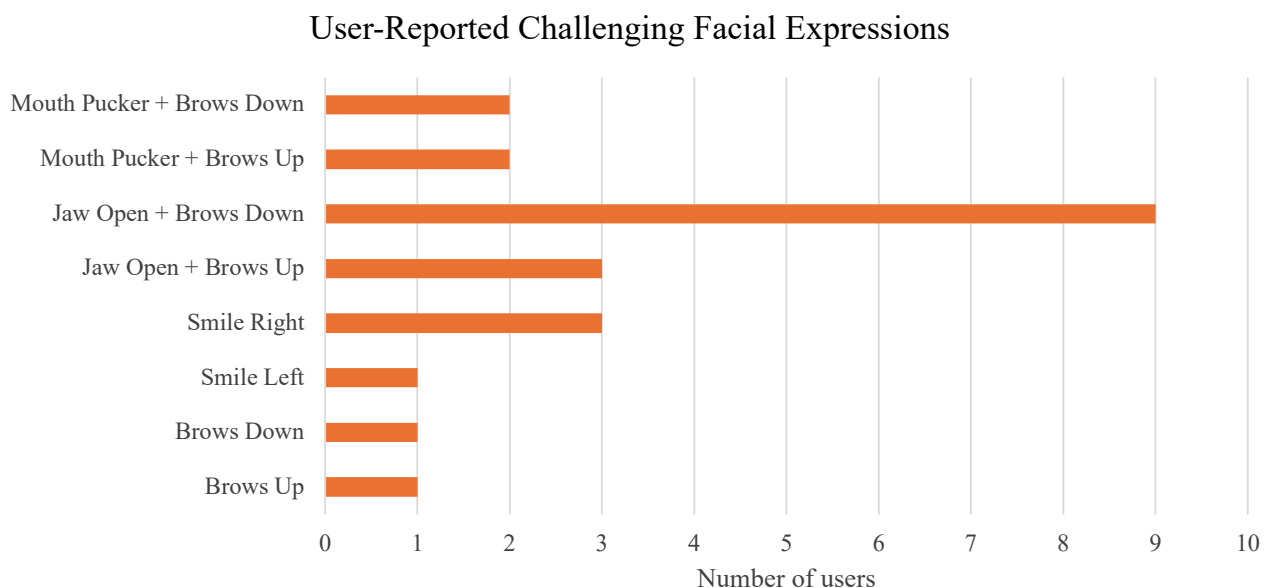


Рис. 4.9. Вирази обличчя, відмічені користувачами як складні у відтворенні (кількість учасників для кожного виразу обличчя)

Відгуки учасників та спостереження. Загалом учасники позитивно оцінили запропонований підхід до взаємодії, відзначивши його потенційну користь у таких сферах як:

- *Асистивні технології* — як безконтактний спосіб взаємодії, система може бути корисною для осіб із порушеннями рухливості верхніх кінцівок.
- *Засіб для тренування м'язів обличчя* — учасники вказали на можливість адаптації системи до реабілітаційних програм або фізіотерапевтичних застосувань.
- *Допоміжні гарячі клавіші* — деякі учасники запропонували використовувати вирази обличчя як швидкі команди (shortcuts), які можуть доповнювати традиційні методи введення.

Отримані результати свідчать про високий потенціал запропонованого підходу, однак виявлені індивідуальні відмінності у складності виконання окремих дій вказують на доцільність подальшого вдосконалення системи з урахуванням потреб користувачів.

4.5.6 Обговорення

Проведене дослідження продемонструвало зручність використання та ефективність засвоєння запропонованого підходу до взаємодії на основі виразів обличчя. Результати підтверджують наявність вираженого ефекту навчання — учасники поступово зменшували час виконання завдань у наступних спробах, що узгоджується з принципами зручності використання, зокрема з положеннями стандарту ISO 9241-11, який підкреслює важливість врахування навчального ефекту в системах людино-комп'ютерної взаємодії.

Однак у третій та четвертій спробах деякі учасники демонстрували тимчасове зниження продуктивності. За їх словами, причинами були м'язова втома та зниження концентрації через «надмірну впевненість» після кількох успішних виконань.

Ці результати свідчать, що хоча запропонована система є інтуїтивно зрозумілою та відносно простою для освоєння, фактори фізичної втоми та когнітивного навантаження необхідно враховувати при проєктуванні довготривалих або інтенсивних сценаріїв взаємодії.

Переваги. До основних переваг системи можна віднести її доступність — для роботи використовується лише стандартна веб-камера, що забезпечує низький поріг входу та не потребує спеціалізованого обладнання. Це відповідає принципам інклюзивного дизайну.

Іншою перевагою є гнучкість системи: вона дозволяє використовувати будь-які комбінації мімічних форм змішування для визначення довільних користувацьких дій, що відкриває широкі можливості для адаптації системи під конкретні потреби користувачів та забезпечує великий простір для індивідуальних налаштувань.

Учасники експерименту також запропонували декілька перспективних напрямків застосування системи: асистивні технології — для допомоги людям із порушеною рухливістю верхніх кінцівок, засіб для тренування м'язів обличчя, а також як додатковий засіб взаємодії у поєднанні із традиційними методами введення.

Обмеження та виклики. Попри виявлені переваги, дослідження також виявило низку важливих викликів. Одним із ключових обмежень є варіативність точності розпізнавання, зумовлена індивідуальними анатомічними особливостями обличчя та відмінностями у характері мімічних рухів. Це призводить до того, що система не завжди однаково ефективно інтерпретує дії різних користувачів. Результати експерименту підкреслюють необхідність впровадження механізму персоналізованого калібрування порогових значень для підвищення точності розпізнавання.

Ще одним викликом є вплив зовнішніх умов, зокрема освітлення, яке може суттєво впливати на якість розпізнавання виразів обличчя. Це вказує на потребу у

більш стійких алгоритмах, здатних адаптуватися до змін довкілля без втрати точності.

Дослідження також виявило труднощі у виконанні складних, багатокомпонентних виразів обличчя, зокрема таких, що потребують одночасної координації кількох м'язів. Наприклад, вираз «Jaw Open + Brows Down» виявився складним для більшості учасників, тоді як «Brows Up» та «Brows Down» були сприйняті як найбільш прості у виконанні.

Ще одним обмеженням є фізична втома м'язів обличчя при тривалому використанні системи. За відгуками деяких учасників, постійна активація мимічних дій викликала втому, що може знижувати ефективність взаємодії під час тривалих сеансів. Це є типовим викликом для систем, які базуються на мимічному введенні.

Загалом, отримані результати узгоджуються з існуючими дослідженнями та підкреслюють важливість персоналізації взаємодії, що дозволяє системі краще адаптуватися до індивідуальних особливостей користувачів.

Перспективи подальших досліджень. Подальші дослідження можуть бути зосереджені на усуненні виявлених обмежень системи. Зокрема, впровадження механізму калібрування для персоналізованого визначення порогових значень має потенціал суттєво підвищити точність розпізнавання, ефективність роботи та загальну доступність системи для ширшого кола користувачів. Впровадження такого механізму узгоджується з сучасними підходами до проектування адаптивних інтерфейсів, орієнтованих на користувача.

Серед інших перспективних напрямів подальших досліджень — підвищення стійкості системи до змін зовнішніх умов, зокрема освітлення, а також оптимізація складних мимічних дій з метою покращення їх розпізнавання та зменшення когнітивного й фізичного навантаження на користувачів.

Оскільки запропонований підхід демонструє потенціал, він потребує подальшої перевірки на більшій вибірці користувачів. Масштабніше тестування та

аналіз ефективності підходу в різних контекстах використання дозволить глибше оцінити його надійність, адаптивність і практичну цінність.

Крім того, учасники експерименту запропонували низку потенційних напрямів застосування системи. Зокрема, інтеграція цього методу у гейміфіковані інструменти для тренування м'язів обличчя або його використання в рамках асистивних технологій для покращення доступності цифрових рішень для людей із порушеннями рухливості верхніх кінцівок становить перспективне поле для подальших досліджень та розробок.

4.5.7 Висновки

В цьому розділі дисертаційної роботи було проведено дослідження підходу до взаємодії на основі виразів обличчя як безконтактної альтернативи для виконання точних користувацьких дій. Завдяки використанню стандартного обладнання, зокрема звичайної веб-камери, запропонована система забезпечує низьковартісне та доступне рішення, придатне для різноманітних сценаріїв використання, включно з асистивними технологіями та допоміжними засобами введення.

Оцінювання системи продемонструвало наявність ефекту навчання: середній час виконання завдань зменшився на 51 % між першою та п'ятою спробами $t(15) = 5.99, p < 0.0001$. Учасники також позитивно оцінили зручність використання, зокрема простоту виконання базових дій, таких як «Brows Up» та «Smile Left», на відміну від складних багатокomпонентних виразів, як-от «Jaw Open + Brows Down», які викликали труднощі у більшості учасників ($\chi^2(7) = 18.0, p = 0.012$). Результати дослідження засвідчили інтуїтивність запропонованої системи та її здатність до швидкого засвоєння користувачами, водночас виявивши потребу в її подальшому вдосконаленні.

Зворотний зв'язок від учасників підкреслив потенційну практичну цінність системи, зокрема її можливість бути корисною для людей із порушеннями рухливості верхніх кінцівок. Крім того, учасники відзначили можливість адаптації

системи для гейміфікованих рішень, зокрема як інструмент для тренування м'язів обличчя.

Попри позитивні результати, дослідження виявило низку викликів. Серед них — варіативність точності розпізнавання, зумовлена анатомічними відмінностями користувачів, фізична втома при тривалому використанні, складність виконання двоступінчастих дій (наприклад, «Jaw Open + Brows Down») та чутливість системи до зовнішніх факторів, таких як освітлення. Ці обмеження вказують на потребу подальшого вдосконалення системи, зокрема впровадження механізмів калібрування та адаптації.

У підсумку, запропонований підхід до взаємодії продемонстрував високий потенціал як доступна та гнучка альтернатива для безконтактного керування. Це дослідження робить внесок у розвиток альтернативних підходів взаємодії та підтримує розробку інклюзивних та адаптивних інтерфейсів людино-комп'ютерної взаємодії.

ВИСНОВКИ

Проведений аналіз наукових джерел виявив обмеженість традиційних підходів до аналізу поведінки користувачів, що базуються на кліках, журналах подій та записі сесій без врахування зорової уваги. Виявлені дослідницькі прогалини підтвердили потребу у створенні інтегрованих рішень, які поєднують можливості реєстрації сесій із відстеженням погляду, а також у розробці нових мультимодальних підходів до безконтактної взаємодії на основі погляду, жестів рук і виразів обличчя. Це стало підґрунтям для формулювання і реалізації наукової задачі дослідження.

Розроблена система запису та відтворення веб-сесій із синхронним збором даних погляду продемонструвала високу оцінку користувачів щодо зручності використання (середній бал — 4,5 із 5), забезпечуючи нові можливості для комплексного аналізу поведінки без потреби в зміні коду веб-застосунків. Оцінювання також виявило потребу у вдосконаленні точності відстеження погляду як одного з ключових компонентів системи. Застосування алгоритмів машинного навчання для автоматизованого аналізу поведінкових даних є перспективним напрямом подальших досліджень, що може суттєво розширити аналітичні можливості системи.

Запропонований веб-орієнтований підхід до маніпулювання об'єктами, що поєднує використання погляду для вибору об'єкта та жестів рук для виконання трансформацій, продемонстрував високу ефективність у складних сценаріях взаємодії, зокрема під час одночасного масштабування, обертання та переміщення об'єктів. Результати експериментального дослідження засвідчили, що в таких завданнях запропонований підхід перевершує традиційне введення за допомогою миші, тоді як для простих дій ефективнішим залишається класичний спосіб взаємодії. Використання лише стандартного обладнання значно підвищує доступність запропонованого рішення. Основними обмеженнями залишаються

точність відстеження погляду та стабільність розпізнавання жестів, що визначає напрями подальшого вдосконалення технології.

Запропонований підхід до безконтактної взаємодії на основі розпізнавання виразів обличчя забезпечив доступне та інтуїтивно зрозуміле керування діями за допомогою стандартної веб-камери, продемонструвавши придатність для широкого кола користувачів. Експериментальне дослідження виявило наявність ефекту навчання: середній час виконання завдань зменшився на 51 % між першою та п'ятою спробами, що свідчить про швидку адаптацію користувачів до запропонованого способу взаємодії. Аналіз результатів показав легкість у використанні простих виразів (наприклад, «Brows Up» або «Smile Left») і виявив труднощі з виконанням складних комбінацій (як-от «Jaw Open + Brows Down»), що визначає потенційні напрями вдосконалення системи. Зворотний зв'язок від учасників підтвердив практичну цінність підходу, зокрема для асистивних технологій і гейміфікованих рішень, таких як тренування м'язів обличчя. Водночас ідентифіковано низку обмежень, пов'язаних із фізичною втомою, анатомічними особливостями користувачів та чутливістю розпізнавання до зовнішніх умов, зокрема освітлення. Це підкреслює необхідність впровадження механізмів адаптивного калібрування та оптимізації алгоритмів розпізнавання.

У дисертаційній роботі вперше запропоновано та реалізовано кросплатформний підхід до поєднання запису веб-сесій із синхронізованим відображенням зорової уваги користувача без потреби модифікації коду веб-сайтів. Розроблено систему безконтактного маніпулювання цифровими об'єктами на основі поєднання погляду та жестів рук. Також вперше реалізовано підхід до безконтактної взаємодії у веб-середовищі за допомогою виразів обличчя на основі налаштовуваних комбінацій мімічних форм змішування. Крім того, удосконалено підходи до збору поведінкових даних шляхом інтеграції інформації про події взаємодії, зміни у структурі веб-сторінки та зорову увагу користувача, що створює передумови для подальшого автоматизованого аналізу взаємодії з інтерфейсами.

Перспективи подальших досліджень включають:

- Впровадження моделей машинного навчання для автоматичного аналізу поведінки користувачів, що відкриває можливості для виявлення аномалій, передбачення намірів користувача, а також динамічної адаптації інтерфейсу до контексту дій.
- Покращення точності розпізнавання погляду, жестів і виразів обличчя, що дозволить зменшити кількість помилкових спрацювань та забезпечити стабільнішу взаємодію навіть в умовах змінного освітлення чи різних анатомічних особливостей користувачів.
- Адаптацію систем до індивідуальних особливостей користувачів, зокрема персоналізацію моделей під конкретні патерни поведінки, звички або фізіологічні характеристики (наприклад, асиметрію обличчя чи амплітуду жестів).
- Інтеграцію таких методів у промислові, медичні та освітні середовища, де безконтактна взаємодія є особливо актуальною — наприклад, у стерильних умовах операційної, для навчання людей з порушеннями моторики або у виробничих середовищах з обмеженим доступом до традиційних засобів введення.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- [1] A. Dix, J. Finlay, G. D. Abowd, and R. Beale, *Human-Computer Interaction*, 3rd ed. Pearson Education, 2004.
- [2] M. Schiessl, S. Duda, A. Thölke, and R. Fischer, ‘Eye tracking and its application in usability and media research’, *MMI Interakt.*, vol. 6, 2003, [Online]. Available: <https://api.semanticscholar.org/CorpusID:501543>
- [3] M. Pretorius, A. P. Calitz, and D. van Greunen, ‘The added value of eye tracking in the usability evaluation of a network management tool’, 2005. [Online]. Available: <https://api.semanticscholar.org/CorpusID:9747314>
- [4] S. Murali, T. Walber, C. Schaefer, and S. H. Lim, ‘Enriching Verbal Feedback from Usability Testing: Automatic Linking of Thinking-Aloud Recordings and Stimulus using Eye Tracking and Mouse Data’, *ArXiv*, vol. abs/2307.05171, 2023, [Online]. Available: <https://api.semanticscholar.org/CorpusID:259766058>
- [5] ‘rrweb’. [Online]. Available: <https://github.com/rrweb-io/rrweb>
- [6] A. Poole and L. J. Ball, ‘Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects’, in *Encyclopedia of Human-Computer Interaction*, C. Ghaoui, Ed., Idea Group, 2005, pp. 211–219.
- [7] K. Rayner, ‘Eye movements in reading and information processing: 20 years of research.’, *Psychological Bulletin*, vol. 124, no. 3, pp. 372–422, 1998, doi: 10.1037/0033-2909.124.3.372.
- [8] C. Ehmke and S. M. Wilson, ‘Identifying web usability problems from eye-tracking data’, in *British Computer Society Conference on Human-Computer Interaction*, 2007. [Online]. Available: <https://api.semanticscholar.org/CorpusID:1427673>
- [9] J. Wang, P. Antonenko, M. Celepkolu, Y. Jimenez, E. Fieldman, and A. Fieldman, ‘Exploring Relationships Between Eye Tracking and Traditional Usability Testing Data’, *International Journal of Human–Computer Interaction*, vol. 35, no. 6, pp. 483–494, Apr. 2019, doi: 10.1080/10447318.2018.1464776.

[10] O. Gordieiev, V. Kharchenko, O. Illiashenko, O. Morozova, and M. Gasanov, 'Concept of Using Eye Tracking Technology to Assess and Ensure Cybersecurity, Functional Safety and Usability', *IJSSE*, vol. 11, no. 4, pp. 361–367, Aug. 2021, doi: 10.18280/ijsse.110409.

[11] P. A. Punde, M. E. Jadhav, and R. R. Manza, 'A study of eye tracking technology and its applications', in *2017 1st International Conference on Intelligent Systems and Information Management (ICISIM)*, IEEE, Oct. 2017, pp. 86–90. doi: 10.1109/ICISIM.2017.8122153.

[12] V. Cantoni and M. Porta, 'Eye tracking as a computer input and interaction method', in *Proceedings of the 15th International Conference on Computer Systems and Technologies*, ACM, Jun. 2014, pp. 1–12. doi: 10.1145/2659532.2659592.

[13] J. Š. Novák, J. Masner, P. Benda, P. Šimek, and V. Merunka, 'Eye Tracking, Usability, and User Experience: A Systematic Review', *International Journal of Human-Computer Interaction*, 2023, doi: 10.1080/10447318.2023.2221600.

[14] P. Tarnowski, M. Kołodziej, A. Majkowski, and R. J. Rak, 'Eye-Tracking Analysis for Emotion Recognition', *Computational Intelligence and Neuroscience*, vol. 2020, pp. 1–13, Sep. 2020, doi: 10.1155/2020/2909267.

[15] L. A. Casado-Aranda, J. Sánchez-Fernández, and J. Á. Ibáñez-Zapata, 'Evaluating Communication Effectiveness Through Eye Tracking: Benefits, State of the Art, and Unresolved Questions', *International Journal of Business Communication*, vol. 60, no. 1, pp. 24–61, Jan. 2023, doi: 10.1177/2329488419893746.

[16] A. Papoutsaki, P. Sangkloy, J. Laskey, N. Daskalova, J. Huang, and J. Hays, 'WebGazer: Scalable Webcam Eye Tracking Using User Interactions', presented at the International Joint Conference on Artificial Intelligence, Jul. 2016. Accessed: Oct. 10, 2024. [Online]. Available: <https://www.semanticscholar.org/paper/WebGazer%3A-Scalable-Webcam-Eye-Tracking-Using-User-Papoutsaki-Sangkloy/73fc8e9b1faf45855ccee197f094ca3c05afe1c>

[17] A. Voßkühler, V. Nordmeier, L. Kuchinke, and A. M. Jacobs, ‘OGAMA (Open Gaze and Mouse Analyzer): Open-source software designed to analyze eye and mouse movements in slideshow study designs’, *Behavior Research Methods*, vol. 40, no. 4, pp. 1150–1162, Nov. 2008, doi: 10.3758/BRM.40.4.1150.

[18] H. Sogo, ‘GazeParser: an open-source and multiplatform library for low-cost eye tracking and analysis’, *Behavior Research Methods*, vol. 45, no. 3, pp. 684–695, Sep. 2013, doi: 10.3758/s13428-012-0286-x.

[19] A. Steffan *et al.*, ‘Validation of an open source, remote web-based eye-tracking method (WebGazer) for research in early childhood’, *Infancy*, vol. 29, no. 1, pp. 31–55, Jan. 2024, doi: 10.1111/infa.12564.

[20] S. Hutt, A. Wong, A. Papoutsaki, R. S. Baker, J. I. Gold, and C. Mills, ‘Webcam-based eye tracking to detect mind wandering and comprehension errors’, *Behavior Research Methods*, vol. 56, no. 1, pp. 1–17, Jan. 2024, doi: 10.3758/s13428-022-02040-x.

[21] ‘GazeRecorder’. [Online]. Available: <https://gazerecorder.com>

[22] T. R. Shaffer, J. L. Wise, B. M. Walters, S. C. Müller, M. Falcone, and B. Sharif, ‘iTrace: enabling eye tracking on software artifacts within the IDE to support software engineering tasks’, in *Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering*, ACM, Aug. 2015, pp. 954–957. doi: 10.1145/2786805.2803188.

[23] T. Vuletic, A. Duffy, L. Hay, C. McTeague, G. Campbell, and M. Greal, ‘Systematic literature review of hand gestures used in human computer interaction interfaces’, *International Journal of Human-Computer Studies*, vol. 129, pp. 74–94, Sep. 2019, doi: 10.1016/j.ijhcs.2019.03.011.

[24] D. Gavgiotaki, S. Ntoa, G. Margetis, K. C. Apostolakis, and C. Stephanidis, ‘Gesture-based Interaction for AR Systems: A Short Review’, *Proceedings of the 16th International Conference on Pervasive Technologies Related to Assistive Environments*, pp. 284–292, Jul. 2023, doi: 10.1145/3594806.3594815.

- [25] J. P. Wachs, M. Kölsch, H. Stern, and Y. Edan, ‘Vision-based hand-gesture applications’, *Commun. ACM*, vol. 54, no. 2, pp. 60–71, Feb. 2011, doi: 10.1145/1897816.1897838.
- [26] I. Virag, L. Stoicu-Tivadar, and M. Crisan-Vida, ‘Gesture-based interaction in medical interfaces’, *2016 IEEE 11th International Symposium on Applied Computational Intelligence and Informatics (SACI)*, pp. 519–523, May 2016, doi: 10.1109/SACI.2016.7507339.
- [27] A. Widmer, R. Schaer, D. Markonis, and H. Müller, ‘Gesture Interaction for Content--based Medical Image Retrieval’, *Proceedings of International Conference on Multimedia Retrieval*, pp. 503–506, Apr. 2014, doi: 10.1145/2578726.2578804.
- [28] S. Zelinskyi and Y. Boyko, ‘Hands as Pointers: A Touchless Gesture-Based Approach for Web Objects Manipulation’, presented at the XI Всеукраїнська науково-практична конференція студентів, аспірантів та молодих вчених, Kyiv, Ukraine, 2024, pp. 308–310.
- [29] F. Argelaguet and C. Andujar, ‘A survey of 3D object selection techniques for virtual environments’, *Computers & Graphics*, vol. 37, no. 3, pp. 121–136, May 2013, doi: 10.1016/j.cag.2012.12.003.
- [30] C. Ware and H. H. Mikaelian, ‘An evaluation of an eye tracker as a device for computer input²’, in *Proceedings of the SIGCHI/GI Conference on Human Factors in Computing Systems and Graphics Interface*, Toronto Ontario Canada: ACM, May 1986, pp. 183–188. doi: 10.1145/29933.275627.
- [31] D. Slambekova, R. Bailey, and J. Geigel, ‘Gaze and gesture based object manipulation in virtual worlds’, in *Proceedings of the 18th ACM symposium on Virtual reality software and technology*, Toronto Ontario Canada: ACM, Dec. 2012, pp. 203–204. doi: 10.1145/2407336.2407380.
- [32] I. Chatterjee, R. Xiao, and C. Harrison, ‘Gaze+Gesture: Expressive, Precise and Targeted Free-Space Interactions’, in *Proceedings of the 2015 ACM on International*

Conference on Multimodal Interaction, Seattle Washington USA: ACM, Nov. 2015, pp. 131–138. doi: 10.1145/2818346.2820752.

[33] M. N. Lystbæk, P. Rosenberg, K. Pfeuffer, J. E. Grønbaek, and H. Gellersen, ‘Gaze-Hand Alignment: Combining Eye Gaze and Mid-Air Pointing for Interacting with Menus in Augmented Reality’, *Proc. ACM Hum.-Comput. Interact.*, vol. 6, no. ETRA, pp. 1–18, May 2022, doi: 10.1145/3530886.

[34] K. Ryu, J.-J. Lee, and J.-M. Park, ‘GG Interaction: a gaze–grasp pose interaction for 3D virtual object selection’, *J Multimodal User Interfaces*, vol. 13, no. 4, pp. 383–393, Dec. 2019, doi: 10.1007/s12193-019-00305-y.

[35] J. Hales, ‘Interacting with Objects in the Environment by Gaze and Hand Gestures’, 2013. Accessed: Oct. 10, 2024. [Online]. Available: <https://www.semanticscholar.org/paper/Interacting-with-Objects-in-the-Environment-by-Gaze-Hales/ecbbff4c31df702e95a7847f5c68bcbeb7e98dac>

[36] N. L. Reddy, R. Murugeswari, Md. Imran, N. Subhash, N. V. K. Reddy, and N. B. Adarsh, ‘Virtual Mouse using Hand and Eye Gestures’, in *2023 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI)*, Chennai, India: IEEE, Dec. 2023, pp. 1–5. doi: 10.1109/ICDSAAI59313.2023.10452550.

[37] K. Pfeuffer, B. Mayer, D. Mardanbegi, and H. Gellersen, ‘Gaze + pinch interaction in virtual reality’, in *Proceedings of the 5th Symposium on Spatial User Interaction*, Brighton United Kingdom: ACM, Oct. 2017, pp. 99–108. doi: 10.1145/3131277.3132180.

[38] K. Pfeuffer, ‘Design Principles & Issues for Gaze and Pinch Interaction’, Jan. 18, 2024, *arXiv*: arXiv:2401.10948. Accessed: Oct. 10, 2024. [Online]. Available: <http://arxiv.org/abs/2401.10948>

[39] B. Anand, B. B. Navathe, S. Velusamy, H. Kannan, A. Sharma, and V. Gopalakrishnan, ‘Beyond touch: Natural interactions using facial expressions’, *2012 IEEE Consumer Communications and Networking Conference (CCNC)*, pp. 255–259, Jan. 2012, doi: 10.1109/CCNC.2012.6181097.

[40] Q. Chen, M. D. Cordea, E. M. Petriu, A. R. V. Koczy, and T. E. Whalen, ‘Human Computer interaction for smart environment applications using hand gestures and facial expressions’, *IJAMC*, vol. 3, no. 1/2, p. 95, 2009, doi: 10.1504/IJAMC.2009.026854.

[41] S. Ren, ‘Computer Vision for Facial Analysis Using Human-Computer Interaction Models’, *J. Inter. Net.*, vol. 22, no. Supp03, p. 2144005, May 2022, doi: 10.1142/S0219265921440059.

[42] Y. Shi, Z. Zhang, K. Huang, W. Ma, and S. Tu, ‘Human-computer interaction based on face feature localization’, *Journal of Visual Communication and Image Representation*, vol. 70, p. 102740, Jul. 2020, doi: 10.1016/j.jvcir.2019.102740.

[43] M. K. Chowdary, T. N. Nguyen, and D. J. Hemanth, ‘Deep learning-based facial emotion recognition for human–computer interaction applications’, *Neural Comput & Applic*, vol. 35, no. 32, pp. 23311–23328, Nov. 2023, doi: 10.1007/s00521-021-06012-8.

[44] A. Samara, L. Galway, R. Bond, and H. Wang, ‘Affective state detection via facial expression analysis within a human–computer interaction context’, *J Ambient Intell Human Comput*, vol. 10, no. 6, pp. 2175–2184, Jun. 2019, doi: 10.1007/s12652-017-0636-8.

[45] A. Condegni, W. Wang, and R. Li, ‘A Digital Human System with Realistic Facial Expressions for Friendly Human-Machine Interaction’, D.-S. Huang, P. Premaratne, B. Jin, B. Qu, K.-H. Jo, and A. Hussain, Eds., in *Lecture Notes in Computer Science*, vol. 14086. Singapore: Springer Nature Singapore, 2023, pp. 787–798. doi: 10.1007/978-981-99-4755-3_68.

[46] F. French *et al.*, ‘Expressive Interaction Design Using Facial Muscles as Controllers’, *MTI*, vol. 6, no. 9, p. 78, Sep. 2022, doi: 10.3390/mti6090078.

[47] H. Gjoreski *et al.*, ‘OCOsense Glasses – Monitoring Facial Gestures and Expressions for Augmented Human-Computer Interaction: OCOsense Glasses for Monitoring Facial Gestures and Expressions’, *Extended Abstracts of the 2023 CHI*

Conference on Human Factors in Computing Systems, pp. 1–4, Apr. 2023, doi: 10.1145/3544549.3583918.

[48] H. Mosquera, H. Loaiza, S. Nope, and A. Restrepo, ‘Identifying facial gestures to emulate a mouse: Control application in a web browser’, *2016 XXI Symposium on Signal Processing, Images and Artificial Vision (STSIVA)*, pp. 1–6, Aug. 2016, doi: 10.1109/STSIVA.2016.7743345.

[49] A. Dongre, R. Pinto, A. Patkar, and M. Lopes, ‘Computer Cursor Control Using Eye and Face Gestures’, *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pp. 1–6, Jul. 2020, doi: 10.1109/ICCCNT49239.2020.9225311.

[50] A. Kumar Raja, C. Sugandhi, G. Nymish, N. Sai Havish, and M. Rashmi, ‘Face Gesture Based Virtual Mouse Using Mediapipe’, *2023 IEEE 8th International Conference for Convergence in Technology (I2CT)*, pp. 1–6, Apr. 2023, doi: 10.1109/I2CT57861.2023.10126453.

[51] D. Rozado, J. Niu, and M. Lochner, ‘Fast Human-Computer Interaction by Combining Gaze Pointing and Face Gestures’, *ACM Trans. Access. Comput.*, vol. 10, no. 3, pp. 1–18, Aug. 2017, doi: 10.1145/3075301.

[52] A. Matos, V. Filipe, and P. Couto, ‘Human-Computer Interaction Based on Facial Expression Recognition: A Case Study in Degenerative Neuromuscular Disease’, *Proceedings of the 7th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion*, pp. 8–12, Dec. 2016, doi: 10.1145/3019943.3019945.

[53] L. R. Kalabarige, K. A. Abhilash, K. A. Trivedi, and M. Dathatreya, ‘Facial Landmark-based Cursor Control and Speech-to-Text System for Paralyzed Individuals’, *2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)*, pp. 849–856, Mar. 2023, doi: 10.1109/ICSCDS56580.2023.10104936.

[54] A. Hlybovets, S. Khmel, and O. Pyechkurova, ‘Usage of Neuro-Computer Interfaces for Collecting and Storing Information About Web Resource Users’, *NRPCOMP*, vol. 6, pp. 48–56, Mar. 2024, doi: 10.18523/2617-3808.2023.6.48-56.

[55] S. Zelinskyi and Y. Boyko, ‘Integrating session recording and eye-tracking: development and evaluation of a Chrome extension for user behavior analysis’, *reks*, vol. 2024, no. 3, pp. 38–54, Aug. 2024, doi: 10.32620/reks.2024.3.03.

[56] O. Gordieiev, V. Kharchenko, D. Gordieieva, I. Kondius, and N. Lishchyna, ‘Area of Interest Based Assessment of Software Interface Usability for Human-Computer Interaction Using Eye-Tracking.’, in *IntellITSIS*, 2022, pp. 119–128. Accessed: Apr. 22, 2025. [Online]. Available: <https://ceur-ws.org/Vol-3156/paper6.pdf>

[57] S. Zelinskyi and Y. Boyko, ‘Exploring gaze-gesture interaction on the web: a comparison with mouse input for object manipulation’, *Information Technology: Computer Science, Software Engineering and Cyber Security*, 2024.

[58] S. Zelinskyi, Y. Boyko, and I. Didmanidze, ‘Evaluating the Usability of Gaze-Gesture Interaction for Object Manipulation on the Web’, presented at the International Scientific-Practical Conference: Modern Challenges and Achievements in Information and Communication Technologies - 2024, 2024.

[59] ‘MediaPipe Hand Landmarker’, Google AI for Developers. Accessed: Jun. 26, 2024. [Online]. Available: https://ai.google.dev/edge/mediapipe/solutions/vision/hand_landmarker

[60] A. Hlybovets and M. Bikchentayev, ‘Recognizing gestures of the ukrainian dactylic alphabet’, *PC&I*, vol. 68, no. 3, pp. 86–100, Jul. 2023, doi: 10.34229/1028-0979-2023-3-9.

[61] S. Zelinskyi and Boyko, Yuriy, ‘Leveraging Large Language Models for Free-Form Voice Command Manipulation of 3D Objects: A Prototype Study’, in *Proceedings of the 50th International Scientific and Practical Conference*, Bodrum, Turkey, 2024, pp. 247–250. doi: 10.52058/50.

[62] J. P. Lewis, K. Anjyo, T. Rhee, M. Zhang, F. Pighin, and Z. Deng, 'Practice and Theory of Blendshape Facial Models', 2014, *The Eurographics Association*. doi: 10.2312/EGST.20141042.

[63] S. Zelinskyi and Y. Boyko, 'Facial Blendshapes for Custom Action: An Experimental Evaluation Using Facial Expressions for Human-Computer Interaction', *MCIT*, no. 7, pp. 174–176, Dec. 2024, doi: 10.31713/MCIT.2024.050.

[64] S. Zelinskyi and Y. Boyko, 'Using Facial Expressions for Custom Actions: Development and Evaluation of a Hands-Free Interaction Method', *CSIT*, no. 4, pp. 116–125, Dec. 2024, doi: 10.31891/csit-2024-4-14.

[65] 'Face landmark detection guide', Google AI for Developers. Accessed: Sep. 06, 2024. [Online]. Available: https://ai.google.dev/edge/mediapipe/solutions/vision/face_landmarker

ДОДАТКИ

Додаток А

Список праць здобувача за темою дисертації

Статті у наукових фахових видання України:

1. Zelinskyi, S., & Boyko, Y. (2024). Integrating session recording and eye-tracking: development and evaluation of a Chrome extension for user behavior analysis. *Radioelectronic and Computer Systems*, 2024(3), 38–54. <https://doi.org/10.32620/reks.2024.3.03>
2. Zelinskyi, S., & Boyko, Y. (2024). Exploring Gaze-Gesture Interaction on the Web: A Comparison with Mouse Input for Object Manipulation. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 3, 33–42. <https://doi.org/10.32782/it/2024-3-4>
3. Zelinskyi, S., & Boyko, Y. (2024). Using Facial Expressions for Custom Actions: Development and Evaluation of a Hands-Free Interaction Method. *Computer Systems and Information Technologies*, 4, 116–125. <https://doi.org/10.31891/csit-2024-4-14>

Праці, які засвідчують апробацію матеріалів дисертації:

1. Zelinskyi, S., Boyko, Y., & Didmanidze, I. (2024). Evaluating the Usability of Gaze-Gesture Interaction for Object Manipulation on the Web. *International Scientific-Practical Conference Modern Challenges and Achievements in Information and Communication Technologies – 2024*
2. Zelinskyi, S., Boyko, Y. (2024). Hands as Pointers: A Touchless Gesture-Based Approach for Web Objects Manipulation. *XI Всеукраїнська науково-практична конференція студентів, аспірантів та молодих вчених*, 308-310.
3. Zelinskyi, S., & Boyko, Y. (2024). Facial Blendshapes for Custom Action: An Experimental Evaluation Using Facial Expressions for Human-Computer

Interaction. *Modeling, Control and Information Technologies*, 7, 174–176.
<https://doi.org/10.31713/mcit.2024.050>

4. Zelinskyi, S., & Boyko, Y. (2024). Leveraging Large Language Models for Free-Form Voice Command Manipulation of 3D Objects: A Prototype Study. *Proceedings of the 50th International Scientific and Practical Conference*, 247–250. <https://doi.org/10.52058/50>