

КВАЛІФІКАЦІЙНА РОБОТА
освітній ступінь – бакалавр

Мультимодальні системи: дослідження інтеграції різноманітних сенсорних модальностей, таких як звукова та візуальна, для можливого покращення взаємодії користувачів у мультимедійних програмах

Науковий керівник – **Афонін Андрій Олександрович**

Виконала: студентка 4-го року навчання,
Освітньої програми «Комп'ютерні науки», 122
Станіславська Катерина Євгенівна

2024

Постановка задачі

1	Дослідити модальності доступні для взаємодії з персональним комп'ютером
2	Проаналізувати характеристики та інтеграцію мультимодальних систем
3	Розробити програму для мультимодальної взаємодії користувача з комп'ютером <ul style="list-style-type: none">○ Інтегрувати розпізнавання жестів○ Інтегрувати розпізнавання голосу○ Додати звуковий зворотній зв'язок від комп'ютера○ Імплементувати додаткові застосування модальностей○ Проаналізувати отримані результати

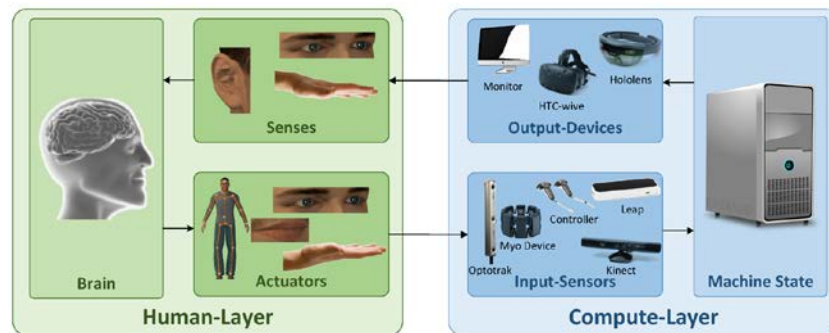
Стан проблеми

Прогрес у користувацькому досвіді:

Мультимодальні системи уже покращують взаємодію з користувачами, особливо в сферах віртуальної та доповненої реальності.

Застосування в технологіях: від смартфонів до VR-симуляторів та безконтактної взаємодії.

Нові можливості та потенціал: Нові можливості для компенсації сенсорних порушень у людей.



Мета дослідження: Розширення можливостей користувачів взаємодіяти з комп'ютером через мультимодальні системи.

Зміст роботи

Аналіз предметної області

- Огляд взаємодії з інтерфейсом комп'ютера, як приклад мультимодальної системи
- Характеристика та інтеграція мультимодальної взаємодії
- Основні аспекти дизайну мультимодальних систем

Програмний застосунок

- Візуальна модальність у вигляді:
 - Жестової взаємодії
 - Керування поглядом
- Звукова модальність у вигляді:
 - Розпізнавання мови на вхід
 - Звукового зворотного зв'язку від системи

Зміст програми

Реалізовано такі модальності як **візуальна** та **звукова**, як на ввід так і на вивід на додачу до традиційної тактильної модальності.

Модальність	Дії на вхід	Зворотний зв'язок
Візуальна	Жести + погляд	Зміни елементів інтерфейсу
Звукова	Мовні команди	Системні звуки у відповідь на дії користувача
Тактильна	Традиційне використання мишки/трекпаду та клавіатури	Вібрації у відповідь на деякі дії на трекпаді

Модуль розпізнавання та відслідковування рук **HandModule** за використання набору бібліотек MediaPipe

process_frame

Обробляє отримане зображення з камери та повертає системні координати x, y, z кожної лендмарки долоні

draw_hands

Додає схематичні лінії та точки за поверх зображення долоні

get_landmarks_list

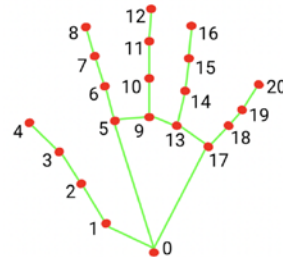
Повертає координати x, y, z на основі піксельних значень кожної лендмарки долоні

landmarks_distance

Знаходить відстань між двома заданими лендмарками

fingers_up

Повертає масив із п'яти значень 0 або 1, відповідно до кожного з пальців. 1 є сигналом про те, що відповідний палець піднятий, 0 - не піднятий.



- | | |
|-----------------------|-----------------------|
| 0. WRIST | 11. MIDDLE_FINGER_DIP |
| 1. THUMB_CMC | 12. MIDDLE_FINGER_TIP |
| 2. THUMB_MCP | 13. RING_FINGER_MCP |
| 3. THUMB_IP | 14. RING_FINGER_PIP |
| 4. THUMB_TIP | 15. RING_FINGER_DIP |
| 5. INDEX_FINGER_MCP | 16. RING_FINGER_TIP |
| 6. INDEX_FINGER_PIP | 17. PINKY_MCP |
| 7. INDEX_FINGER_DIP | 18. PINKY_PIP |
| 8. INDEX_FINGER_TIP | 19. PINKY_DIP |
| 9. MIDDLE_FINGER_MCP | 20. PINKY_TIP |
| 10. MIDDLE_FINGER_PIP | |

Основний модуль

Розпізнавання жестів та виконання відповідних команд за використання функцій з модуля **HandModule**

Наприклад, якщо

- Піднято середній, безіменний пальці та мізинець (*fingers_up*)
- Відстань між вказівним і великим пальцями менше 13 одиниць (*landmarks_distance*)



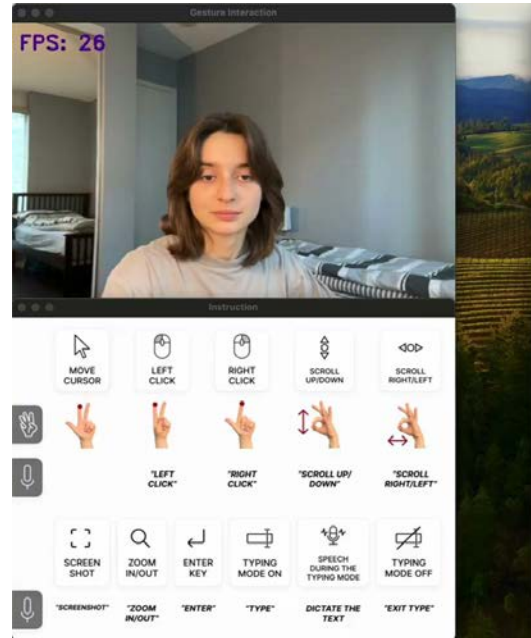
То, програма розуміє, що це жест для прокручування і далі виконується команда.

Модальність	Введення	Виведення	
	Для людини, приклади	Пристрій на вхід, приклади	Зворотний зв'язок, приклади
Візуальна	Вираз обличчя, рух очима, рух губ, рух тіла, включаючи жести, тощо	Камера	Інформація у формі тексту, графіки, відео тощо
Звукова	Мова, несловний звук (наприклад, наспівування мелодії)	Мікрофон	Записаний звук
Тактильна	Нагніснення клавіш, мшишка, сенсорних панелей або екранів, трекпаду; використання цифрового стилуса	Клавіатура, мишка, цифровий стилус, трекпад, джойстики, сенсорний дисплей або панель і тд	Різного виду вібрації, електронні улазці

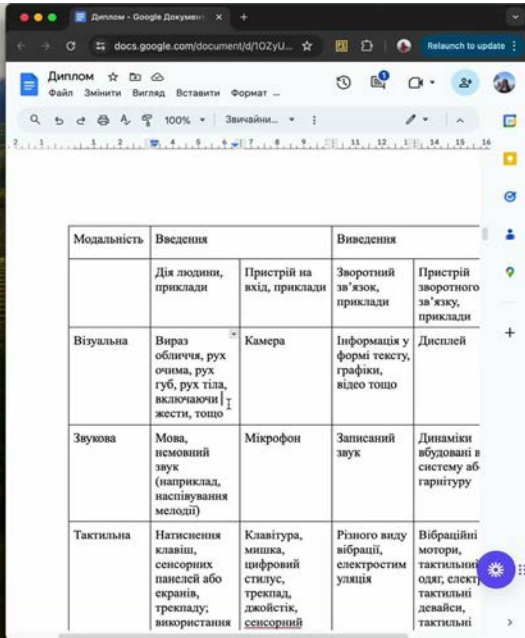
Основний модуль

Розпізнавання голосових команд за допомогою VOSK API

- Команди, що дублюють жестове керування, наприклад клік
- Додаткові команди, наприклад скріншот та масштабування
- Можливість набору тексту голосом



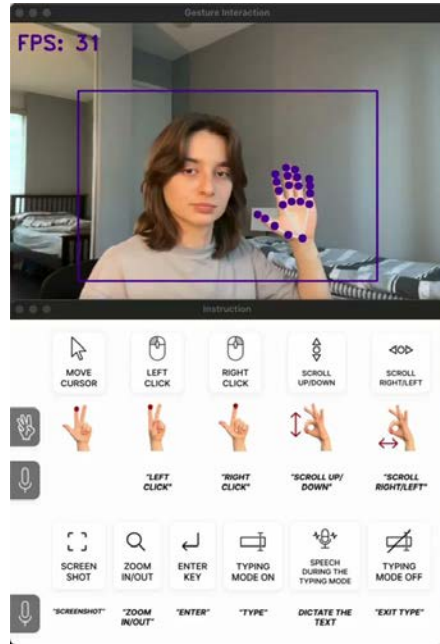
Модальність	Введення	Виведення		
	Дія людини, приклади	Пристрій на вхід, приклади	Зворотний зв'язок, приклади	Пристрій зворотного зв'язку, приклади
Візуальна	Вираз обличчя, рух очима, рух губ, рух тіла, включаючи жести, тощо	Камера	Інформація у формі тексту, графіки, відео тощо	Дисплей
Звукова	Мова, немовний звук (наприклад, наспівування мелодії)	Мікрофон	Записаний звук	Динаміки вбудовані в систему або гарнітуру
Тактильна	Натиснення клавіш, сенсорних панелей або екранів, трекпад; використання	Клавіатура, мишка, сенсорний екран, трекпад, джойстик, сенсорний	Різного виду вібрації, електричним уліця	Вібраційні мотори, тактильний одяг, слект тактильні девайси, тактильні



Основний модуль

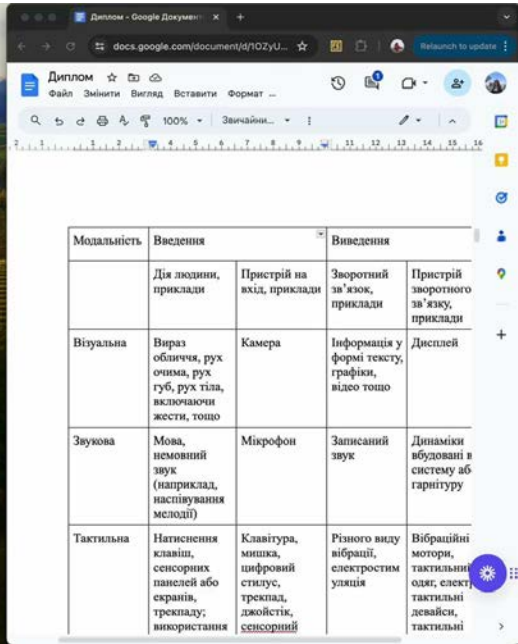
Додавання звукового зворотного зв'язку за допомогою бібліотеки playsound

Одночасно з виконанням жестових команд або вводу команд голосом надається звуковий зворотний зв'язок для деяких команд.



The screenshot shows a video feed of a woman with a purple dot grid overlay on her hand, indicating gesture recognition. Below the video is a control panel with various gesture icons and their corresponding actions:

- MOVE CURSOR
- LEFT CLICK
- RIGHT CLICK
- SCROLL UP/DOWN
- SCROLL RIGHT/LEFT
- "LEFT CLICK"
- "RIGHT CLICK"
- "SCROLL UP/ DOWN"
- "SCROLL RIGHT/LEFT"
- SCREEN SHOT
- ZOOM IN/OUT
- ENTER KEY
- TYPING MODE ON
- SPEECH DURING THE TYPING MODE
- TYPING MODE OFF
- "SCREENSHOT"
- "ZOOM IN/OUT"
- "ENTER"
- "TYPE"
- DICTATE THE TEXT
- "EXIT TYPE"



The screenshot shows a Google Docs document with a table containing gesture interaction data:

Модальність	Введення	Виведення
	Дія людини, приклади	Пристрій на вхід, приклади
Візуальна	Вираз обличчя, рух очима, рух губ, рух тіла, влючаючі жести, тощо	Зворотний зв'язок, приклади
Звукова	Мова, немовний звук (наприклад, наступання мелодії)	Мікрофон
Тактильна	Написання клавіш, сенсорних панелей або екранів, трекпаду, використання	Клавiтура, мишка, цифровий стилус, трекпад, джойстик, сенсорний
		Зворотний зв'язок, приклади
		Інформація у формі тексту, графіки, відео тощо
		Дисплей
		Записаний звук
		Динаміки вбудовані в систему або гарнітуру
		Різного виду вібрації, електрострим уляція
		Вібраційні мотори, тактильні одяг, електр тактильні ледайси, тактильні

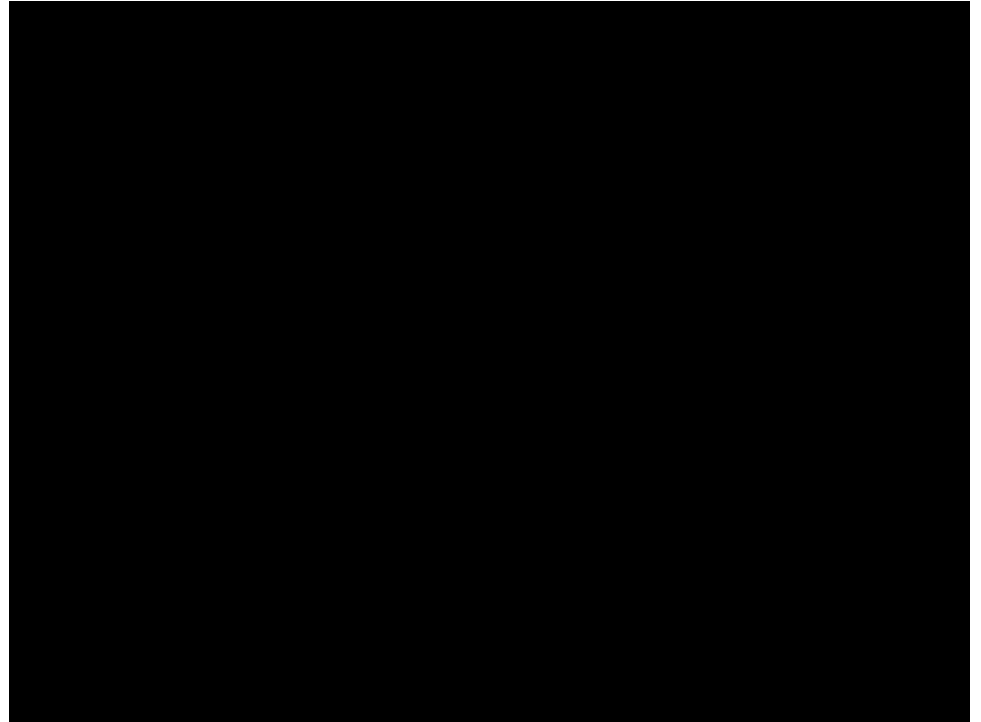
Додатковий експеримент

Розпізнавання погляду та виконання відповідних кліку ЛКМ

- *Move_cursor* виконується лендмарками правого ока
- Клік – за допомогою лівого

Недоліки:

- Неможливість виконання багатьох команд виключно очима
- Відсутність точності за мінімальних рухів голови



Результати

Програма **розширює можливості взаємодії** з комп'ютером, доповнюючи традиційну тактильну модальність миші та клавіатури візальною та звуковою модальностями

Інтеграція модальностей **зменшує зовнішнє когнітивне навантаження.**

Підвищує доречне навантаження, сприяючи кращому залученню користувачів.

Створення **інклюзивного цифрового середовища.**

Недоліки та потенціал до покращення

Неточність управління жестами і поглядом та розпізнавання команд.

Звукові помилки можуть виникати через зовнішні шуми.

Необхідно мінімізувати фізичні зусилля для виконання команд для покращення доступності

Дякую за увагу!