

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Національний університет «Запорізька політехніка»



Факультет комп'ютерних наук та технологій
Кафедра «Комп'ютерні системи та мережі»

МІНЯЙЛО ПАВЛО ІВАНОВИЧ
Група КНТ-612м

КОМП'ЮТЕРНА СИСТЕМА ДИСТАНЦІЙНОГО
КЕРУВАННЯ ЗА ДОПОМОГОЮ ГОЛОСУ

АВТОРЕФЕРАТ

магістерської роботи на здобуття освітньо-кваліфікаційного
рівня «магістр» 123 «Комп'ютерна інженерія»
освітньої програми «Спеціалізовані комп'ютерні системи»

2023 р.

Магістерська робота є рукопис.

Робота виконана в Національному університеті «Запорізька політехніка», на кафедрі комп'ютерних систем та мереж

Керівник

кандидат технічних наук, доцент
Касьян Костянтин Миколайович,
Національний університет «Запорізька
політехніка», доцент кафедри
комп'ютерних систем та мереж

**Офіційний
рецензент:**

Малий Олександр Юрійович,
к.т.н., доцент кафедри «Інформаційні
технології електронних засобів» НУ
«Запорізька політехніка»

Захист відбудеться "20" грудня 2023р.

Секретар екзаменаційної комісії, старший викладач
кафедри комп'ютерних систем та мереж Н.А. Куликовська

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. Сучасний світ неможливо уявити без впливу комп'ютерних технологій, які стали ключовою частиною повсякденного життя. Ці технології не тільки вдосконалюються з часом, але й перетворюють спосіб, яким спілкуємося та взаємодіємо з пристроями, такими як смарт-домашні системи та голосові помічники. Вони пропонують широкий спектр можливостей, від створення локальних мереж до розвитку швидкісної передачі даних та прогресивного програмного забезпечення.

У цьому світлі, існує значний інтерес до розвитку як централізованих, так і децентралізованих систем, які пропонують різні підходи до обробки та передачі інформації. Централізовані системи забезпечують надійність та єдність управління, тоді як децентралізовані системи пропонують гнучкість та відсутність залежності від центрального сервера, що може зменшити витрати і збільшити ефективність.

Мета і завдання дослідження. є розробка комп'ютерної системи для розпізнавання мови, призначеної для використання в "розумних" колонках. Цей проект передбачає створення структурної схеми та розробку з'єднань між елементами системи, що дозволить їй підключатися до мережі і відповідно взаємодіяти з користувачами.

Об'єктом дослідження – процеси голосового керування пристроями.

Предмет дослідження – методи та алгоритми розпізнавання та обробки мовних сигналів.

Методи дослідження об'єктно-орієнтоване програмування апаратні платформи STM32, Raspberry Pi, мови програмування Python, C++, API Google.

Наукова новизна отриманих результатів:

Застосовано інтегрований підхід до обробки природної мови поєднуючий хмарні сервіси та мікроконтролери.

Практичне значення отриманих результатів:

Розроблений програмно-апаратний комплекс дає змогу здійснювати голосове керування розумною колонкою.

Структура та обсяг роботи. Магістерська робота складається зі вступу, чотирьох розділів, висновків, списку використаних джерел, додатку. Основна частина містить 70 сторінок, 22 рисунки і 11 таблиці, список використаних джерел зі 42 найменування.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У першому розділі аналіз ринку показав, що сучасні розумні колонки стають все більш інтегрованими в повсякденне життя людей. Це вимагає постійного удосконалення технологічних характеристик, включаючи якість звуку, дизайн, інтерфейс користувача та інтеграцію з іншими розумними пристроями. Інновації в цій галузі повинні орієнтуватися на підвищення зручності використання та розширення функціональності.

Сучасні споживачі вимагають, щоб розумні колонки були оснащені голосовим управлінням та інтеграцією з широким спектром голосових асистентів. Це підкреслює необхідність розробки інтелектуальних інтерфейсів, що можуть вплинути на користувацький досвід.

На основі проведеного аналізу та оцінки поточного стану ринку, стає зрозумілою важливість створення комп'ютерної системи, яка забезпечує високу точність розпізнавання мови та є зручною у використанні для кінцевого користувача. Така система має включати сучасні алгоритми обробки звуку та інтелектуальні можливості для підвищення ефективності та зручності використання.

Таблиця 1 – Аналіз портативних колонок

Параметр	JBL Boombox [11]	Defender Enjoy [12]
Виробники	виробляються компанією JBL, відомим американським брендом аудіотехніки, який спеціалізується на виготовленні високоякісних гучномовців і	виробляється компанією Defender, яка є менш відомою, але пропонує доступніші варіанти аудіообладнання

	аудіоаксесуарів.	
Якість звуку	відомі своєю високою якістю звуку, зокрема, чіткими високими та глибокими басами	може мати менш виразний звуковий профіль порівняно з JBL, але все ж пропонує прийнятну якість для своєї цінової категорії.
Потужність та гучність	має більш високу гучність і потужність, що робить її ідеальною для використання на відкритому повітрі або великих приміщеннях.	має меншу потужність порівняно
Автономність роботи	пропонують тривалу роботу на одному заряді, забезпечуючи довгий час використання без підзарядки.	може мати менший час автономної роботи через меншу ємність акумулятора.
Функціональність	часто оснащуються додатковими функціями, такими як водонепроникність, підключення до декількох джерел одночасно, та вбудованими мікрофонами для голосових дзвінків.	може не мати деяких з цих функцій, але зазвичай пропонує базові можливості підключення.
Портативність	є великою та важчою, але досі портативною	більш компактні та легкі
Вартість	Вище середнього	Вище середнього

У розділі другому було спроектовано систему.

Для забезпечення ефективного використання системи користувачами, необхідно впровадити функції, які дозволять легко отримувати інформацію з пристроїв, виконувати поставлені завдання та забезпечувати безперебійну роботу. Важливим аспектом є також надання користувачам можливості регулювати

гучність та призупиняти роботу пристрою у випадку виникнення проблем.

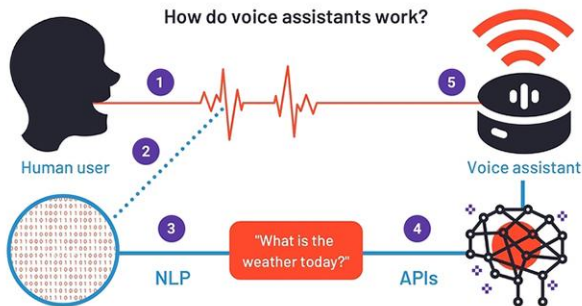


Рисунок 1 – Інтеграція із іншими пристроями

Система повинна інтегрувати технології голосового управління для забезпечення більшої зручності користувачів, дозволяючи їм керувати пристроями за допомогою голосових команд. Це зменшить необхідність фізичної взаємодії з пристроями та підвищить загальний комфорт використання.

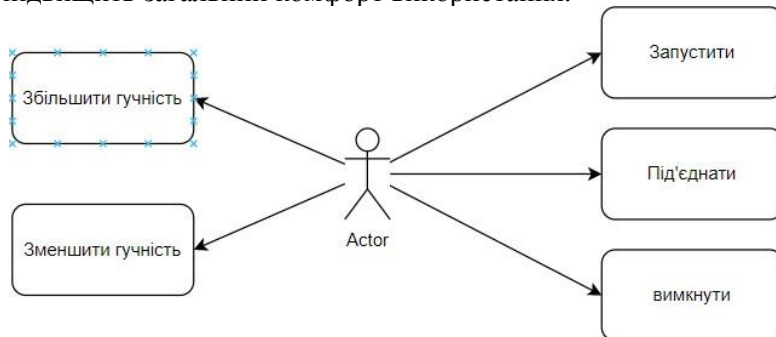


Рисунок 2 – Взаємодія із системою

Проведено аналіз мікроконтролерів. Порівняння показує, що STM32F103C8T6 має досить конкурентоспроможні характеристики в своєму класі, особливо зважаючи на його доступність та популярність. Однак, інші мікроконтролери можуть пропонувати більше пам'яті, вищу тактову частоту або додаткові периферійні інтерфейси, що може бути критично для певних застосувань.

STM32CubeIDE об'єднує можливості налаштування та створення проектів від STM32CubeMX, пропонуючи таким чином

інструмент "все в одному", що економить час на встановлення та в роботі.

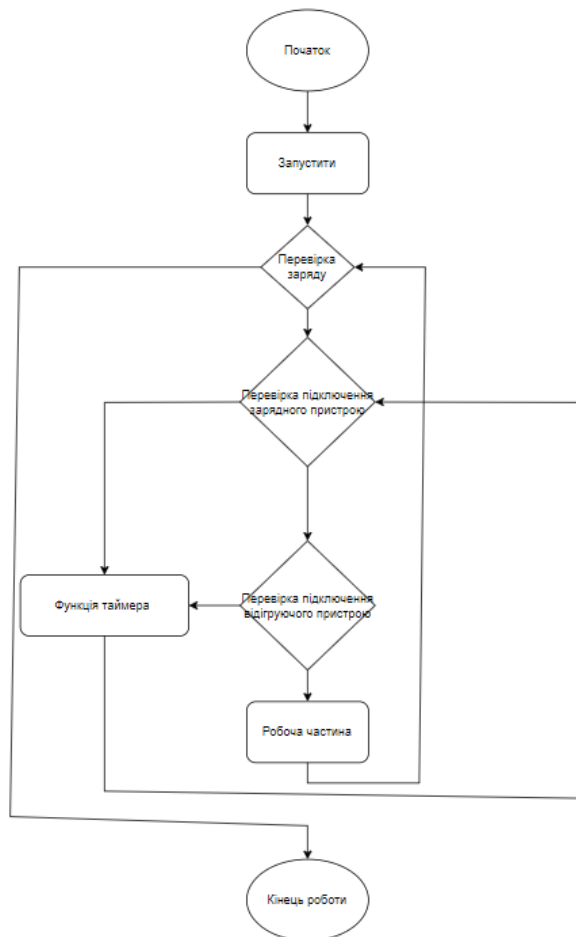


Рисунок 4 – Блок-схема роботи розумною колонки

Google Speech API, відомий як Google Cloud Speech-to-Text, є одним з провідних сервісів у галузі розпізнавання голосу. Для порівняння, можна взяти такі відомі сервіси, як IBM Watson Speech to Text, Microsoft Azure Speech Service та Amazon Transcribe. Основні характеристики, які слід враховувати при порівнянні, включають точність розпізнавання, підтримку мов, можливість налаштування (наприклад, для розпізнавання специфічної

термінології), ціноутворення, а також інтеграцію з іншими сервісами та платформами. Google Speech API показав кращі показники для вибору його в якості обробника мовлення.

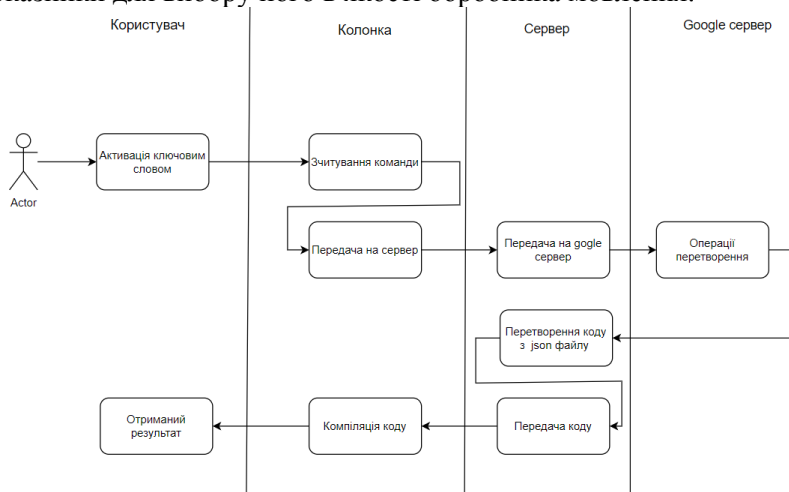


Рисунок 4 – Діаграма взаємодії компонентів системи

У третьому розділі розробка системи полягала у створенні ефективної взаємодії з сервісом Google Speech API для конвертації аудіофайлів у текст. Цей процес передбачав налагодження зв'язку між системою та сервісом через окремий сервер, відповідальний за обробку вхідних та вихідних запитів. Важливість цього підходу полягає у забезпеченні надійної та ефективного обробки аудіо даних.

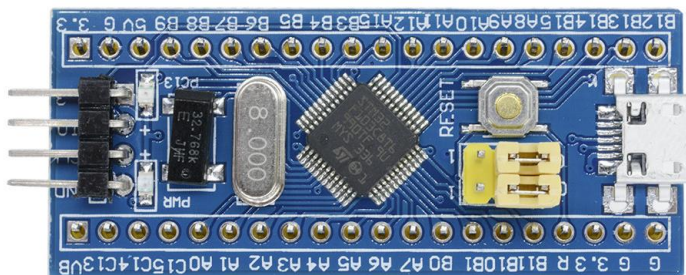


Рисунок 5 – Контролер системи

Для успішної інтеграції з Google Speech API необхідно було розглянути декілька критичних аспектів. Перш за все, потрібно

було детально вивчити процес інтеграції, включаючи авторизацію, формати даних, які використовуються, та оптимізацію процесу конвертації аудіо в текст. Особливу увагу слід звернути на вибір мови програмування та механізмів обробки запитів для сервера, який взаємодіє з API.

Особливу увагу у проєкті приділено питанням безпеки та конфіденційності. Важливо забезпечити захист даних та конфіденційності інформації, що передається між системою та Google Speech API, з використанням сучасних стандартів шифрування та передачі даних. Це гарантує, що особиста інформація користувачів залишається захищеною від несанкціонованого доступу.

У четвертому розділі розроблюваний сервіс зосереджується на значному вдосконаленні технології розпізнавання мови, включаючи використання інноваційних алгоритмів нейронних мереж та машинного навчання сучасних бібліотек. Це сприяє не лише підвищенню точності розпізнавання мови, але й забезпечує більш плавне та природне взаємодію з розумними колонками. Важливою особливістю сервісу є його здатність розпізнавати регіональні діалекти та акценти, що робить його доступним для широкої аудиторії та покращує розуміння команд користувачів з різних регіонів.

Таблиця 2 – Експериментальне дослідження роботи

Експеримент	Точність розпізнавання Google	Точність розпізнавання IBM
1	94%	92%
2	93%	90%
3	92%	91%
4	95%	93%
5	91%	92%

Наукова новизна проєкту полягає у впровадженні інтегрованого підходу до обробки природної мови. Цей підхід дозволяє системі не тільки розпізнавати слова, але й розуміти контекст та нюанси мовлення, що значно підвищує ефективність взаємодії з користувачами. Система здатна адаптуватися до

індивідуальних особливостей мовлення кожного користувача, забезпечуючи гнучкість та персоналізацію взаємодії.

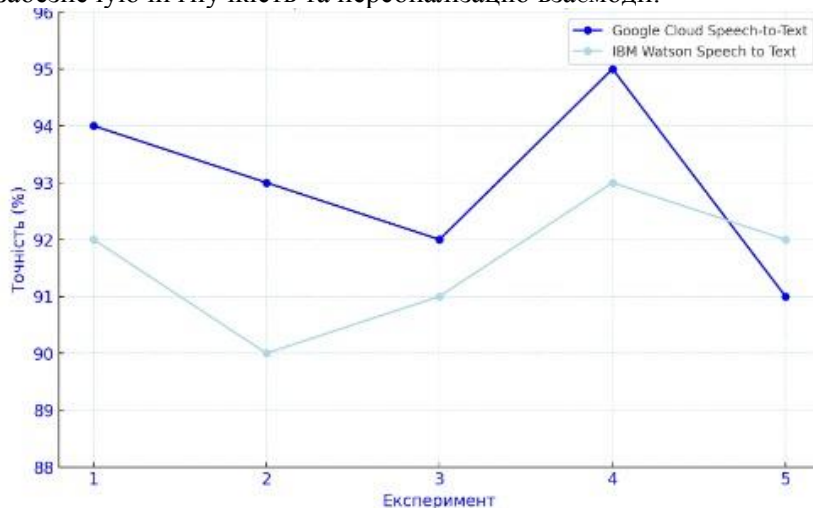


Рисунок 6 – Результати дослідження

З практичного погляду, розробка цієї системи розпізнавання мови відкриває можливості для значного поліпшення взаємодії з "розумними" колонками. Аналізуючи дані експериментів, можна зробити висновок, що сервіс Google Cloud Speech-to-Text має вищу точність розпізнавання мови порівняно з сервісом IBM Watson Speech to Text. У кожному з п'яти проведених експериментів сервіс Google показав або такі самі, або кращі результати точності. Це свідчить про те, що в цілому алгоритми, які використовує Google для розпізнавання мовлення, можуть бути більш ефективними у даному контексті.

ВИСНОВКИ

На основі проведеного аналізу та оцінки поточного стану ринку, стає зрозумілою важливість створення комп'ютерної системи, яка забезпечує високу точність розпізнавання мови та є зручною у використанні для кінцевого користувача. Така система має включати сучасні алгоритми обробки звуку та інтелектуальні

можливості для підвищення ефективності та зручності використання.

Розроблюваний сервіс зосереджується на значному вдосконаленні технології розпізнавання мови, включаючи використання інноваційних алгоритмів нейронних мереж та машинного навчання сучасних бібліотек. Це сприяє не лише підвищенню точності розпізнавання мови, але й забезпечує більш плавне та природне взаємодію з розумними колонками. Важливою особливістю сервісу є його здатність розпізнавати регіональні діалекти та акценти, що робить його доступним для широкої аудиторії та покращує розуміння команд користувачів з різних регіонів.

Наукова новизна проекту полягає у впровадженні інтегрованого підходу до обробки природної мови. Цей підхід дозволяє системі не тільки розпізнавати слова, але й розуміти контекст та нюанси мовлення, що значно підвищує ефективність взаємодії з користувачами. Система здатна адаптуватися до індивідуальних особливостей мовлення кожного користувача, забезпечуючи гнучкість та персоналізацію взаємодії.

З практичного погляду, розробка цієї системи розпізнавання мови відкриває можливості для значного поліпшення взаємодії з "розумними" колонками. Аналізуючи дані експериментів, можна зробити висновок, що сервіс Google Cloud Speech-to-Text має вищу точність розпізнавання мови порівняно з сервісом IBM Watson Speech to Text. У кожному з п'яти проведених експериментів сервіс Google показав або такі самі, або кращі результати точності. Це свідчить про те, що в цілому алгоритми, які використовує Google для розпізнавання мовлення, можуть бути більш ефективними у даному контексті.