МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ІВАНА ФРАНКА

Факультет прикладної математики та інформатики
(повне найменування назва факультету)

Кафедра дискретного аналізу та інтелектуальних систем
(повна назва кафедри)

Магістерська робота

Керування деякими прикладними програмами ОС Windows, використовуючи голосовий інтерфейс(MS Word, MS Power Point та інші)

Виконав: студент групи ПМіМ-23с
спеціальності
122 «Комп'ютерні науки»
(шифр і назва спеціальності)

(підпис) Павлик В.Р (прізвище та ініціали)

Керівник (підпис) Позднякова І.В. (прізвище та ініціали)

Рецензент (підпис) Бернакевич І.Є. (прізвище та ініціали)

Львів – 2022

ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ІВАНА ФРАНКА

Факультет Прикладної математики та інформатики _____
Кафедра Дискретного аналізу та інтелектуальних систем _____
Спеціальність 122 «Комп'ютерні науки» _____
(шифр і назва)

«ЗАТВЕРДЖУЮ»
Завідувач кафедри Притула М.М. _____

"31 "серпня 2022 року

З А В Д А Н Н Я

НА МАГІСТЕРСЬКУ РОБОТУ СТУДЕНТУ

Павлику Владиславу Руслановичу
(прізвище, ім'я, по батькові)

1 .Тема роботи «Керування деякими прикладними програмами ОС Windows, використовуючи голосовий інтерфейс(MS Word, MS Power Point та інші) »

керівник роботи Позднякова Інна Володимирівна, канд. фіз.-мат. наук,
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)
затверджені Вченою радою факультету від **" 13" вересня 2022 року №15**

2. Строк подання студентом роботи **12.12.2022 р.** _____

3. Вихідні дані до роботи документація по C#,Windows Forms, Windows Speech Recognition. Середовище програмування -Visual Studio 2022

4. Зміст магістерської роботи (перелік питань, які потрібно розробити)
   1. Опис використаних технологій.
   2. Програмна реалізація.
   3. Функціонал програми.

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень)
   1. Архітектура проекту.
   2. Опис допоміжних бібліотек.
   3. Демонстрація роботи програми.

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв

7. Дата видачі завдання **31 серпня 2022 р.**

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів магістерської роботи	Строк виконання етапів роботи	Примітка
1.	Пошук та вивчення необхідних технологій	01.09.2022 - 12.09.2022	
2.	Пошук та вивчення необхідної архітектури програмного забезпечення	13.09.2022 – 22.09.2022	
3.	Розробка дизайну	23.09.2022 – 28.09.2022	
4.	Розробка логіки	29.09.2022 – 10.10.2022	
5.	Поєднання логіки і дизайну	11.10.2022 – 24.10.2022	
6.	Тестування	25.10.2022 – 27.10.2022	
7.	Виправлення помилок/багів	28.10.2022 – 04.11.2022	
8.	Повторне тестування і внесення остаточних правок	05.11.2022 – 13.11.2022	
9.	Оформлення магістерської роботи	06.11.2022 – 28.11.2022	

Студент _____ **Павлик В.Р.**
(підпис) (прізвище та ініціали)

Керівник роботи _____ **Позднякова І.В.**
(підпис) (прізвище та ініціали)

РЕФЕРАТ

Магістерська робота складається із змісту, вступу, чотирьох розділів, висновку і списку використаних джерел. Загальний об’єм 38 сторінок, список джерел містить двадцять чотири найменування, робота ілюстрована рисунками.

У даній роботі було реалізовано додаток, за допомогою якого можна було б давати голосові команди різним прикладним програмам ОС Windows, з використанням голосового інтерфейсу.

Програма може керувати такими програмами як MS Word, MS Power Point, Spotify, Google Chrome, голосовий пошук в пошуковій системі Google та інші. Також реалізовано службу підтримки програми, яку реалізовано в вигляді Google Forms. Всі заповнені та залишені заявки , я отримую як сповіщення в особисті повідомлення електронної пошти.

Ключові слова: голос, голосовий інтерфейс, голосові команди, мова, Windows Forms, Windows Speech Recognition.

ЗМІСТ

Вступ ……....6
Розділ 1 ……9
Технічний аналіз завдання .....……..9
Сфери мовних технологій .....……..9
Злите мовлення .....…….10
Спільнокореневі слова .....…….....11
Проблеми автоматичного розпізнавання мови 12
Розділ 2 ……..14
Огляд існуючих систем керування розпізнаванням мовної інформації та методів вирішення 14
Проблематика 14
Застосування нейронних мереж для розпізнавання мови 15
Алгоритми зворотного розповсюдження 16
Розділ 3 ……..18
ГОЛОСОВЕ КЕРУВАННЯ ТЕХНІЧНИМИ ПРИСТРОЯМИ……18
Огляд пристроїв для зчитування аудіо сигналів……18
Підбір пристрою для розробки аудіо системи ……23
Розділ 4 ……..27
Елементи розробки додатку ..……....27
Необхідні елементи для створення додатку ……27
Основні інструменти створення вигляду додатку……...28
Функціонування методів та алгоритмів розв’язання……..28
Програмна реалізація.……....29
Функціонал додатку……...29
ВИСНОВОК……..35
СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ.……36

ВСТУП

В теперішній час наукова спільнота вкладає гігантську кількість грошей в розвиток ноу-хау та науково-дослідницькі розробки для вирішення проблем автоматичного розпізнавання і розуміння мови(розмови). В першу чергу це стимулюється практичними вимогами(потребами), пов'язаними з створенням систем воєнного і комерційного призначення. Якщо не говорити про перше, можна зазначити, що тільки в європейській спільноті об'єм продажів систем громадянського призначення складає декілька мільярдів доларів. При цьому потрібно звернути увагу на те, що в практичному використанні відсутні системи, які по не зрозумілим причинам вважаються вершиною розвитку систем автоматичного розпізнавання мови. Це системи, які можна назвати демонстраційними і ті, які 50 років тому називались «фонетичний друкуючий прилад». Їхньою метою є перевід мови в відповідний письмовий текст.

Якщо розглядати класичну схему «наука – технології – практичні системи», то перш за все, потрібно визначити ті умови, в яких буде працювати практична система автоматичного розпізнавання чи розуміння мови. Найбільш серйозні проблеми, які виникають:

- легковажний користувач;
- випадковий набір слів, жаргон, суржик, діалектизми;
- наявність акустичних перешкод;
- наявність мовних шумів, перешкод;
- проблеми з вимовою чи гучністю вимови користувача;

З іншого боку потрібно визначити важливість задачі, її наукову фунтаментальність, зв'язок з іншими областями знань. При цьому необхідно враховувати стан науково-промислового потенціалу, та його можливості. Відома думка, що правильно поставлене завдання – це вже половина рішення.

Зупинимось на можливостях і недоліках відповідних систем автоматичного розпізнавання мови (які на теперішній час, маю можливість розпізнавати сотні і навіть тисячі слів з точністю до 98%).

Від користувача потребується заздалегідь налаштувати систему на свій голос. Це займає від десятків хвилин до декількох годин наговорювання тексту в мікрофон.

Так як слова, які вимовляються в добре вимовленому тексті, виявляються плаваючими в океані омонімів, тому кількість помилок(словесних) збільшується приблизно в 5 раз. Поспішне відслідковування таких помилок, окрім випадків виникнення безглуздих текстів, вже є великою проблемою. Прилад корекції помилок в більшості демонстраційних системах погано налагоджений (відточений).

Були згадки, про те, що навіть в добре організованому спонтанно вимовленому тексті, ймовірність правильного розпізнавання мови , не перевищує 1/3.

І напевно останнє, але не менш важливе, що час обробки введенего відрізку мови, в таких системах може займати хвилини, а то й більше.

Все вище сказане говорить, про те, що в якості кінцевої цілі, запропоновані демонстраційні системи «мова-текст» навряд чи представляють інтерес. Це не виключає можливості використовування їх в якості полігону, для оцінки наукових ідей та відкриттів, але в цьому випадку повинні чітко викладатись ті моделі, які закладаються в дані системи автоматичного розпізнавання мови і яким чином повинна перевірятись їх практична перспектива. Таким чином, ми переходимо на протилежний кінець тріади «практичні систем – мовні технології – мовна наука».

Метою магістерської роботи є написати програму по голосових командах, для керування презентаціями.

Відповідно до мети роботи розглянемо такі **завдання**:

- проаналізувати технічний аналіз завдання;
- дослідити проблеми автоматичного розпізнавання мови;
- здійснити огляд існуючих систем керування розпізнаванням мовної інформації;

- ознайомитись з проблематикою систем керування розпізнавання мовної інформації;
- вивчити алгоритми зворотного розповсюдження;
- ознайомитись з застосуванням нейронних мереж та прихованих Марковських моделей для розпізнавання мови;
- розробити програмну реалізацію щодо керування презентаціями та іншими додатками, за допомогою голосових команд;
- описати програмну реалізацію голосових команд.

Ціллю моєї магістерської роботи є написання програми, за допомогою, якої можна б було давати тривіальні(не дуже тривіальні) команди, за допомогою голосу.

РОЗДІЛ 1. ТЕХНІЧНИЙ АНАЛІЗ ЗАВДАННЯ

1.1 Сфери мовних технологій

На даний час робота з розпізнаванням мови не тільки не втратила актуальність, а й розвивається широким фронтом, знаходячи для себе велику кількість сфер для практичного застосування. Тепер можна виділити 4 порівняно ізольованих напрямки в області(сфері) мовних технологій: [1]

1. Розпізнавання мови – перетворення мовного акустичного сигналу в ланцюжок символів, слів. Ці системи можна охарактеризувати по списку параметрів. Перш за все це об’єм словника: малий об’єм до 20 слів, великий – тисячі і десятки тисяч. Кількість дикторів: від одного і до будь якої кількості. Стиль вимови: від ізольованих команд до повноцінного мовлення та від читання і до спонтанної мови. Коефіцієнт розгалуження, це величина, яка визначає кількість гіпотез на кожному кроці розпізнавання: від малих величин (< 10÷15) до великих (>100÷200). Відношення сигнал/шум від великих (>30 дБ) до низьких (<10 дБ). Якість каналів зв’язку: від високоякісного мікрофону до телефонного каналу. Якість роботи систем розпізнавання мови, переважно характеризуються якістю та надійністю розпізнавання слів, і відсотком помилок.

2. Визначення особистості, того хто говорить. Ці системи поділяються на два класи: верифікація того хто говорить (підтвердження його особистості), та ідентифікація того хто говорить(визначення його особистості із заздалегідь обмеженої кількості людей). Обидва ці класи в майбутньому можуть бути розділені на залежні від тексту та незалежні від тексту. Наступний характерний параметр – це об’єм фрази-пароля. Два інших (як і в розпізнаванні мови): відношення сигнал/шум і якість каналу сигналу зв’язку. Якість роботи систем верифікації/ідентифікації, того хто говорить характеризується двома величинами: ймовірність не розпізнання «свого» диктора і ймовірність прийняття «чужого» диктора за свого [4].

3. Синтез мови. Практично існує два класи:

1) Відтворення записаного в тій чи іншій формі обмеженого числа повідомлень;

2) Синтез мови по тексту. Синтезатори характеризуються по наступним параметрам: розбірливість(словесна чи складова), природність звучання та стійкість до перешкод.[3]

4. Компресія мови. Основна (та єдина) класифікуюча ознака цієї системи, це ступінь компресії: від низької (32-16 кбіт/сек) до високої (1200-2400 кбіт/сек). Якість роботи систем компресії мови, характеризується перш за все, розбірливістю компресованої мови. Додатковими характеристиками дуже важливими є впізнаваність голосу, того хто говорить і можливість визначення стресового стану, того хто говорить.

1.2 Злите мовлення

Для розпізнання слів в злитому мовленні апробовані два різних підходи. В першому випадку при глобальному підході слово, яке необхідно розпізнати, порівнюється з кожним іншим словом словника. Під час порівняння використовується, як правило, спектральне представлення кожного слова. Серед різних методів даного типу хороші результати показав метод динамічного програмування.

В іншому випадку при аналітичному підході кожне слово або група слів спочатку сегментується на менші частинки. Сегментами є спільно-складові або фонемо-подібні одиниці. Це дозволяє проводити розпізнавання на складовому або на фонетичному рівні і одночасно зберігати в пам’яті параметри (тривалість, енергію і так далі), які відносяться до просодії та будуть корисними в подальшому. Сегментація може буде побудована на знаходженні голосних, які часто розташовуються біля максимуму інтегративної енергії спектру. Під час такого підходу першим критерієм сегментації є зміна енергії в часі. Деякі приголосні, наприклад m, n, l, інколи володіють такою ж енергією, як і голосні. Тому необхідний ввід додаткових параметрів для обчислення наявності голосного звуку в кожному раніше виявленому сегменті.

Для ідентифікації приголосних, як правило, проводиться розділ розривних і нерозривних приголосних. Досягається це шляхом виявлення паузи(змички),

яка відповідає змиканню перед реалізацією розриву. Завдання ускладнюється для позиції початку висловлювання, де порівняно просто визначається змичка тільки для дзвінких розривних приголосних. Після виявлення кожної категорії звуків, переважно користуються впорядкованими правилами, які базуються на інформації, яка залежить від акустичного і фонетичного контексту. В злитому мовленні фонетична реалізація якогось конкретного висловлювання залежить від декількох факторів, включаючи діалект, швидкість вимови мов, манери вимови диктора та інші.[3]

1.3 Спільнокореневі слова

Основні ознаки розпізнання спільнокореневих слів – ієрархічна багатоярусна структура і контроль кожного ярусу за допомогою відповідних граматик, символи якої є розпливчастими лінгвістичними змінними величинами.

Стратегія розпізнання заснована на групуванні одиниць мови в широкі фонетичні класи, за якими йде класифікація на більш детальні групи.

Під час розпізнавання злитого мовлення виникають труднощі: розпізнавання злитого мовлення набагато складніше, ніж розпізнавання окремо сказаних слів, насамперед, внаслідок неявних меж (границь) між словами. В результаті важко визначити початок і кінець відповідності між фонемним ланцюжком слова з словника і фонемним ланцюжком, який розпізнається. Система акустико-фонетичного аналізу злитої мови, зазвичай розглядає, як частину загальної системи по автоматичному розпізнаванні.

Попередня сегментація і класифікація звукових елементів включає в себе визначення подібних голосних, розривних приголосних, пауз. Завдання сегментації, розглядається як завдання ділення мовного потоку на функціонально важливі частинки, які вирішуються по різному. При розробці систем розпізнання мови, враховується важливість першої ступені обробки акустичного сигналу, що пов’язана з роботою акустичного процесора. Процес автоматичної сегментації неперервно пов’язаний з маркуванням звукової послідовності. Розробка автоматизованої сегментації і маркування викликана

необхідністю залучення великої акустико-фонетичної бази даних і прагненням до об’єктивізації мовного аналізу [5].

1.4 Проблеми автоматичного розпізнавання мови

Завдання машинного розпізнавання мови привертає увагу спеціалістів вже дуже давно. Проте, просунутися в цьому напрямку вдалось відносно недалеко. Процес розпізнавання мови представляє собою перетворення акустичного сигналу, отриманого від мікрофону, в послідовність слів. Отриманий набір гіпотез ланцюгів слів далі використовується для розуміння мови [4,6].

При цьому виникає ряд проблем. По-перше, людина зазвичай не робить паузи між словами, а під час злитої вимови до завдання розпізнавання додається ще завдання виділення слів з потоку мови, що є більш важким завданням. Виникає необхідність виділяти односкладові слова – а саме з ними і пов’язано найбільше помилок, вже існуючих систем. Можна вимагати, щоб людина вимовляла слова по одному, роблячи достатньо тривалі паузи або щоб кожне наступне слово вимовлялось після звукового сигналу. Але даний підхід не зручний і може бути використаний тільки для подання простих команд.[7]

Наступна проблема – відмінність голосів, діалектів, дикцій, вікових відмінностей, емоційний та фізичний стан диктора. Значний вплив вносить акустичний аспект, такий як заміна мікрофона, розташування мікрофона відносно рота, акустична обстановка в приміщенні.[9]

Саме через ці і багато інших проблем до повного вирішення завдання розпізнавання мови як і раніше дуже далеко. Існує два режими роботи, які істотно відрізняються один від одного: з налаштуванням на голос певного диктора і без такого налаштування. Розміри словника під час роботи з налаштування на диктора (speaker-dependent) в теперішній час можуть досягати декількох чи багатьох тисяч слів злитого мовлення. Процедура налаштування на диктора виглядає наступним чином: диктор читає певний текст спеціальним чином, комп’ютер розпізнає слова і видає варіант розпізнавання. Диктор помічає

помилки і читає текст знову. Після декількох таких ітерацій процес сходиться, і комп’ютер знаходиться в стані розпізнати мову [4,8 ,9,10,11,12].

Нарешті, останній і найважчий для реалізації, але і найперспективніший режим роботи – розпізнавання без настройки на диктора. При цьому гарантується, що система розпізнає будь-яке слово, яке входить в словник, ким би воно не вимовлялось. Тут, як правило, словники нараховують невелику кількість слів (переважно не більше двох десятків) і існують для відносно невеликого числа мов (близько тридцяти – тридцяти п’яти).

Створення словника для розпізнавання мови без налаштування на голос потребує великих затрат. Для вирішення цього завдання розробникам приходиться опитувати велике число носіїв мови, виділяти певні загальні елементи мови, усереднювати їх – і все це для того, щоб забезпечити розпізнавання десяти-двадцяти слів. Найчастіше словник без настройки на голос користувача потребує роздільної вимови слів. Для цілого ряду речень, цього виявляється достатньо [4,8,9 ,10].

РОЗДІЛ 2. ОГЛЯД ІСНУЮЧИХ СИСТЕМ КЕРУВАННЯ РОЗПІЗНАВАННЯМ МОВНОЇ ІНФОРМАЦІЇ ТА МЕТОДІВ ВИРІШЕННЯ

2.1 Проблематика

Розпізнавання та породження (синтез) мови комп’ютером є, безумовно, важливою проблемою. Десятиліттями вчені і інженери шукали способи, які б дозволили б людям спілкуватись з комп’ютером так само як між собою, а не заставляли б людину пристосовуватись під спосіб спілкування, який прийнятний для машини. Багато зроблено, але напевно, і на сьогоднішній день, можна вважати, що питання й близько не закрите, хоча саме в останній час були досягнуті значні успіхи: уже багато років голосові команди є однією з можливих опцій програмного забезпечення комп’ютерів, поява функцій розпізнавання мови уже звичне діло в переліку текстових процесорів, системи розпізнавання мови, працюють там, де потребується допомога довідкових послуг і в системах безпеки. А також в наших мобільних телефонах, найбільш популярні на даний момент це: Siri (IPhone, IPad, MacBook і більша частина пристроїв, які працюють на операційній системі IOS), Google Now або як його ще називають Google Assistant(працює в більшості продукції компанії Google), Amazon Alexa(продукція компанії Amazon) та декілька інших.

Питання цифрової обробки сигналів, це окремі області математичної статистики, штучного інтелекту (теорії нейронних мереж), пов’язані з розробкою двигунів і застосувань і породженням мови.

Питання оброки мови є, головною частиною дисциплін, іменованих цифровою обробкою і розпізнанням образів.

Методи цифрової обробки сигналів зазвичай здійснюють перетворення, очистку і трансформацію звукового сигналу в цифровий формат даних і інші представлення, які можуть безпосередньо оброблятися системою розпізнавання мови. Ці завдання включають, також фільтрацію шумових сигналів, які домішуються до звуку під час передачі акустичних сигналів від сприймаючих приладів (мікрофонів) або по мережі. Методи розпізнання образів

використовують при виділення і розпізнанні окремих слів або речень мовного потоку, або в деяких випадках для ідентифікації особи, яка говорить.

Крім того, системи розпізнавання і синтезу мови включають питання лінгвістики, в якій закладені фундаментальні концепції і принципи розпізнавання мови і розуміння мови.

Два основні підходи вирішення задач розпізнавання голосового повідомлення [9].

2.2 Застосування нейронних мереж для розпізнавання мови

Введення в нейронні мережі:

Штучна нейронна мережа – це математична модель, а також пристрої паралельних обчислювань, які представляють собою систему з’єднаних і взаємодіючих між собою простих процесорів(штучних нейронів). Як математична модель штучна нейронна мережа, представляє собою приватний випадок методів розпізнавання образів або дискримінантного аналізу.

Такі процесори зазвичай досить прості, особливо в порівнянні з процесорами, які використовують в персональних комп’ютерах чи ноутбуках.

Кожен процесор подібної мережі має справу тільки з сигналами, які він періодично отримує, і з сигналами, які він періодично посилає іншим процесорам. Однак, будучи з’єднаний в достатньо велику мережу з керованою взаємодією, такі локально прості процесори разом спроможні виконувати досить важкі завдання.

Поняття виникло під час вивчення процесів, які протікають в мозку, вивчення роздумів чи мислення та при спробі змоделювати ці процеси. Отримані моделі називаються штучними нейронними мережами (ШНМ).

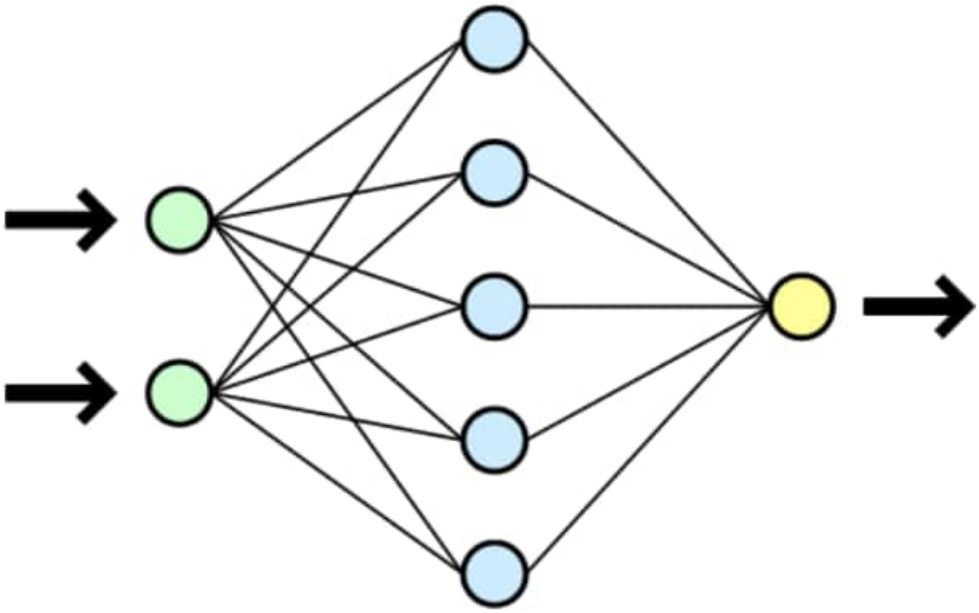

Рис 2.1 – схема простої нейронної мережі.

Зеленим позначено вхідні елементи, а жовтим – вихідні.

Нейронні мережі не програмуються в звичному сенсі цього слова, вони навчаються. Можливість навчання – одна з головних переваг нейронних мереж над традиційними алгоритмами. Технічно навчання заключається в знаходженні коефіцієнтів зв’язків між нейронами. В процесі навчання нейронна мережа, здатна виявляти складні залежності між вхідними і вихідними даними, а також виконувати узагальнення. Це означає, що в випадку успішного навчання, мережа може поверний вірний результат на підставі даних, які були відсутні в початковій вибірці [10,11].

2.3 Алгоритми зворотного розповсюдження

Складніше справа з багатошаровими мережами, так як від самого початку невідомі бажані виході шарів мережі (за виключенням останнього) і їх неможливо навчити, керуючись тільки величиною помилок на виході мережі, так як це було з одношаровою мережею.

Найбільш прийнятним варіантом вирішення проблеми стала ідея розповсюдження сигналу помилки від виходу мережі до її входу, шар за шаром. Алгоритми, які реалізують навчання мережі по цій схемі, отримали назву алгоритмів зворотного розповсюдження.

Алгоритм потребує диференційності активаційної функції на всій осі абсцис. З цієї причини, функція одинарного стрибку не може використовуватися і в якості стискаючої функції [24].

РОЗДІЛ 3. ГОЛОСОВЕ КЕРУВАННЯ ТЕХНІЧНИМИ ПРИСТРОЯМИ

3.1 Огляд Пристроїв для зчитування аудіо сигналів

Перші записи звуків проходили методом безпосереднього нарізання запису: за допомогою рогоподібних мікрофонів вібрація повітря передавалась на голку грамофона, яка і виконувала нарізання цих коливань на поверхні воскового циліндра, який обертався. Теперішні мікрофони перетворять коливання повітря в електричний сигнал, і хоча ця теорія лежить в основі роботи всіх мікрофонів, їх різновиди полягають в процесах, які виконують ці перетворення. Все різноманіття існуючих моделей зводиться до трьох основних типів мікрофонів.[18]

Динамічний мікрофон.

Цей вид мікрофонів користується особливою популярністю серед вокалістів. В порівнянні з рештою типів, динамічні мікрофони володіють цілим списком переваг: вони порівняно дешеві і в той же час надійні, можуть успішно працювати в областях високого звукового тиску. Стійкість до гучних звуків дозволяє використовувати ці мікрофони для зняття звуків ударів. Принцип його роботи нагадує роботу гучномовця, але в зворотному режимі. Як і в гучномовці, для проведення перетворення тут застосовуються коливання котушки індуктивності. Як видно на рис.3.1,

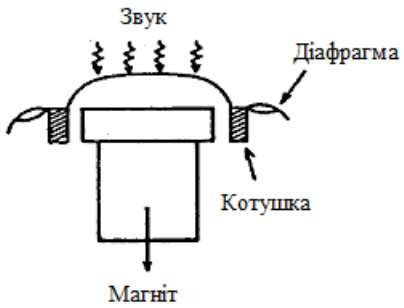

рис. 3.1. Структура динамічного мікрофона

Котушка індуктивності, розміщена навколо магнітного осердя, виконує коливальні рухи під впливом тиску повітряних хвиль, які відходять від джерела звуку. Генерація електричного току виникає кожен раз, коли відбувається

переміщення котушки (таке явище називається законом Флемінга). Таким чином, акустична енергія, яка потрапляє в мікрофон перетворюється в електричну, при цьому форма електричного струму відповідає характеру звукових коливань. Якщо провести посилення даного струму і відправити його до гучномовця, то можна почути посилений звук.

Особливості внутрішнього приладу динамічних мікрофонів обмежує кількість можливих типів діаграм спрямованості: такі мікрофони можуть бути тільки односпрямованими чи всеспрямованими. Окрім того, інерційність котушки, яка переміщується в магнітному полі з дротом обмежує частотний діапазон, який сприймається даними мікрофонами. На сьогоднішній день різними фірмами розроблено велику кількість різних динамічних мікрофонів вартістю від одиниць до декількох тисяч доларів, причому переважна кількість моделей спеціально орієнтовані на конкретне застосування.

Конденсаторний мікрофон.

Конденсаторні мікрофони являють собою більш вдосконалений тип, аніж динамічні. Принцип їх роботи заснований на зміні електричної ємності двох пластин, під час зміни відстані між ними. Одна з цих пластин жорстко закріплена в корпусі мікрофона, а інша (тонка) виступає в ролі мембрани. Коливання мембрани, викликані коливанням повітря, приводять до зміни ємності. Для роботи такого мікрофону на його пластини необхідно подати електричну напругу. Для цього переважно, використовується напруга в 48 Вольт, яка подається на мікрофон безпосередньо з посилювача чи мікшерного пульту.

Дані мікрофони переважно використовуються професіоналами, тому що, вони забезпечують більш високу якість звучання, але є досить дорогим типом мікрофонів, через складність в виробництві. Але з моменту їх появи, вони отримали чималу популярність і стали поширеною моделлю в музикальному світі. Принцип побудови конденсаторного мікрофона показаний на рис. 3.2.

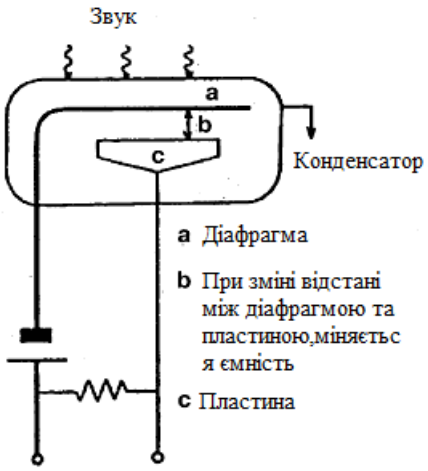

рис. 3.2. Структура конденсаторного мікрофона

Під час надходження електричного струму на діафрагму і пластину, останні утворюють конденсатор, коли звукові коливання виробляють тиск на діафрагму, вони змушують її переміщуватися, що призводить до зміни відстані між діафрагмою і пластиною, тим самим, змінюючи ємність між ними. Ці зміни відповідають формі сигналу. Після чого сигнал посилюється. Особливістю конденсаторного мікрофона є те, що він використовує електричний струм. Інша його характеристика полягає в застосуванні легкої діафрагми, замість важких котушок динамічних мікрофонів, що означає велику чутливість до звуку. Причому діафрагма конденсаторного мікрофона може бути реалізованою, як завгодно тонкою, що дозволяє домогтися відчутно більш високої чутливості і здатності до відтворення високих частот. Тому частотний діапазон даних мікрофонів значно ширший. Крім того, конденсаторні мікрофони можуть мати практично будь-яку, діаграму спрямованості. Звук, який лунає з конденсаторного мікрофону, відрізняється особливою витонченістю, навіть під час різких звукових впливів. Завдяки цих характеристикам конденсаторні мікрофони часто використовують, всюди, де пред’являються високі вимоги до якості і точності відтворення складових звуку.

Електретні мікрофони.

Принцип дії електретних мікрофонів аналогічний принципу дії конденсаторних, з тією відмінністю, що для їх роботи не потрібно зовнішнє джерело живлення. Мембрана таких мікрофонів отримує електричний заряд в

процесі виробництва, і для її харчування досить невеликої напруги (зазвичай близько 1,5 вольт), яка забезпечується в встановленій в мікрофоні батареї.

У порівнянні з конденсаторними, мембрана електретних мікрофонів значно товща, тому їх чутливість і частотні характеристики дещо гірші.

З точки зору спрямованості існують три основних типи мікрофонів - всеспрямовані, односпрямовані і двоспрямовані .

Всеспрямований мікрофон має однаковий вихідний рівень при будь-якому напрямку. Він покриває всі 360 градусів. Всеспрямований мікрофон вловлює максимальну кількість просторових звуків. При концертному застосуванні всенаправленний мікрофон повинен бути розташований дуже близько до джерела звуку, щоб був правильний баланс між безпосереднім і просторовим звуком. До того ж, ми не можемо відвернути всенаправленний мікрофон в сторону від непотрібних джерел звуку, таких як портали, що може викликати заведення (ефект зворотного зв`язку).

Односпрямований мікрофон найбільш чутливий до звуку, що приходить з одного напрямку, і менш чутливий до решти звуків. Типовою для таких мікрофонів є кардіоїдна характеристика (діаграма має форму серця). При ній найбільша чутливість досягається в напрямку вздовж осі мікрофона (0 градусів), а найменша - в протилежному (180 градусів). Ефективний кут роботи мікрофона становить 130 градусів, тобто по 65 градусів в будь-який бік від осі перед мікрофоном. Таким чином кардіоїдний мікрофон вловлює близько третини просторових звуків у порівнянні з всеспрямованим. Однонаправлені мікрофони відокремлюють потрібний прямий звук від непотрібних бічних і просторових.[23]

Двоспрямовані мікрофони добре приймають сигнали , які поступають з переду так і ззаду. На рис. 3.3 показані схеми спрямованості цих трьох типів мікрофонів.

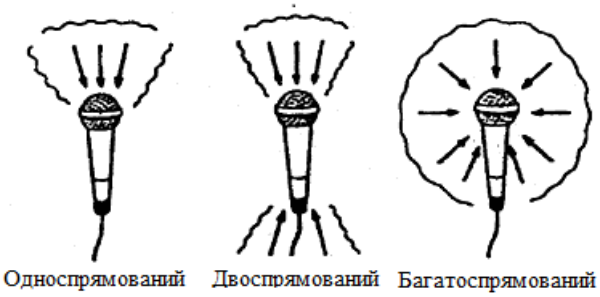

рис. 3.3. Спрямованість мікрофонів.

Кут зняття звуку.

Під кутом зняття звуку (рис. 3.4) мається на увазі зона можливого розташування джерела звукового сигналу, всередині якого не спостерігається значної втрати ефективності мікрофону.

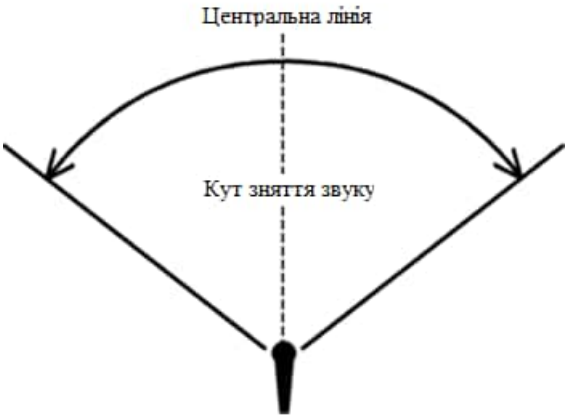

рис. 3.4. Схема зняття звуку

Збалансовані і незбалансовані мікрофони.

Терміни «збалансований» та «незбалансований» показують як посилається сигнал. Основне завдання застосування збалансованого мікрофону – в знищенні електричного шуму. По цій причині, саме збалансований мікрофон зазвичай застосовується при професійних записах.

Кабель для збалансованого мікрофону (рис. 3.5) складається з трьох провідників, які називають: «червоний», «білий» та «екран».

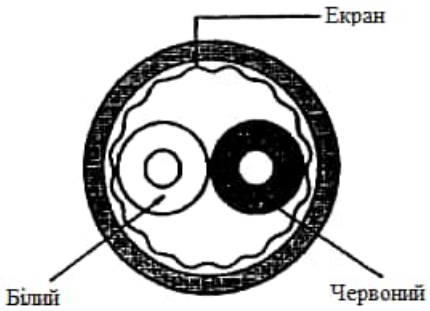

рис. 1.5. Переріз кабелю для збалансованого мікрофону.

Екран покриває червоний і білий дроти, захищаючи їх від небажаного зовнішнього електричного струму. Особливо важливо при використанні довгого кабелю для мікрофона. Для збалансованого мікрофона використовуються виходи, які мають трьох штиркову контактну групу.

Незбалансовані мікрофони використовують в власних з'єднаннях тільки два провідника, один з яких білий і екранний одночасно. Ці мікрофони цілком задовольняють різні вимоги, але тільки при невеликій довжині кабелю. В якості виходу для них використовується стандартний 1/4″ телефонний штекер.

3.2 Підбір пристрою для розробки аудіо системи.

Під час вивчення інструкції по експлуатації мікрофону ви, напевно зустрічали таке поняття як «опір» або «імпеданс». Це електротехнічний термін, який показує ступінь опору перемінного струму, який виміряється в Омах [Ом]. З точки зору опору відрізняють високоомні (10-50 кОм) та низькоомні (50-600 Ом) мікрофони. Якщо вихідний опір мікрофона не відповідає вхідному опору підсилювача, це може привести до зниження відношення корисного сигналу до шуму. На це потрібно звертати увагу. Хоча ці два опори не повинні бути практично однаковими, мікрофонний опір повинен знаходитись в допустимих межах величини опору підсилювача. Якщо підсилювач володіє низькоомним входом, то вхідний опір мікрофона повинен бути таким ж. Якщо ви під'єднуєте високоомний мікрофон до низькоомному входу підсилювача, то відбудеться зниження потужності сигналу.

Перевага практичного використання низькоомного мікрофона полягає в його низькій чутливості до шуму, навіть при довгих кальних з'єднаннях.

Вибір мікрофона.

Під час вибору мікрофону потрібно враховувати, як весь перелік його технічних характеристик, так і умови запису, тому конкретні рекомендації дати доволі важко. Проте загальні правила вибору мікрофону, все ж існують.

Ненаправлений мікрофон можна використовувати під час запису голосу і звуків в сильно приглушеному приміщенні. Його ж слід використовувати для передачі загальної акустичної обстановки під час багато мікрофонного запису.

Односторонньо направлений мікрофон з характеристикою типу «кардіоїда» бажано використовувати під час запису в приміщенні з великою кількістю звукових відображень. Використовують його и в випадку, коли в приміщенні, де проводять запис, проникають сторонні шуми. Мікрофон слід встановлювати тильною стороною до джерела звукових перешкод. Такий мікрофон рекомендується використовувати при широкому фронті розміщення виконавців. Даний мікрофон використовують при малоймовірній в аматорських умовах багато мікрофонного запису для чіткого розділення груп звуків, а також при розміщенні джерела звуку близько до мікрофону, щоб знизити низькочастотні спотворення, притаманні в цьому випадку ненаправленому і двохсторонньо направленому мікрофонам.

Двохсторонньо направлений мікрофон з діаграмою типу «вісімка» слід, використовувати при записі в заглушеному приміщенні, коли необхідно збільшити відносний рівень перевідбитих сигналів, а також при записі окремих звуків і голосів для виділення низьких частот в умовах близького розміщення, джерел звуку біля мікрофону. Використовують такий мікрофон і в тому випадку, коли необхідно відсторонитися від направлених джерел шуму. Для цього мікрофон орієнтують нульовою зоною чутливості до джерела шуму.

Для остаточного вибору потрібно визначитись в якому звуковому полі використовуватиметься мікрофон.

Пряме звукове поле.

Поняття «пряме звукове поле» описує той випадок, коли звук досягає мікрофон, не відбиваючись попередньо від стін, стелі, підлоги чи інших поверхонь (рис. 3.6.).

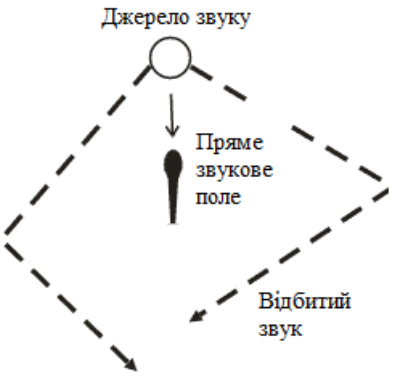

рис. 3.6 Пряме звукове поле і відбитий звук.

Якщо мікрофон знаходиться в прямому звуковому полі, то ви можете направити його нульовою зоною чутливості до джерела небажаного шуму. Даний прийом допомагає суттєво скоротити ефект зворотного зв’язку і уникнути витоку звука.

Відбите звукове поле.

Поняття «відбите звукове поле» описує ситуацію, коли звук, перед тим, як достигнути мікрофона, відбивається від стіни, стелі, підлоги або іншої поверхні(рис 3.7).

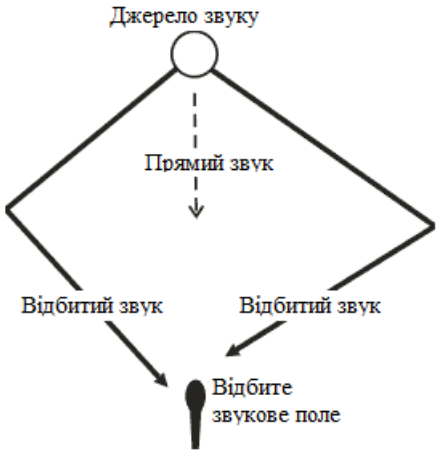

Рис. 3.7 Відбите звукове поле.

В даному випадку, для повного контролю над небажаним шумом недостатньо маніпуляцій з нульовою зоною чутливості мікрофону. Проте і в

відбитому звуковому полі однонаправлені мікрофони забезпечують більш надійний захисти від ефекту зворотного зв’язку сторонніх сигналів в порівнянні з багатонаправленим. Числовою мірилом ефективності мікрофону в відбитому звуком полі є коефіцієнт спрямованості (directivity index), який находиться в зворотній залежності від кута зняття звуку. Значення коефіцієнта для мікрофонів з різними діаграмами спрямованості, зображені в табл. 3.1:

Табл. 3.1

Діаграми спрямованості.

Спрямованість мікрофону	Коефіцієнт спрямованості	Відкидання шуму (дБ)
Всенаправленний	1	0 дБ
Кардіоїдний	1.7	4.8 дБ
Суперкардіоїдний	1.9	5.7 дБ

Дані в таблиці свідчать про те, що коефіцієнт спрямованості суперкардіоїдного мікрофона в 1.9 рази перевищує відповідний показник для всенаправленого мікрофона. В реальних умовах це виражається в наступному: в відбитому звуковому полі суперкардіоїдний мікрофон знімає на 5.7 дБ менше реверберацийних шумів.[19,20,21,22]

РОЗДІЛ 4. ЕЛЕМЕНТИ РОЗРОБКИ ДОДАТКУ

4.1 Необхідні елементи для створення додатку

Середовище розробки

Для створення додатку, використовувалось інтегроване середовище розробки програмного забезпечення та інших інструментальних засобів Microsoft Visual Studio. Саме в цьому середовищі можна створити програми з графічним інтерфейсом.

Для створення графічних інтерфейсів за допомогою платформи .Net(мова програмування C#) застосовуються різні технології - Windows Forms, WPF, додатки для магазину Windows Store (для ОС Windows 8 / 8.1 / 10). Однак, найбільш простою і зручною платформою досі залишається Windows Forms або форми, що і були використані у даному проекті. Windows Forms дають можливість легко та зрозуміло описати та реалізувати якийсь інтерфейс.

Основні компоненти платформи:

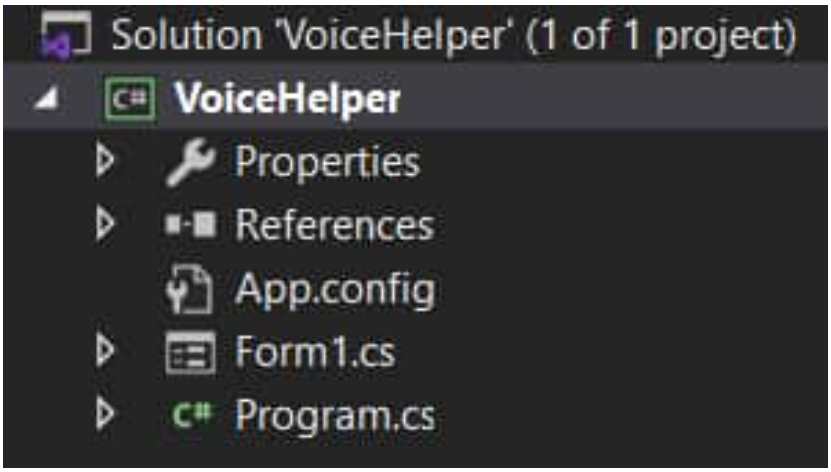

Рис. 4.1 Базова структура додатку

App.config – файл конфігурації(зберігає налаштування додатку)

References – зберігаються посилання, які містять відомості необхідні Visual Studio для пошуку компонентів або служб [14].

Form1.cs – код форми(програмна логіка-функціонал) [15].

4.2 Основні інструменти створення вигляду додатку

Вигляд та його функціонування

```
this.textBox1.Location = new System.Drawing.Point(53, 62);
this.textBox1.Multiline = true;
this.textBox1.Name = "textBox1";
this.textBox1.Size = new System.Drawing.Size(133, 121);
this.textBox1.TabIndex = 0;
//
// textBox2
//
this.textBox2.Location = new System.Drawing.Point(245, 62);
this.textBox2.Multiline = true;
this.textBox2.Name = "textBox2";
this.textBox2.Size = new System.Drawing.Size(133, 121);
this.textBox2.TabIndex = 1;
```

Рис 4.2. Елемент для роботи з текстом

Text.Box – це елемент призначений для роботи з текстом (рис. 4.2) (ввід, вивід. Редагування, читання тексту).[16]

```
MessageBox.Show("No internet connection");
```

Рис. 4.3 Елемент виводу повідомлення

Message box –це елемент для виводу повідомлення (рис 4.3).[17]

Отже, саме за допомогою об’єднання цих елементів, в їх різних інтерпретаціях, з елементарними операторами мови C#, отримуємо класи та функції, які забезпечують функціонування вигляду даного додатку, тобто ввід, вивід, обробка інформації.

4.3 Функціонування методів та алгоритмів розв’язання

Form1.cs, файл в якому показана реалізація даного додатку.

Основні функції та методи:

KillProg – метод, для закриття певного файлу за допомогою (далі використовується в голосових командах)

Say – метод, для подальшого керування/виводу відповідей додатку.

rec_SpeachRecognized – метод, для вводу команд та відповідей на них.

restart – для запуску програми, в подальшому викликається в rec_SpeachRecognized для запуску певної програми(програму можна змінювати).

text_field – поле в якому пише текст(команди), який хочемо додати до файлу.

add_to_file – реалізація кнопки, за допомогою, якої додаємо текст(команду) написану у text_field у файл, який містить команди.

open_file – реалізація кнопки, за допомогою, якої відкриваємо файл у якому містяться всі команди

support form – реалізація кнопки, за допомогою, якої ми переходимо у Google Form, в якій можна залишити фідбек про програму або повідомити про проблему чи баг

4.4 Програмна реалізація

Умови коректного функціонування.

Додаток надає змогу за допомогою голосових команд, виконувати певні дії(прості відповіді, маніпуляції з самим додатком, з текстом, з вікном програми керування деякими іншими програмами(Spotify, PowerPoint, Google Chrome), відкриття посилань та пошуку в пошуковій системі). Для цього потрібен файл **commands.txt** в якому прописані, всі команди, які повинен виконувати додаток. Доступ до файлу **commands.txt** виконаний за допомогою інтерфейсу програми. Додавання нових слів(команд) доступне в інтерфейсі додатку. Для розуміння та синтезу мови, використовувалась програма Windows Speech Recognition, яка розроблена Microsoft для розпізнавання мови, що дозволяє голосовим командам керувати інтерфейсом. І обов’язковим також є мікрофон, для подачі голосових команд, додатку.

4.5. Функціонал додатку

Інтерфейс, команди та відповіді реалізовано англійською мовою, для універсальності.

Розглянемо роботу додатку за допомогою знімків екрану та короткого опису дій.

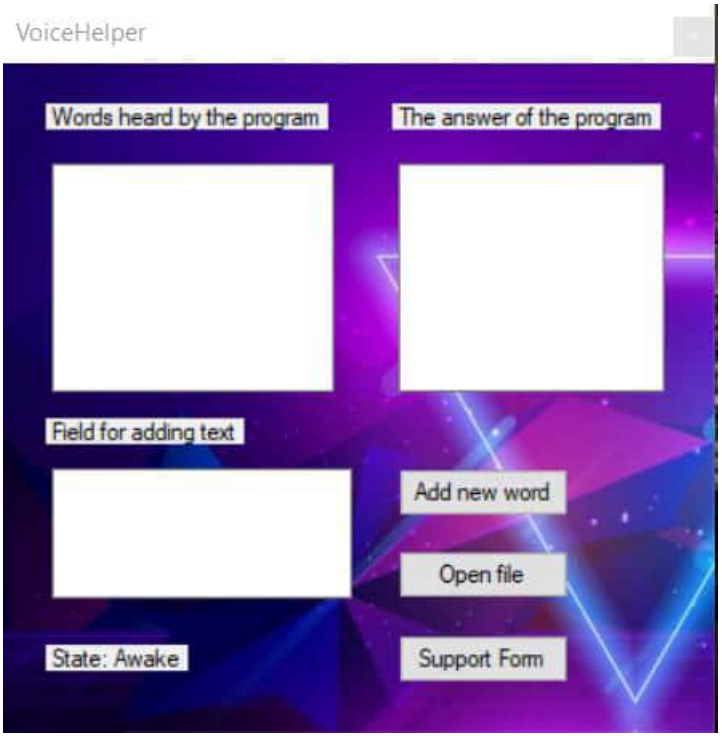

Рис. 4.4. Головне вікно додатку.

На рис. 4.4 показано головне вікно додатку. Все досить просто та зрозуміло, але я досить коротко розкажу про все.

Поле, яке називається «Words heard by the program» - поле в якому відображаються слова(команди), які програма почула від нас та синтезувала їх. Кожне нове слово (команда) додаватиметься в новий рядок.

Поле, яке називається «The answer of the program» - поле в якому ми бачимо відповідь програми на нашу команду, вказану в полі «Words heard by the program». Відповідь ми також чуємо. Можна сказати, що в полі «The answer of the program» дублюється сказане програмою, але тільки в тому випадку, якщо для команди продумана відповідь, чи дія з відповіддю, а не просто дія.

Поле, яке називається «Field for adding text» - поле в якому ми пишемо слово, слова чи словосполучення, які хочемо додати до файлу, якому містяться всі команди та слова.

«State:Awake» або «State: Sleep» - ярлик, який показує статус програми. В програмі реалізовані два статуси: Awake-активний(за замовчуванням) та Sleep-неактивний. Коли додаток знаходиться в статусі Awake, то він готовий приймати та виконувати команди і відповідно в статусі Sleep – не готовий, додаток тільки

«чує» та відображає наші слова в полі «Words heard by the program», але нічого не виконує.

Кнопка «Add new word» - кнопка в якій реалізовано додавання слів, тексту, словосполучень, які написані в полі «Field for adding text» до файлу.

Кнопка «Open file» - кнопка в якій реалізовано, відкривання файлу, в якому містяться всі команди та слова, фрази та словосполучення.

```
list.Add(File.ReadAllLines(@"D:\Education\KursovaVoiceHelper\Program\VoiceHelper commands\commands.txt"));
```

Рис. 4.5. Зчитування з файлу

На рис. 4.5 зображено зчитування з файлу всіх доступних команд.

Кнопка «Support Form» - кнопка в якій реалізовано, відкривання посилання, в якому містяться технічна підтримка додатку, яка реалізована в Google Forms.

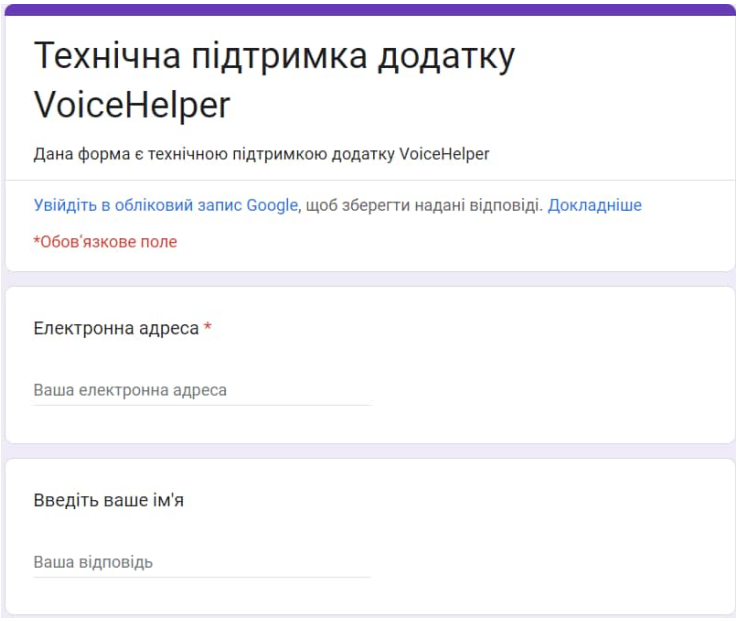

Рис. 4.6 технічна підтримка яка реалізована в Google Forms

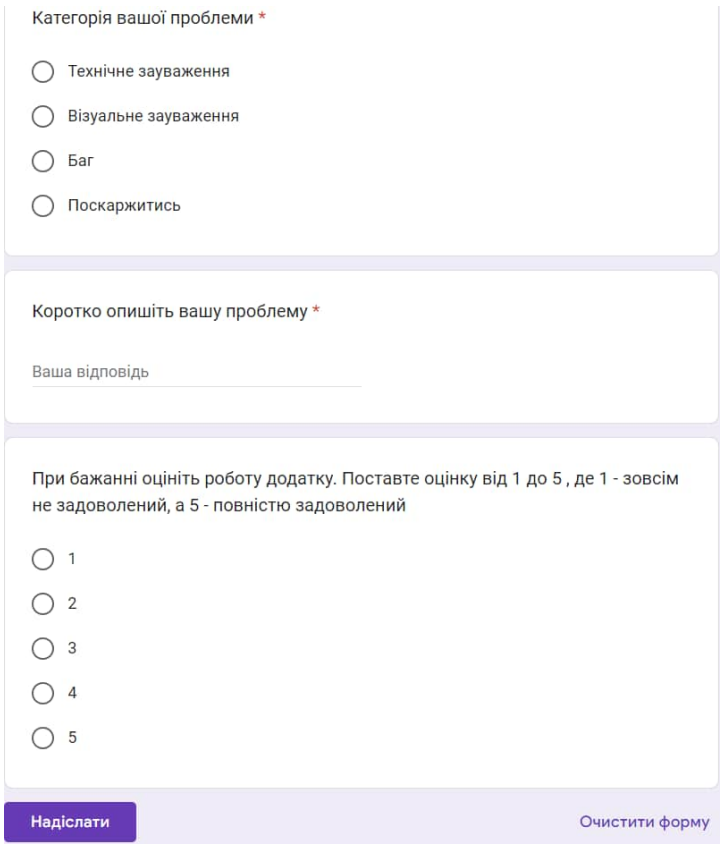

Рис. 4.7 технічна підтримка, яка реалізована в Google Forms

На рис. 4.6 – 4.7 зображена технічна підтримка додатку. Технічна підтримка створена у вигляді Google Forms, де створені запитання, після заповнення яких, я матиму змогу реагувати на зауваження чи побажання та виправляти їх. Всі відповіді на мою Google Form, я отримую у вигляді сповіщення на мою електронну адресу.

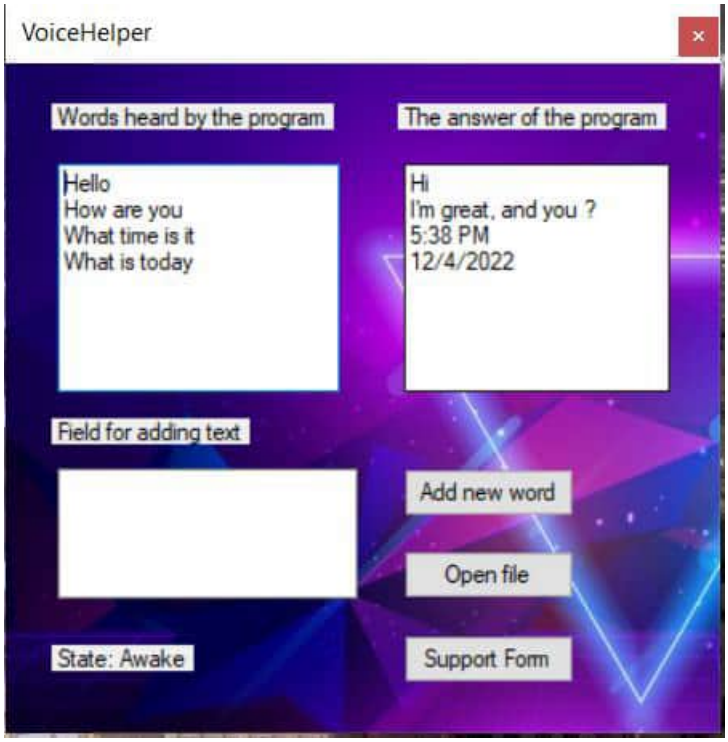

Рис. 4.8 Приклад роботи додатку №1

На рис. 4.8,зображено приклад роботи додатку, у вигляді зчитування моїх команд, та простих відповідей на них.

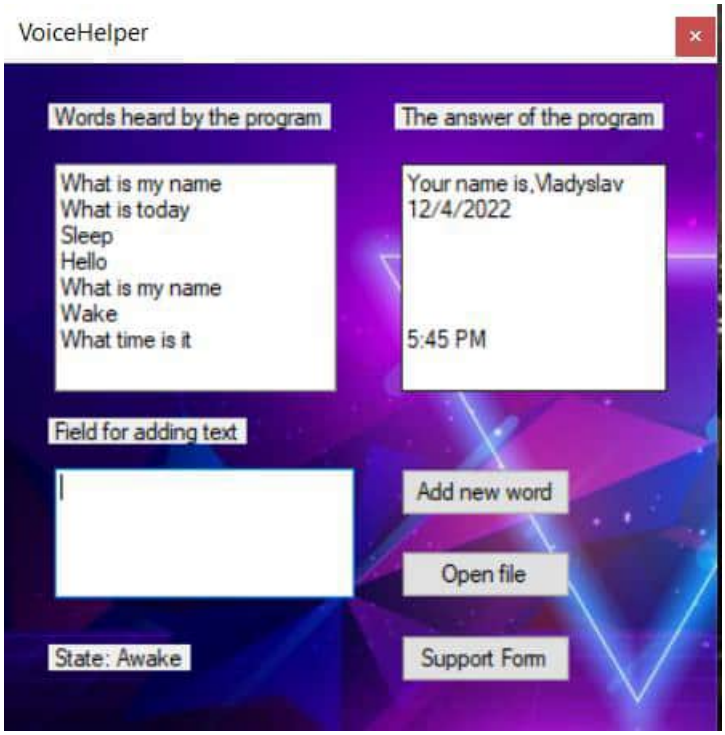

Рис. 4.9. Приклад роботи додатку №2

На рис. 4.9, зображено приклад роботи додатку, а саме його функція(статус) “Sleep” і “Awake”. При запуску додатку статус автоматично стоїть “Awake”, що означає, що програма розпізнає мої команди та виконує їх.

Після голосової команди “Sleep”, статус програми змінюється на “Sleep”, що означає, що програма чує та розпізнає мої команди, але нічого не виконує. Статус можна повернути назад до “Awake” , за допомогою команди “Wake”, після чого програма працюватиме у звичному режимі

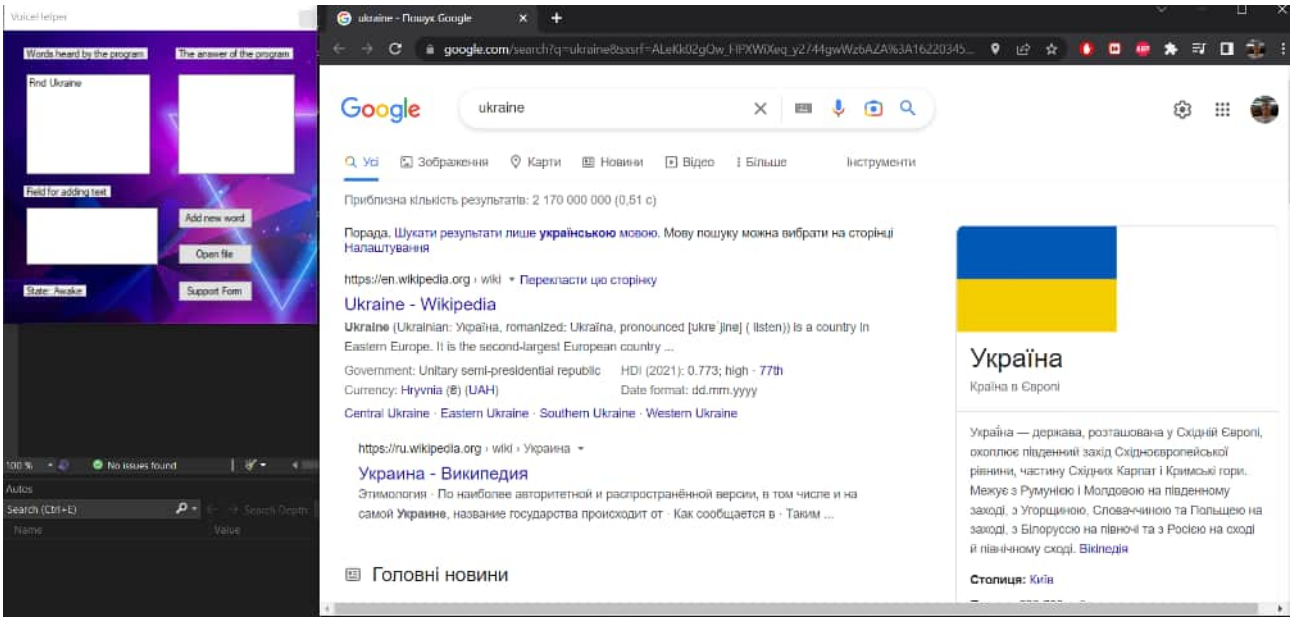

Рис. 4.10. Приклад роботи додатку №3

На рис. 4.10 зображено, приклад роботи додатку, а саме голосовий пошук в пошуковій системі Google. Пошук відбувався по ключовій команді “Find Ukraine”. Слово Ukraine в даному випадку виступає, як звичайний приклад, на місці цього слова може будь яке інше зі всіх слів, які є додані до мого словника.

На рис. 4.8-4.10, зображено далеко не весь функціонал. Роботу більшості команд нереально показати на знімках екрану, тому більш детальні приклади роботи програми, будуть представлені у коротеньких відео та додані до презентації.

ВИСНОВОК

Метою мого завдання була реалізація додатку, за допомогою якого можна було б керувати прикладними додатками ОС Windows (MS Word, MS Power Point, Google Chrome, Spotify та інші) з використанням голосового інтерфейсу.

Реалізована можливість пошуку в пошуковій системі - https://google.com.ua

Додаток має як переваги так і недоліки. Головними перевагами я вважаю, це його коректна робота, виконання поставлених завдань та простота інтерфейсу. Основні недоліки на мою думку наступні: відносно мала кількість команд, сприйняття тільки англійської ,час виконання, якість розпізнавання слів. Якість розпізнавання слів та сприйняття тільки англійської мови це особливість програми Windows Speech Recognition, яку я використовував для зчитування та синтезування слів.

Моя програма може бути досить корисною, для певного типу користувачів, а саме для таких, які звикли та яким легше користуватись голосовим інтерфейсом, замість того, щоб виконувати ці є маніпуляції за допомогою кліків.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Пітер Норвіг, Стюарт Расселл, “Штучний інтелект: сучасний підхід”,1995

2. Maurice Herlihy, “The Art of Multiprocessor Programming”, 2012

3. Dong Yu “Automatic Speech Recognition: A Deep Learning Approach (Signals and Communication Technology) 2015th Edition”, 2015

4. Dong Yu, “Automatic Speech Recognition A Deep Learning Approach”,2014

5. **“**General Phonetics: For Missionaries and Students of Languages (Classic Reprint)” Hardcover, 2018

6. Mota Soares, “Computional Mechanics: Solids, Structures and Coupled Problems”,2006

7. Farid Golnaraghi, “Automatic Control Systems 9 th Edition”, 2006

8. Oviatt, S. L. Multimodal interfaces. In The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, J. Jacko and A. Sears, Eds. Lawrence Erlbaum Assoc. Mahwah, NJ, chap.14, 2003. — pp. 286–304.

9. P. Juell and R. Marsh, " A hierarchical neural network for human face detection," Pattern Recog. 29, pp. 781-787, 1996.

10. S.-H. Lin, S.-Y. Kung, and L.-J. Lin, " Face recognition/detection by probabilistic decision-based neural network ," IEEE Trans. Neural Networks 8, pp. 114-132., 1997.

11. H. A. Rowley, S. Baluja, and T. Kanade, " Neural network-based face detection ," IEEE Trans. Pattern Anal. Mach. Intell. 20, pp. 23-38., January 1998

12. Mark Gales. “Application of Hidden Markov Models in Speech Recognition (Foundations and Trends in Signal Processing)”, 2008

13.Wikipedia: [Електронний ресурс] – Режим доступу : https://uk.wikipedia.org/wiki/%D0%93%D0%BE%D0%BB%D0%BE%D0%B2%D

0%BD%D0%B0_%D1%81%D1%82%D0%BE%D1%80%D1%96%D0%BD%D0%BA%D0%B0.

14. K.-R. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf. " An introduction to kernel-based learning algorithms ," IEEE Transactions on Neural Networks, 12(2), pp. 181-201, 2001.

15. Jochen Maydt and Rainer Lienhart. " Face Detection with Support Vector Machines and a Very Large Set of Linear Features ," IEEE ICME 2002, Lousanne, Switzerland, pp. xx-yy, Aug. 2002

16. D. Roth, M.-H. Yang, and N. Ahuja, " A SNoW-based face detector ," in Advances in Neural Information Processing Systems 12 (NIPS 12), MIT Press, Cambridge, MA, pp. 855-861, 2000.

17. D. Roth, " The SNoW Learning Architecture ," Technical Report UIUCDCS-R-99-2102, UIUC Computer Science Department, 1999.

18. K.-R. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf. " An introduction to kernel-based learning algorithms ," IEEE Transactions on Neural Networks, 12(2), pp. 181-201, 2001.

19. Керування посиланнями в проекті :[Електронний ресурс] – Режим доступу:https://docs.microsoft.com/ru-ru/visualstudio/ide/managing-references-in-a-project?view=vs-2019.

20. Основи форм(WF) : [Електронний ресурс] – Режим доступу : https://metanit.com/sharp/windowsforms/2.1.php .

21. Робота з текстом: [Електронний ресурс] – Режим доступу : https://metanit.com/sharp/windowsforms/4.3.php .

22. Робота з повідомленнями: [Електронний ресурс] – Режим доступу : https://metanit.com/sharp/windowsforms/4.19.php.

23. Винцюк Т.К. “Аналіз,розпізнавання та інтерпретація мовних сигналів”, 1987.

24. Алгоритм зворотного розповсюдження [Електронний ресурс] – Режим доступу:https://uk.wikipedia.org/wiki/%D0%9C%D0%B5%D1%82%D0%BE%D0

%B4_%D0%B7%D0%B2%D0%BE%D1%80%D0%BE%D1%82%D0%BD%D0%BE%D0%B3%D0%BE_%D0%BF%D0%BE%D1%88%D0%B8%D1%80%D0%B5%D0%BD%D0%BD%D1%8F_%D0%BF%D0%BE%D0%BC%D0%B8%D0%BB%D0%BA%D0%B8