# ДОДАТОК А

Графічний матеріал кваліфікаційної роботи

1

**ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ РАДІОЕЛЕКТРОНІКИ**
**КАФЕДРА ЕОМ**

**Кваліфікаційна робота**

**Другий рівень (магістр)**

# МЕТОДИ ОБРОБКИ ГНУЧКИХ ГОЛОСОВИХ ЗАПИТІВ В АВТОМАТИЗОВАНИХ СИСТЕМАХ КЕРУВАННЯ ЗАВДАННЯМИ

**Автор**
Давидов Я.А.
ст. гр. КСМм-23-1

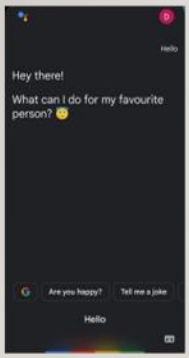
**Керівник**
Барковська О.Ю.
доц. каф. ЕОМ

2

## ОГЛЯД ПРОБЛЕМНОЇ ОБЛАСТІ

3

## АНАЛІЗ ІСНУЮЧИХ РІШЕНЬ




Google Assistant


Apple Siri

Microsoft Cortana


Amazon Alexa

4

## МЕТА І ЗАДАЧІ КВАЛІФІКАЦІЙНОЇ РОБОТИ

***Мета роботи полягає у розробці ефективних методів обробки гнучких голосових запитів для автоматизованих систем керування завданнями.***

– огляд методів розпізнавання голосових команд;
– оцінка ефективності голосового управління на основі існуючих рішень;
– розробка архітектури системи розпізнавання мовлення із гнучкими запитами;
– проведення експериментів оцінки продуктивності ResNet у залежності від рівня шуму та методів виділення ключових ознак;
– дослідження впливу використання направлених в сторону джерела звуку мікрофона на точність детектування голосових команд;
– аналіз результатів експериментів.

5

## АКТУАЛЬНІСТЬ ОБРАНОЇ ТЕМИ


| ID | Основне слово | Синоніми | Категорія |
|---|---|---|---|
| 1 | Хліб | Батон, булка, випічка… | Продукти |
| 2 | Ковбаса | Салямі, сосиски… | Продукти |
| 3 | Макарони | Вермішель, паста… | Продукти |
| 4 | Олія | Масло, Соняшникова… | Продукти |
| 5 | Десерт | Торт, тістечко, печиво… | Продукти |
| 6 | Фрукти | Ягоди, цитрусові, плоди… | Продукти |
| 7 | Вода | Мінералка, газованка… | Напої |
| 8 | Кава | Еспресо, капучино… | Напої |
| 9 | Корм для тварин | Корм, їжа для котів, вологий корм… | Напої |
| 10 | Пакет | Сумка, кульок… | Товари |
| 11 | Доставка | Привезти, доставити… | Послуга |
| 12 | Знижка | Акція, розпродаж, вигода… | Обслуговування |
| 13 | Кошик | Корзина, кошелка, кошик для покупок… | Обслуговування |

Приклад бази даних синонімів на прикладі Супермаркету

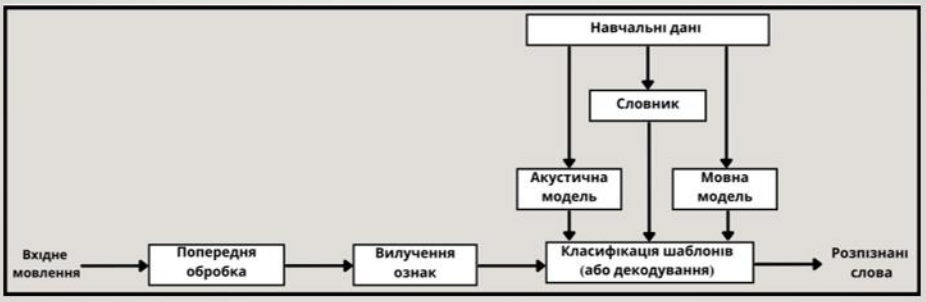
6

## ГНУЧКІ ГОЛОСОВІ ЗАПИТИ

Гнучкі голосові запити – це голосові команди, які формулюються користувачем у довільній або нестандартній формі, з використанням різних слів, фраз і мовних конструкцій для вираження одного й того ж наміру. Вони відрізняються від чітких та шаблонних команд тим, що не мають фіксованої структури й можуть містити варіативні елементи мовлення.

7

## ПРИНЦИПИ РОБОТИ ТРАДИЦІЙНИХ РІШЕНЬ

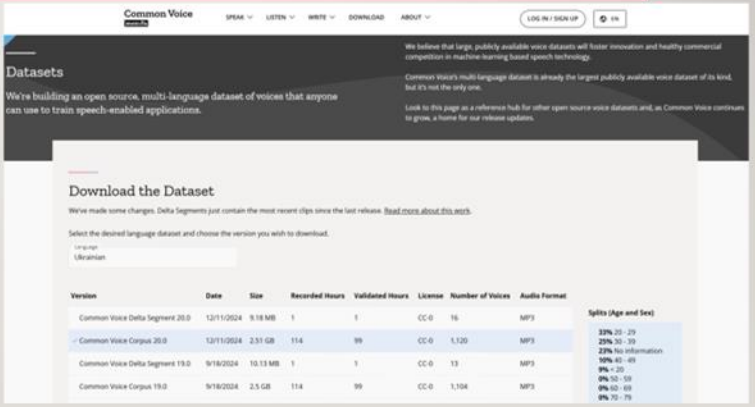



Компоненти існуючих систем розпізнавання мовлення

8

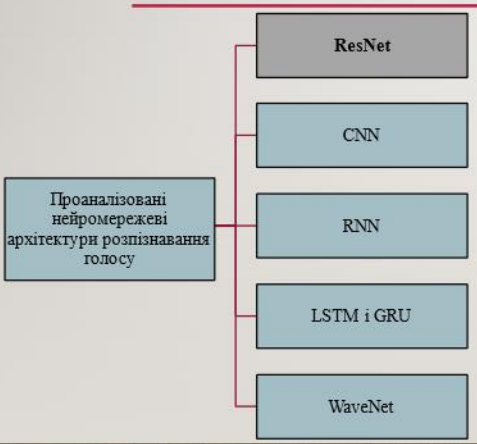
## ОБРАНИЙ ДАТАСЕТ

У версії 20.0 Mozilla Common Voice (MCV) загальний обсяг мовних даних становить 33 150 годин, із яких 22 108 годин були підтверджені спільнотою шляхом краудсорсингової перевірки якості. Корпус містить в собі записи 133 мов, включаючи українську.



9

## ВИБІР НЕЙРОМЕРЕЖЕВОЇ АРХІТЕКТУРИ




Обрана архітектура:

– 18-шаровий ResNet;

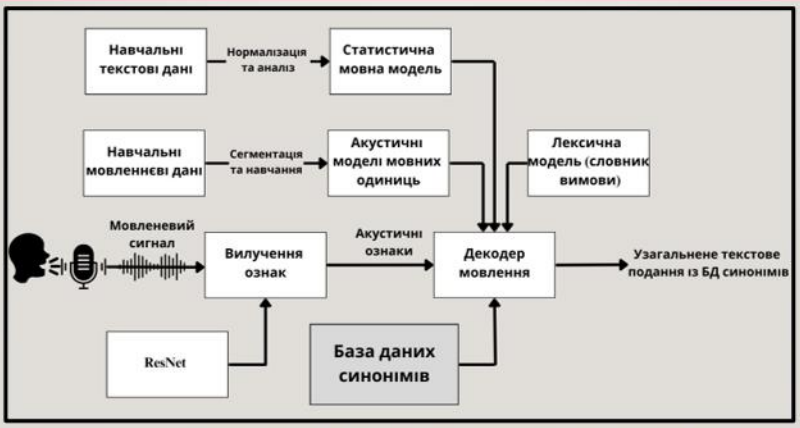
– 3x3 згортка з 64 фільтрами та пулінг для зменшення розмірності;

– блоки згорток із фільтрами (64, 128, 256, 512), підвибіркою (stride = 2) і dropout для регуляризації;

– глобальний усереднений пулінгом і повнозв'язний шар для класифікації.

10

## ЗАПРОПОНОВАНА МОДЕЛЬ РОЗПІЗНАВАННЯ ГОЛОСОВИХ ЗАПИТІВ




Архітектура запропонованої системи розпізнавання гнучких голосових запитів

11

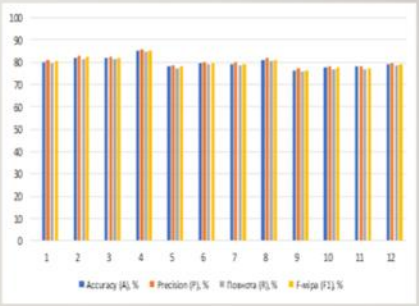
## МЕТОДОЛОГІЯ ОЦІНКИ ВПЛИВУ РІВНЯ ЗАШУМЛЕНОСТІ, МЕТОДУ ВИЗНАЧЕННЯ КЛЮЧОВИХ ОЗНАК, МЕТОДІВ ФІЛЬТРАЦІЇ НА ТОЧНІСТЬ РОЗПІЗНАВАННЯ МОВЛЕННЯ

| № експерименту | Рівень зашумленості | Метод виділення ознак | Метод фільтрації шуму | Мета експерименту |
|---|---|---|---|---|
| Експерименти №1-24 | Низький (SNR ≥ 30 дБ) / Високий (SNR ≤ 15 дБ) | Mel-Spectrogram | Без фільтрів | Оцінка базової точності розпізнавання голосової команди за умов низької зашумленості та використання MFCC |
| | | | Low-pass фільтр | |
| | | | High-pass фільтр | |
| | | | Комбінований (low-pass + high-pass) | |
| | | STFT | Без фільтрів | Оцінка базової точності розпізнавання голосової команди за умов низької зашумленості та використання STFT |
| | | | Low-pass фільтр | |
| | | | High-pass фільтр | |
| | | | Комбінований (low-pass + high-pass) | |
| | | CQT | Без фільтрів | Оцінка базової точності розпізнавання голосової команди за умов низької зашумленості та використання CQT |
| | | | Low-pass фільтр | |
| | | | High-pass фільтр | |
| | | | Комбінований (low-pass + high-pass) | |

12

## ОТРИМАНІ РЕЗУЛЬТАТИ ЕКСПЕРИМЕНТІВ 1-12 (НЕНАПРАВЛЕНІ МІКРОФОНИ, SNR ≥ 30 ДБ )

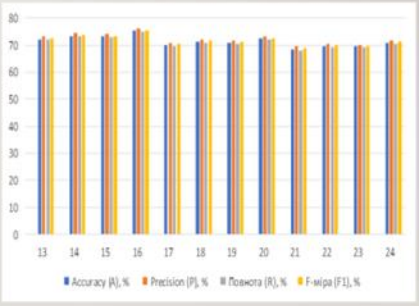
| Номер експерименту | Метод вилучення ознак | Метод фільтрації шуму | Accuracy (A), % | Precision (P), % | Повнота (R), % | F-міра (F1), % |
|---|---|---|---|---|---|---|
| 1 | Mel-Spectrogram | Без-фільтрів | 80,21 | 81,10 | 79,90 | 80,49 |
| 2 | Mel-Spectrogram | Low-pass | 82,10 | 82,90 | 81,70 | 82,29 |
| 3 | Mel-Spectrogram | High-pass | 81,80 | 82,60 | 81,40 | 81,99 |
| 4 | Mel-Spectrogram | Комбінований | 85,10 | 85,90 | 84,80 | 85,34 |
| 5 | STFT | Без-фільтрів | 78,10 | 78,90 | 77,50 | 78,19 |
| 6 | STFT | Low-pass | 79,50 | 80,30 | 79,00 | 79,60 |
| 7 | STFT | High-pass | 79,20 | 80,00 | 78,70 | 79,34 |
| 8 | STFT | Комбінований | 81,10 | 81,90 | 80,60 | 81,19 |
| 9 | CQT | Без-фільтрів | 76,50 | 77,20 | 75,80 | 76,49 |
| 10 | CQT | Low-pass | 77,80 | 78,40 | 77,00 | 77,73 |
| 11 | CQT | High-pass | 78,10 | 78,10 | 76,80 | 77,44 |
| 12 | CQT | Комбінований | 79,10 | 79,80 | 78,50 | 79,14 |



13

## ОТРИМАНІ РЕЗУЛЬТАТИ ЕКСПЕРИМЕНТІВ 13-24 (НЕНАПРАВЛЕНІ МІКРОФОНИ, SNR ≤ 15 ДБ )

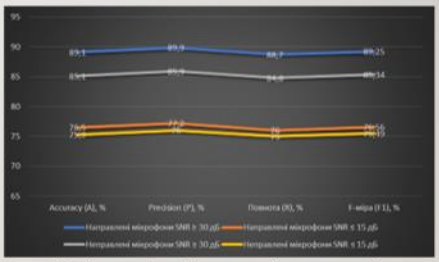
| Номер експерименту | Метод вилучення ознак | Метод фільтрації шуму | Accuracy (A), % | Precision (P), % | Повнота (R), % | F-міра (F1), % |
|---|---|---|---|---|---|---|
| 1 | Mel-Spectrogram | Без-фільтрів | 72,21 | 73,10 | 71,90 | 72,49 |
| 2 | Mel-Spectrogram | Low-pass | 73,50 | 74,40 | 73,20 | 73,70 |
| 3 | Mel-Spectrogram | High-pass | 73,20 | 74,10 | 72,80 | 73,45 |
| 4 | Mel-Spectrogram | Комбінований | 75,30 | 76,00 | 75,00 | 75,49 |
| 5 | STFT | Без-фільтрів | 70,10 | 71,00 | 69,80 | 70,39 |
| 6 | STFT | Low-pass | 71,30 | 72,20 | 71,00 | 71,50 |
| 7 | STFT | High-pass | 71,00 | 71,80 | 70,60 | 71,20 |
| 8 | STFT | Комбінований | 72,50 | 73,30 | 72,10 | 72,63 |
| 9 | CQT | Без-фільтрів | 68,50 | 69,40 | 68,10 | 68,66 |
| 10 | CQT | Low-pass | 69,70 | 70,60 | 69,30 | 69,86 |
| 11 | CQT | High-pass | 69,40 | 70,20 | 69,00 | 69,60 |
| 12 | CQT | Комбінований | 70,90 | 71,80 | 70,50 | 71,06 |




14

## МЕТОДОЛОГІЯ ОЦІНКИ ВПЛИВУ ВИКОРИСТАННЯ НАПРАВЛЕНИХ МІКРОФОНІВ НА ТОЧНІСТЬ РОЗПІЗНАВАННЯ МОВЛЕННЯ

| Номер експерименту | Рівень зашумленості | Обраний пайплайн обробки аудіоряду | Accuracy (A), % | Precision (P), % | Повнота (R), % | F-міра (F1), % |
|---|---|---|---|---|---|---|
| 25 | Низький (SNR ≥ 30 дБ) | MFCC ⇩ комбінований фільтр (low-pass + high-pass) | 89,10 | 89,90 | 88,70 | 89,25 |
| 26 | Високий (SNR ≤ 15 дБ) | MFCC ⇩ комбінований фільтр (low-pass + high-pass) | 76,50 | 77,20 | 76,00 | 76,56 |

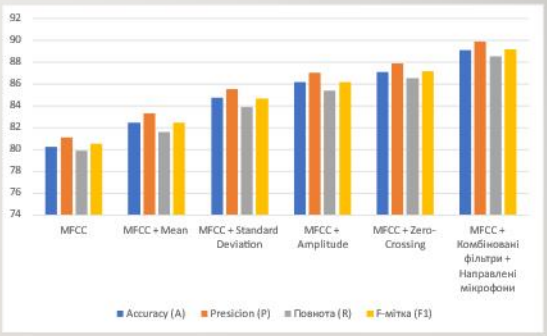



Порівняльний графік результатів направлених і ненаправлених мікрофонів

15

## ДЕМОНСТРАЦІЯ ПОКРАЩЕНЬ ЗАПРОПОНОВАНОГО ПІДХОДУ

| Вилучення ознак | Accuracy (A), % | Presicion (P), % | Повнота (R), % | F-міра (F1), % |
|---|---|---|---|---|
| MFCC | 80,21 | 81,10 | 79,90 | 80,49 |
| MFCC + Mean | 82,45 | 83,30 | 81,60 | 82,44 |
| MFCC + Standard Deviation | 84,75 | 85,50 | 83,90 | 84,69 |
| MFCC + Amplitude | 86,20 | 87,00 | 85,40 | 86,19 |
| MFCC + Zero-Crossing | 87,10 | 87,90 | 86,50 | 87,19 |
| MFCC + Комбіновані фільтри + Направлені мікрофони | 89,10 | 89,90 | 88,50 | 89,20 |




16

## ВИСНОВКИ

У процесі виконання кваліфікаційної роботи було розроблено методи обробки гнучких голосових запитів для автоматизованих систем керування завданнями. Розглянуто методи розпізнавання голосових команд такі як ResNet, CNN, RNN, WaveNet, LSTM і GRU. Виявлено що ResNet дає найвищу точність розпізнавання мовлення.

Запропоновано архітектуру системи, яка відрізняється від існуючих словником синонімів, що дозволяє підвищити ефективність роботи системи в умовах реального використання. Проведено експерименти які показують залежність продуктивності нейромережевої моделі ResNet від рівня шуму та методів виділення ключових ознак аудіосигналу. Результати свідчать, що використання вдосконалених підходів, таких як комбіновані фільтри суттєво покращує accuracy, presicion, повноту та F-міру навіть за умов високого рівня зашумленості.

Досліджено вплив використання направлених в сторону джерела звуку мікрофона, який показав суттєві покращення в порівнянні з ненаправленими мікрофонами точність зросла з 85,10% до 89,10% при SNR ≥ 30 дБ та з 75,30% до 76,50 % при SNR ≤ 15 дБ. Інші метрики також продемонстрували позитивну динаміку.

# 17 АПРОБАЦІЯ ОТРИМАНИХ РЕЗУЛЬТАТІВ

- Olesia Barkovska, Ihor Velykodnyi, Oleksii Liashenko, Ihor Ivanisenko, Yaroslav Davydov "Study of the architectural features of the ResNet neural network model for solving the task of speaker recognition" 2024 IEEE 5th KhPI Week on Advanced Technology. IEEE Conference, October 7 – 11, 2024, Kharkiv, Ukraine
- Давидов Я.А., Барковська О.Ю Методи обробки гнучких голосових запитів в автоматизованих системах керування завданнями // Проблеми інформатизації : XII міжнародна науково-технічна конференція. - 21-22 листопада 2024. –с.73. doi: https://doi.org/10.32620/PI.24.t2