

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ  
імені ІГОРЯ СІКОРСЬКОГО»

ФАКУЛЬТЕТ ЕЛЕКТРОНИКИ

(повна назва інституту/факультету)

КОНСТРУЮВАННЯ ЕЛЕКТРОННО-ОБЧИСЛЮВАЛЬНОЇ АПАРАТУРИ

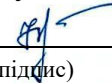
(повна назва кафедри)

«На правах рукопису»

УДК 004.35

«До захисту допущено»

Завідувач кафедри

 О. М. Лисенко  
(підпис) (ініціали, прізвище)

“\_15\_”\_січня\_\_\_2023 р.

**Магістерська дисертація**

зі спеціальності (спеціалізації) 172 – Телекомунікації та радіотехніка


(код і назва спеціальності)

на тему: Дослідження голосової активності в пристроях з низьким енергоспоживанням

Виконав: студент II курсу, групи ДК-21мп

(шифр групи)

Демиденко Олексій Анатолійович  
(прізвище, ім'я, по батькові)

  
(підпис)

Науковий керівник

к.т.н., доц. Лебедев Д. Ю.  
(посада, науковий ступінь, вчене звання, прізвище та ініціали)

  
(підпис)

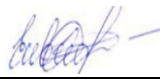
Консультант

\_\_\_\_\_  
(назва розділу) (науковий ступінь, вчене звання, прізвище, ініціали)

\_\_\_\_\_  
(підпис)

Рецензент к.т.н., доц. Іванько К.О.

(посада, науковий ступінь, вчене звання, науковий ступінь, прізвище та ініціали)

  
(підпис)

Засвідчую, що у цій магістерській дисертації  
немає запозичень з праць інших авторів без  
відповідних посилань.

Студент   
(підпис)

Київ – 2023 року

**Національний технічний університет України  
«Київський політехнічний інститут  
імені Ігоря Сікорського»**

Інститут/факультет електроніки

(повна назва)

Кафедра конструювання електронно-обчислювальної апаратури

(повна назва)

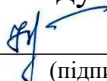
Рівень вищої освіти – другий (магістерський) за освітньо-професійною програмою  
«Інформаційно-обчислювальні засоби радіоелектронних систем»

Спеціальність (спеціалізація) 172 – Телекомунікації та радіотехніка

(код і назва)

**ЗАТВЕРДЖУЮ**

Завідувач кафедри

  
(підпис)

О.М. Лисенко

(ініціали, прізвище)

«15» січня 2023р.

**ЗАВДАННЯ**

**на магістерську дисертацію**

студенту Демиденко Олексію Анатолійовичу

(прізвище, ім'я, по батькові)

1. Тема дисертації Дослідження голосової активності в пристроях з  
низьким енергоспоживанням

науковий керівник дисертації

Лебедев Д. Ю., к.т.н., доц.,

(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затвержені наказом по університету від «9» Листопада 2023р. № 5218-с

2. Строк подання студентом дисертації 4 січня 2023 року

3. Об'єкт дослідження: Процеси, що використовуються у дослідженнях голосової активності у системах з низьким енергоспоживанням.

4. Предмет дослідження: Методи дослідження голосової активності у пристроях з низьким енергоспоживанням.

5. Перелік завдань, які потрібно розробити: 1. Аналіз існуючих патентних рішень, наукових статей та літератури на задану тему. 2. Детальне дослідження методу дослідження голосової активності у системах з малим енергоспоживанням.

3. Розробка системи, що реалізовує даний метод. 4. Перевірка системи шляхом моделювання у програмному комплексі. 5. Розроблення стартап проекту.

6. Перелік графічного (ілюстративного) матеріалу Презентація у форматі PowerPoint

7. Орієнтовний перелік публікацій : 1 публікація.

8. Консультанти розділів дисертації

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

9. Дата видачі завдання 01.02.2023р.

### КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів виконання магістерської дисертації	Строк виконання етапів магістерської дисертації	Примітка
1	Аналіз матеріалів за об'єктом дослідження	13.04.23 – 25.06.23	Виконано
2	Розробка методу та системи детектування голосової активності в системах з низьким енергоспоживанням	26.06.23 – 31.09.23	Виконано
3	Проведення моделювання проекту	1.10.23 – 01.12.23	Виконано
4	Створення стартап-проекту	02.12.23 – 15.12.23	Виконано
5	Оформлення дисертації	1.12.23 – 3.01.24	Виконано

Студент

  
(підпис)

Демиденко О. А.

(ініціали, прізвище)

Науковий керівник дисертації

  
(підпис)

Лебедев Д. Ю.

(ініціали, прізвище)

## РЕФЕРАТ

Магістерська дисертація складається з 97 сторінок, в якій міститься 31 рисунок, 17 таблиць, використано 22 джерела.

**Актуальність.** Сучасний розвиток технологій та розширене застосування систем, мобільних пристроїв та Інтернету речей вимагають ефективних та енергоефективних методів взаємодії з користувачем. Вивчення голосової активності дозволяє розробляти енергоефективні алгоритми, які сприяють тривалому автономному функціонуванню у багатьох системах. Дана сфера досліджень виявляється ключовою для мобільних пристроїв та технологій. Дослідження у цій області спрямоване на розробку алгоритмів, що оптимізують енергоспоживання та підвищують зручність використання. У наш час через швидкий розвиток методів машинного навчання нові та швидкі алгоритми є критичною потребою особливо при роботі у середовищі з обмеженим енергоспоживанням.

**Зв'язок роботи з науковими програмами, планами, темами.** Дисертаційне дослідження проводилося відповідно до наукових напрямків кафедри конструювання електронно-обчислювальної апаратури КПІ ім. Ігоря Сікорського, а також пріоритетного напрямку розвитку науки і техніки України "Інформаційні та комунікаційні технології".

**Метою дисертаційної роботи** є дослідження ефективності реалізації методів детектування голосової активності у системах з малим енергоспоживанням. Під цим мається на увазі дослідження методів та більш точна їх оцінка.

Для досягнення мети в роботі вирішено наступні **задачі**:

- Аналіз існуючих рішень та наукових статей за темою дослідження.
- Пошук та аналіз інформації про основні параметри, що використовуються та потрібні при детектуванні голосової активності, яка дозволяє оцінювати та

краще усвідомлювати вже існуючі методи на предмет можливих покращень у реалізації алгоритмів та їх результату.

- Розробка удосконаленого методу й системи та демонстрація моделювання.
- Розроблення стартап-проєкту.

**Об'єктом дослідження** є процес детектування голосової активності у системах з низьким енергоспоживанням.

**Предметом дослідження** є методи та алгоритми для детектування голосової активності у системах з низьким енергоспоживанням та їх реалізація

**Методи дослідження:** у роботі було використано декілька методів, основні з них – метод прихованої моделі Маркова, модель рекурентної, згорткової нейронної мережі. Проведено моделювання методу з використанням рекурентної нейронної мережі у середовищі Matlab з різними фоновими шумами при різних параметрах відношення рівня сигналу до шуму.

**Наукова новизна** визначається результатами аналізу ефективності існуючих методів детектування голосової активності у системах з малим енергоспоживанням. Шляхом порівняльного аналізу обґрунтовано вибір для використання в якості базового при детектуванні голосової активності в системах з низьким енергоспоживанням методу рекурсивних нейронних мереж, що дозволило провести дослідження його ефективності у розробленій системі та провести моделювання.

**Практичне значення роботи** визначається створеними алгоритмічними рішеннями та системою забезпечення детектування голосової активності.

**Апробація результатів дослідження:** проведено тестування методу та розробленої системи у програмному середовищі.

**Публікації:** за матеріалами досліджень опубліковано 1 друковану статтю у фаховому виданні України (категорія Б) [22].

**Ключові слова:** детектування голосової активності, системи з малим енергоспоживанням, нейронна мережа.

## ABSTRACT

The master's thesis consists of 97 pages, which contains 31 figures, 17 tables, and uses 22 sources.

**Topicality.** The modern development of technologies and the expanded use of systems, mobile devices and the Internet of Things require effective and energy-efficient methods of interaction with the user. The study of voice activity allows the development of energy-efficient algorithms that contribute to long-term autonomous functioning in many systems. This area of research is key for mobile devices and technology. Research in this area is aimed at developing algorithms that optimize energy consumption and increase usability. Nowadays, due to the rapid development of machine learning methods, new and fast algorithms are a critical need especially when operating in a power-constrained environment.

**Connection of work with scientific programs, plans, topics.** Dissertation research was conducted in accordance with the scientific directions of activity of the DECED department, as well as the priority direction of the development of science and technology of Ukraine "Information and communication technologies".

**The purpose of the research** is to investigate the effectiveness of the methods of voice activity detection in systems with low energy consumption. This means the study of methods and their more accurate evaluation.

The following **tasks** are achieved by the following methods:

- Analysis of existing solutions and scientific articles on the research topic.
- Search and analysis of information on the key parameters used and required for voice activity detection, which allows to evaluate and better aware of already existing methods for possible improvements in the implementation of algorithms and their results
- Development of an improved method and system and demonstration of modeling.

- Development of a startup project.

**The object of the research.** The process of detecting voice activity in systems with low energy consumption.

**The subject of the research.** Methods and algorithms for detecting voice activity in low-energy systems and their implementation.

**Research Methods.** Several methods were used in the work, the main ones are the hidden Markov model method, the Gaussian model, the support vector method, the recurrent, convolutional neural network model. The method was simulated using a recurrent neural network in the Matlab environment with various background noises at various signal-to-noise ratio parameters.

**Scientific novelty.** Is determined by the results of the analysis of the effectiveness of existing methods of detecting voice activity in systems with low energy consumption by in-depth evaluation of each method. By means of a comparative analysis, the choice to use the recursive neural network method as a basic method for detecting voice activity in low-energy systems was justified, which made it possible to conduct a study of its effectiveness in the developed system and conduct simulations..

**Practical importance.** Determined by the created algorithmic solutions and the system for ensuring the detection of voice activity with low energy consumption.

**Approbation of the research results:** Testing of the method and developed system has been conducted in a software environment.

**Publications:** 1 printed article was published based on research materials in the professional edition of Ukraine (category B) [22]

**Keywords:** detection of voice activity, systems with low energy consumption, neural network.

## ЗМІСТ

Перелік умовних скорочень і термінів.....	10
ВСТУП.....	11
РОЗДІЛ 1. АНАЛІЗ ЛІТЕРАТУРНИХ ДЖЕРЕЛ ПО ТЕМІ ДОСЛІДЖЕННЯ.....	13
1.1. Детектування голосової активності.....	13
1.2. Історія розвитку детектування голосової активності.....	14
1.3. Основні методи дослідження голосової активності.....	17
1.3.1. Змішана Гаусівська модель.....	19
1.3.2. Метод опорних векторів.....	22
1.3.3. Приховані моделі Маркова.....	27
1.3.4. Згорткові нейронні мережі.....	31
1.3.5. Рекурентні та рекурсивні нейронні мережі.....	34
Висновки до розділу 1.....	37
РОЗДІЛ 2. РОЗРОБКА АЛГОРИТМУ РОБОТИ СИСТЕМИ.....	39
2.1 Розгляд основних вимог до системи.....	39
2.2 Основна структура системи.....	42
2.2.1 Опис компонентів системи.....	42
2.3 Принцип роботи ключових компонентів системи.....	43
2.4 Використання у системах з низьким енергоспоживанням.....	48
Висновки до розділу 2.....	51
РОЗДІЛ 3. МОДЕЛЮВАННЯ ТА РОЗРОБКА СИСТЕМИ ДЕТЕКТУВАННЯ ГОЛОСОВОЇ АКТИВНОСТІ.....	52
3.1 Програмна сфера розробки.....	52
3.2 Алгоритм роботи та модель програми.....	53
3.3 Моделювання результатів обробки.....	59



Висновки до розділу 3.....	62
РОЗДІЛ 4. РОЗРОБКА СТАРТАП-ПРОЄКТУ .....	63
4.1. Опис ідеї проекту “Дослідження голосової активності у системах з низьким енергоспоживанням” .....	63
4.2 Технологічний аудит ідеї проекту.....	64
4.3 Аналіз ринкових можливостей запуску стартап-проекту .....	65
4.4 Розроблення ринкової стратегії проекту .....	69
4.5 Розроблення маркетингової програми стартап-проекту.....	72
Висновки до розділу 4.....	74
Загальні висновки .....	76
Список використаних джерел .....	77
ДОДАТОК А. Лістинг програми.....	81
ДОДАТОК Б. Копія публікації за темою дисертаційної роботи.....	85

## Перелік умовних позначень, скорочень, термінів

VAD – Voice activity detection (Детектування голосової активності)

RNN – Recurrent neural network (Рекурентна нейронна система)

CNN – Convolutional neural network (Згорткова нейронна система)

GMM – Gaussian mixture models (Змішані Гаусівські моделі)

HMM – Hidden Markov models (Приховані Марківські моделі)

LSTM – Long-short term memory networks (Мережі довгострокових пам'яті)

EM – Expectation maximization (Алгоритм очікування максимізації)

BPTT – Backpropagation through time (Зворотнє розповсюдження у часі)

SWOT – Strength&Weak&Opportunities&Troubles (Матриці аналізу сильних та слабких загроз та можливостей)

ПЗ – програмне забезпечення

SNR – Signal to Noise ratio (Відношення сигнал/шум)

## ВСТУП

**Актуальність теми.** Методи виявлення голосової активності стають все більш важливими в різних застосунках обробки аудіосигналів через зростаючий попит на ефективні комунікаційні системи, пристрої, де виконується керування голосом.

Методи служать фундаментальним елементом в обробці аудіосигналів, ідентифікуючи та виокремлюючи сегменти мовлення в заданому аудіосигналі. Важливість використання точних методів охоплює різні сфери, впливаючи на ефективність різноманітних застосунків. Вони сильно впливають на оптимізацію комунікаційних каналів, розрізняючи між сегментами мовлення та тиші. Це підвищує використання пропускну здатності, особливо в телекомунікаційних системах, що призводить до більш ефективного передавання даних. Також сучасні методи є сильно адаптовані до зменшення шуму у передачі звуку. Ідентифікація сегментів мовлення та виключення тиші або періодів шуму поліпшує співвідношення сигнал-шум, особливо в середовищах зі змінними рівнями фонового шуму. У застосунках Інтернету речей (IoT) сприяють оптимізації споживання енергії.

Активація компонентів обробки голосу лише під час виявлення мовлення дозволяє пристроям інтернету речей економити енергію та продовжувати термін служби батареї у пристроях. Актуальність методів детекції голосу охоплює різноманітні застосунки, починаючи від оптимізації комунікаційних систем і закінчуючи покращенням користувацького досвіду у пристроях, що керовані голосом. Постійні дослідження та розвиток у даній області є критичними для задоволення вимог сучасних технологій, забезпечуючи ефективність, точність та поліпшене споживання енергії. Оскільки область обробки аудіоданих продовжує розвиватися, важливість методів у формуванні майбутнього голосових застосунків є необхідною.

**Мета і задачі дослідження.** Дослідження ефективності реалізації методів дослідження голосової активності у системах з малим енергоспоживанням. Під цим мається на увазі дослідження методів та більш точна їх оцінка.

**Об'єкти дослідження.** Різні методи детектування голосової активності у системах з низьким енергоспоживанням.

**Предмет дослідження.** Алгоритми та їх реалізація для детектування голосової активності у системах з низьким енергоспоживанням.

**Наукова новизна одержаних результатів.** Визначається удосконаленням існуючих методів детектування голосової активності у системах з малим енергоспоживанням шляхом глибокої оцінки кожного методу, параметрів що використовуються для роботи алгоритмів та детальним їх описом.

**Апробація результатів дисертації.** Проведено тестування методу та розробленої системи у програмному середовищі.

**Публікації:** за матеріалами досліджень опубліковано 1 друковану статтю у фаховому виданні України (категорія Б) [22].

**Ключові слова.** Детектування голосової активності, системи з малим енергоспоживанням, нейронна мережа.

## **РОЗДІЛ 1. АНАЛІЗ ЛІТЕРАТУРНИХ ДЖЕРЕЛ ПО ТЕМІ ДОСЛІДЖЕННЯ.**

### **1.1 Детектування голосової активності**

Виявлення мовленнєвої активності є дуже важливим і ефективним процесом у виявленні та обробці мови і сигналів. Метою цього процесу є кодування та розпізнавання мовних сегментів. Ця діяльність є важливою та використовується для відключення процесу розпізнавання при виявленні в мовному сигналі ділянок з мовленням або, навпаки, ділянок без мовлення або з дуже низьким рівнем мовлення. Основними критеріями виявлення мовної активності є точність, адаптивність і здатність працювати без інформації про присутній шум.

Перевагами використання методів і алгоритмів виявлення мовної активності є економія ресурсів, що передаються по використовуваному мовному каналу, а також збільшення економії пам'яті та інших ресурсів комп'ютерної системи. Якщо в передачі мови є помилки, можуть бути внесені шуми, які в подальшому можуть змінити правильну інтерпретацію результатів. Особливо для вимірювань в телекомунікаціях та інших сферах критичної інфраструктури точність передачі мови є особливо важливою, незалежно від умов роботи системи.

З цієї причини було розроблено низку різних методів, кожен з яких компенсує інший різними характеристиками, низькою затримкою, апаратною реалізацією тощо. Деякі алгоритми можуть навіть розпізнавати емоційний стан людини, мова якої розпізнається.

Однак, при всіх своїх перевагах, ці методи також мають ряд недоліків, таких як:

- Слабкий рівень боротьби з шумом при збільшенні голосового потоку.
- Необхідність простих алгоритмів
- Необхідність зменшення апаратних затрат при детектуванні голосу.

Для загальної картини варто розглянути поняття голосової активності та її основних критеріїв.

Голосова активність визначається наявністю звуків людського голосу у аудіосигналі, таких як мова, спів, сміх або інші звуки, що створюються голосовим апаратом людини. Для виявлення голосової активності застосовують алгоритми, які аналізують аудіосигнал, шукаючи характеристики, що відрізняють голосові сигнали від інших звуків.

Голосову активність можна використовувати для різних цілей, таких як розпізнавання мовлення, оптимізація голосового зв'язку, шумозаглушення та використання в біометрії. Наприклад, алгоритми виявлення голосової активності застосовуються для виділення моментів мовлення перед передачею сигналу на систему розпізнавання мовлення або для зменшення пропускну здатності в голосовому зв'язку.

Для вимірювання голосової активності використовують різні одиниці вимірювання, кожна з яких має свої характеристики та застосування:

- Децибели. Найпоширеніша одиниця для вимірювання гучності звуку, використовується для порівняння рівня звуку з еталонним значенням.
- Відношення сигнал/шум (SNR). Вимірює співвідношення між сигналом (голосом) та небажаним шумом в децибелах. Високе SNR свідчить про кращу якість звуку.
- Частка голосу. Визначає відсоток часу, протягом якого в аудіосигналі є мова. Використовується в системах розпізнавання мовлення та голосового зв'язку

## **1.2 Історія розвитку детектування голосової активності**

Так, як було зазначено – виявлення мовлення є дуже важливою частиною нашого сьогоденного життя, тож варто розглянути історію виникнення та дослідження.

У 1952 році з'явився перший пристрій для розпізнавання мови, який був спроможний визначати цифри, вимовлені людиною. У 1962 році на виставці комп'ютерних технологій в Нью-Йорку був представлений пристрій IBM Shoebox. У 1963 році інженерами корпорації Sperry були представлені невеличкі пристрої

для розпізнавання з використанням волоконно-оптичного запам'ятовуючого пристрою, відомі як "Септрони". Ці пристрої виконували послідовні дії за командами, вимовленими оператором, і знаходили застосування у різних галузях, включаючи провідний зв'язок, військові технології, авіацію та автоматизовані системи управління [1].

У 1983 році був представлений інтерактивний комплекс "розумної авіоніки" для атакуючих гелікоптерів "Апач". Цей комплекс мав здатність розпізнавати команди та запити пілота, конвертувати їх у сигнали управління для бортового обладнання і взаємодіяти з пілотом голосом, надаючи інформацію щодо можливості виконання поставлених завдань [1].

У 1990-их роках виникли перші комерційні програми, що становили інновацію для того періоду. Однією з перших таких програм був Dragon Naturally Speaking, який здобув популярність серед людей із травмою руки, оскільки це ускладнює процес набору великого обсягу тексту. Цей додаток перетворював голос користувача в текст.

В початку 2000-их років спостерігався пік досліджень у галузі голосової активності, що спричинився спонсоруванням двох програм розпізнавання мовлення Міністерством оборони Сполучених Штатів Америки. Ці програми включали в себе Ефективне доступне багаторазове перетворення мовлення в текст (EARS) та Глобальне використання автономної мови (GALE).

З початку 2005 року Агентство національної безпеки Сполучених Штатів Америки активно впроваджувало технології розпізнавання мови. Це дозволило аналітикам ефективно проводити пошук ключових фрагментів у великому обсязі записаних діалогів. Усі записи індексувалися для забезпечення зручного доступу до необхідних фрагментів. З того часу, отримавши успішні результати випробувань, вивчення різних перспектив та розвиток нових алгоритмів та методів зайняли увагу науковців.

На початку 21-ого століття переважна більшість методів була заснована на перевіреному методі прихованих моделей Маркова у поєднанні з нейромережами прямого зв'язку, однак зараз існує безліч підходів, наприклад:

- Використання рекурентних нейронних мереж
- Приховані Марковські Мережі
- Нейронні мережі інших типів
- Тимчасові динамічні алгоритми
- Алгоритм наскрізного автоматичного розпізнавання мови

Оскільки історія виникнення була розглянута, варто розглянути основні засоби застосування та методи досліджень.

У сьогоденні водіїв детектування голосових сигналів є однією з найзручніших опцій у автомобілі, оскільки прості голосові команди можуть ініціювати мобільні дзвінки, керування мультимедійною системою у автомобілі, управління системою автоматичного паркування. Деякі сучасні автомобілі пропонують розпізнавання природньою мовою, замість певного набору голосових команд. Це дозволяє водієві використовувати цілі речення замість набору вже певних голосових команд, це зручно тим, що користувач може не запам'ятовувати команди.

Якщо брати до уваги медичну сферу – застосування є дуже широким, наприклад заповнення медичної документації. Постачальник медичних послуг контактує з системою розпізнавання мовлення, ідентифіковані слова миттєво відображаються. З іншого боку відкладене розпізнавання мовлення являє собою передачу даних у цифрову систему. Дані оброблюються за допомогою механізму, а отриманий результат разом з вихідним голосовим файлом надсилається до особи, що заповнює.

У військовій справі методи голосового детектування з фільтрацією є особливо важливими, оскільки проблема акустичного шуму є дуже серйозною, що для пілотів літаків та гелікоптерів, навіть для пілотів гелікоптерів, оскільки вони зазвичай не використовують маску для зменшення акустичного шуму через



мікрофон навушників, було винайдено голосові програми для керування станціями та навігаційними системами.

Для телекомунікацій це важливо, оскільки майже всі контакт-центри, системи телефонії використовують детектування голосу. На даний час, оскільки технології розвиваються, у зв'язку з розвитком процесорів, що влаштовані у портативні мобільні пристрої детектування голосової активності стало ще більш практичним та простим процесом. Мова здебільшого використовується для створення чи налаштування певних команд у налаштуваннях самого пристрою як частина його інтерфейсу.

Але як на мене одна з найяскравіше виражених сфер застосування – взаємодія з людьми з вадами. Спеціально розроблене програмне забезпечення для субтитрування на певних подіях може бути використаним для людей з проблемами слуху чи зору.

Також можна привести приклад того, як це допомагає людям, що мають обмежені властивості кінцівок чи повністю глухим людям – використання сервісів голосової пошти в зрозумілий текст чи ретрансляція будь-якого сигналу.

### **1.3 Основні методи дослідження голосової активності**

Вивчення активності голосу охоплює застосування різноманітних методів і технік для аналізу аудіосигналів і визначення наявності чи відсутності голосу у конкретний момент часу. Основні методики:

1. Енергетичний поріг - аудіосигнал розбивається на короткі фрейми, при цьому для кожного фрейму визначається його енергія. Встановлення порогу енергії служить для розрізнення голосових і не-голосових фреймів, і фрейми з енергією вище порогу розглядаються як голосові. Дана техніка є простою у реалізації, але лише у чистих умовах. Завдяки простоті відбувається швидка реакція на рівень голосу, що і робить дану техніку швидкою та практичною, де час реакції є важливим. Але має дуже суттєвий недолік – чутливість до

шуму, тобто при великому рівні шуму може випадково детектувати активність голосу.

2. Частота перетину нуля - вимірювання частоти, з якою аудіосигнал перетинає нульову лінію. Значна частота перетину нуля може служити індикатором наявності голосу, оскільки голосові сигнали часто виявляють коливання.
3. Кепстральні коефіцієнти - вилучення характеристик, які ефективно відображають спектральні особливості аудіосигналу. Цей підхід широко застосовується в акустичному моделюванні та розпізнаванні мови для аналізу спектральних особливостей голосу.
4. Метод кореляції - аналіз кореляції між аудіосигналами та конкретними шаблонами голосу чи не-голосу. Цей метод використовується для визначення ступеня схожості із заданими голосовими шаблонами. Даний метод є відносно простим, що і зробило його одним з основних у використанні при різних ступенях складності задач, чутливий до змін у аудіосигналі та не вимагає використання великого обсягу даних для тренування. Проте є вразливим до шуму та не є ефективним у довгостроковій обробці сигналу, через закладену відсутність врахування параметрів сигналу, наприклад різку зміну гучності.
5. Машинне навчання - використання алгоритмів класифікації та нейронних мереж для автоматичного визначення голосової активності. Моделі навчаються на обширних обсягах аудіоданих для розпізнавання відмінностей між голосовими та не-голосовими сигналами. Метод поширений за свою можливість адаптації до певного середовища дослідження та можливість навчатися на базі реальних даних, є дуже точним, оскільки нейронні мережі забезпечують можливість навчання на реальних даних. Але потребує велику кількість даних для ефективного навчання, оскільки через малу кількість точне навчання може бути проблематичним.
6. Динамічні порогові та адаптивні методики - використання алгоритмів, які можуть адаптувати пороги в реальному часі в залежності від змін рівня шуму та інших умов.

7. Глибинне навчання - застосування глибоких нейронних мереж для автоматичного вилучення ознак та визначення голосової активності. Такі моделі здатні автоматично визначати складні патерни у звукових сигналах. Основна перевага у тому, що мережі можуть автоматично вибирати та виділяти ключові критерії з сигналів, через це і метод має дуже велику точність, особливо при довготривалому тренуванні. Через це і може адаптуватися до будь-яких умов проведення дослідження, таких як дуже часта швидка зміна рівня гучності сигналу. Проте має суттєвий недолік як і більшість систем – потреба у великих обсягах даних для навчання та вибагливість до обчислювальних ресурсів.

В наступних підрозділах будуть описані основні методи та їх особливості для дослідження голосової активності.

### **1.3.1 Змішана Гаусівська модель**

В галузі машинного навчання та аналізу даних часто виникає необхідність визначати закономірності та кластери у великих наборах даних. Проте традиційні алгоритми кластеризації мають обмеження у визначенні кластерів з різними формами та розмірами. Саме тут на допомогу приходять моделі змішаних Гаусівських розподілів (Gaussian Mixture Models).

Моделі змішаних Гаусівських розподілів є типом алгоритмів машинного навчання, які використовуються для класифікації даних на різні категорії на основі ймовірнісного розподілу. Моделі можна використовувати в різних галузях, включаючи фінанси та маркетинг.

Гаусівські змішані моделі є важливим інструментом для ідентифікації кластерів у наборах даних, де границі кластерів можуть бути неоднозначними. Крім того, моделі можуть використовуватися для оцінки ймовірності того, що нова точка даних належить до кожного кластеру. Важливою рисою є їх стійкість до викидів, що означає їхню здатність надавати точні результати навіть у випадку, коли деякі

точки даних не чітко вписуються в будь-який конкретний кластер. Ця гнучкість та стійкість робить модель потужним та гнучким інструментом для кластеризації.

GMM може розглядатися як ймовірнісна модель, в якій для кожної групи використовуються Гаусівські розподіли з визначеними їх середніми та коваріаційними матрицями. Два основні компоненти - це вектори середніх ( $\mu$ ) та матриці коваріацій ( $\Sigma$ ). Важливо відзначити, що Гаусівський розподіл, також відомий як нормальний розподіл, є неперервним розподілом ймовірностей.

Модель GMM являє собою суму  $M$  компонент:

$$p(x|\lambda) = \sum_{i=1}^M \omega_i p_i(x), \quad (1.1)$$

де  $\lambda$  – модель диктора,  $\omega_i$  – коефіцієнти ваги компонентів моделі,  $p_i(x)$  – функція густини розподілу складових моделі.

Кожний компонент являє  $D$ -мірну Гаусівську густину розподілу:

$$p(x) = \frac{1}{(2\pi)^{D/2} |\mathbb{R}_i|^{1/2}} \exp \left\{ -\frac{1}{2} (x - \mu_i)^T (\mathbb{R}_i)^{-1} (x - \mu_i) \right\}, \quad (1.2)$$

де  $D$  – розмірність простору ознак,  $\mu_i$  – вектор математичного сподівання,  $\mathbb{R}_i$  – коваріаційна матриця розмірністю  $D \times D$ .

Отже, для створення моделі необхідно визначити вектори середніх, матриці коваріації та вагові коефіцієнти компонентів. Зазвичай для оцінки параметрів використовують метод максимуму подібності або метод максимуму апостеріорної ймовірності.

У змішаних Гаусівських Моделях метод очікування-максимізації є потужним інструментом для оцінки параметрів. Очікування позначається як  $E$ , а максимізація - як  $M$ . Очікування використовується для знаходження параметрів Гаусівських розподілів, які представляють кожен компонент змішаної моделі. Максимізація позначається як  $M$ , і вона визначає, чи можливе додавання нових точок даних.

Метод очікування-максимізації є двоетапним ітеративним алгоритмом. Він чергується між виконанням етапу очікування, на якому ми обчислюємо очікування для кожної точки даних за поточними оцінками параметрів, і потім максимізуємо їх для створення нового Гаусівського розподілу. Далі йде етап максимізації, де ми оновлюємо середні Гаусівські значення на основі оцінки максимальної ймовірності. Метод EM починається ініціалізацією параметрів моделі, а потім вдосконалює ці оцінки. На кожній ітерації етап очікування обчислює очікування функції відносно поточних параметрів. Це очікування потім використовується для максимізації ймовірності на етапі максимізації. Весь процес повторюється до досягнення збіжності. Нижче представлено зображення, що ілюструє двоетапний ітеративний аспект алгоритму.

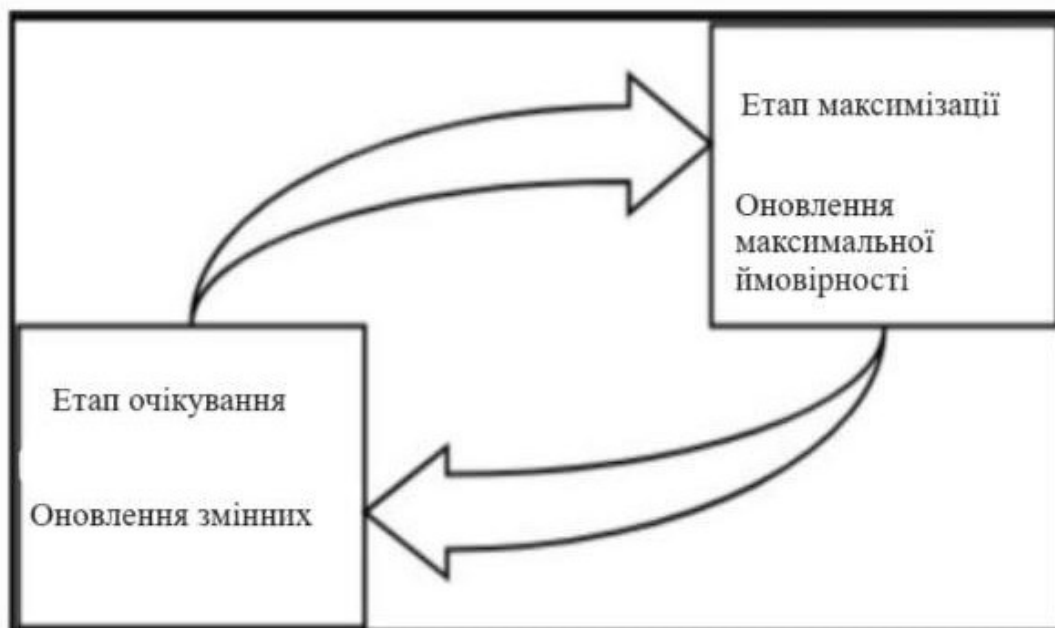


Рисунок 1.1 Ілюстративне зображення EM алгоритму

Отже, Гаусівські моделі мають багато переваг та можуть бути активно використані у:

- Випадках, коли дані генеруються змішаними Гаусівськими розподілами, коли існує невизначеність щодо правильної кількості кластерів та коли кластери

мають різні форми. У кожному з цих випадків використання моделі може допомогти покращити точність результатів. Наприклад, коли дані генеруються змішаними розподілами, використання моделі може краще визначити основні закономірності даних. Крім того, коли існує невизначеність щодо правильної кількості кластерів, використання моделі зменшує рівень помилок.

- У випадку аналізу часових рядів моделі можуть бути використані для вивчення зв'язку між волатильністю, тенденціями та шумом, що допомагає прогнозувати майбутні ціни стока. Один кластер може представляти тенденцію в часовому ряді, тоді як інший може містити шум і волатильність від інших факторів, таких як сезонність чи зовнішні події, які впливають на ціну стока.
- Моделі можуть генерувати синтетичні точки даних, які схожі на оригінальні і також можуть використовуватися для розширення обсягу даних.

### **1.3.2 Метод опорних векторів**

Даний метод є алгоритмом машинного навчання, що використовується для класифікації об'єктів. Основа даного методу – алгоритм максимізації певної математичної функції відносно вже наявного набору даних [2]. Основними поняттями даного методу є:

- Віддаляюча гіперплощина
- Гіперплощина максимальної межі
- М'яка межа
- Функція ядра

Відділяюча гіперплощина є математичним об'єктом, який розділяє класи об'єктів з однаковими ознаками. Наприклад, це можна побачити на рис. 4, де у тривимірному просторі площина відокремлює кульки світлого кольору від темних кульок.

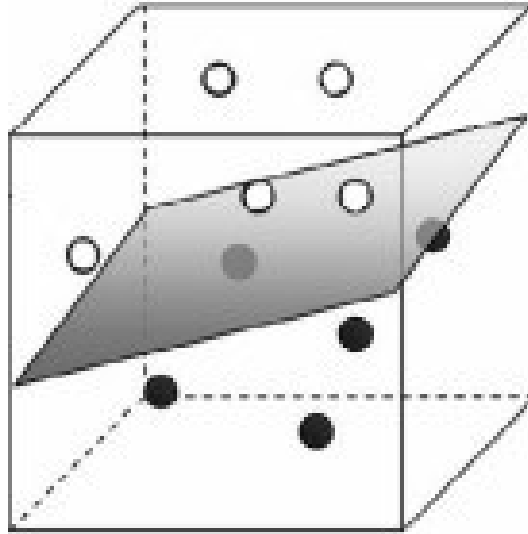


Рисунок 1.2 Віддаляюча площина [2]

Отже, цей процес математично можна розглядати в контексті вимірів, що значно перевищують третій. Термін, що вказує на лінію, розміщену між елементами різних класів, називається багатовимірною гіперплощиною. Способи створення такої гіперплощини за допомогою методу опорних векторів не є єдиними. Все завжди залежить від багатьох можливостей розташування цієї гіперплощини. Даний метод відрізняється від інших методів гіперплощинної класифікації тим, що він надає можливість вибору оптимального розташування гіперплощини. Гіперплощина вибирається так, щоб знаходитися на максимальній відстані від елементів кожного класу, тобто у центрі певної зони, яка відокремлює ці елементи один від одного. Основна ідея полягає в другому ключовому аспекті - гіперплощині максимальної межі. Об'єкти, що класифікуються, не завжди можуть бути розділені гіперплощиною. У реальних системах будуть наявними похибки в даних, які видно на рисунку 1.3

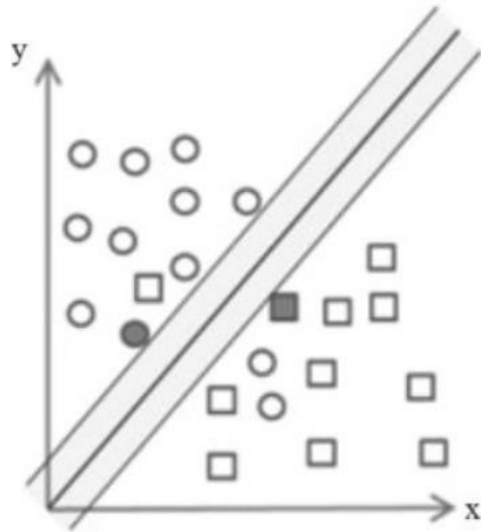


Рисунок 1.3 Похибки при класифікації об'єктів [2]

Об'єкти, які піддаються класифікації, можуть мати лінійне розподілення лише в окремих випадках і в більшості випадків вони мають не лінійну структуру. Для вирішення цієї проблеми використовують функції ядра, які трансформують дані з низько вимірного простору у багатовимірний. Якщо вибрати відповідну функцію ядра, об'єкти можуть бути лінійно розділені гіперплощиною у багатовимірному просторі. Таким чином, функції ядра діють як засіб приведення в простір вищої розмірності.

Розглянемо основні математичні залежності, які використовуються у методі. Кожен об'єкт класифікації розглядається як вектор у  $n$ -вимірному просторі. Кожна координата вектору відображає конкретну ознаку, і її значення визначається тим, наскільки ця ознака виражена у даному об'єкті. Значення, яке більше, свідчить про більшу виразність відповідної ознаки в об'єкті.

Об'єкти, що класифікуються, можуть бути поділені лінійно лише в окремих випадках. Щоб знову зробити задачу лінійно поділеною і мати можливість застосовувати розглянуті вище лагранжіани, необхідно перетворити дані в простір вищого порядку, де можливе лінійне розділення об'єктів гіперплощиною. Цей процес називається спрямленням простору або мапінгом даних, що ілюстрований на рисунку. 1.4



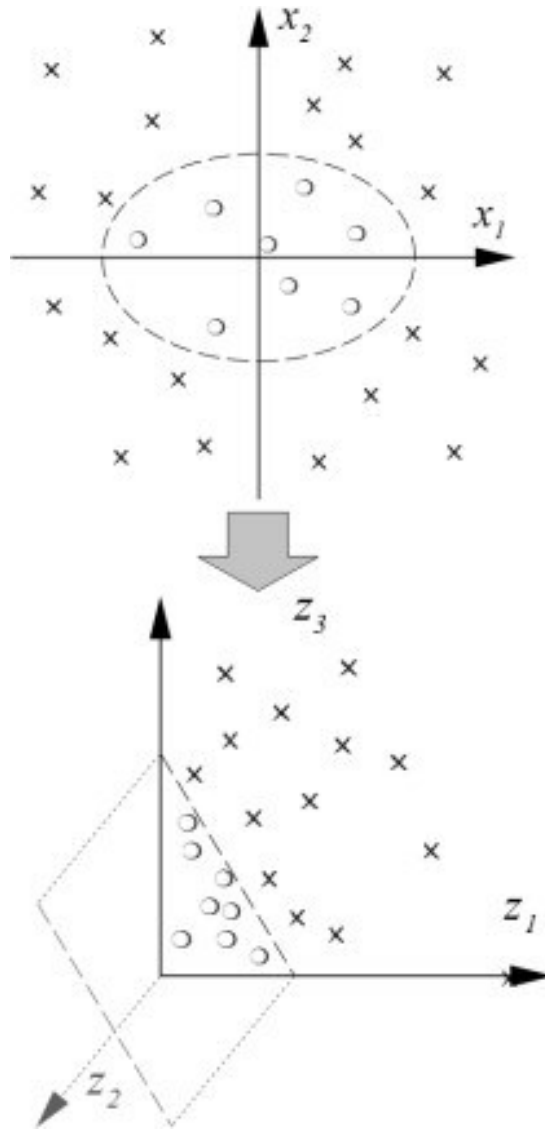


Рисунок.1.4 Приклад мапінгу даних [2]

Мапінг даних, позначений символом  $\Phi$ , здійснює відображення інформації з вихідного гіперпростору  $R^d$  до евклідового простору за допомогою функцій. Таким чином, для зроблення задачі лінійно розділюемою, потрібно застосовувати перетворення до кожної точки і працювати з відповідними їм  $\Phi(x)$ . Логіка цього розподілення ілюструється на рисунку 1.5.

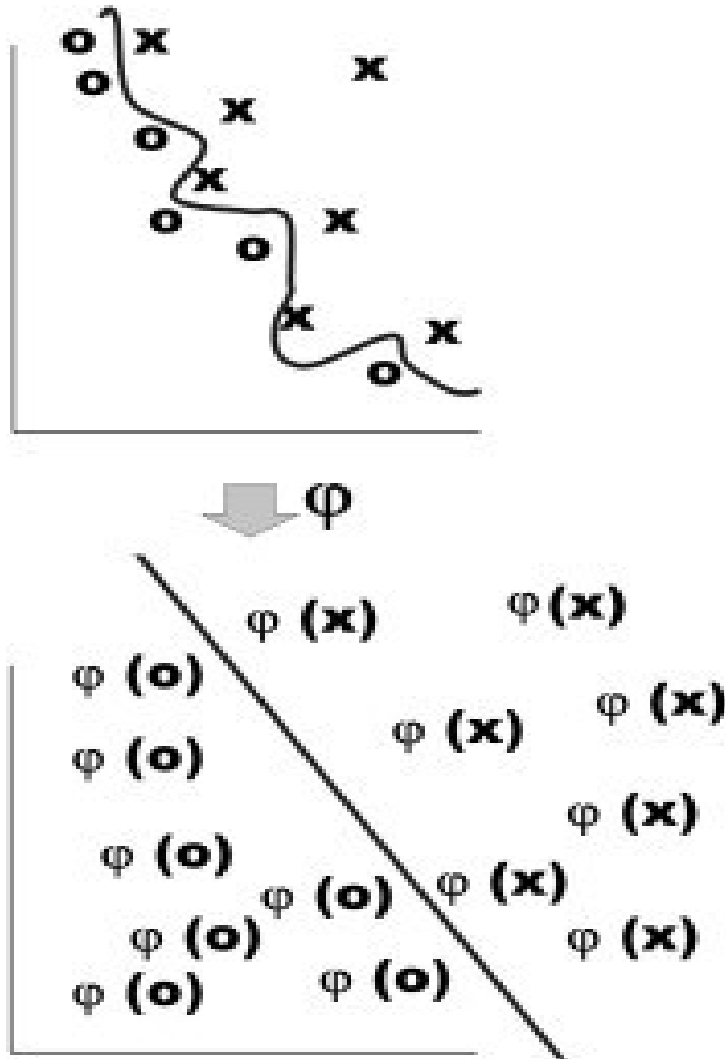


Рисунок.1.5 Ілюстрація  $\Phi(x)$  до вихідних даних

Гарним прикладом ініціалізації даного методу є робота [3] у якій описаний даний метод. Суть методу полягає у наступних етапах:

1. На етапі попередньої обробки алгоритм навчання витягає компактне представлення мовленнєвого сигналу за допомогою обчислення кепстральних коефіцієнтів. Це компактне представлення сприяє обчислювальній ефективності та надійності навчання та тестування методу опорних векторів, якщо сигнал вже представлений у формі кепстральних коефіцієнтів. Цей крок особливо корисний в умовах шумного середовища, де ймовірність помилкових класифікацій велика.

2. На етапі навчання алгоритм обчислює оптимальну гіперплощину та відповідні опорні вектори після відображення даних у високовимірній простір.
3. На тестовому етапі модель, створена на етапі навчання, перевіряється на точність на тестових даних.

### 1.3.3 Приховані моделі Маркова

Прихована модель Маркова – Модель прихованого Марківського процесу (НММ) є ймовірнісною моделлю, яка імітує поведінку процесу, схожого на Марківський ланцюг, але із невідомими параметрами та спостережуваним в умовах шуму. Дана модель здобула популярність, переважно завдяки успішним застосуванням у розпізнаванні мови, а також вона використовується в таких галузях, як розпізнавання письма, розпізнавання жестів.

У класичній Марковській моделі стан видимий для спостерігача, тому ймовірності переходів є єдиним параметром. В протипагу цьому, в прихованій Марковській моделі ми спостерігаємо лише змінні, які впливають на даний стан. Кожен стан має ймовірнісний розподіл серед всіх можливих вихідних значень. Таким чином, послідовність символів, згенерована прихованою Марківською моделлю, надає інформацію про послідовності станів. У діаграмі переходів прихованої Марківської моделі  $x$  позначає приховані стани,  $y$  - спостережувані результати,  $a$  - ймовірності переходів,  $b$  - ймовірність результату.

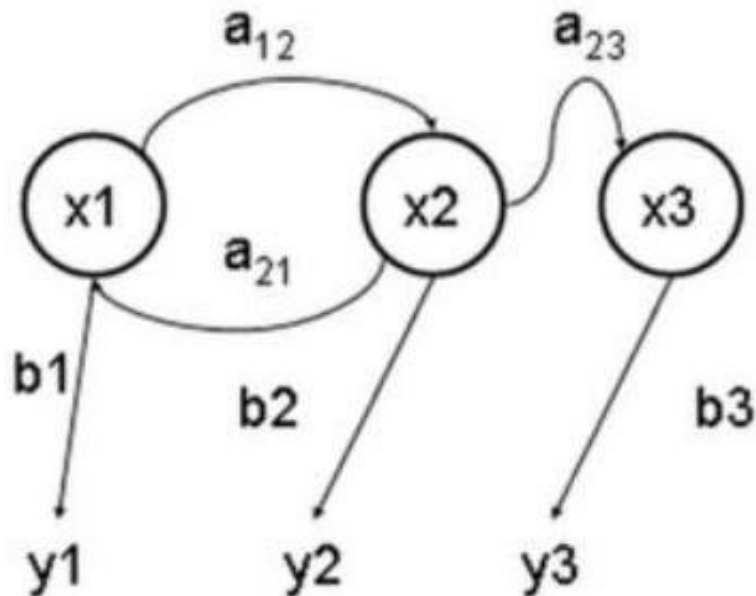


Рисунок. 1.6 Діаграма переходів у моделі

Робота з прихованими Марковськими моделями передбачає виконання двох етапів: на першому етапі проводиться навчання, що включає визначення параметрів моделі; на другому етапі здійснюється декодування, тобто визначення ймовірності того, що спостережувана послідовність векторів була згенерована вказаною моделлю.

Процес навчання прихованої Марковської моделі включає в себе визначення параметрів на основі навчального набору прикладів. Ці параметри охоплюють матрицю ймовірностей переходів між станами та параметри Гаусівських моделей (математичне очікування, матриця коваріації) для кожного стану. Для вирішення цих завдань зазвичай використовують два ітераційних алгоритми: алгоритм forward – backward та алгоритм Баума-Велча.

Алгоритм forward-backward це алгоритм динамічного програмування, який використовує передачу повідомлень. Він дозволяє обчислити відфільтровані та згладжені точки, які можна використовувати для проведення вивчення, оцінки максимальної ймовірності, класифікації послідовності, виявлення аномалій та кластеризації на основі моделі.

Алгоритм Баума-Велча є алгоритмом очікування-максимізації (EM) для прихованих Марківських моделей. Він передбачає оцінку параметрів методом максимальної схожості. Оскільки обчислення суми по всім можливим параметрам є складним з точки зору обчислювального часу, для оцінки параметрів моделі використовується алгоритм EM. Алгоритм передбачає обчислення ймовірностей за допомогою алгоритму forward-backward.

Розглянемо приховані моделі Маркова згідно прикладу статті, що розглядається [4], оскільки це є однією з найбільш поширених моделей. Існують два підходи до використання моделі виявлення мовлення/паузи. Перший ґрунтується на роздільному навчанні моделі і наступному тестуванні — це підхід з виявленням адаптації. Після навчання параметри моделі залишаються незмінними, і модель стає нечутливою до змін середовища — це підхід виявлення без адаптації. В обох процедурах потрібно вирішити, який тип Маркової моделі використовувати. Було обрано неперервні моделі із трьома станами. Важлива також обробка сигналу — для кожного кадру розраховуються Кепстральні коефіцієнти. Використовуються дві моделі: для мовчання і моделі мовлення.

Згідно з роботою [22] розглянемо алгоритм без адаптації. Він складається з:

- Ініціалізації
- Навчання
- Обрахування критеріальної функції
- Розрахунку
- Виявлення сегментів з тишею та мовленням

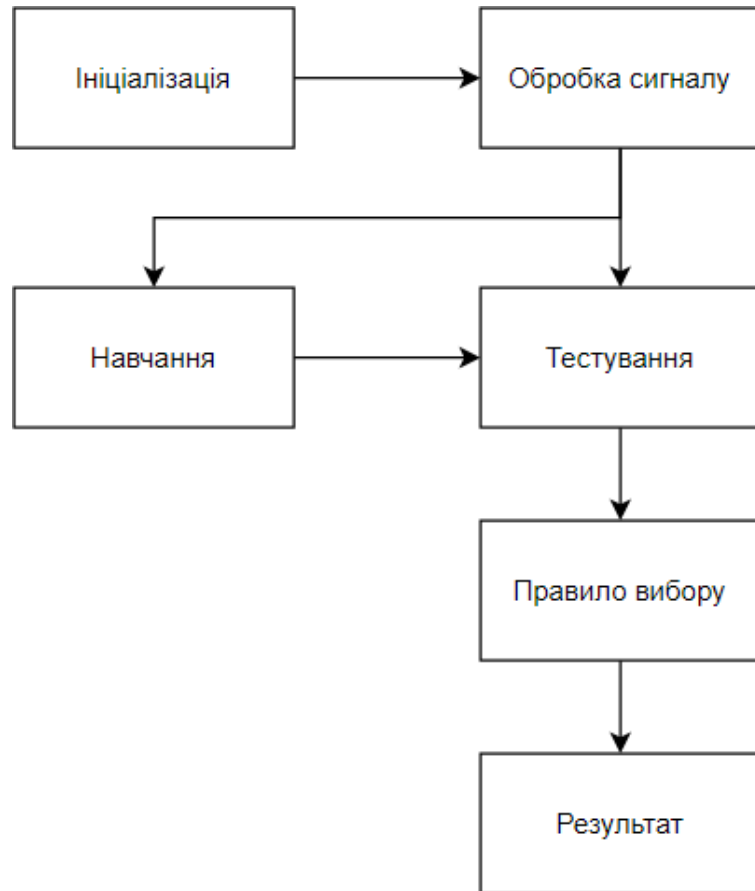


Рисунок 1.7 Схема детектора голосової активності без адаптації [4], [22]

Оскільки у даному матеріалі розглядається робота детектора з адаптацією, то варто його розглянути.

Алгоритм системи є наступним:

- Ініціалізація
- Навчання
- Оновлення мовчазної моделі
- Критерії адаптації
- Розрахунок порогового значення
- Оновлення моделі згідно зміни порогового значення
- Виявлення мовленнєвих сегментів
- Фінальна обробка

Також маємо блок-схему вже самого принципу роботи детектора

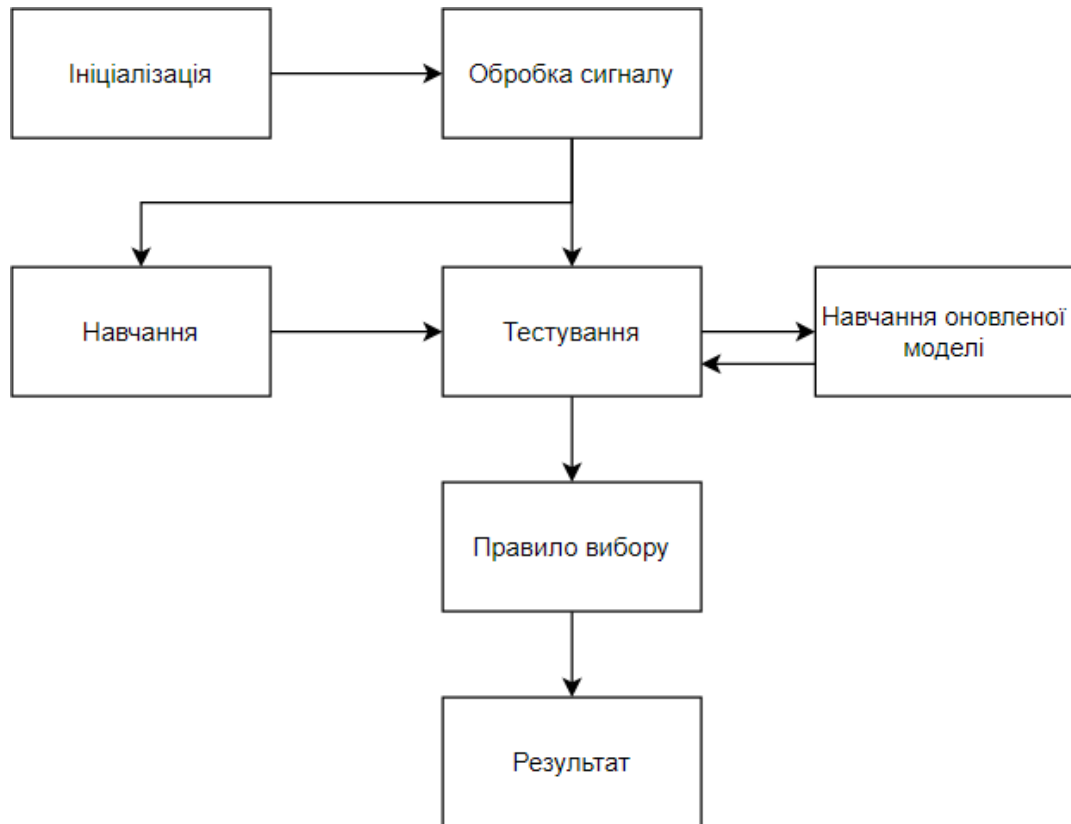


Рисунок 1.8 Схема алгоритму роботи детектора мовлення з адаптацією [4], [22]

### 1.3.4 Згорткові нейронні мережі

Згорткова нейронна мережа використовує процес згортки для одночасного зменшення обсягу інформації, що зберігається у пам'яті та кращого розпізнавання зображень високої роздільної здатності. Ця мережа виділяє ключові ознаки зображення, такі як ребра, контури або межі. На наступних етапах обробки вона розпізнає повторювані фрагменти текстур, які можуть об'єднатися у фрагменти зображення. Кожен шар нейронної мережі застосовує власне перетворення.

Ці нейронні мережі відрізняються від інших тим, що вони використовуються переважно для обробки зображень, іноді для аудіо та інших типів вхідних даних. Зазвичай ці мережі використовують "сканування", що обробляє дані частинами.

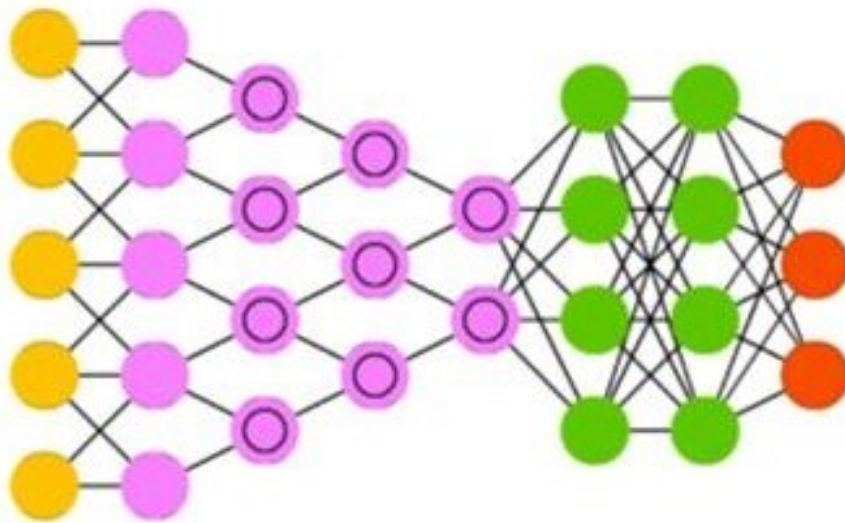


Рисунок 1.9 Архітектура згорткової нейронної мережі [5]

Архітектура згорткових мереж може бути порівняна з принципом дії зорової кори мозку. В зоровій корі існують невеликі області клітин, які реагують лише на певні візуальні шаблони в полі зору. Наприклад, деякі клітини реагують на вертикальні лінії, інші на горизонтальні або діагональні. Ці клітини, організовані у вигляді візуальних стрижнів, утворюють візуальне сприйняття разом. Цей підхід із спеціалізованими компонентами, які вирішують конкретні завдання, аналогічний до клітин зорової кори, можна застосовувати і до згорткових нейронних мереж.

Отже, важливо, щоб мережа навчилася виявляти такі ознаки, формуючи їх з базових ліній. Початковий етап може включати розділення зображення на блоки пікселів та визначення домінуючої лінії в кожному блоку. Це дозволяє отримати масиви ліній як прості ознаки об'єкта на зображенні. Дана операція називається операцією згортки.

Згорткова нейронна мережа використовує процес згортки для одночасного зменшення обсягу інформації у пам'яті та кращого розпізнавання зображень високої роздільної здатності. На наступних етапах обробки вона розпізнає повторювані фрагменти текстур, які можуть об'єднатися у фрагменти зображення. Кожен шар



нейронної мережі застосовує власне перетворення, що дозволяє правильно класифікувати та виділяти шукану інформацію з зображення чи потоку голосу.

Для прикладу варто розглянути статтю [5], де розглядається саме такий метод визначення голосової активності.

Суть роботи даної системи відображається на блок-схемі запропонованої моделі зображена на рис. Мел-спектр зашумленого вхідного сигналу використовується як вхід для моделі. Спочатку він проходить через згорткову нейронну мережу зберігаючи при цьому просторовий розмір вхідного сигналу, CNN враховує локальну просторову інформацію вхідного спектру і виводить вбудовування для кожного кадру, далі вихід кадрів подається на вхід механізму уваги.

Створена чотиришарова рекурентна нейронна мережа з пакетною нормалізацією та максимальним пулінгом між кожними зі шарів згорткового фільтра. Операція кадрової розгортки виконується для виходу CNN. Залишкові частоти та канали розгортаються для кожного кадру, при цьому зберігаючи розмір кадру. Потім використовується кадровий FC-шар, який перетворює кожне перебудування. Потім послідовно застосовується механізм уваги. Вихід механізму уваги також є послідовністю, але, на відміну від вхідної послідовності, вихідна послідовність містить контекстуальну інформацію зі всього вхідного сигналу. Послідовність виходу проходить через кадровий FC-шар, який визначає наявність мовлення у кожному кадрі.

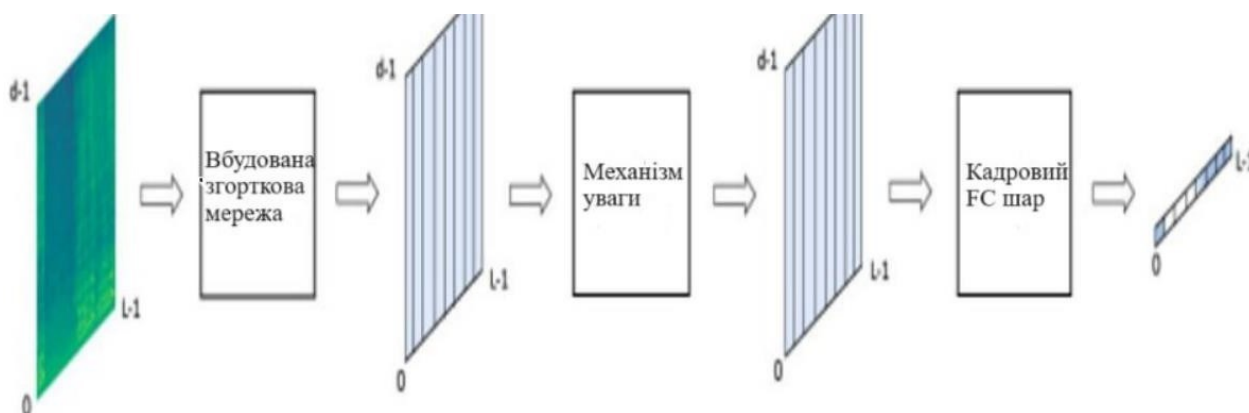


Рисунок 1.10 Приблизна архітектура мережі [5]

### 1.3.5 Рекурентні та Рекурсивні нейронні мережі

Варто розглянути цей клас мереж, оскільки вони у більшості випадків використовуються у різних алгоритмах детектування голосової активності.

Рекурентні нейронні мережі (RNN) відрізняються тим, що вони можуть обробляти послідовні дані, враховуючи контекст попередніх значень. На відміну від звичайних нейронних мереж, які обробляють дані у фіксованому порядку, RNN можуть працювати з входами змінної довжини та здійснювати висновки, що залежать від попередніх входів. Рекурентні нейронні мережі широко використовуються для обробки природної мови, дозволяючи не лише генерувати текст, а й виконувати завдання, такі як переклад. Навчаючись на парних реченнях, складених двома різними мовами, RNN обробляє пропозиції по одній, створюючи вихідне речення, яке на кожному кроці залежить від вхідної інформації. Це дозволяє RNN правильно перекладати навіть складні тексти, оскільки вона ураховує попередні входи та виходи, що сприяє розумінню контексту.

Головна ідея полягає в тому, що вихід RNN на певному кроці залежить від його попередніх входів та виходів. На кожному етапі мережа отримує новий вхід та використовує попередній стан для обчислення вихідного значення. Мережа має внутрішні стани, які визначають її пам'ять або внутрішній стан в конкретний момент часу. Це дозволяє мережі "запам'ятовувати" інформацію з попередніх кроків. RNN можна уявити як мережу, що працює з входами у послідовності часу. Функція передачі визначає, як інформація передається від одного кроку до наступного. Це може бути просто копіювання стану або складніші операції.

Проте у даних мережах є поширена проблема - проблема зникання градієнта, де інформація може втрачатися з часом через функцію активації, що призводить до подібного ефекту, що спостерігається в глибоких мережах прямого поширення. Хоча це може здатися не таким серйозним, оскільки це впливає тільки на ваги, а не на стани нейронів, ваги мають ключове значення для збереження інформації про минуле. Якщо ваги стають дуже великими або малими, інформація про минулі стани стає менш інформативною. RNN знаходять застосування в різних галузях, і навіть

дані, які не мають часової залежності (наприклад, зображення або рядки тексту), можуть бути подані у вигляді послідовності, де ваги використовуються для попереднього елемента послідовності, а не для того, що трапився певний час тому назад.

Рекурсивні нейронні мережі (RNNs) - це група нейронних мереж, спеціально призначених для роботи з послідовними даними, де важливий контекст і залежності від часу. Основна відмінність між рекурсивними та звичайними нейронними мережами полягає у їхній здатності обробляти вхідні дані з послідовністю. Сигнали, що передають звук, можуть бути перетворені у послідовності за часом або спектрограми для подальшого аналізу. Замість використання стандартних згорткових шарів, розглядається можливість використання рекурсивних шарів, які дозволяють враховувати контекст часових залежностей. Архітектури з рекурсією можуть бути більш ефективними в адаптації до динамічних змін в аудіосигналі з плином часу. Вони можуть успішно враховувати довгострокові залежності, що є важливим у вивченні голосових шаблонів. Можливість обміну інформацією між кроками може допомагати аналізувати динамічні характеристики аудіосигналу. Однак рекурсивні мережі можуть бути обчислювально-витратними, особливо при роботі з об'ємними аудіоданими, і може виникати виклик у навчанні на велику кількість вхідних даних через обмежені ресурси.

Основні відмінності цих двох мереж полягають у наступному:

Рекурентні нейромережі (RNN) є важливим класом нейромереж. Вони тісно пов'язані із рекурсивними нейромережами. Рекурентні нейромережі призначені для моделювання часових послідовностей і знаходять застосування в області обробки природної мови, оскільки дані, пов'язані з мовою є послідовними за своєю природою. Зазвичай організовані у ланцюжкові структури, рекурентні нейромережі мають спільну значимість на всій довжині ланцюга, що забезпечує постійну розмірність.

Рекурсивні нейромережі працюють з ієрархічними структурами даних через їх деревоподібне розташування. Кожен вузол у дереві має фіксовану кількість

дочірніх вузлів, що дозволяє виконувати рекурсивні операції та використовувати одні й ті ж ваги на кожному кроці.

За ефективністю рекурсивні мережі перевершують прямі мережі. Важливо відзначити, що рекурентні мережі виявляють часове відновлення, що означає, що рекурсивні мережі можна розглядати як узагальнення рекурентних мереж.

Гарним прикладом використання цих методів є стаття [6] та [22] де описані базові принципи детектування.

Ця стаття описує використання рекурентної нейронної мережі, яка навчається через Конекціоністську часову класифікацію, для прямого перетворення вхідного мовлення в символи. Також розглядається її застосування для моделювання мови, що виявляє значно кращі можливості порівняно з статистичними алгоритмами. Важливо відзначити, що нейронні мережі вимагають великої кількості операцій множення та додавання.

Ваги для повністю зв'язаних шарів або рекурентних нейронних мереж використовуються лише один раз при отриманні, тому їх звернення до них має дуже низьку часову локальність. Зроблені спроби зменшити розмір параметрів за допомогою квантування.

Алгоритм розпізнавання мови, представлений у й статті [6] складається з рекурсивної нейронної мережі для акустичного моделювання, рекурсивної нейронної мережі для моделювання мови на рівні символів та статистичного моделювання на рівні слів. Мережа генерує ймовірності символів, аналізуючи кожен кадр вхідного мовлення. Результатом роботи є ймовірності наступних символів, а статистична мережа показує ймовірності наступних слів. Отримана інформація інтегрується для пошуку найкращої гіпотези за допомогою алгоритму пошуку.

Акустична модель має глибоку мережеву структуру і навчається через онлайн-алгоритм. Хоча деякі останні алгоритми розпізнавання мови на основі

рекурсивної нейронної мережі використовують двонаправлену структуру для покращення продуктивності розпізнавання.

На рисунку показано алгоритм роботи

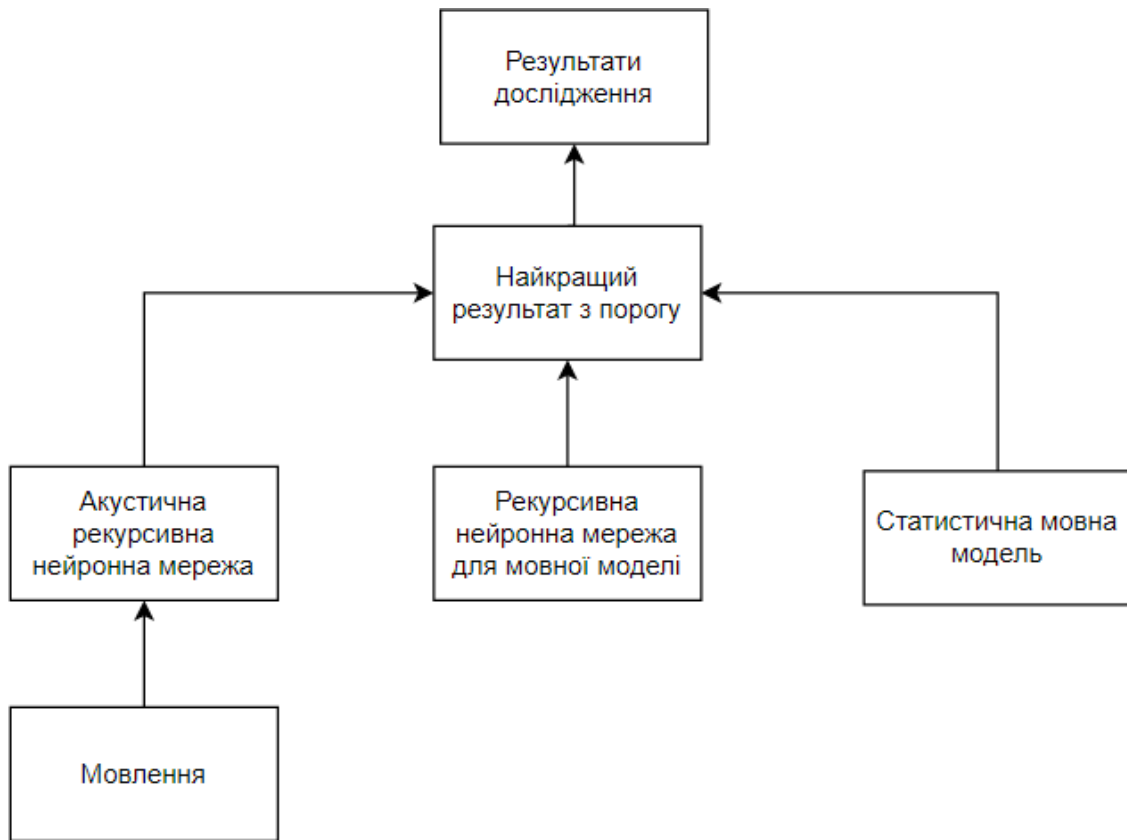


Рисунок 1.11 Структурна схема алгоритму роботи детектора голосу [6], [22]

## Висновки до розділу 1

Виявлення голосової активності є важливим етапом у реалізації різних технологій, таких як системи розпізнавання мови. Застосування алгоритмів обробки сигналів і штучних нейронних мереж дозволяє точно визначати моменти активності голосу, підвищуючи ефективність та зручність взаємодії з голосовими інтерфейсами. У розділі було розглянуто основні методи для детектування голосової активності, їх недоліки та переваги. Такими були: змішані Гаусівські моделі, приховані моделі Маркова, метод опорних векторів, згорткові, рекурентні

та рекурсивні нейронні мережі. Показано спосіб розділення або так званого «мапінгу» даних. Роз'яснено спосіб роботи детектора голосової активності через блок-схему. Було зроблено висновок про ефективність та найбільш оптимальну роботу прихованих моделей Маркова та змішаних Гаусівських моделей. Роль нейронних мереж у детектуванні голосової активності та їх ефективність. Продемонстрована архітектура нейронних мереж для деталізації принципу їх роботи.

## РОЗДІЛ 2. РОЗРОБКА АЛГОРИТМУ РОБОТИ СИСТЕМИ

У даному розділі буде детально розглянута система та її зміни. Так, як розглядається детектування голосової активності використовуючи нейронні мережі варто більш детально описати принцип та особливості роботи даного алгоритму.

### 2.1 Розгляд основних вимог до системи

Побудова системи виявлення голосової активності за допомогою нейронних мереж включає в себе кілька ключових вимог. Подані аспекти які варто врахувати:

1. Попередня обробка – дуже важливий процес. Являє собою видобуття ознак з аудіо-сигналів, часто застосовуючи техніки такі як мел-частотні кепстральні коефіцієнти.

2. Набір даних - різноманітний та репрезентативний набір аудіо-зразків, що містять як мовлення, так і не-мовлення. Цей набір даних повинен охоплювати різних дикторів, акценти та умови оточуючого середовища

3. Архітектура самої мережі - вибір відповідної архітектури нейронної мережі. Рекурентні нейронні мережі (RNN), мережі довгострокової пам'яті (LSTM) або згорткові нейронні мережі (CNN) часто використовуються для завдань, пов'язаних з мовленням. Розгляд можливості використання попередньо навчених моделей або архітектур, призначених для обробки аудіо.

4. Розподіл даних для навчання - розділення набору даних на тренувальний, валідаційний та тестовий набори для точної оцінки продуктивності моделі.

5. Регуляризація - впровадження технік регуляризації, таких як відсів, для запобігання перенавчанню, особливо якщо обмежена кількість маркованих даних.

Оскільки мова зайшла про кепстральні коефіцієнти варто пояснити їх сенс.

Оскільки обробка сигналів та виявлення частотних характеристик мовлення є важливим кроком у будь-якій системі детектування мови, голосовий сигнал має бути перетвореним у тип параметричного представлення даних для подальшої обробки використовується метод мел-частотних кепстральних коефіцієнтів, який і

є найпопулярнішим. Суть даного методу полягає у тому, що він змінює голос на частотні спектри трикутників, які знаходяться на певному інтервалі частот як низьких так і високих для отримання фонетично важливих характеристик мовлення. Для прикладу розподілу на коефіцієнти на рис.2.1 зображена форма слова “Yes”

Можна побачити МКЧП на прикладі слова “Yes” на рис. 2.2

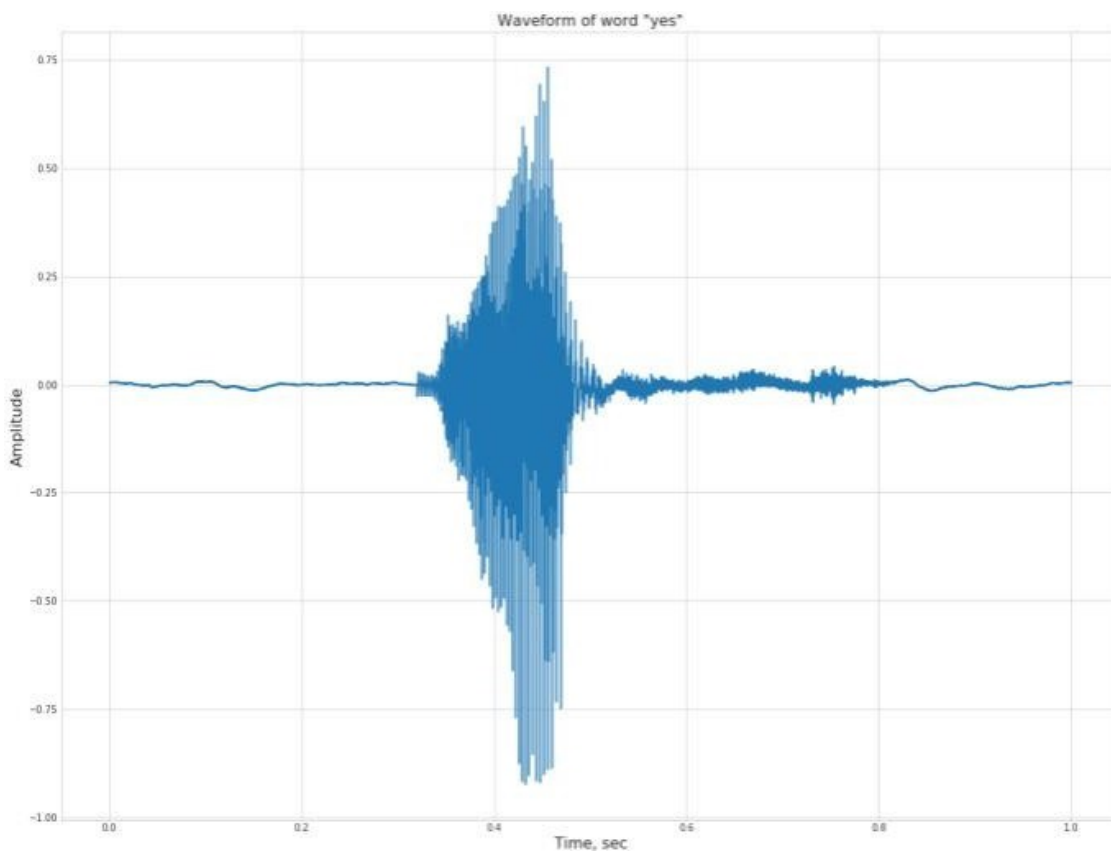


Рисунок 2.1 форма слова “Yes” [7]



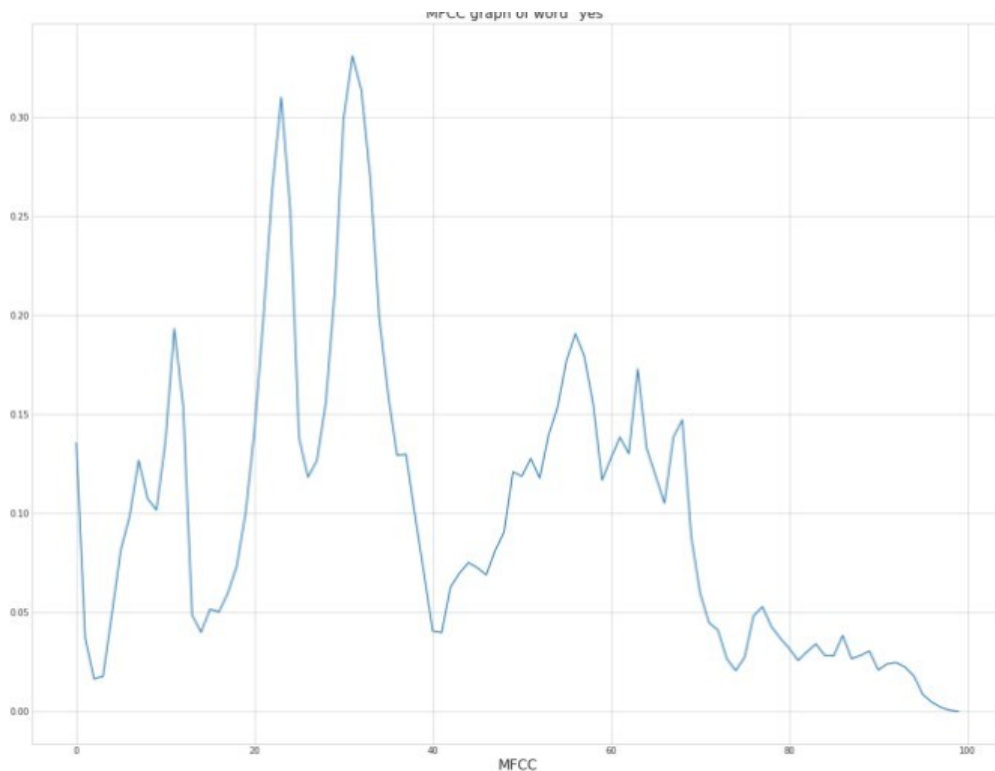


Рисунок 2.2 Графік МЧКП слова “Yes” [7]

Значимість кепстральних коефіцієнтів є дуже великою, оскільки це є попередня обробка сигналу, на рис 2.3 можна побачити даний процес

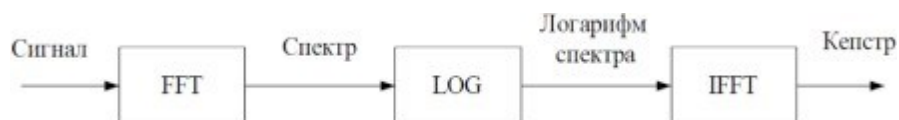


Рисунок 2.3 Процес отримання кепстру з сигналу

Спочатку відбувається перетворення Фур'є, згодом відбувається процедура логарифмування, вона спрямована на те, щоб згладити від спектру, що аналізується та зменшити кількість паразитних піків. Сигнал перетворюється на фрейми та згодом відбувається обернене перетворення Фур'є звідти і отримується сам кепстр.

## 2.2 Основна структура системи

Варто зазначити, що загальна структура системи детектування голосової активності яка проектується на рис. 2.4



Рисунок 2.4 Блок-схема алгоритму детектора мовлення

З блок-схеми можна зробити висновок, що одним з основних блоків є саме первинна обробка сигналу, оскільки нейронні мережі для визначення голосової активності можуть працювати лише з спектрами сигналу, а не з первинним його представленням.

### 2.2.1 Опис компонентів системи

Згідно рис. 2.4 дана система складається з декількох компонентів таких як:

- Блок первинної обробки сигналу
- Вилучення ознак мовлення
- Класифікація мовлення
- Розпізнавання мовлення

Попередня обробка мовлення – відіграє важливу роль в усуненні непотрібних джерел варіацій. Зазвичай цей етап включає приглушення шумів, прибирання відлуння та вирівнювання висоти голосу, що дозволяє значно поліпшити точність розпізнавання мовлення.

Вилучення ознак мовлення - мова та інтонація кожної людини відрізняються через різні особливості, які вони вбудовують у своє мовлення. Теоретично слід мати можливість ідентифікації мови на основі сигналу, але через значні варіації у мовленні виникає необхідність зменшити ці варіації. Це можна зробити за допомогою виділення певних ознак.

Класифікація мовлення - системи, що використовуються для вилучення прихованої інформації з вхідних обробних даних сигналів з використанням математичних операцій.

Розпізнавання мовлення – останній крок у процесі роботи алгоритму. Після успішного виконання минулих кроків відбувається саме розпізнавання, у даному випадку використовується підходом штучного інтелекту.

### **2.3 Принцип роботи ключових компонентів системи**

У випадку даної системи початкова обробка сигналу відбувається за допомогою комп'ютера. Для перетворення звукових хвиль в текст спочатку використовується взяття вибірок сигналу за допомогою теореми Найквіста, оскільки звукові хвилі є неперервним сигналом. Хоча можна безпосередньо вводити сигнал у нейронну мережу, проводиться попередня обробка сигналу для отримання кращих результатів та точних передбачень сказаних слів. Попередня обробка полягає в розділенні обширного вибіркового сигналу на невеликі фрагменти. Отриманий у результаті цього обробленого цифрового сигналу тепер вводиться у рекурентну нейронну мережу, яка є основною моделлю розпізнавання мовлення для передбачення.

Процес відбору та попередньої обробки даних є ключовим етапом у розробці системи розпізнавання мовлення на текст. Цей етап має значущий вплив на продуктивність та інші ключові характеристики системи. Звукові хвилі є одновимірними та мають єдине значення у кожен момент часу, залежно від величини хвилі. Для перетворення цієї звукової хвилі в числові дані проводиться запис величини хвилі в рівномірно розташованих точках – цей процес називається вибірка. На рис. 2.5 відображено результат вибірки.

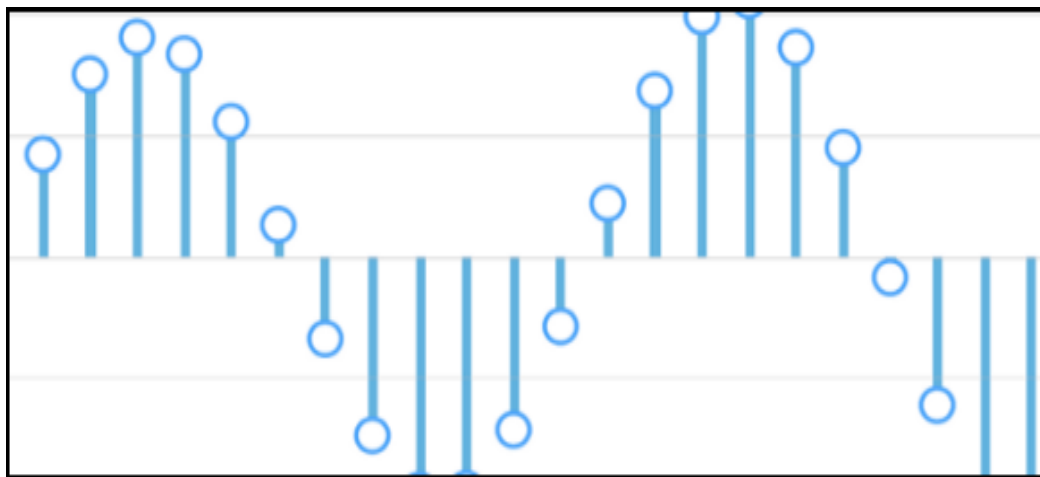


Рисунок 2.5 вибірка мовленнєвого сигналу

Математичні методи використовуються для точного відтворення оригінальної звукової хвилі з розміщеними вибірками, забезпечуючи, що частота вибірки дорівнює або принаймні вдвічі більше найвищої частоти, на яку вона записана. Далі ці дані вводяться у мережу. Попередня обробка передбачає розбиття вибіркових даних на групи. Зазвичай це передбачає групування звукової хвилі протягом певних інтервалів часу. Спільні процеси вибірки та попередньої обробки називаються конвертацією звуку.

Далі відбувається процес вилучення ознак. Для цього використовуються Мел Частотні Кепстральні Коефіцієнти, вони представляють собою стандартний метод вилучення ознак, тобто передбачає використання частотної області, яка ґрунтується на Мел-шкалі, що відтворює враження чуттєвого сприйняття людського вуха. У порівнянні з характеристиками часової області, вони більш точні, оскільки вони

належать до категорії характеристик частотної області. Незважаючи на їхню точність, важливо враховувати чутливість до шуму, оскільки це суттєво залежить від спектральної форми. Для подолання цього обмеження можна використовувати техніки, що використовують періодичність мовних сигналів.

У процесі класифікації мовлення здебільшого використовується Прихована модель Маркова - це найбільш загально визнаний метод розпізнавання шаблонів у мовленні. Вони вважаються безпечнішими та мають стійку математичну базу порівняно з іншими підходами, такими як підходи на основі шаблонів та знань.

У цьому методі модель системи розглядається як Марківський процес з прихованими станами. Мовлення розбивається на менші звукові одиниці, кожна з яких представляє стан. У простіших моделях Маркова стани чітко визначаються для користувача, а ймовірності переходу між станами є параметрами. Марківські Моделі відомі своїм застосуванням у навчанні з підкріплення та розпізнаванні образів, таких як мовлення та почерк. Умовно Марківський процес відображений на рисунку 2.6

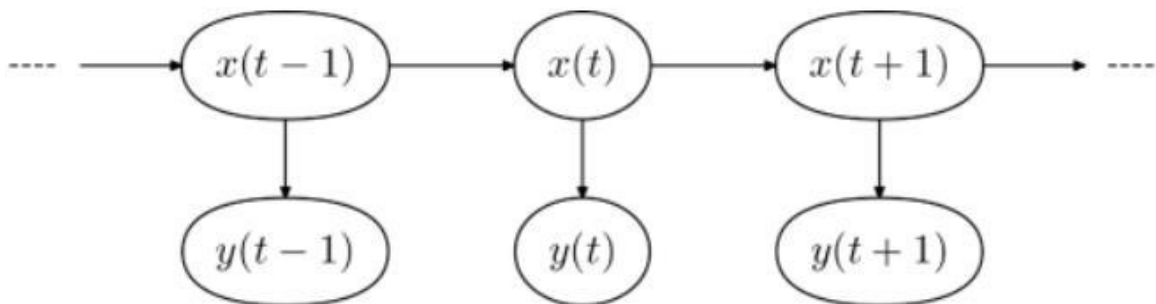


Рисунок 2.6 Умовне відображення Марківського процесу

Останнім етапом відбувається найважливіший фінальний етап – розпізнавання.



Рисунок 2.7 Схеми обробки мови [8]

Рисунок 2.7 ілюструє використання рекурентних нейронних мереж (RNN) для розпізнавання мовлення. Аудіосигнал, який надходить на вхід, легко обробляти, оскільки його подають на глибоку нейронну мережу. Після подачі невеликих аудіофрагментів до мережі, вона визначає літеру, яка відповідає вимовленому звуку. Рекурентна нейронна мережа - це мережа, яка має пам'ять і визначає майбутні прогнози. Це тому, що при прогнозуванні однієї літери це впливає на ймовірність прогнозування наступної літери.

Можна розглянути на прикладі слова «Мумбай». Якщо вже вимовили "Мум", ймовірність того, що наступне висловлення буде "Бай" для завершення слова "Мумбай" - досить велика. Ймовірність сказати щось невимовне, наприклад "АВС", після слова "Мум", набагато нижча. Таким чином, наявність пам'яті про попередні прогнози підвищує здатність нашої мережі робити більш точні прогнози в майбутньому. Проста рекурентна нейронна мережа працює на подібному принципі.

Наприклад, людина може ходити на роботу однією дорогою одного дня, іншою наступного, а згодом третьою, потім повернутися до першого та другого варіанту шляху, повторюючи цей шаблон. В цьому контексті вхід залежить від попередніх виходів, і послідовність повторюється через певний час.

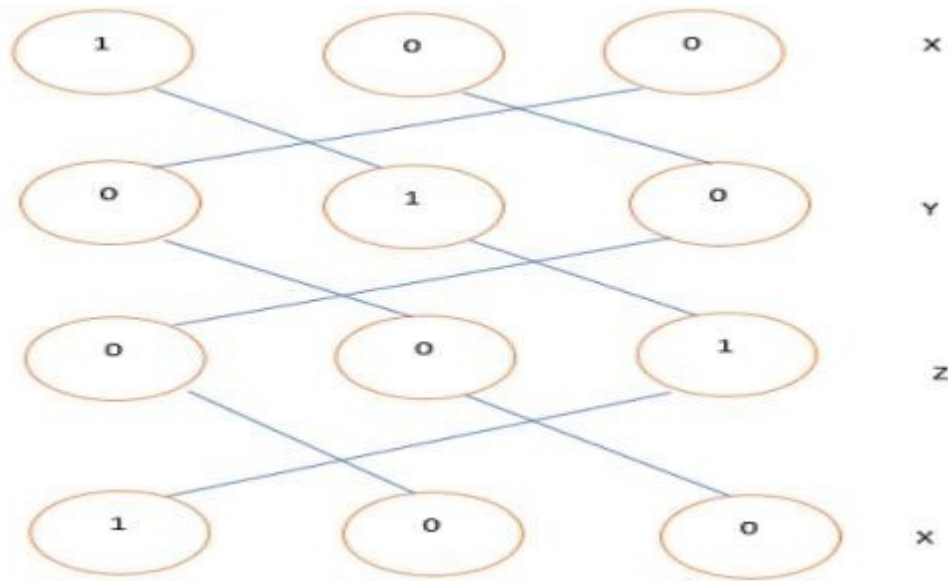


Рисунок 2.8 відображення Рекурентної нейронної мережі [8]

Найчастіше використовуваними варіантами рекурентних нейронних мереж є LSTM (довгострокова короткочасна пам'ять) це є основними блоками для кількох шарів мережі. Як вже відомо, нейронна мережа стикається з проблемою довгострокових залежностей, яку можна усунути за допомогою нової її форми, відомої як LSTM. Всі Рекурентні мережі складаються з повторюючої структури, але у випадку LSTM структура трошки відрізняється. Вона складається з чотирьох компонентів. Основний елемент у LSTM - це клітинний стан, через який проходить будь-яка інформація. Також можна додавати чи вилучати будь-яку інформацію за потреби за допомогою воріт. Є три типи кластерів: кластери вводу, кластери виведення та кластери забуття. Функції цих кластерів полягають в захисті або контролі над клітинним станом. Мережа LSTM складається з сигмоїдальної функції, яка має лише два значення на виході: вона може передати все, що є на вході, або ні. Тому LSTM знайшли застосування в новіших версіях програмного забезпечення для розпізнавання мовлення.

Для навчання рекурентної нейронної мережі використовується алгоритм зворотного розповсюдження помилки (backpropagation), але з підходом, який відрізняється від традиційного. Причина відмінності полягає в тому, що на кожному кроці часу параметри, які використовуються, спільні для всієї мережі. Обчислений градієнт на кроці виходу залежить не лише від поточного кроку часу, але й від

обчислень попередніх кроків. Цей метод обчислень відомий як "Зворотне розповсюдження у часі"). Одним з недоліків навчання нейронної мережі за допомогою алгоритму зворотного розповсюдження є проблема довгострокових залежностей, також відома як проблема зникнення або вибухання градієнту. Для вирішення таких недоліків використовуються певні техніки, і тип мережі LSTM.

## **2.4 Використання у системах з низьким енергоспоживанням**

Варто розглянути процес енергоефективності використання методу на прикладі двох мікроконтролерів. Для ефективного використання нейронних мереж на таких мікроконтролерах необхідно провести оптимізацію, що включає зменшення як кількості шарів, так і кількості нейронів у відповідних шарах. При розробці вбудованих рішень, заснованих на глибокому навчанні, важливо знаходити баланс між швидкістю прогнозування, його точністю, витратою енергії та обсягом пам'яті мікроконтролера.

Для оцінки можна взяти мікроконтролери ARM Cortex-M, вироблених STM, зокрема серії F. Серія F охоплює широкий спектр родин мікроконтролерів за часом виконання, розміром пам'яті, обробкою даних та можливостями передачі, тоді як серія H надає високу продуктивність, безпеку та мультимедійні можливості. Мікроконтролери серії L призначені для застосувань з надзвичайно низьким споживанням енергії в енергоефективних вбудованих системах. Варто порівняти мікроконтролер ARM Cortex-M з Arduino Nano.



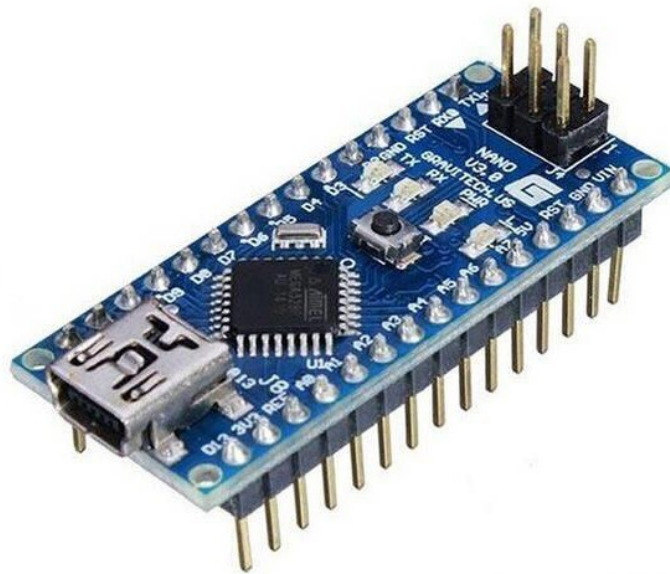


Рисунок 2.9 Зображення мікроконтролера Arduino Nano

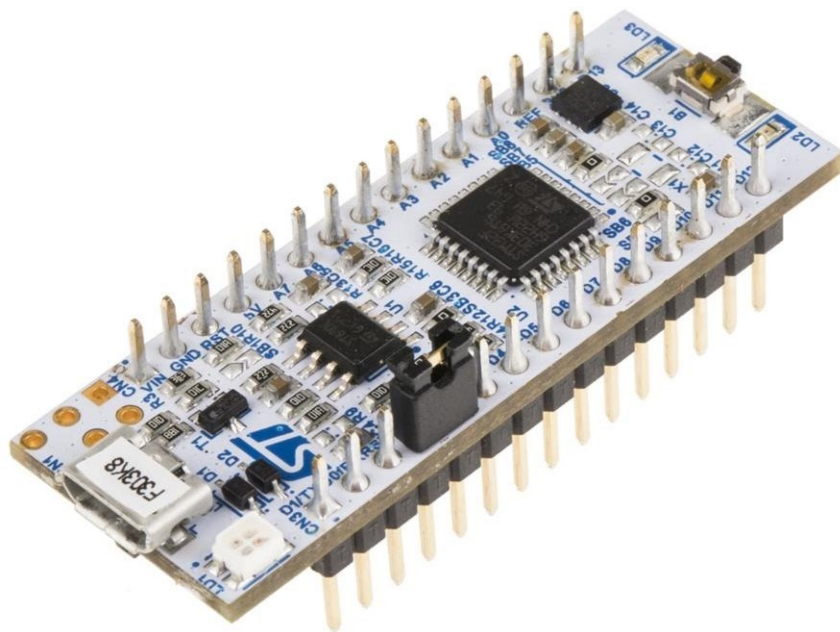


Рисунок 2.10 Мікроконтролер STM32F серії

Використання Arduino Nano в нейронних мережах включає кілька важливих аспектів через обмежені обчислювальні ресурси мікроконтролера.

1. Обмеження пам'яті - має обмежену програмну пам'ять (32 КБ) та ОЗП (2 КБ)
2. Обчислювальна потужність - має відносно низьку швидкість обробки (16 МГц). Складні обчислення нейронних мереж можуть виконуватися повільно. Варто розглянути можливість вивантаження обчислень на більш потужний пристрій за необхідності
3. Підтримка бібліотек для навчання нейронної мережі – є можливість використання вже готових бібліотек, що оптимізовані саме для цього мікроконтролера – TensorFlow, наприклад.
4. Вимоги роботи у реальному часі – варто оцінити чи є важливою обробка у режимі реального часу для програми, оскільки через обмежену швидкість обробки даний мікроконтролер може бути неефективним та енергозатратним.

Тож, можна зробити висновок, що Arduino Nano є найбільш оптимальним для нескладних задач нейронної мережі та простих додатків, хоча позиціонує себе як енергоефективна платформа.

Розглянемо мікроконтролер STM32. Використання мікроконтролерів STM32 у нейронних мережах передбачає конкретні аспекти через обмежені обчислювальні ресурси мікроконтролера.

1. Пам'ять – у даних контролерів є обмеження як і флеш так і оперативної пам'яті, хоча вона є більш розширеною у порівнянні з Arduino
2. Потужність обчислень – даний тип контролерів є швидкодіючим для обчислень, хоча все рівно варто обирати не складні типи нейронних мереж, через велику енергозатратність при обчисленнях.
3. Бібліотеки для навчання мережі – є вже вбудовані бібліотеки машинного навчання, оптимізовані для вбудованих систем, такі як TensorFlow Lite для мікроконтролерів або CMSIS-NN.

4. Вимоги до роботи у реальному часі – є оптимізованими до роботи у реальному часі, оскільки мають відносно велику швидкість обробки.

Тож, можна зробити висновок, що мікроконтролери STM32 – є оптимальним як і для складних так і для простих задач, в особливості нейронних мереж.

Також їх дуже сильною перевагою є те, що Мікроконтролери STM32 відомі своєю низькою споживаною енергією. Ця функція є дуже важливою створення енергоефективних вбудованих рішень.

## **Висновки до розділу 2**

На основі розглянутих в першому розділі методів детектування голосової активності було виконано проектування основного алгоритму роботи розробленої системи. Розглянуто основні правила проектування даних систем та основні вимоги для правильного функціонування. Пояснено сенс існування кепстральних коефіцієнтів, які є важливими для правильної обробки сигналу, з якого детектується мовлення. Зображено структуру системи детектування голосової активності у вигляді блок-схеми, детально описано принцип та алгоритм її роботи. Описано склад компонентів системи та принцип їх роботи. Схематично пояснено обробку мови за допомогою рекурентних нейронних мереж, яка використовується у моделюванні системи. Детально описано основні переваги та недоліки використання системи даного типу на платформі мікроконтролера ARM Cortex-M та Arduino Nano.

## **РОЗДІЛ 3. МОДЕЛЮВАННЯ ТА РОЗРОБКА СИСТЕМИ ДЕТЕКТУВАННЯ ГОЛОСОВОЇ АКТИВНОСТІ**

У даному розділі буде описана практична частина розробки системи детектування голосової активності. Виявлення голосової активності вважається важливим елементом багатьох систем, таких як системи автоматичного розпізнавання мови та ідентифікації динаміків. Особливо складно визначати голосову активність в умовах низького відношення сигнал/шум (SNR), коли мовлення перекривається шумом, тому для дослідження використовується програмне забезпечення Matlab.

### **3.1 Програмна сфера розробки**

Matlab – потужним середовищем для числових обчислень та розробки алгоритмів. Розроблений компанією MathWorks, цей інструмент став невід'ємним у наукових дослідженнях та інженерній розробці завдяки своїм можливостям.

Однією з визначальних рис середовища є спрощений підхід до операцій над матрицями та векторами, що робить його ідеальним для вирішення задач лінійної алгебри. Широкий спектр функцій і інструментів дозволяє швидко розробляти та впроваджувати алгоритми. Крім того, надає розширені можливості для обробки та аналізу даних, візуалізації результатів та розробки графічних інтерфейсів.

З роками Matlab став невід'ємним інструментом у різних галузях, включаючи інженерію, фізику, фінанси та біологію. З його простим синтаксисом та обширними можливостями Matlab залишається популярним серед дослідників та інженерів, допомагаючи їм розв'язувати складні числові обчислювальні задачі та проводити дослідження.

Основні концепції даного програмного забезпечення:

- Аналіз даних. Надані потужні інструменти для аналізу даних, статистичного аналізу та обробки сигналів. Існує багато наборів і функцій.
- Розробка алгоритмів. Matlab надає платформу для розробки та впровадження алгоритмів, що робить його цінним для дослідників, інженерів і вчених.
- Розробка застосунків. Програмне забезпечення дозволяє користувачам створювати автономні застосунки, роблячи його універсальним для створення власних рішень.
- Simulink. Він продуктом-супутником до програмного забезпечення, який надає графічне середовище для моделювання, симуляції та аналізу динамічних систем.

Дане програмне середовище має широкі можливості для відображення векторів та матриць у вигляді графіків, а також анотування та друку цих графіків. Він включає функції високого рівня для візуалізації двовимірних та тривимірних даних, обробки зображень, анімації та графіки презентацій. Також він має функції низького рівня, які дозволяють повністю налаштовувати вигляд графіки та створювати повноцінні графічні інтерфейси, що і є необхідним для перевірки роботи алгоритму.

### **3.2 Алгоритм роботи та модель програми**

Перед розглядом безпосередньо програми та алгоритму варто зазначити дані, що ми використовуємо – вони складаються з трьох основних категорій аудіозразків: мовлення, шум, а також з комбінації мовних сигналів з шумом. Мовлення включає аудіозаписи з мовою, що охоплюють різні теми та мовні моделі. Аудіозразки категорії "шум" представляють реальні акустичні середовища. Крім того, комбінування мови та звуків шуму створює нові аудіосигнали для тестування алгоритму з метою переконання, що мовлення правильно вилучається, ідентифікується та виявляється. Ці аудіофайли охоплюють широкий спектр мовних

моделей і лінгвістичного контенту, що робить їх цінними для тренування та тестування алгоритмів виявлення голосової активності.

Щоб підготувати набір даних для оцінки системи виявлення голосової активності, виконується кілька процедур з обробки набору даних, щоб забезпечити правильність обробки, особливо щодо змішування мовлення та шуму. Оскільки аудіодані було зібрано з різних джерел, виконується обробка для забезпечення однакової частоти дискретизації аудіофайлів для забезпечення сумісності, також аудіофайли обмежуються однією тривалістю. Проведено змішування сигналів, комбінуючи аудіосигнали для імітації реальних сценаріїв.

Спершу встановлюються правила для шуму та мовлення, щоб не отримувалися помилкові сигнали, одночасно забезпечуючи можливість виявлення мовлення в аудіосигналі. Важливо знайти баланс, щоб виявляти людське мовлення, уникаючи помилкових виявлень шуму.

Для правильної роботи програми створена функція `pr_energy`. Ця функція обробляє змішаний сигнал за допомогою полосової фільтрації для зменшення фонового шуму та покращення точності. Та використовує одношарову нейронну мережу для навчання. Використовується фільтр Баттерворта 4-го порядку із полосою пропускання від 300 Гц до 1500 Гц.

```
fpass = [300 1500];  
[b, a] = butter(4, fpass/(fs/2));  
signal = filter(b, a, signal);
```

Рисунок 3.1 Фрагмент коду створення фільтру Баттерворта

Цей діапазон був обраний через його стабільні характеристики продуктивності під час експериментів та акценту на мовленні людини, яке, як правило, потрапляє в цей діапазон частот. Фільтр відіграє важливу роль у функціонуванні системи. Він ефективно видаляв небажаний фоновий шум, зберігаючи при цьому важливі компоненти мовлення.

Використовується нейронна мережа, що детектує голос на кожному з відрізків вхідного сигналу. На вхід мережі подаються вхідний сигнал, на яких відбувається детектування голосу на виході мережі буде вже двійковий результат, тобто 0 - голосу немає, 1 –голос присутній.

```
% початок циклу зі згорткою
net = patternnet([20 10]);
net.trainParam.epochs = 100; % кількість ітерацій у мережі
```

Рисунок 3.2 Фрагмент коду створення нейронної мережі

Для попередньої обробки голосового сигналу використовуємо цикл згортки. Для початку ми маємо оголосити змінні для нього:

```
frame_size = round((win*10^-3)*fs);
signal(1:fs) = 0;
signal(end-fs:end) = 0;
len = length(signal);

voice_bin_len = ceil(len/frame_size);
voice_bin = zeros(1, voice_bin_len);
snr = zeros(1, voice_bin_len);
count = 1;
```

Рисунок 3.3 Фрагмент коду оголошення змінних для згортки

Де вказано розмір вікна, повну кількість елементів, та що головне - розмір результуючого двійкового масиву де зберігатимуться двійкові значення, тобто значення чи є голос, чи ні.

Далі відбувається власне цикл згортки. Суть циклу полягає у наступному:

1. Перевірка на те, чи дійшло вікно до кінця масиву
2. Процес вибірки сигналу
3. Обрахунок поточної вибірки сигналу
4. Перетворення Фур'є
5. Знаходження початкових та кінцевих індексів у частотному діапазоні

6. Вибірка результуючих значень
7. Включення динамічних індексів
8. Знаходимо кожне значення відношення сигналу до шуму
9. Детектування голосової активності за допомогою нейронної мережі
10. Ітерація до наступного фрейму

```
for i = 1:round(frame_size):round(len)
    edge = (frame_size + i) - 1;
    if edge > len
        edge = len;
    end
    frame = signal(i:edge);
    ts = 1/fs;
    Lfft = 1024;
    freq_axis = (0:Lfft/2-1)/(Lfft*ts);
    frame_size = max(1, round((win*10^-3)*fs));
    S = fft(double(frame), frame_size);
    S = S(1:floor(length(S)/2));
    locations = find(freq_axis >= 195);
    begin_index = locations(1);

    locations = find(freq_axis >= 3843);
    end_index = locations(1);
    S(1:begin_index) = 0.0;
    S(end_index:end) = 0.0;
    S = abs(S);
    S_squared = S.^2;
    Et = sum(S_squared);
```



```

S_sorted = sort(S_squared);

idx_start = round(0.10 * length(S_sorted));
idx_end = round(0.20 * length(S_sorted));
Np = (1/100.0)*sum(S_sorted(idx_start:idx_end)) + .00001;
L = 1;

upper_idx = length(S_sorted);

for j = upper_idx:-1:1
    if sum(S_sorted(j:upper_idx)) >= Et*.4;
        L = j;
        break;
    end
end
if L == upper_idx
    L = L - 1;
end
Sp = (1.0/(upper_idx-L))*sum(S_sorted(L:upper_idx)) + 1e-10;

snr(count) = (Sp / (Np + 1e-10));
voice_bin(count) = (Sp/Np) > thresh + 90;
count = count + 1;

```

Рисунок 3.4 Фрагмент коду циклу згортки

Після циклу згортки ми вже отримали сигнал, надалі варто порахувати його енергію. Для даних обрахунків створюється окремий масив, де записується значення енергії кожного фрейму. Працює це наступним чином – створюється масив, визначається його розмір, знаходиться квадратне значення сигналу та перевіряється чи не дійшло вікно до кінця масиву після успішної перевірки відбувається вже розрахунок конкретного значення та перехід до наступних фреймів. Перед цим копіюється результат детектування у новий масив, щоб повернути його при виході з функції.

```

e_len = ceil(len/frame_size); % розмір масиву
energy = zeros(1, e_len); % масив для значень енергії
signal = signal.^2; % знаходимо квадрат сигналу
count = 1; % змінна для ітерації
% перевірка на те чи не дійшло вікно до кінця масиву
for i = 1:frame_size:len
    edge = (i + frame_size) - 1;
    if edge > len
        edge = len;
    end
    % розраховуємо значення енергії
    pow2 = sum(signal(i:edge));
    pow = sqrt(pow2);
    energy(count) = pow;
    count = count + 1; % наступна ітерація

```

Рисунок 3.5 Фрагмент коду визначення енергії

Також у програмі використовується основний файл signalProcessing, у якому відбувається основна робота з сигналами. Спочатку ми читаємо сигнали, у яких детектуємо голосову активність, тобто сигнал з голосом та шум.

Далі відбувається процедура ресемплування за необхідністю, це є важливим кроком перед початком обробки, оскільки алгоритм може працювати лише з сигналами однієї частоти дискретизації.

```

if fs1 ~= fs2
    audioSignal2 = resample(audioSignal2, fs1, fs2);
end
fs = fs1;

```

Рисунок 3.6 Фрагмент коду процедури ресемплування

Далі сигнали вирівнюються по довжині, оскільки це потрібно для правильної роботи алгоритму.

```
len1 = length(audioSignal1);  
len2 = length(audioSignal2);  
minLen = min(len1, len2);  
  
audioSignal1 = audioSignal1(1:minLen);  
audioSignal2 = audioSignal2(1:minLen);
```

Рисунок 3.7 Фрагмент коду вирівнювання довжини сигналу

У подальшому відбувається змішування сигналів та їх нормалізація, щоб сигнал був у визначеному діапазоні. Також варто зазначити, що для точної детекції рівень сигнал/шуму (SNR) приблизно рівний 90 дБ, це пов'язано з частотою людського голосу.

Після змішування та нормалізації сигналу відбувається виклик функції `pr_energy_sad`, де відбувається обробка та видається вже кінцевий результат.

Обробка різноманітних аудіосигналів, зокрема тих із рівнем 90 дБ, сприяли точності та стійкості системи виявлення голосової активності. Ці параметри дозволили системі точно визначати інтервали мовлення серед різноманітних джерел шуму. Довжина вікна забезпечила ефективний аналіз аудіосегментів, гарантуючи виявлення сегментів мовлення з високою точністю. Поріг сигнал-шум служив надійним критерієм для відокремлення мовлення від шуму, покращуючи продуктивність системи в умовах складних акустичних середовищ.

### **3.3 Моделювання результатів обробки**

Для експериментального дослідження було обрано два сигнали – звуки мовлення людини та звук удару молотка, як шуму.

Результат детектування є наступним:

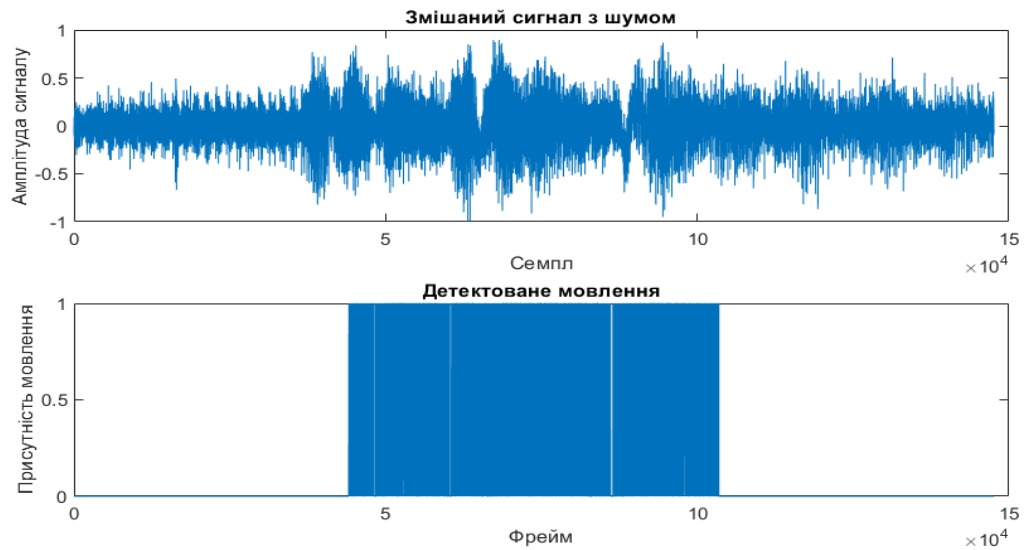


Рисунок 3.8 Результат детектування мовлення з шумом удару молотка

Як бачимо, мовлення детектується на відрізку з  $5 \cdot 10^4$  по  $10 \cdot 10^4$  семпл.

Для наглядності дослідження оберемо інші сигнали – те ж саме мовлення людини та звук сирени поліцейського автомобіля.

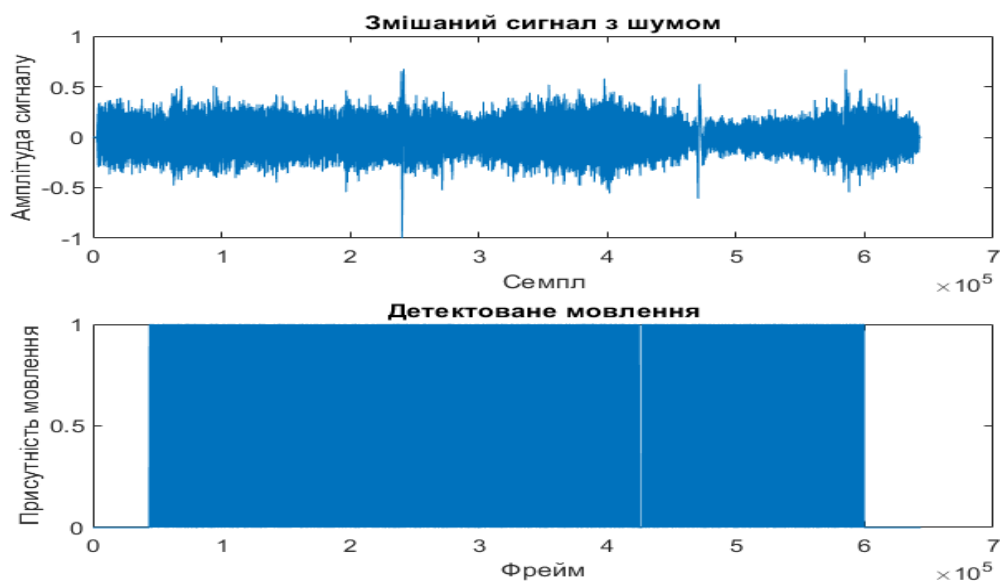


Рисунок 3.9 Результат детектування мовлення з шумом сирени автомобіля

По результату можна стверджувати, що мовлення детектується майже на всьому проміжку аудіозапису.

До прикладу візьмемо той самий аудіофайл та дуже сильний сильний шум, що був згенерований генератором випадкового шуму.

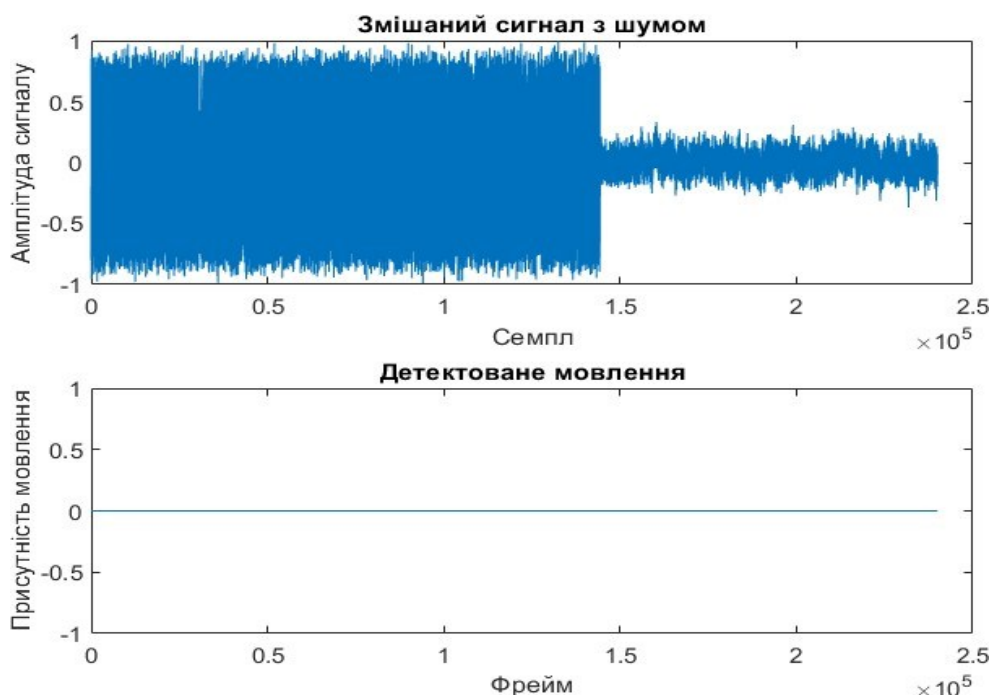


Рисунок 3.10 результат детектування сигналу з дуже сильним шумом

У результаті мовлення не детектується.

Оскільки експериментально було проведено детектування звуків при різному рівні SNR, варто додати таблицю залежності спрацювання детекції від рівня співвідношення сигнал/шуму даного алгоритму та для порівняння інших відомих алгоритмів.

Шум	Рівень SNR			% детектування голосу іншим алгоритмом з використанням нейронної мережі без використання фільтру Баттерворта
	-10 dB	10 dB	30 dB	
Стук молотка	10%	55%	80%	
Поліцейська сирена	10%	50%	70%	
Генератор шуму	15%	40%	80%	

Шум	Рівень SNR			% детектування голосу у даному алгоритмі
	-10 dB	10 dB	30 dB	
Стук молотка	10,2%	56,2%	81,7%	
Поліцейська сирена	10,2%	51,8%	71,6%	
Генератор шуму	15,4%	41,3%	82%	

Отже, можна стверджувати що спроектований алгоритм з використанням полосового фільтру являється кращим за вже відомі.

### **Висновки до розділу 3**

У даному розділі описано моделювання роботи системи за допомогою програмного середовища Matlab. Описано основні фрагменти програми – тобто функцію `nr_energy` та `SignalProcessing`, детально описаний принцип роботи блоків кожної з програм, за результатами роботи програми створені графіки для відображення. Пояснено використання такого компоненту як фільтр Баттерворта у програмі та також пояснено про використання операції згортки. Можна стверджувати, що результати моделювання є хорошими, оскільки майже у всіх випадках, крім випадку дуже сильного шуму детектується мовлення. Також було проведено експериментальне дослідження зміни рівня відношення сигналу до шуму відповідно до кожного зі звуків. Експериментальна реалізація даного методу детектування голосової активності на базі мікроконтролера ST32F Series є частиною майбутніх досліджень.

## РОЗДІЛ 4. РОЗРОБКА СТАРТАП ПРОЄКТУ

### 4.1 Опис ідеї проекту “Дослідження голосової активності у системах з низьким енергоспоживанням”

Необхідно послідовно проаналізувати та подати у вигляді таблиці:

- зміст ідеї (що пропонується);
- можливі напрямки застосування;
- основні вигоди, що може отримати користувач
- чим відрізняється від існуючих аналогів та замінників;

Таблиця 4.1 Опис ідеї стартап-проекту

<i>Зміст ідеї</i>	<i>Напрямки Застосування</i>	<i>Вигода для кінцевого користувача</i>
Удосконалення алгоритмів детектування голосової активності у системах з низьким енергоспоживанням	Визначення голосової активності у системах з низьким енергоспоживанням	Підтвердження впевненості щодо покращення використання детектування у даному типі систем

Аналіз потенційних техніко-економічних переваг ідеї (чим відрізняється від існуючих аналогів та замінників) порівняно із пропозиціями конкурентів передбачає:

- Визначення переліку техніко-економічних властивостей та характеристик ідеї (орієнтований можливий перелік властивостей та характеристик подано у додатку А);
- Визначення попереднього кола конкурентів (проектів-конкурентів) або товарів-замінників чи товарів-аналогів, що вже існують на ринку, та проводиться збір інформації щодо значень техніко-економічних показників для ідеї власного проекту та проектів-конкурентів відповідно до

визначеного вище переліку.

- Проводиться порівняльний аналіз показників: для власної ідеї визначаються показники, що мають а) гірші значення (W, слабкі); б) аналогічні (N, нейтральні) значення; в) кращі значення (S, сильні) (табл. 4.2).

Таблиця 4.2 Визначення сильних, слабких та нейтральних характеристик ідей проекту

№ п/ п	Техніко- економічні характеристики ідеї	(потенційні) товари/концепції і конкурентів		W (слабк а сторона)	N (нейтральн а сторона)	S (сильна сторона)
		Мій проект	Конкуре нт1			
1.	Покращений контроль якості	Високий	Середній			+
2.	Потреба ручного контролю	Середня	Висока	+		
3.	Автоматичний збір багатьох показників ефективності	Висока	Середня			+

Визначений перелік слабких, сильних та нейтральних характеристик та властивостей ідеї потенційного товару є підґрунтям для формування його конкурентоспроможності. Проект – конкурентоспроможний.

#### 4.2 Технологічний аудит ідеї проекту

В межах даного підрозділу необхідно провести аудит технології, за допомогою якої можна реалізувати ідею проекту (технології створення



товару). Визначення технологічної здійсненності ідеї проекту передбачає аналіз таких складових (таблиця 4.3):

- за якою технологією буде виготовлено товар згідно ідеї проекту?
- чи існують такі технології, чи їх потрібно розробити/добробити?
- чи доступні такі технології авторам проекту?

*Таблиця 4.3 Технологічна здійсненність ідеї проекту*

<i>№ n/n</i>	<i>Ідея проекту</i>	<i>Технології її реалізації</i>	<i>Наявність технологій</i>	<i>Доступність технологій</i>
	Дослідження Голосової активності у системах з малим енергоспоживання м	Автоматичний обрахунок, рішення	Технології наявні, але не розголошуються на широкий загаль	Доступні

### **4.3 Аналіз ринкових можливостей запуску стартап-проекту**

Визначення ринкових можливостей, які можна використати під час ринкового впровадження проекту, та ринкових загроз, які можуть перешкодити реалізації проекту, дозволяє спланувати напрями розвитку проекту із урахуванням стану ринкового середовища, потреб потенційних клієнтів та пропозицій проектів-конкурентів.

Визначаються потенційні групи клієнтів, їх характеристики, та формується орієнтовний перелік вимог до товару для кожної групи (таблиця 4.4).

Таблиця 4.4 Характеристика потенційних клієнтів стартап-проекту

<i>№ n/n</i>	<i>Фактор</i>	<i>Зміст загрози</i>	<i>Можлива реакція компанії</i>
1	Збій у програмному забезпеченні	Знеможливлення подальшої роботи системи	Очікування відновлення, або пошук інших постачальників послуг
2	Поява затримок у детектуванні	Зниження швидкості	Очікування відновлення, або пошук інших постачальників послуг

Після визначення потенційних груп клієнтів проводиться аналіз ринкового середовища: складаються таблиці факторів, що сприяють ринковому впровадженню проекту, та факторів, що йому перешкоджають (таблиці 4.5-4.6).

Таблиця 4.5 фактори загроз

<i>№ n/n</i>	<i>Фактор</i>	<i>Зміст загрози</i>	<i>Можлива реакція компанії</i>
1	Збій у програмному забезпеченні	Знеможливлення подальшої роботи системи	Очікування відновлення, або пошук інших постачальників послуг
2	Поява затримок у детектуванні	Зниження швидкості	Очікування відновлення, або пошук інших постачальників послуг

Табл. 4.6 фактори можливостей

<i>№ n/n</i>	<i>Фактор</i>	<i>Зміст можливості</i>	<i>Можлива реакція компанії</i>
1	Надання розширених послуг у дослідженні	Розширення ринку споживачів	Співпраця з ширшим колом компаній
2	Надання різного рівня <u>обсяжності</u> обслуговування	Розширення ринку споживачів	Створення відповідних підписок

Надалі проводиться аналіз пропозиції: визначаються загальні риси конкуренції на ринку (таблиця 4.7).

*Таблиця 4.7 Ступеневий аналіз конкуренції на ринку*

<i>Особливості конкурентного середовища</i>	<i>В чому проявляється дана характеристика</i>	<i>Вплив на діяльність підприємства (можливі дії компанії, щоб бути конкурентоспроможною)</i>
1. Тип конкуренції – чиста	Компанії не впливають на загальний стан на ринку зв'язку	Постійний аналіз ситуації
2. За рівнем конкурентної боротьби – національний	На національному ринку присутні компанії у даній галузі	Перспективна охоплення ринку даної сфери
3. За галузевою ознакою – внутрішньогалузева	Застосування лише в галузі Дослідження голосової активності	Співпраця та подальший розвиток з іншими компаніями
4. Конкуренція за видами товарів – товарно-видова	Конкуренція можлива лише з існуючими методиками досліджень голосової активності	Удосконалення методів <u>детекції</u> голосової активності

На основі характеристик ідеї проекту (таблиця 4.2), вимог споживачів до товару (таблиця 4.4) та факторів маркетингового середовища (таблиця 4.6 - 4.7) визначається та обґрунтовується перелік факторів конкурентоспроможності. Аналіз оформлюється у таблиці 4.8

Таблиця 4.8 Обґрунтування факторів конкурентноспроможності

№ n/n	Фактор конкурентноспроможності	Обґрунтування (наведення чинників, що роблять фактор для порівняння конкурентних проектів значущим)
1	Ресурси знань	Сума наукової, технічної та ринкової інформації
2	Комерційна діяльність	Надання послуг якісного детектування голосової активності
3	Система розроблення і впровадження нововведень	Прогнозування погіршення швидкості, оцінка якості детектування мови, як результату

Фінальним етапом ринкового аналізу можливостей впровадження проекту є складання SWOT-аналізу (матриці аналізу сильних (Strength) та слабких (Weak) сторін, загроз (Troubles) та можливостей (Opportunities) (таблиця 4.9) на основі виділених ринкових загроз та можливостей.

Перелік ринкових загроз та ринкових можливостей складається на основі аналізу факторів загроз та факторів можливостей маркетингового середовища. Ринкові загрози та ринкові можливості є наслідками (прогнозованими результатами) впливу факторів, і, на відміну від них, ще не є реалізованими на ринку та мають певну ймовірність здійснення. Наприклад: зниження доходів потенційних споживачів – фактор загрози, на основі якого можна зробити прогноз щодо посилення значущості цінового фактору при виборі товару та відповідно, – цінової конкуренції (а це вже – ринкова загроза).

Таблиця 4.9 SWOT аналіз стартап-проекту

<p><b>Сильні сторони:</b></p> <ul style="list-style-type: none"> <li>- ведення прогнозування погіршення якості детектування</li> <li>- використання сучасних технологій</li> </ul>	<p><b>Слабкі сторони:</b></p> <ul style="list-style-type: none"> <li>- апаратні витрати</li> <li>- неналагоджені канали реалізації</li> <li>- маркетингові дослідження</li> </ul>
<p><b>Можливості:</b></p> <ul style="list-style-type: none"> <li>- Надання послуг розширеного спектру досліджень</li> <li>- Надання послуг захищеного використання сервісу</li> </ul>	<p><b>Загрози:</b></p> <ul style="list-style-type: none"> <li>- Ризики програмних збоїв</li> <li>- Поява затримок передачі при детектуванні</li> </ul>

На основі SWOT-аналізу розробляються альтернативи ринкової поведінки (перелік заходів) для виведення стартап-проекту на ринок та орієнтовний оптимальний час їх ринкової реалізації з огляду на потенційні проекти конкурентів, що можуть бути виведені на ринок. Визначені альтернативи аналізуються з точки зору строків та ймовірності отримання ресурсів (таблиця 4.10).

Таблиця 4.10 Альтернативи ринкового впровадження стартап-проекту

№ п/п	Альтернатива (орієнтовний комплекс заходів) ринкової поведінки	Імовірність отримання ресурсів	Строки реалізації
1	Надання більш широкого спектру послуг при детектуванні голосової активності	Потребує змін у програмній частині	Через більший обсяг з'являється потреба у більшій кількості часу
2	Зниження на перших порах ціни на послуги	Потребує тих самих ресурсів	З'являється потреба у більшій кількості часу, потрібна буде сильна маркетингова кампанія

За необхідності рекомендовано обрати другу альтернативу ринкового впровадження стартап-проекту.

#### 4.4 Розроблення ринкової стратегії проекту

Розроблення ринкової стратегії першим кроком передбачає визначення стратегії охоплення ринку: опис цільових груп потенційних споживачів (таблиця 4.11).

Таблиця 4.11 вибір цільових груп потенційних споживачів

<i>№ n/ n</i>	<i>Опис профілю цільової групи потенційних клієнтів</i>	<i>Готовність споживачів сприйняти продукт</i>	<i>Орієнтовний попит в межах цільової групи (сегменту)</i>	<i>Інтенсивність конкуренції в сегменті</i>	<i>Простота входу у сегмент</i>
1	Корпоративний сектор	Висока	Високий	Високий	Висока
2	Приватний сектор	Низька	Середній	Середня	Висока
Які цільові групи обрано: першу з огляду на орієнтовний попит в межах цільової групи.					

За результатами аналізу потенційних груп споживачів (сегментів) обираються цільові групи, для яких пропонуватиметься товар, та визначають стратегію охоплення ринку:

- якщо компанія зосереджується на одному сегменті – вона обирає стратегію концентрованого маркетингу;
- якщо працює із кількома сегментами, розробляючи для них окремо програми ринкового впливу – вона використовує стратегію диференційованого маркетингу;
- якщо компанія працює із всім ринком, пропонуючи стандартизовану програму (включно із характеристиками товару/послуги) – вона використовує масовий маркетинг.

Для роботи в обраних сегментах ринку необхідно сформувати базову стратегію розвитку (таблиця 4.12).

Таблиця 4.12 визначення базової стратегії розвитку

<i>№ п/ п</i>	<i>Обрана альтернатива розвитку проекту</i>	<i>Стратегія охоплення ринку</i>	<i>Ключові конкурентоспромо жні позиції відповідно до обраної альтернативи</i>	<i>Базова страт егія розвит ку</i>
1	Зниження на перших порах ціни на послуги	Стратегія диференційованого маркетингу	Знижки, певні програми лояльності	Стратегія диференціації
2	Надання розширеного спектру послуг у детектуванні голосу	Стратегія диференційованого маркетингу	Розширений спектр послуг у порівнянні з іншими	Стратегія диференціації

Наступним кроком є вибір стратегії конкурентної поведінки (таблиця 4.13).

Таблиця 4.13 Визначення базової стратегії конкурентної поведінки

<i>№ п/ п</i>	<i>Чи є проект «першопрохідцем» на ринку?</i>	<i>Чи буде компанія шукати нових споживачів, або забирати існуючих у конкурентів?</i>	<i>Чи буде компанія копіювати основні характеристики товару конкурента, і які?</i>	<i>Стратегія конкурентн ої поведінки</i>
	Ні, Однак це є новий підхід до даної сфери	Буде	Апаратні або програмні рішення	Стратегія експерта

На основі вимог споживачів з обраних сегментів до постачальника (стартап-компанії), а також в залежності від обраної базової стратегії розвитку (таблиця 4.12) та стратегії конкурентної поведінки (таблиця 4.13) розробляється стратегія позиціонування (таблиця 4.14), що полягає у формуванні ринкової позиції (комплексу асоціацій), за яким споживачі мають ідентифікувати торгівельну марку/проект.

*Таблиця 4.14 Визначення стратегії позиціонування*

<i>№ n/ n</i>	<i>Вимоги до товару цільової аудиторії</i>	<i>Базова стратегія розвитку</i>	<i>Ключові конкуренто- спроможні позиції власного стартап- проекту</i>	<i>Вибір асоціацій, які мають сформувати комплексну позицію власного проекту (три ключових)</i>
1	Простота реалізації	Стратегія диференціації	Система орієнтована на легку роботу для користувача	Простота використання
2	Надання розширеного спектру послуг	Стратегія диференціації	Розділення надання рівнів послуг у відповідності до вартості	Постійна підтримка рівня якості

Результатом виконання підрозділу є узгоджена система рішень щодо ринкової поведінки стартап-компанії, яка визначатиме напрями роботи стартап-компанії на ринку.

#### **4.5 Розроблення маркетингової програми стартап-проекту**

Першим кроком є формування маркетингової концепції товару, який отримає споживач. Для цього у таблиці 4.15 потрібно підсумувати результати попереднього аналізу конкурентоспроможності товару.



Таблиця 4.15 Визначення ключових переваг концепції потенційного товару

<i>№ n/n</i>	<i>Потреба</i>	<i>Вигода, яку пропонує товар</i>	<i>Ключові переваги перед конкурентами (існуючі або такі, що потрібно створити</i>
1	Низька ціна на послуги дослідження голосової активності	Робота з сучасними та простими методами	Зменшення вартості досліджень
2	Висока якість результатів дослідження	Робота з сучасними методами	Отримання якісних результатів
3	Підтримка рівня обслуговування	Урахування статистики погіршення якості	Прогнозування збоїв у роботі

Наступним кроком є визначення оптимальної системи збуту, в межах якого приймається рішення (таблиця 4.16):

- проводити збут власними силами або залучати сторонніх посередників (власна або залучена система збуту);
- вибір та обґрунтування оптимальної глибини каналу збуту;
- вибір та обґрунтування виду посередників.

Таблиця 4.16 Формування системи збуту

<i>№ n/n</i>	<i>Специфіка закупівельної поведінки цільових клієнтів</i>	<i>Функції збуту, які має виконувати постачальник товару</i>	<i>Глибина каналу збуту</i>	<i>Оптимальна система збуту</i>
1	Великі компанії, що працюють у полі досліджень голосової активності	Продаж послуг	80%	Власна
2	Малий бізнес та приватне використання	Продаж послуг	20%	Власна

Останньою складовою маркетингової програми є розроблення концепції маркетингових комунікацій, що спирається на попередньо обрану основу для позиціонування, визначену специфіку поведінки клієнтів (таблиця 4.17).

*Таблиця 4.17 Концепція маркетингових комунікацій*

<i>№ п/ п</i>	<i>Специфіка поведінки цільових клієнтів</i>	<i>Канали комунікацій, якими користуют ься цільові клієнти</i>	<i>Ключові позиції, обрані для позиціонуван ня</i>	<i>Завдання рекламного повідомлення</i>
1	Поведінка вимоги якісних результатів	Інтернет	Використовують ся найсучасніші алгоритми	Донести переваги використання до будь-якого цільового клієнта
2	Поведінка вимоги швидкості у проведенні досліджень	Інтернет	Використання найсучасніших методик та алгоритмів	

Результатом є ринкова (маркетингова) програма, що включає в себе концепції товару, збуту, просування, спирається на цінності та потреби потенційних клієнтів, конкурентні переваги ідеї, стан та динаміку ринкового середовища, в межах якого буде впроваджено проект.

#### **Висновки до розділу 4**

У даному розділі була сформована загальна ідея стартап проекту «Дослідження голосової активності у системах з низьким енергоспоживанням» визначені сильні, слабкі та нейтральні сторони. Дана сфера є і буде завжди актуальною, оскільки щодня створюються нові методи та алгоритми. Якість виконання досліджень постійно потрібно підтримувати на тому рівні, що може бути

прописаною у договорі та навіть покращувати задля збільшення конкурентної спроможності. Як на мене, можливість постійного розвитку алгоритмів та впровадження належного рівня підтримки є основним ключем до успіху та використання продукту у більшості компаній.

Розроблено:

- опис ідеї стартап-проекту;
- технологічну здійсненність ідеї проекту;
- аналіз ринкових можливостей запуску стартап-проекту (характеристика потенційних клієнтів стартап-проекту, фактори загроз, фактори можливостей, ступеневий аналіз конкуренції на ринку, обґрунтування факторів конкурентоспроможності, SWOT-аналіз стартап-проекту, альтернативи ринкового впровадження стартап-проекту);
- ринкову стратегію проекту (вибір цільових груп потенційних споживачів, визначення базової стратегії розвитку, визначення базової стратегії конкурентної поведінки, визначення стратегії позиціонування);
- маркетингову програму стартап-проекту (визначення ключових переваг концепції потенційного товару, формування системи збуту, концепція маркетингових комунікацій)

Результатом розробки як вже і зазначено, є методика контролю якості детектування, яка допомагає отримувати позитивний зворотній зв'язок від користувачів, так як будь які проблеми з продуктом – привід задуматися над тим, щоб стати потенційним клієнтом конкурента.

Областю застосування даного проекту є багато сфер, наприклад наукова діяльність, приватні дослідження або великі компанії, які використовуються отримані дані для подальших розробки.

## ЗАГАЛЬНІ ВИСНОВКИ

У даній роботі вирішено актуальну та важливу задачу з дослідження ефективності методів детектування голосової активності у системах з малим енергоспоживанням. Під час виконання роботи отримано наступні результати:

1. Проведено аналіз різних методів детектування, їх основні переваги та недоліки у розробці системи для детектування голосової активності. Детально описано основні алгоритми, що мають широке розповсюдження у сфері детектування мовлення. Пояснено основну їх основну суть та наведено приклади з їх використанням. Також наведено інформацію про історію детектування голосової активності.

2. Проведено детальний опис та аналіз методу детектування голосової активності з використанням рекурентних нейронних мереж. Розібрано складові даного методу, особливості роботи, функціонування, продемонстровано блок-схему алгоритму, детально розписано кожен її пункт. Враховуючи всі переваги та недоліки, в якості основного методу, який буде застосовуватись при проєктуванні системи для детектування голосової активності була обрана модель з використанням рекурентної нейронної мережі.

3. На основі розглянутого методу та алгоритму у попередньому розділі було проведено симуляцію роботи системи у програмному середовищі Matlab. Було детально описано принцип роботи програми та кожного її фрагменту, описано основні дані та параметри системи. Експериментальна реалізація даного методу детектування голосової активності на базі мікроконтролера ST32F Series є частиною майбутніх досліджень.

4. Розроблено початковий стартап-проект системи, у результаті якого можна стверджувати, що систему можна сміливо виводити на ринок, не зважаючи на величезну конкуренцію у даній сфері.

## ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

- [1] Speech Recognition, Wikipedia. *Wikipedia, the free encyclopedia*. URL: [https://en.wikipedia.org/wiki/Speech\\_recognition](https://en.wikipedia.org/wiki/Speech_recognition) (дата звернення 20.10.2023)
- [2] ) О. І. ШЕРЕМЕТ, О. В. САДОВОЙ, Метод опорних векторів (SVM), Дніпродзержинський державний технічний університет. URL: <https://www.dstu.dp.ua/Portal/Data/74/72/3st13-17.pdf> (дата звернення 20.10.2023)
- [3] O. Gauci, C.J. Debono, E.Gatt, P. Micallef (2011), The Application of Support Vector Machine for Speech Classification Department of Communications and Computer Engineering University of Malta. URL: <https://www.um.edu.mt/library/oar/bitstream/123456789/22573/1/Proceedings%20of%20CSAW'06%20-%20A11.pdf> (дата звернення 20.10.2023)
- [4] Petr Pollak, Jiri Tatarinov, Hidden Markov Models in voice activity detection, Czech Technical University. URL: [https://www.researchgate.net/publication/229046740\\_Hidden\\_Markov\\_Models\\_in\\_voice\\_activity\\_detection](https://www.researchgate.net/publication/229046740_Hidden_Markov_Models_in_voice_activity_detection) (дата звернення 20.10.2023)
- [5] Amit Sofer Shlomo E. Chazan (2017) , CNN self-attention voice activity detector. URL: <https://arxiv.org/pdf/2203.02944.pdf> (дата звернення 20.10.2023)
- [6] Minjae Lee, Kyuyeon Hwang, Jinhwan Park, Sungwook Choi, Sungho Shin and Wonyong Sung (2016), FPGA-based Low-power Speech Recognition with Recurrent Neural Networks, 2016 IEEE International Workshop on Signal Processing Systems, URL: <https://ieeexplore.ieee.org/document/7780102> (дата звернення 20.10.2023)
- [7] Аксенов О.Д., Логин В.М. (2019), Метод Мел-Частотных кепстральных коэффициентов в задаче распознавания речи. URL:

[https://libeldoc.bsuir.by/bitstream/123456789/36578/1/Aksenov\\_Metod.pdf](https://libeldoc.bsuir.by/bitstream/123456789/36578/1/Aksenov_Metod.pdf) ( дата звернення ( 25.11.2023)

[8] Understanding the Principles of Recursive Neural Networks: A Generative Approach to Tackle Model Complexity Alejandro China Departmento de Física Fundamental, Facultad de Ciencias UNED, Madrid Spain. URL: <https://arxiv.org/ftp/arxiv/papers/0911/0911.3298.pdf> (дата звернення 25.11.2023)

[9] Zhan Shen<sup>1</sup>, Jianguo Wei<sup>1</sup>, Wenhuan Lu, and Jianwu Dang (2016), Voice Activity Detection Based on Sequential Gaussian Mixture Model with Maximum Likelihood Criterion. 10th International Symposium on Chinese Spoken Language Processing(ISCSP) 2016. Tianjin, China. URL: <https://www.semanticscholar.org/paper/Voice-activity-detection-based-on-sequential-model-Shen-Wei/b51bbff7af3d26f28f7e876246599094cc89d360> (дата звернення 25.11.2023)

[10] Farzane Tajdini, Mohsen Piri (December 2022), Deep neural network learning for speech recognition, Conference: 9th National Congress of Electrical and Computer Engineering of IranAt: Iran-Ardabi. URL: [https://www.researchgate.net/publication/366086450\\_Deep\\_neural\\_network\\_learning\\_for\\_speech\\_recognition](https://www.researchgate.net/publication/366086450_Deep_neural_network_learning_for_speech_recognition) (дата звернення 20.10.2023)

[11] Ido Ariav and Israel Cohen (2019), An End-to-End Multimodal Voice Activity Detection Using WaveNet Encoder and Residual Networks. IEEE Journal of Selected Topics in Signal Processing 2019. URL: <https://israelcohen.com/wp-content/uploads/2019/03/08649655.pdf> (дата звернення 20.10.2023)

[12] Aditya Amberkar, Parikshit Awasarmol, Gaurav Deshmukh, Piyush Dave (2018), Speech Recognition using Recurrent Neural Networks, 2018 IEEE International Conference on Current Trends toward Converging Technologies, Coimbatore, India. URL: <https://ieeexplore.ieee.org/document/8551185> (дата звернення 20.10.2023)

[13] Ajitesh Kumar (2017), Gaussian Mixture Models: What are they & when to use URL: <https://vitalflux.com/gaussian-mixture-models-what-are-they-when-to-use/> (дата звернення 20.10.2023)

[14] J. David Vanich (2016), Zebra crosswalk detection assisted by neural networks URL: <https://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=2726&context=theses> (дата звернення 20.10.2023)

[15] R. Cheng, K. Wang, K. Yang, N. Long, J. Bai, and D. Liu. Real-time pedestrian crossing lights detection algorithm for the visually impaired. Multimedia Tools and Applications, 77. URL: [https://www.researchgate.net/publication/321835374\\_Real-time\\_pedestrian\\_crossing\\_lights\\_detection\\_algorithm\\_for\\_the\\_visually\\_impaired](https://www.researchgate.net/publication/321835374_Real-time_pedestrian_crossing_lights_detection_algorithm_for_the_visually_impaired) (дата звернення 20.10.2023)

[16] Manan Vyas (2014), A GAUSSIAN MIXTURE MODEL BASED SPEECH RECOGNITION SYSTEM USING MATLAB B.E Electronics, University of Mumbai. URL: <https://aircconline.com/sipij/V4N4/4413sipij09.pdf> (дата звернення 25.11.2023)

[17] Alex Graves, Supervised Sequence Labelling with Recurrent Neural Networks. URL : <https://www.cs.toronto.edu/~graves/preprint.pdf> (дата звернення 20.12.2023)

[18] Aliaksei KOLESAN, Dmitrij ŠEŠOK (2020), Voice Activation Systems for Embedded Devices: Systematic Literature Review Aliaksei KOLESAN, Dmitrij ŠEŠOK, Department of Information Technologies, Vilnius Gediminas Technical University URL: <https://content.iospress.com/download/informatica/infor398?id=informatica%2Finfo398> (дата звернення 15.12.2023)

[19] Charu C. Aggarwal (2018), Neural Networks and Deep Learning. URL: <https://link.springer.com/book/10.1007/978-3-319-94463-0> (дата звернення 29.10.2023)

[20] Taiba Majid Wani; Teddy Surya Gunawan; Syed Asif Ahmad Qadri; Mira Kartiwi; Eliathamby Ambikairaja (2021), Comprehensive Review of Speech Emotion Recognition Systems URL: <https://ieeexplore.ieee.org/abstract/document/9383000> (дата звенення 25.11.2023)

[21] РОЗРОБЛЕННЯ СТАРТАП-ПРОЕКТУ. Методичні рекомендації до виконання розділу магістерських дисертацій для студентів інженерних спеціальностей. Київ, НТУУ «КПІ ім. Ігоря Сікорського», 2016, 28 с.

[22] Демиденко О.А., Лебедев Д.Ю. (2023), ДОСЛІДЖЕННЯ ГОЛОСОВОЇ АКТИВНОСТІ В СИСТЕМАХ З НИЗЬКИМ ЕНЕРГОСПОЖИВАННЯМ, Вчені записки Таврійського національного університету імені В. І. Вернадського. Серія: Технічні науки, Vol 34, №6



## ДОДАТОК А. ЛІСТИНГ ПРОГРАМИ

### np\_energy.m

```
function [ret, energy, snr] = np_energy(signal, fs, win, thresh)
% фільтруємо вхідний сигнал від більшої частини шумів
    %fpass = [300 1500]; % смуга пропускання фільтру в Гц
    %[b, a] = butter(4, fpass/(fs/2)); % смуговий фільтр
    баттерворта 4-го порядку
    %signal = filter(b, a, signal); % фільтрація вхідного
    сигналу

% хедер циклу для згорткового циклу
    frame_size = round((win*10^-3)*fs); % розмір вікна в
    кількості елементів
    signal(1:fs) = 0;
    signal(end-fs:end) = 0;
    len = length(signal); % повний розмір сигналу
    (кількість елементів)
% розмір результуючого двійкового масиву де зберігатимуться
    двійкові значення (голос відсутній, голос присутній)
    voice_bin_len = ceil(len/frame_size);
    voice_bin = zeros(1, voice_bin_len); % сам двійковий
    масив
    snr = zeros(1, voice_bin_len); % масив зі значеннями
    відношення сигналу до шуму
    count = 1; % змінна для циклу
% початок циклу зі згорткою
    %net = patternnet([20 10]);
    %net.trainParam.epochs = 100; % кількість ітерацій у
    мережі
    % Тренування нейронної мережі
    %net = train(net, features, targets);
    for i = 1:round(frame_size):round(len)
        edge = (frame_size + i) - 1; % перевірка на те чи
        не дійшло вікно до кінця масиву
        if edge > len
            edge = len;
        end
        frame = signal(i:edge); % поточна вибірка (фрейм)
        сигналу
        ts = 1/fs; % період дискретизації
        Lfft = 1024; % кількість точок для перетворення
        фур'є
```

```

    freq_axis = (0:Lfft/2-1)/(Lfft*ts); % масив
значень вісі абцис (X) -> частотний діапазон для Фур'є
    frame_size = max(1, round((win*10^-3)*fs)); %
розмір поточної вибірки фрейму
    S = fft(double(frame), frame_size); % результат
перетворення Фур'є
    S = S(1:floor(length(S)/2));
    locations = find(freq_axis >= 195); % знаходимо
потрібний початковий індекс в частотному діапазоні
    begin_index = locations(1);
    % знаходимо потрібний кінцевий індекс в частотному
діапазоні
    locations = find(freq_axis >= 3843);
    end_index = locations(1);
    S(1:begin_index) = 0.0; % зануляємо все за межами
частотного діапазону
    S(end_index:end) = 0.0;
    S = abs(S); % вибираємо результуючі значення по
модулю
    S_squared = S.^2; % знаходимо девіацію
    Et = sum(S_squared);

    S_sorted = sort(S_squared); % відсортовуємо
значення сигналу
    % включення динамічних індексів
    idx_start = round(0.10 * length(S_sorted));
    idx_end = round(0.20 * length(S_sorted));
    Np = (1/100.0)*sum(S_sorted(idx_start:idx_end)) +
.00001;
    L = 1;
    % Dynamically determine the upper index based on
actual size
    upper_idx = length(S_sorted);

    % Ensure the loop does not attempt to access beyond
array bounds
    for j = upper_idx:-1:1
        if sum(S_sorted(j:upper_idx)) >= Et*.4;
            L = j;
            break;
        end
    end
    if L == upper_idx
        L = L - 1;
    end
    % Updated to use upper_idx instead of 512
    Sp = (1.0/(upper_idx-L))*sum(S_sorted(L:upper_idx))
+ 1e-10;

```

```

        % output = net(dataset);
        snr(count) = (Sp / (Np + 1e-10)); % знаходимо кожне
значення відношення сигналу до шуму
        voice_bin(count) = (Sp/Np) > thresh + 90; %
детектуємо голос
        count = count + 1; % ітеруємося до наступного
шматка (фрейму) сигналу
    end

    % копіюємо результат детекції голосу в новий масив щоб
повернути його при виході з функції
    ret = zeros(1, voice_bin_len*frame_size);
    for k = 1:frame_size
        ret(k:frame_size:end) = voice_bin;
    end

        % знаходимо енергію сигналу
    e_len = ceil(len/frame_size); % розмір масиву
    energy = zeros(1, e_len); % масив для значень енергії
    signal = signal.^2; % знаходимо квадрат сигналу
    count = 1; % змінна для ітерації
    % перевірка на те чи не дійшло вікно до кінця масиву
    for i = 1:frame_size:len
        edge = (i + frame_size) - 1;
        if edge > len
            edge = len;
        end
        % розраховуємо значення енергії
        pow2 = sum(signal(i:edge));
        pow = sqrt(pow2);
        energy(count) = pow;
        count = count + 1; % наступна ітерація
    end
end
end

```

## signalProcessing.m

```

% зчитування аудіосигналів
[audioSignal1, fs1] = audioread('cafe.wav');
[audioSignal2, fs2] = audioread('noise_1.wav');

% ресемпсування за необхідності
if fs1 ~= fs2
    audioSignal2 = resample(audioSignal2, fs1, fs2);
end
fs = fs1;
% вирівнювання довжин
len1 = length(audioSignal1);

```

```

len2 = length(audioSignal2);
minLen = min(len1, len2);

audioSignal1 = audioSignal1(1:minLen);
audioSignal2 = audioSignal2(1:minLen);

% змішування сигналів
mixedSignalF = audioSignal1 + audioSignal2';

%Нормалізація сигналу
mixedSignalF = mixedSignalF / max(abs(mixedSignalF));

% створення нового аудіофайлу змішаного сигналу
audiowrite('mixed_signalX.wav', mixedSignalF, fs1);

win = .31; % ширина вікна
thresh = 90; % SNR

% Виклик функції
[detectedSpeech, energy, snr] = np_energy_sad(mixedSignalF,
fs, win, thresh);

% візуалізація
figure(1)
subplot(2,1,1);
% plot(filteredSignal);
plot(mixedSignalF);
title('Змішаний сигнал з шумом');
xlabel('Семпл');
ylabel('Амплітуда сигналу');

subplot(2,1,2)
plot(detectedSpeech);
title('Детектоване мовлення');
xlabel('Фрейм');
ylabel('Присутність мовлення');

% зчитування
[audioSignal, fs] = audioread('noise_1.wav');
[detectedSpeech, energy, snr] = np_energy_sad(audioSignal,
fs, win, thresh);

```

## **ДОДАТОК Б. Копія публікації за темою дисертаційної роботи**

Стаття буде опублікована у журналі: « Вчені записки Таврійського національного університету імені В. І. Вернадського. Серія: Технічні науки » у Томі №34, №6, 2023

Радіотехніка та телекомунікації

УДК 520.6.05

*Демиденко О.А.*

Студент 2го курсу магістратури кафедри конструювання електронно-обчислювальної апаратури,

Національного технічного університету України

«Київський політехнічний інститут імені Ігоря Сікорського»

*Лебедев Д.Ю.*

кандидат технічних наук, доцент,

доцент кафедри конструювання електронно обчислювальної апаратури

Національного технічного університету України

«Київський політехнічний інститут імені Ігоря Сікорського»

## **ДОСЛІДЖЕННЯ ГОЛОСОВОЇ АКТИВНОСТІ В СИСТЕМАХ З НИЗЬКИМ ЕНЕРГОСПОЖИВАННЯМ**

*В даній роботі буде наведено опис сучасних рішень та методів для дослідження голосової активності. Голосова активність в системах з низьким енергоспоживанням є динамічним полем досліджень, де вчені та розробники постійно працюють над вдосконаленням алгоритмів, апаратно-програмних рішень та методів передачі даних. Вона визначається балансом між точністю розпізнавання та обробкою команд і збереженням енергії для*

максимально довгого функціонування пристроїв на обмежених джерелах живлення. Ці особливості роблять цю область досліджень важливою для розвитку енергоефективних та мобільних технологічних рішень. У тексті роботи проаналізовані вже існуючі методи та рішення, що активно використовуються у реальних системах такі, як приховані моделі Маркова, рекурентній нейронні мережі, згорткові нейронні мережі. Розглянуті переваги та недоліки методів виявлення голосової активності, детально описано кожен з них. В роботі розглядаються ключові аспекти виявлення та обробки голосової активності у зазначених системах, а також пропонуються підходи до підвищення ефективності роботи алгоритмів детектування голосової активності при обмежених ресурсах.

На основі проведеного аналізу, запропоновані оптимальні технічні рішення та розроблена блок-схема для даних рішень, яка дозволить оптимізацію процесів у майбутньому при дослідженні голосової активності, створено детальний опис системи

Обґрунтовано розширення функціональних можливостей системи розпізнавання голосової активності та сформовані напрями для їх подальшого вдосконалення. Розробка систем вимагає поєднання передових технологій у сферах обробки сигналів, машинного навчання, оптимізації алгоритмів та використання енергоефективних компонентів оскільки вони є ключовим фактором роботи у багатьох секторах, особливо у сферах де точна та швидка передача даних є необхідністю, тому дослідження відкривають шлях до нових можливостей та глобальних змін у певних галузях, де обмеження є визначим фактором.

**Ключові слова:** голосова активність, системи з низьким енергоспоживанням, дослідження голосової активності, нейронна мережа, штучний інтелект

**Постановка проблеми.** У сучасному світі голосова активність стає все більш важливою складовою інтерфейсів для взаємодії з технічними

пристроями та системами. Проте, існуючі системи часто стикаються із проблемами, пов'язаними із високим енергоспоживанням, що впливає на їхню ефективність та придатність до використання в умовах обмеженого живлення. Для вирішення цих проблем необхідно ретельно вивчати та оптимізувати голосові системи з фокусом на низькому енергоспоживанні.

Однією з ключових проблем є точність розпізнавання мови при обмежених обчислювальних потужностях та обмежених ресурсах енергії. Традиційні алгоритми розпізнавання мови можуть бути вимогливими до обчислень, що призводить до підвищеного споживання енергії. Отже, існує потреба в розробці та оптимізації алгоритмів, спроможних працювати на пристроях з обмеженим живленням, забезпечуючи при цьому надійну та швидку реакцію на голосові команди.

Ще однією проблемою є розробка низькоенергетичних апаратних засобів для обробки голосового сигналу. Стандартні методи є об'ємними та неоптимізованими під потреби пристроїв з обмеженим живленням. Необхідно дослідження нових технологій та методів, що дозволять знизити витрати енергії при обробці голосового сигналу, забезпечуючи при цьому високу якість розпізнавання.

Крім того, ефективність голосових систем в умовах обмеженого живлення є критичною для їхнього успішного впровадження. Розуміння та управління енергоспоживанням, а також розробка стратегій для підтримання працездатності систем при обмежених ресурсах, є важливими аспектами дослідження.

**Аналіз останніх досліджень і публікацій.** У світі існують різні технології дослідження голосової активності.

Спершу варто зазначити, системи що використовують сучасні алгоритми штучного інтелекту, наприклад як зазначено у [1]. Використання штучного інтелекту який може бути покладений у дослідженнях, є використання глибокого навчання для оптимізації голосових інтерфейсів та

алгоритмів розпізнавання мови з точки зору енергоефективності. Також можуть бути проведені дослідження щодо розробки нових технологій апаратного забезпечення, таких як енергоефективні мікросхеми та датчики, придатних для використання в голосових системах на пристроях з обмеженим живленням. Також подібна концепція була розглянута у роботі [2], де також для дослідження голосової активності була використана навчена нейронна мережа.

Розгляд даної роботи є дуже актуальним, оскільки більшість сучасних систем працюють з використанням штучного інтелекту. Хоча для дослідження активності можуть бути використані інші методи, такі як: енергетичний, де перевіряється рівень «енергії» аудіосигналу для визначення наявності мовлення. Тут загалом використаний підхід, заснований на пороговому значенні, коли «енергія» сигналу порівнюється з попередньо визначеним пороговим значенням. Якщо «енергія» перевищує поріг, сегмент класифікується як мова. Або, наприклад, статистичні методи суть яких полягає у тому, щоб відрізнити мовні сегменти від немовних. У цих методах зазвичай використовуються такі методи, як швидкість перетину нуля, короткочасні спектральні характеристики та оцінка висоти тону.

Аналіз більшості останніх робіт вказав на те, що найбільш оптимальний та простий метод для детектування голосової активності є метод машинного навчання, оскільки алгоритми використовують контрольовані або неконтрольовані підходи до навчання моделей, які можуть класифікувати мовні та немовні сегменти. Ці моделі навчаються на анотованих наборах даних і можуть фіксувати складні моделі та зв'язки в аудіосигналі.

**Постановка технічного завдання.** На основі проведеного аналізу визначено параметри, що будуть покращуватися:

1. Розробити систему, яка буде детектувати голосову активність та буде використовувати оптимальні методи детектування з огляду на енергоспоживання.



2. Проаналізувати систему на правильність роботи, створити зручну та сучасну нейромережу, яка буде мати простий інтерфейс та зручні параметри досліджень.

**Виклад основного матеріалу дослідження.** Проаналізувавши частину наявних рішень пропонується детальний огляд та порівняння наявних та найбільш актуальних методів. Спершу варто розглянути метод з використанням нейронної мережі. Суть алгоритмів у тому, що вони включають полегшені моделі. Ці алгоритми вивчають анотовані навчальні дані для класифікації мовних і немовленнєвих сегментів. Можуть бути використані такі моделі, як Gaussian Mixture Models (GMM), Hidden Markov Models (HMM) або легкі архітектури нейронних мереж, такі як згорткові нейронні мережі (CNN) або рекурентні нейронні мережі (RNN).

Розглянемо приховані моделі Маркова, оскільки це є однією з найпоширеніших моделей.

Існує два підходи до використання виявлення мовлення та паузи моделі. Перший базується на роздільному навчанні моделі і наступне тестування – підхід з виявленням адаптації. Після навчання не зміняться її параметри та модель буде нечутливою до змін середовища – підхід виявлення без адаптації.

В обох процедурах ми повинні вирішити, який тип Маркової моделі використовувати. Було обрано неперервні моделі із 3-станами. Важлива також обробка сигналу - для кожного кадру ми розраховували кепстральні коефіцієнти. Використовуються дві моделі: мовчання і модель мовлення. Це

найпростіший варіант виявлення голосової активності з використанням моделей Маркова, які хоч і не дуже використовуються у сьогоденні, проте є актуальними.

Алгоритм без адаптації. Складається з наступних кроків:

1. Ініціалізація

2. Навчання. Моделі навчаються за допомогою вручну позначених даних.
3. Критеріальна функція. Після навчання на навчальному комплексі переходимо до тестування.
4. Розрахунок.
5. Виявлення сегментів з тишею та мовленням .

Також наведена блок-схема самого алгоритму на рисунку 1.

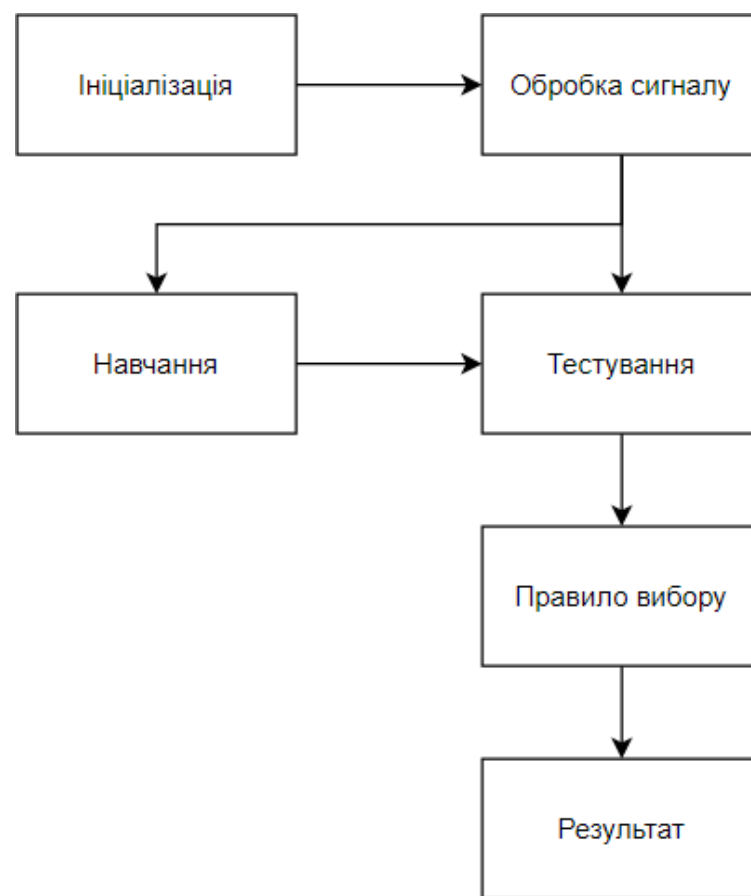


Рисунок 1 – Алгоритм роботи детектора голосової активності без адаптації

Алгоритм роботи детектора з адаптацією:

1. Ініціалізація. Задається випадковим чином початкові моделі мовчазної і мовної моделей.
2. Навчання.

3. Оновлення мовчазної моделі. Приспускається, що на тестовому наборі сигналів є мовчання на початку кожного сигналу. Оновлюється модель використовуючи ці початкові кадри.
4. Критерії адаптації.
5. Розрахунок порогу значень для вибірки.
6. Оновлення моделі.
7. Виявлення сегментів з тишею та мовленням.
8. Подальша обробка. Оскільки вивід виявленого мовлення зазвичай містить неправильні рішення через коливання фону шумових характеристик, можемо згладити результат за допомогою медіанної фільтрації.

Також наведена блок-схема самого алгоритму на рисунку 2.

Також варто зазначити про методи з використанням згорткових та рекурентних мереж.

Останнім часом активно досліджуються алгоритми розпізнавання мови на основі повністю нейронних рекурентних мереж. Рекурентна нейронна мережа наскрізно навчається за допомогою конекціоністської часової класифікації для безпосереднього перетворення вхідного мовлення в символи. Вона також використовується для моделювання мови яка демонструє набагато кращі можливості, ніж триграмові статистичні алгоритми. Нейронні мережі вимагають багато операцій множення та додавання, проте вони є найбільш розповсюдженими та оптимальним з точки зору сьогоденних задач та їх умов реалізації.

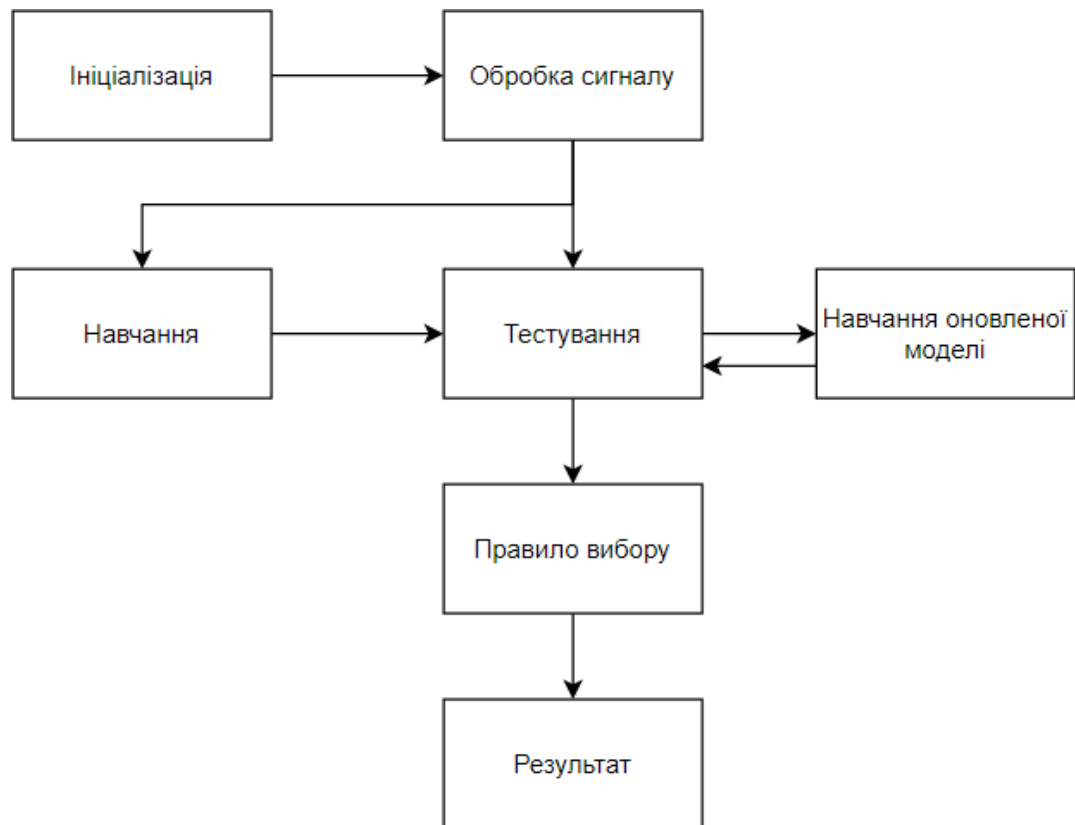


Рисунок 2 - Схема детектора голосової активності з адаптацією.

Ваги для повністю зв'язаних шарів або рекурентних нейронних мереж використовуються лише один раз при отриманні, тому їх звернення до них мають дуже низьку часову локальність. Існують спроби зменшити розмір параметрів за рахунок спроби зменшити розмір параметрів шляхом квантування.

Алгоритм розпізнавання мови, реалізований у цій роботі складається з рекурсивної нейронної мережі для акустичного моделювання, рекурсивної нейронної мережі для моделювання мови на рівні символів та статистичного моделювання на рівні слів. Використовується алгоритм і генерує ймовірності символів, аналізуючи кожен кадр вхідного мовлення. Мережа виводить ймовірності наступних символів, тоді як статистична мережа показує ймовірності наступних слів. Інформація, отримана з цих трьох модулів,

інтегрується для пошуку найкращої гіпотези за допомогою алгоритму пошуку.

Акустична модель має глибоку мережеву структуру і наскрізно навчається за допомогою онлайн-алгоритму. Хоча деякі нещодавні наскрізні алгоритми розпізнавання мови на основі рекурсивної нейронної мережі розпізнавання мови використовують двонаправлену структуру для покращення продуктивності розпізнавання.

На рисунку 3 наведена схема алгоритму описана вище.

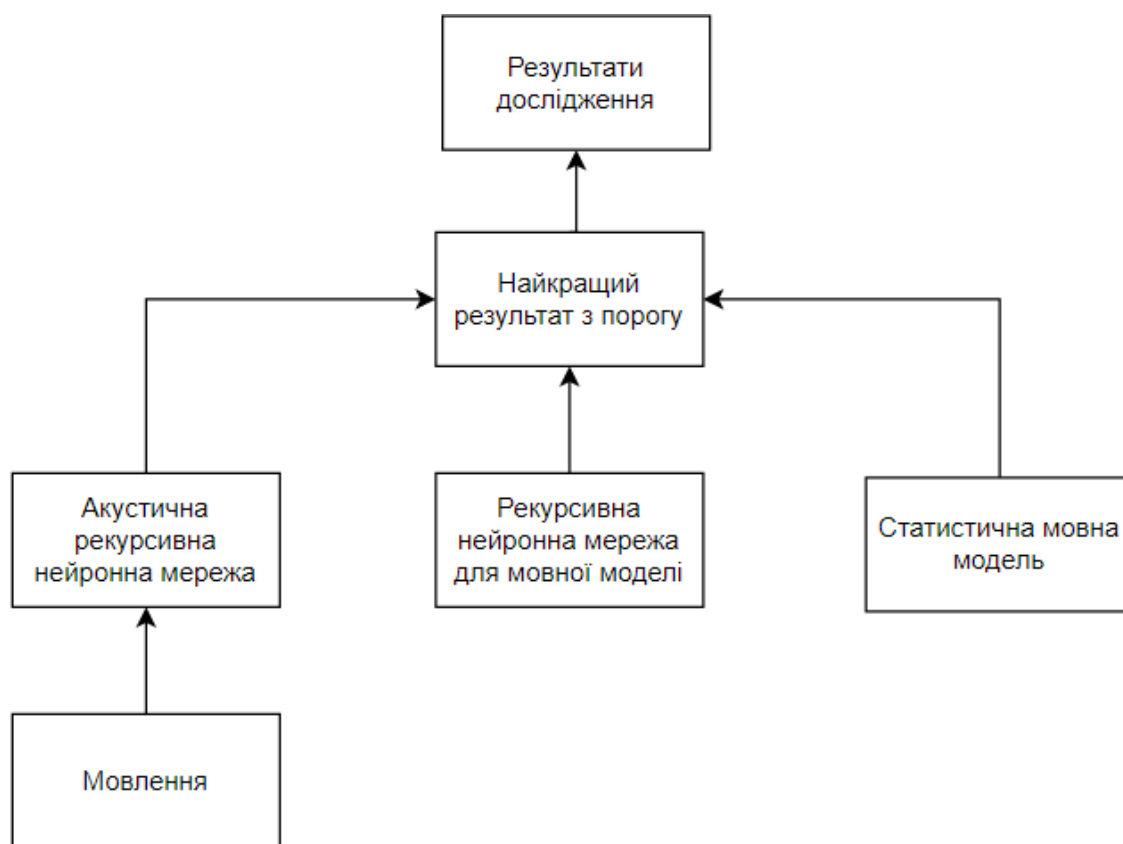


Рисунок 3 – Алгоритм роботи на основі нейронних мереж

Для прикладку порівняння розглянемо алгоритм виявлення голосової активності на основі енергії спектру

1. Скомпонування аудіо. Розділення вхідного сигналу на кадри.

Загальна тривалість кадру знаходиться в діапазоні від 10 до 30 мілісекунд.

2. Обчислення енергії. Обчислення енергії кожного кадру. Енергію кадру можна обчислити шляхом підсумовування квадратів значень звукових зразків у цьому кадрі.

3. Встановлення значення порогу. Визначення порогового значення, щоб відрізнити мову від мовчання. Цей поріг можна встановити на основі характеристик звукового сигналу та рівня шуму.

4. Виявлення мови. Порівняння енергії кожного кадру з порогом. Якщо енергія кадру перевищує порогове значення, то мовлення, якщо ні - мовчання.

У підсумку можна сказати, що метод енергії є простим, зручним та легким для використання у базових додатках, проте метод з використанням нейронних мереж є більш складним, але дозволяє використати більшу адаптивність, продуктивність, особливо у складних умовах експерименту, проте для вірного результату дослідження потрібні навчальні дані та обчислювальні ресурси. Вибір між цими методами часто залежить від конкретних вимог програми та наявних ресурсів. У багатьох випадках гібридний підхід, що поєднує сильні сторони обох методів, може бути корисним.

**Висновки.** Проведено аналіз наявних рішень для дослідження голосової активності в ході якого встановлено, що не існує єдиного рішення, що наразі вирішує всі проблеми дослідження голосової активності.

Наведені приклади систем демонструють вирішення для конкретних випадків. На основі порівняльного аналізу визначені основні параметри, якими системи різняться. Напрямоком подальших досліджень є реалізація системи голосової активності у системах з низьким енергоспоживанням та її оптимізація, враховуючи вже наведені дані та певні умови для конкретних вимог системи.

### Список літератури:

1. Всесвітня організація охорони здоров'я (2021). Deafness and hearing loss. <https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss>
2. K. Hwang and W. Sung, "Single stream parallelization of generalized LSTM-like RNNs on a GPU," in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2015
3. K. Hwang and W. Sung, "Sequence to sequence training of ctc-rnns with partial windowing," in International Conference on Machine Learning (ICML), 2016
4. F. Vesperini, P. Vecchiotti, E. Principi, S. Squartini, and F. Piazza, "Deep neural networks for multi-room voice activity detection: Advancements and comparative evaluation," in 2016 International Joint Conference on Neural Networks (IJCNN). IEEE, 2016
5. G. Gelly and J.-L. Gauvain, "Optimization of RNN-based speech activity detection," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 26, no. 3, 2018
6. B. Liu, Z. Wang, S. Guo, H. Yu, Y. Gong, J. Yang, and L. Shi, "An energy-efficient voice activity detector using deep neural networks and approximate computing," Microelectronics Journal, 2019
7. Minjae Lee, Kyuyeon Hwang, Jinhwan Park, "FPGA-based Low-power Speech Recognition with Recurrent Neural Networks" , 2016 IEEE International Workshop on Signal Processing Systems
8. T. N. Sainath, O. Vinyals, A. Senior, and H. Sak, "Convolutional, long short-term memory, fully connected deep neural networks," in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), Apr. 2015

## **Demydenko O., Lebedev D. VOICE ACTIVITY DETECTION IN LOW-POWER SYSTEMS**

*This article will provide a description of modern solutions and methods for voice activity research. Voice activity in low-power systems is a dynamic field of research, where scientists and developers are constantly working on improving algorithms, hardware and software solutions, and data transmission methods. It is determined by the balance between the accuracy of command recognition and processing and energy conservation for the longest operation of devices on limited power sources. These features make this area of research important for the development of energy-efficient and mobile technological solutions. The text of the work analyzes already existing methods and solutions that are actively used in real systems, such as hidden Markov models, recurrent neural networks, convolutional neural networks. Advantages and disadvantages of voice activity detection methods are considered, each of them is described in detail. The paper examines the key aspects of voice activity detection and processing in these systems, as well as offers approaches to increase the effectiveness of voice activity detection algorithms with limited resources.*

*On the basis of the conducted analysis, optimal technical solutions were proposed and a block diagram was developed for these solutions, which will allow optimization of processes in the future in the study of voice activity, a detailed description of the system was created*

*The expansion of the functional capabilities of the voice activity recognition system is substantiated and directions for their further improvement are formed. The development of systems requires a combination of advanced technologies in the fields of signal processing, machine learning, optimization of algorithms and the use of energy-efficient components, as they are a key factor in the operation of many sectors, especially in areas where accurate and fast data transmission is a necessity, so research opens the way to new opportunities and global changes in certain industries where restrictions are a determining factor.*



**Keywords:** *voice activity, low-power systems, voice activity research, neural network, artificial intelligence*