

УДК 520.6.05

DOI <https://doi.org/10.32782/2663-5941/2023.6/04>**Демиденко О.А.**Національний технічний університет України  
«Київський політехнічний інститут імені Ігоря Сікорського»**Лебедев Д.Ю.**Національний технічний університет України  
«Київський політехнічний інститут імені Ігоря Сікорського»

## ДОСЛІДЖЕННЯ ГОЛОСОВОЇ АКТИВНОСТІ В СИСТЕМАХ З НИЗЬКИМ ЕНЕРГОСПОЖИВАННЯМ

В даній роботі наведено опис сучасних рішень та методів для дослідження голосової активності, які використовуються в автономних системах з низьким енергоспоживанням. 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 даній роботі 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 енергетичних 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 багатьох галузях життєдіяльності людей, 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. Тра-

диційні 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, короточасні 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 немовленневих 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.

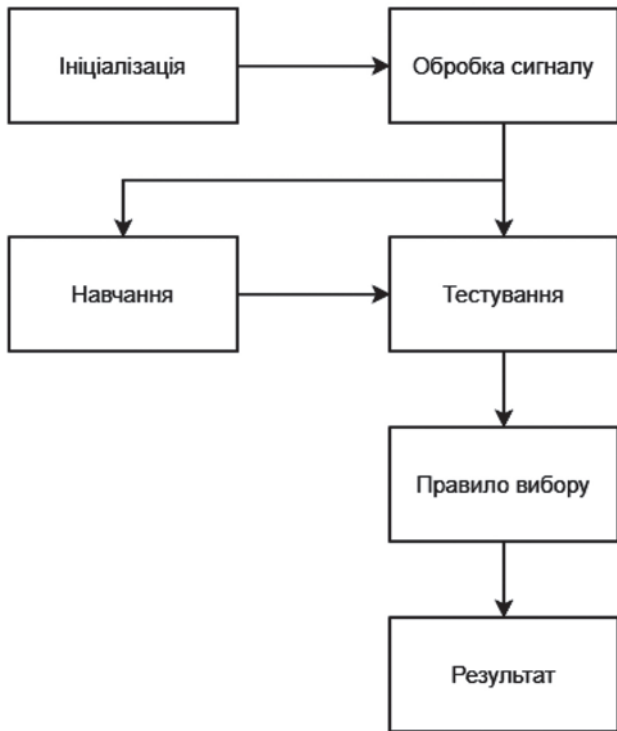


Рис. 1. Алгоритм роботи детектора голосової активності без адаптації

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 допо-

могою коннекціоністської 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.

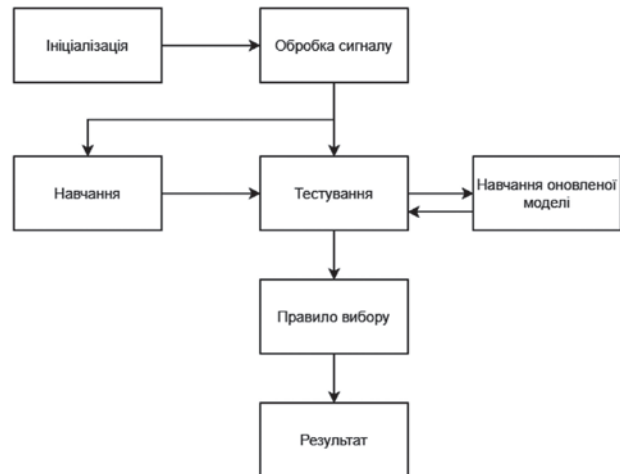


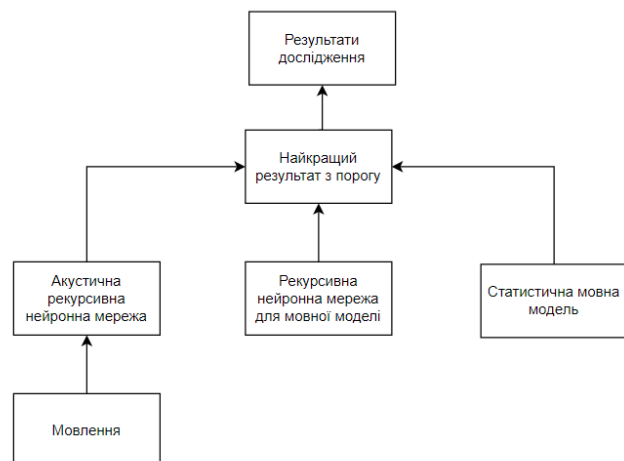
Рис. 2. Схема детектора голосової активності з адаптацією

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.



**Рис. 3. Алгоритм роботи на основі нейронних мереж**

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

1. Скомпоювання 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.A., Lebedev D.Yu. 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.*

**Key words:** *voice activity, low-power systems, voice activity research, neural network, artificial intelligence.*