

DOI: <https://doi.org/10.36910/6775-2524-0560-2025-61-10>

УДК: 004.02

Дідус Андрій Володимирович, аспірант

<https://orcid.org/0009-0004-2235-6742>

Терейковський Ігор Анатолійович, д.т.н., професор

<https://orcid.org/0000-0003-4621-9668>

Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна

МЕТОД ПОБУДОВИ ЗАСОБІВ РОЗПІЗНАВАННЯ КЛЮЧОВИХ СЛІВ У МАЛОРЕСУРСНИХ КОМП'ЮТЕРНИХ СИСТЕМАХ

Дідус А. В., Терейковський І. А. Метод побудови засобів розпізнавання ключових слів у малоресурсних комп'ютерних системах. Розпізнавання ключових слів у малоресурсних автономних системах, таких як наземні дрони, стикається з фундаментальним компромісом між точністю та обчислювальною ефективністю. У статті представлено метод побудови засобів розпізнавання, що вирішує цю проблему шляхом оптимізації класичних підходів, а не застосування ресурсоемних нейронних мереж. В основі методу лежить принцип пріоритизації інформативності ознак, реалізований через механізм зваженого акустичного фінгерпринтингу: мел-кепстральні коефіцієнти (MFCC) зважуються, агрегуються та перетворюються на компактні рядкові «відбитки», які потім порівнюються за допомогою відстані Левенштейна. Експериментальна валідація методу на прикладі системи розпізнавання 100 команд для дрона продемонструвала високу ефективність: досягнуто F1-міру 0.92 в ідеальних умовах та 0.78 при співвідношенні сигнал/шум 5 дБ. Порівняльний аналіз показав, що розроблений підхід значно перевершує базові класичні аналоги та є ефективною автономною альтернативою хмарним сервісам. Таким чином, запропонований метод дозволяє створювати високоточні та обчислювально легкі системи розпізнавання ключових слів, повністю адаптовані для роботи на периферійних пристроях без доступу до мережі.

Ключові слова: розпізнавання ключових слів, системи з обмеженими ресурсами, метод побудови, акустичний фінгерпринтинг, зважування ознак, відстань Левенштейна.

Didus A., Tereikovskiy I. A Method for Constructing Keyword Recognition Tools in Low-Resource Computer Systems. Keyword Spotting (KWS) in low-resource autonomous systems, such as ground drones, faces a fundamental trade-off between accuracy and computational efficiency. This paper presents a method for constructing recognition systems that resolves this issue by optimizing classical approaches rather than employing resource-intensive neural networks. The method is based on the principle of prioritizing feature informativeness, implemented through a weighted acoustic fingerprinting mechanism: Mel-Frequency Cepstral Coefficients (MFCCs) are weighted, aggregated, and transformed into compact string "fingerprints," which are then compared using the Levenshtein distance. Experimental validation of the method, based on a system recognizing 100 drone commands, demonstrated high efficacy: an F1-score of 0.92 was achieved under ideal conditions and 0.78 at a 5dB Signal-to-Noise Ratio. The comparative analysis showed that the developed approach significantly outperforms baseline classical analogues and serves as an effective autonomous alternative to cloud services. Thus, the proposed method enables the creation of highly accurate and computationally lightweight keyword spotting systems, fully adapted for operation on edge devices without network access.

Keywords: Keyword Spotting, Low-Resource Systems, Method of Construction, Acoustic Fingerprinting, Feature Weighting, Levenshtein Distance, MFCC.

Постановка проблеми. Ефективне розпізнавання ключових слів у голосовому сигналі є фундаментальною технологією для сучасних людино-машинних інтерфейсів, особливо в контексті автономних та вбудованих систем, таких як наземні дрони. Попередні дослідження підтвердили, що ключовим викликом у цій сфері залишається досягнення оптимального компромісу між точністю розпізнавання та обчислювальною ефективністю. У той час як глибокі нейронні мережі демонструють високу точність, їх ресурсоемність робить їх практично непридатними для пристроїв, що функціонують в умовах жорстких апаратних обмежень та без стабільного доступу до хмарних сервісів. [1-6]

Аналіз ефективності та формалізація процесу розпізнавання показали, що класичні алгоритми, такі як приховані марковські моделі (НММ) та динамічна часова деформація (DTW), є перспективною альтернативою для малоресурсних систем. Проте ці традиційні підходи часто базуються на припущенні про рівнозначність усіх акустичних ознак, що обмежує їхню роздільну здатність та стійкість до шумів. Існує наукова проблема, яка полягає у відсутності методу, що дозволив би підвищити точність класичних систем розпізнавання ключових слів не за рахунок ускладнення моделі, а шляхом більш глибокого аналізу та диференційованого врахування інформативності самих акустичних ознак. [1-5]

Таким чином, постає актуальна науково-прикладна задача розробки та валідації нової моделі розпізнавання, яка б, спираючись на класичні принципи, інтегрувала механізми вибіркового аналізу

ознак для створення систем, що є одночасно точними, робастними та обчислювально ефективними для розгортання на периферійних пристроях.

Аналіз останніх досліджень і публікацій. Сучасні дослідження в галузі розпізнавання ключових слів демонструють домінування підходів, що базуються на глибокому навчанні. Провідну роль займають архітектури згорткових (CNN) та рекурентних (RNN) нейронних мереж, а також більш сучасні моделі на основі механізмів уваги, зокрема трансформери. Ці моделі демонструють найвищу точність розпізнавання, особливо при роботі з великими наборами даних, що робить їх стандартом для хмарних сервісів та потужних обчислювальних систем. [1-7]

Водночас, висока обчислювальна складність та значні вимоги до пам'яті глибоких нейронних мереж суттєво обмежують їх застосування у спеціалізованих комп'ютерних системах з обмеженими ресурсами, таких як бортові комп'ютери автономних пристроїв. Хоча існують методи оптимізації, як-от дистиляція знань чи використання спеціалізованих архітектур (напр., DS-CNN), вони все ще потребують значних зусиль для налаштування та не завжди забезпечують необхідний рівень продуктивності в режимі реального часу. [4-7]

Як альтернатива розглядаються класичні алгоритми, що довели свою ефективність та надійність. Фундаментальні підходи, що базуються на прихованих марковських моделях (HMM) та методах динамічного програмування, як-от динамічна часова деформація (DTW), залишаються актуальними завдяки своїй обчислювальній простоті та низьким вимогам до ресурсів. Попередні оцінки ефективності підтверджують, що саме ці методи є перспективними для фонових процесів розпізнавання на малопотужних пристроях. [5-9]

Проте, традиційні реалізації класичних методів мають суттєве обмеження: вони часто розглядають усі виділені акустичні ознаки, наприклад, усі мел-кепстральні коефіцієнти (MFCC), як рівнозначні. Такий підхід ігнорує той факт, що різні компоненти ознак несуть різну кількість інформації про фонетичний зміст слова, що знижує роздільну здатність методу, особливо в умовах зашумлених середовищ. Формалізація процесу розпізнавання ключових слів вказує, що етап виділення та моделювання ознак є критичною точкою для потенційної оптимізації.

Таким чином, у науковій літературі існує невирішена проблема: відсутність формалізованого методу, який би поєднував обчислювальну легкість класичних алгоритмів із більш глибоким аналізом інформативності акустичних ознак, характерним для складних моделей. Ця робота спрямована на заповнення цієї прогалини шляхом розробки та валідації нового підходу, що базується на диференційованому зважуванні ознак, їх перетворенні у рядкові "відбитки" та подальшому метричному порівнянні. [5]

Мета статті. Це розроблення методу побудови розпізнавання ключових слів у малоресурсних комп'ютерних системах за рахунок адаптації класичних методів аналізу акустичних ознак, що дозволяє забезпечити високу точність розпізнавання при менших обчислювальних потужностях.

Виклад основного матеріалу дослідження. Метод побудови засобів розпізнавання ключових слів, що розглядається, є узагальненням та систематизацією модульної архітектури моделі побудови засобів розпізнавання ключових слів, представленої в попередніх дослідженнях. Процес побудови згідно з даним методом декомпозується на послідовні етапи (рис. 1). [9-14, 17]

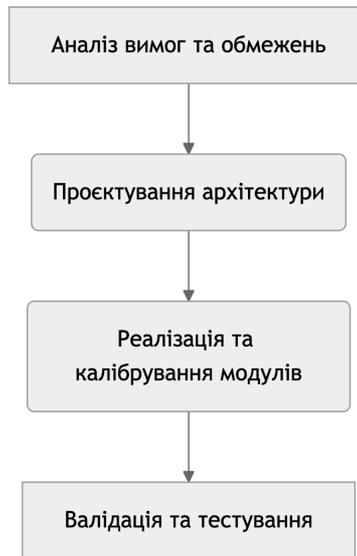


Рис. 1. Загальні етапи методу побудови засобів розпізнавання ключових слів

Метод базується на трьох фундаментальних принципах проектування.

Модульність – архітектура повинна бути декомпонована на функціонально незалежні модулі, зокрема модуль параметризації сигналу та модуль класифікації. Це забезпечує гнучкість у налаштуванні та можливість незалежної оптимізації кожного компонента. Загальна структура, що впливає з цього принципу, показана на рис. 2.

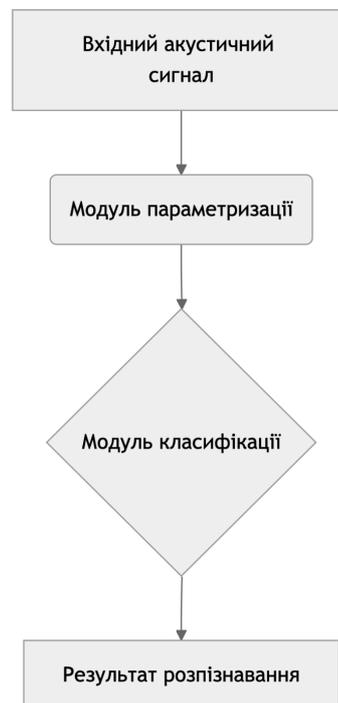


Рис. 2. Структура модулів засобів розпізнавання ключових слів

На рисунку 2 представлена узагальнена структура засобів розпізнавання, що впливає з принципу модульності. Процес розпізнавання декомпоновано на два основні функціональні модулі: модуль параметризації та модуль класифікації. Призначенням першого модуля є перетворення вхідного акустичного сигналу в набір інформативних числових ознак. Другий модуль, у свою чергу, відповідає за аналіз цих ознак та прийняття фінального рішення щодо наявності ключового слова. Така декомпозиція дозволяє забезпечити гнучкість у налаштуванні та можливість незалежної оптимізації кожного компонента системи.

Пріоритезація інформативності ознак - цей принцип передбачає відхід від припущення про рівнозначність акустичних параметрів. Вводиться етап зважування, на якому відбувається підсилення найбільш фонетично значущих компонентів у векторі ознак. Це реалізується через механізм зваженого акустичного фінгерпринтингу, де послідовність векторів ознак M перетворюється на компактний рядковий «відбиток» F з урахуванням вектора вагових коефіцієнтів W :

$$F = \text{Serialize} \left(Q \left(\frac{1}{T} \sum_{t=1}^T M_t \odot W \right) \right), \quad (1)$$

де:

- F – фінальний рядковий «відбиток» (fingerprint), що є компактним представленням голосової команди.
- M – матриця акустичних ознак розмірністю $T \times N$.
- T – кількість часових кадрів (фреймів) у аналізованому сегменті мовлення.
- N – розмірність вектора ознак (наприклад, кількість мел-кепстральних коефіцієнтів).
- W – вектор вагових коефіцієнтів розмірністю N , який визначає пріоритет кожної акустичної ознаки.
- \odot – оператор поелементного множення (добуток Адамара).
- Q – функція квантування, яка перетворює усереднені неперервні значення ознак у дискретні символи або рівні.
- Serialize – функція, що об'єднує послідовність дискретних символів у єдиний рядок для подальшого метричного порівняння.

Такий підхід відображений у формулі(1) дозволяє підвищити роздільну здатність методу без збільшення обчислювальної складності.

Детермінована метрична класифікація - цей принцип полягає у застосуванні обчислювально простих метрик відстані для прийняття рішення, що є альтернативою ресурсоемним ймовірнісним чи нейромережевим класифікаторам. Вибір слова W зі словника V здійснюється на основі мінімізації відстані (напр., відстані Левенштейна) між вхідним "відбитком" та еталонним:

$$W_{rec} = \text{argmin} \text{Lev}(F_{input}, F_k), \quad (2)$$

де:

- W – розпізнане ключове слово, яке є результатом роботи моделі.
- argmin – оператор, що повертає аргумент (у даному випадку — ключове слово k), при якому функція досягає свого мінімального значення.
- Lev – функція, що обчислює відстань Левенштейна між двома рядковими «відбитками».
- F – «відбиток» для ключового слова.

Формула (2), яка описує метод класифікації забезпечує високу швидкість та передбачуваність роботи системи.

Вибір конкретної технологічної парадигми для розпізнавання ключових слів є ключовим інженерним рішенням, що безпосередньо залежить від операційних вимог та апаратних обмежень цільової системи. Хоча дана робота зосереджена на розробці та валідації нового класичного методу, важливо чітко визначити його місце у ширшому ландшафті доступних технологій. Запропонований нижче алгоритм зображений на рис. 3 формалізує процес прийняття рішень, що дозволяє обґрунтовано вибрати оптимальний підхід, базуючись на пріоритетах конкретної задачі: максимальна ефективність та автономність або максимальна точність. [11]

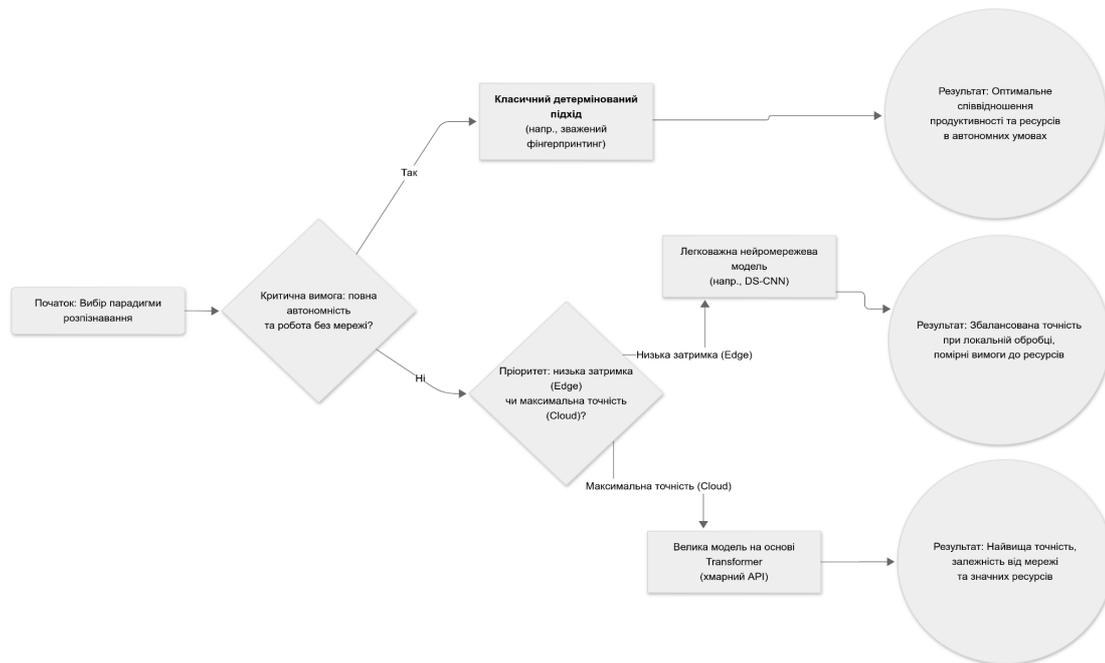


Рис 3. Алгоритм вибору технології розпізнавання ключових слів

Первинним критерієм вибору є вимога до повної операційної автономності. Для систем, де стабільне мережеве з'єднання є недоступним, ненадійним або недопустимим з міркувань безпеки (що є типовим сценарієм для військових застосувань або віддалених автономних платформ), процес прийняття рішень однозначно веде до вибору класичного детермінованого підходу. Запропонований у даній роботі метод, що базується на зваженому акустичному фінгерпринтингу, є прикладом реалізації такої парадигми, оскільки він оптимізований для досягнення максимального співвідношення продуктивності до обчислювальних ресурсів саме в офлайн-умовах.

У випадку, якщо вимога до повної автономності не є критичною, критерій вибору зміщується на компроміс між затримкою та точністю класифікації. Якщо пріоритетом залишається швидка реакція та локальна обробка на пристрої, перевага надається легковажним нейромережевим моделям (наприклад, архітектурам на основі глибинно-роздільних згорток, DS-CNN). Такі моделі забезпечують збалансовану точність, що перевищує класичні аналоги, при збереженні помірних вимог до ресурсів. Натомість, якщо головною метою є досягнення максимально можливої точності, а система може толерувати мережеві затримки, оптимальним вибором стає велика модель на основі архітектури Transformer, що, як правило, використовується через хмарний інтерфейс (API). Ці моделі досягають найвищої точності завдяки величезній кількості параметрів, але ціною залежності від значних обчислювальних ресурсів та стабільного з'єднання. [8-15]

Практична перевірка та валідація запропонованого методу була проведена в рамках експериментальних досліджень для порівняння різних методів розпізнавання ключових слів. Отримані результати, які зображені в таблиці 1 показують порівняння запропонованого методу з трьома іншими ключовими парадигмами: базовим класичним підходом, методом на основі динамічної часової деформації (DTW) та хмарним сервісом розпізнавання мовлення (ASR).

Таблиця 1. Порівняння продуктивності різних моделей розпізнавання ключових слів

Модель	Об'єм пам'яті	Час інференсу	Фактор реального часу (RTF)	F1-міра (чисте аудіо)	F1-міра (шум 5 дБ)
Базова модель	~150 КБ	~3 мс	3	0,75	0,45
Запропонована модель	~250 КБ	~5 мс	5	0,92	0,78
Класичний DTW	~2 МБ	~20 мс	0,02	0,88	0,65
Хмарний ASR	Н/Д (сервер)	~450 мс	0,45	0,97	0,91

Результати, наведені в таблиці, чітко ілюструють ключові компроміси між різними парадигмами розпізнавання ключових слів. Як і очікувалося, хмарний ASR-сервіс демонструє найвищу точність класифікації (F1-Score 0.97), що встановлює верхню межу продуктивності для даної задачі. Однак його практичне застосування для автономних систем є нежиттєздатним через значну затримку (RTF = 0.45) та фундаментальну залежність від мережевого з'єднання, що є неприйнятним для критичних сценаріїв експлуатації. [17]

На протипагу цьому, запропонована модель виявляється оптимальним рішенням для цільового застосування. Вона значно перевершує базову модель, особливо в умовах шуму (різниця в F1-Score становить 33 процентних пункти), що підтверджує ефективність використання зважених динамічних ознак. У порівнянні з класичним DTW, запропонований підхід досягає вищої точності при суттєво меншому обсязі пам'яті (у 8 разів) та вищій швидкості обробки.

Таким чином, експериментальні дані підтверджують, що запропонована модель досягає найкращого балансу, забезпечуючи точність, близьку до найкращих рішень, при збереженні надзвичайно низьких обчислювальних витрат та повної автономності. Це робить її ідеальним кандидатом для розгортання на бортових системах управління з обмеженими ресурсами.

Висновки та перспективи подальшого дослідження. У ході дослідження розроблено та валідовано метод побудови засобів розпізнавання ключових слів, що узагальнює конкретну реалізацію високопродуктивної класичної архітектури. Ключовим внеском роботи є формалізація принципу пріоритетизації інформативності ознак, який реалізовано через механізм зваженого акустичного фінгерпринтингу. Продемонстровано, що такий підхід, що базується на інтелектуальному аналізі ознак, є життєздатною альтернативою екстенсивному нарощуванню обчислювальної складності моделі для досягнення високої точності розпізнавання.

Емпірична валідація методу підтвердила його практичну ефективність. Результати порівняльного аналізу кількісно демонструють, що система, побудована згідно із запропонованими принципами, займає унікальну позицію у співвідношенні "точність-ефективність". Встановлено, що розрив у точності між оптимізованими класичними методами та великомасштабними хмарними моделями може бути суттєво скорочений (F1-міра 0.92 проти 0.97), при цьому зберігається перевага в обчислювальній ефективності на кілька порядків (RTF \approx 0.005 проти 0.450). Отримані дані також дозволяють стверджувати, що саме механізм зважування динамічних ознак є основним фактором, що забезпечує високу робастність системи в умовах сильного шуму.

Перспективи подальших досліджень полягають у розвитку запропонованих принципів. Першочерговим напрямком є вивчення можливості динамічної адаптації вектора вагових коефіцієнтів до змін акустичного середовища. Другим напрямком є дослідження альтернативних метричних просторів для порівняння "відбитків", зокрема, їх проектування у неперервний векторний простір для застосування таких метрик, як косинусна подібність. Третій напрямок може включати теоретичне дослідження асимптотичних меж точності для класичних методів, що базуються на пріоритетизації ознак, у порівнянні з нейромережевими архітектурами.

Список бібліографічного опису

1. Chen, G., Parada, C. and Heigold, G. (2014). Small-footprint keyword spotting using deep neural networks. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, pp. 4087-4091. doi: 10.1109/ICASSP.2014.6854370.
2. Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE, 77(2), pp. 257-286. doi: 10.1109/5.18626.
3. Seo, D., Oh, H.-S. and Jung, Y. (2021). Wav2KWS: Transfer Learning From Speech Representations for Keyword Spotting. IEEE Access, 9, pp. 80682-80691. doi: 10.1109/ACCESS.2021.3078715.
4. Dua, S. et al. (2022). Developing a Speech Recognition System for Recognizing Tonal Speech Signals Using a Convolutional Neural Network. Applied Sciences, 12(12), 6223. <https://doi.org/10.3390/app12126223>.
5. Alharbi, S. et al. (2021). Automatic Speech Recognition: Systematic Literature Review. IEEE Access, 9, pp. 131858-131876. doi: 10.1109/ACCESS.2021.3112535.
6. Oruh, J., Viriri, S. and Adegun, A. (2022). Long Short-Term Memory Recurrent Neural Network for Automatic Speech Recognition. IEEE Access, 10, pp. 30069-30079. doi: 10.1109/ACCESS.2022.3159339.
7. Yang, C.-H. H. et al. (2021). Decentralizing Feature Extraction with Quantum Convolutional Neural Network for Automatic Speech Recognition. ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, pp. 6523-6527. doi: 10.1109/ICASSP39728.2021.9413453.
8. Mahmood, A. and Köse, U. (2021). Speech recognition based on convolutional neural networks and MFCC algorithm. Advances in Artificial Intelligence Research, 1(1), pp. 6-12.
9. O'Shaughnessy, D. (2024). Trends and developments in automatic speech recognition research. Computer Speech & Language, 83, 101538. <https://doi.org/10.1016/j.csl.2023.101538>.
10. Furtuna, T. F. (2008). Dynamic Programming Algorithms in Speech Recognition. Informatica Economica, 12(2), pp.

94-98.

11. Picone, J. (1990). Continuous speech recognition using hidden Markov models. *IEEE ASSP Magazine*, 7(3), pp. 26-41. doi: 10.1109/53.54527.
12. Morwal, S., Jahan, N., and Chopra, D. (2012). Named Entity Recognition using Hidden Markov Model (HMM). *International Journal on Natural Language Computing (IJNLC)*, 1(4), pp. 15-23.
13. Wang, X. et al. (2012). Hidden-Markov-Models-Based Dynamic Hand Gesture Recognition. *Mathematical Problems in Engineering*, 2012, pp. 1-11. <https://doi.org/10.1155/2012/986134>.
14. Slam, W., Li, Y., and Urouvas, N. (2023). Frontier Research on Low-Resource Speech Recognition Technology. *Sensors*, 23(22), 9096. <https://doi.org/10.3390/s23229096>.
15. Leggetter, C. J. and Woodland, P. C. (1995). Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models. *Computer Speech & Language*, 9(2), pp. 171-185. doi: 10.1006/csla.1995.0010.
16. Korchenko, O. et al. (2025). Modular Neural Network Model for Biometric Authentication of Personnel in Critical Infrastructure Facilities Based on Facial Images. *Applied Sciences*, 15(5), 2553. <https://doi.org/10.3390/app15052553>.
17. Дичка, І. А., Терейковський, І. А., Дідус, А. В., Терейковська, Л. О., & Бояринова, Ю. Є. (2023). Оцінка ефективності засобів розпізнавання ключових слів у голосовому сигналі. *Вчені записки Таврійського національного університету імені В.І. Вернадського. Серія: Технічні науки*, 34(3), 1-7. <https://doi.org/10.32782/2663-5941/2023.3.1/19>.

References:

1. Chen, G., Parada, C. and Heigold, G. (2014). Small-footprint keyword spotting using deep neural networks. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, pp. 4087-4091. doi: 10.1109/ICASSP.2014.6854370.
2. Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2), pp. 257-286. doi: 10.1109/5.18626.
3. Seo, D., Oh, H.-S. and Jung, Y. (2021). Wav2KWS: Transfer Learning From Speech Representations for Keyword Spotting. *IEEE Access*, 9, pp. 80682-80691. doi: 10.1109/ACCESS.2021.3078715.
4. Dua, S. et al. (2022). Developing a Speech Recognition System for Recognizing Tonal Speech Signals Using a Convolutional Neural Network. *Applied Sciences*, 12(12), 6223. <https://doi.org/10.3390/app12126223>.
5. Alharbi, S. et al. (2021). Automatic Speech Recognition: Systematic Literature Review. *IEEE Access*, 9, pp. 131858-131876. doi: 10.1109/ACCESS.2021.3112535.
6. Oruh, J., Viriri, S. and Adegun, A. (2022). Long Short-Term Memory Recurrent Neural Network for Automatic Speech Recognition. *IEEE Access*, 10, pp. 30069-30079. doi: 10.1109/ACCESS.2022.3159339.
7. Yang, C.-H. H. et al. (2021). Decentralizing Feature Extraction with Quantum Convolutional Neural Network for Automatic Speech Recognition. *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toronto, ON, Canada, pp. 6523-6527. doi: 10.1109/ICASSP39728.2021.9413453.
8. Mahmood, A. and Köse, U. (2021). Speech recognition based on convolutional neural networks and MFCC algorithm. *Advances in Artificial Intelligence Research*, 1(1), pp. 6-12.
9. O'Shaughnessy, D. (2024). Trends and developments in automatic speech recognition research. *Computer Speech & Language*, 83, 101538. <https://doi.org/10.1016/j.csl.2023.101538>.
10. Furtuna, T. F. (2008). Dynamic Programming Algorithms in Speech Recognition. *Informatica Economica*, 12(2), pp. 94-98.
11. Picone, J. (1990). Continuous speech recognition using hidden Markov models. *IEEE ASSP Magazine*, 7(3), pp. 26-41. doi: 10.1109/53.54527.
12. Morwal, S., Jahan, N., and Chopra, D. (2012). Named Entity Recognition using Hidden Markov Model (HMM). *International Journal on Natural Language Computing (IJNLC)*, 1(4), pp. 15-23.
13. Wang, X. et al. (2012). Hidden-Markov-Models-Based Dynamic Hand Gesture Recognition. *Mathematical Problems in Engineering*, 2012, pp. 1-11. <https://doi.org/10.1155/2012/986134>.
14. Slam, W., Li, Y., and Urouvas, N. (2023). Frontier Research on Low-Resource Speech Recognition Technology. *Sensors*, 23(22), 9096. <https://doi.org/10.3390/s23229096>.
15. Leggetter, C. J. and Woodland, P. C. (1995). Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models. *Computer Speech & Language*, 9(2), pp. 171-185. doi: 10.1006/csla.1995.0010.
16. Korchenko, O. et al. (2025). Modular Neural Network Model for Biometric Authentication of Personnel in Critical Infrastructure Facilities Based on Facial Images. *Applied Sciences*, 15(5), 2553. <https://doi.org/10.3390/app15052553>.
17. Dychka, І. А., Tereikovskiy, І. А., Didus, А. В., Tereikovska, Л. О., & Boiarynova, Yu. Ye. (2023). Evaluation of the effectiveness of keyword recognition tools in a voice signal. *Vcheni zapysky Tavriyskoho natsionalnoho universytetu imeni V.I. Vernadskoho. Serii: Tekhnichni nauky [Scientific notes of V.I. Vernadsky Taurida National University. Series: Technical Sciences]*, 34(3), 1-7. <https://doi.org/10.32782/2663-5941/2023.3.1/19>. (In Ukrainian).