

УДК 004.53

DOI <https://doi.org/10.32782/2663-5941/2023.3.1/19>

Дичка І.А.

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

Терейковський І.А.

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

Дідус А.В.

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

Терейковська Л.О.

Київський національний університет будівництва і архітектури

Бояринова Ю.Є.

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

ОЦІНКА ЕФЕКТИВНОСТІ ЗАСОБІВ РОЗПІЗНАВАННЯ КЛЮЧОВИХ СЛІВ У ГОЛОСОВОМУ СИГНАЛІ

У статті проаналізовано ефективність процесу розпізнавання ключових слів у голосовому сигналі з метою підвищення ефективності та точності в рамках обмежених обчислювальних ресурсів. Розглянуто різні моделі розпізнавання ключових слів у голосових сигналах як з використанням класичних підходів, так і з використанням сучасніших, також сформовано основні критерії ефективності. Такі критерії включають в себе такі характеристики, як обмеження розміру словника, застосування складних обчислень, можливість обробки даних у реальному часі, стійкість до шуму та мовних варіацій, придатність для пристроїв з низьким ресурсом, високу точність і обробку довгих послідовностей. Проведено оцінювання цих критеріїв ефективності для порівняння засобів на базі різних підходів.

Встановлено, що приховані моделі Маркова показують хороші результати для фонових процесів в комп'ютерних системах з низьким ресурсом, тоді як згорткові та рекурентні нейронні мережі мають переваги при використанні більших обчислювальних ресурсів.

Сформовані характеристики дозволяють оцінити різні методи виявлення ключових слів у голосових сигналах і вказують на їх обмеження. Базуючись на отриманих даних, пропонується проводити подальші дослідження у сфері розпізнавання ключових слів в напрямку відповідності системам з обмеженими обчислювальними ресурсами, наприклад, враховуючи досвід засобів, які створені на базі прихованих моделей Маркова.

Окрім цього, стаття розглядає методи та підходи до розпізнавання ключових слів у голосових сигналах в комп'ютерних системах на базі мікроконтролерів, які мають обмежені характеристики в порівнянні з повноцінними комп'ютерними системами.

Проаналізовано використання нейронних мереж із затримкою часу, оптимізовану архітектуру нейронних мереж на мікроконтролерах, методи глибокого навчання та системи рекурентних нейронних мереж. Наявні дослідження сприяють розробці більш точних і ефективних методів виявлення ключових слів, що допомагає поліпшити системи розпізнавання та зробити їх підходящими для тих чи інших задач.

Ключові слова: розпізнавання ключових слів, голосовий сигнал, нейромережеві методи, приховані Марківські моделі, згорткові нейронні мережі, рекурентні нейронні мережі.

Постановка проблеми. У сучасному все більш цифровому світі йде постійний процес покращення взаємодії користувача з комп'ютером. Одним із найбільш перспективних напрямків

покращення такої взаємодії є застосування засобів аналізу голосового потоку. Від голосових помічників, таких як Alexa від Amazon, до автоматизованих систем обслуговування клієнтів,

здатність точно розшифрувати та відповідати на голосові команди чи запити стає все більш важливою. Основна складова даних систем є відомою як визначення ключових слів (KWS). Це область досліджень, яка охоплює області розпізнавання мови, штучного інтелекту та обчислювальної математики. Визначення ключових слів у голосових сигналах – це важлива сфера, яка вимагає високої точності і швидкості, що пов'язано з широким застосуванням даної технології у різних сферах життя. Однак, розпізнавання ключових слів часто накладає значні обмеження на обчислювальні ресурси. Адже розпізнавання ключових слів часто виконується як фоновий процес, що може реалізовуватись паралельно з іншими складними задачами, такими як аналіз даних або виконання інших алгоритмів обробки сигналів. Крім того, застосування моделей та методів розпізнавання ключових слів ставить виклик: як оптимізувати процес розпізнавання ключових слів, щоб він був ефективним і достатньо точним, відповідно до поставлених вимог, у застосуванні з обмеженими ресурсами? Цим обумовлюється актуальність науково-прикладної задачі розробки ефективних засобів розпізнавання ключових слів, адаптованих до застосування на малоресурсних комп'ютерних системах [1].

Аналіз останніх досліджень і публікацій. Виявлення ключових слів у контексті аудіосигналів – це процес ідентифікації конкретних «ключових слів» або фраз у розмовній мові. Ці ключові слова попередньо визначені на основі вимог програми, і система KWS навчена розпізнавати їх серед безлічі варіацій, присутніх у людській мові. Складність завдання виникає через мінливість мовлення, акцентів і вимови, а також потенційну присутність фонового шуму та інших особливостей мовлення [1, 2].

Основними складовими визначення ключових слів є «ключові слова» та «словники». Ключові слова – це слова або фрази, які система навчена ідентифікувати. Фактично, це тригер, на який має відреагувати дана технологія у безперервному потоці вхідного аудіо. Словники ж є комплексними базами даних, які містять фонетичні представлення цих ключових слів. Ці словники можуть бути різними за розмірами, включати різні акценти, мови та навіть нюанси у вимові, що дозволяє системі розпізнавати потрібні ключові слова незалежно від людини, яка говорить чи контексту [1].

Основна роль словників у виявленні ключових слів полягає у наданні фонетичних або текстових

представлень ключових слів, полегшуючи їх ідентифікацію в мові чи тексті. Для виявлення ключових слів використовуються різні типи словників, конкретне застосування яких залежить від вимог поставленого завдання.

1. Словники вимови. Це найпоширеніший тип словників, які використовуються для розпізнавання мовлення. Вони містять фонетичну транскрипцію слів або фраз, забезпечуючи відображення між написаним текстом і його вимовою. Словники вимови є важливими для таких завдань, як виявлення ключових слів у звукових сигналах, коли система має розуміти різні способи вимови слова.

2. Словники мовних моделей. Ці словники містять не лише фонетичні транскрипції, але й інформацію про ймовірність появи слів або фраз разом у послідовності. Їх часто використовують у поєднанні зі словниками вимови, щоб підвищити точність розпізнавання ключових слів.

3. Доменно-спеціалізовані словники. Ці словники розроблено для певних галузей або напрямків, наприклад, медицина, інженерія. Вони містять жаргон, акроніми та галузеву термінологію, яка зазвичай не зустрічається в загальному словнику.

4. Словники з наголосом. Це спеціалізовані словники, які враховують варіації у вимові через різні наголоси. Ці словники дозволяють точніше виявляти ключові слова в голосових даних, де є різні акценти.

5. Шумово-адаптивні словники. Це розширені типи словників, які можуть адаптуватися до різних умов шуму. Вони корисні в реальних програмах визначення ключових слів, де мовні дані можуть бути пошкоджені різними видами фонового шуму.

6. Багатомовні словники. Ці словники містять фонетичну транскрипцію слів або фраз кількома мовами. Вони особливо корисні для визначення ключових слів у багатомовних середовищах.

Кожен із цих типів словників відіграє вирішальну роль у пристосуванні процесу визначення ключових слів до конкретних вимог завдання, підвищуючи загальну точність і ефективність системи [1].

Технології розпізнавання ключових слів в аудіосигналах базуються на різноманітних математичних моделях та методах. Приховані марковські моделі (HMM), методи динамічного програмування, а віднедавна передові архітектури глибокого навчання, такі як згорткові нейронні мережі (CNN) і рекурентні нейронні мережі (RNN), утворюють обчислювальну основу KWS. Ці моделі дозволяють перетворювати, та структурувати

для подальшого аналізу необроблені неструктуровані аудіодані.

Приховані моделі Маркова (НММ) відіграли ключову роль у виявленні ключових слів, фіксуючи ймовірності різних фонетичних станів і переходів, віддзеркалюючи природний розвиток мовлення. Їх статистичний характер дозволяє ефективно моделювати послідовні характеристики мовлення.

Динамічне програмування доповнює НММ, визначаючи найбільш імовірну послідовність фонетичних станів, що веде до спостережуваних моделей мовлення. Ця синергія забезпечує ефективне виявлення ключових слів у безперервних голосових сигналах.

Розробки [3, 4, 5, 6] останніх років в області нейронних мереж, зокрема згорткових нейронних мереж (CNN) і рекурентних нейронних мереж (RNN), значно вдосконалили визначення ключових слів. CNN чудово ідентифікують чіткі фонетичні шаблони, тоді як RNN, особливо їх варіант довгострокової короткочасної пам'яті (LSTM), вміло обробляють послідовні дані, інкапсулюючи часові залежності в мові.

Загалом НММ та методи динамічного програмування забезпечують міцну основу для виявлення ключових слів, але останні досягнення в області нейронних мереж часто можуть давати ще більшу точність, проте часто втрачає у швидкодії та кількості використаних обчислювальних ресурсів. Постійний розвиток у даній сфер дає змогу покращувати здатність виявляти ключові слова в аудіосигналах, підвищуючи характеристики даної технології.

Однак в проаналізованій літературі не знайдено обґрунтованого переліку критеріїв оцінки ефективності засобів розпізнавання ключових слів в голосовому сигналі користувачів комп'ютерних систем, що не значно звужує можливості визначення перспективних шляхів вдосконалення означених засобів.

Мета статті формування переліку критеріїв ефективності комп'ютерних засобів розпізнавання ключових слів в голосовому сигналі користувача комп'ютерної системи.

Виклад основного матеріалу. Формування переліку критеріїв ефективності комп'ютерних засобів розпізнавання ключових слів в голосовому сигналі користувача комп'ютерної системи, базувалось на запропонованому в [2] підході, що передбачає співвіднесення таких критеріїв з вимогами до засобів розпізнавання, визначених за результатами аналізу науково-прикладних робіт в області розпізнавання ключових слів.

Стаття [3] представляє метод для виявлення ключових слів за допомогою нейронної мережі з часовою затримкою (TDNN). Ваги мережі ініціалізуються через передавання навчання зі схожого завдання розпізнавання мови з великим словником (LVCSR). Запропонована архітектура TDNN перевершує базову модель `snp-one-fstride4`, досягаючи значно нижчого рівня помилкового відхилення на реалістичних даних, зменшуючи обчислювальну складність на 50%. Дослідження також показує, що застосування пропуску кадрів може значно зменшити кількість множень без істотного впливу на точність, що робить систему підходящою для пристроїв з низьким споживанням енергії в режимі безперервного слухання.

Задекларовано, що підхід до розпізнавання ключових слів на базі TDNN має ряд переваг:

1. TDNN дозволяє ефективно обробляти послідовність даних і враховувати темпоральні залежності в аудіосигналах, що підвищує точність розпізнавання ключових слів.

2. Ініціалізація ваг мережі через передавання навчання зі схожого завдання LVCSR сприяє кращому загальному представленню ознак і пришвидшенню процесу навчання.

3. TDNN показує високу адаптивність до різних даних та завдань, забезпечуючи зниження помилкового відхилення та покращення точності розпізнавання.

4. Застосування пропуску кадрів дозволяє зменшити кількість множень без істотного впливу на точність, що робить систему підходящою для пристроїв з низьким споживанням енергії в режимі безперервного слухання.

5. TDNN може масштабуватися для різних розмірів наборів даних, кількості ключових слів та рівня складності завдань, завдяки своїй гнучкості та архітектурі.

6. TDNN може бути легко інтегрована з іншими методами глибокого навчання, такими як механізми уваги та трансформери, щоб отримати ще кращі результати в розпізнаванні мови та виявленні ключових слів.

У дослідженні [4] розглядаються проблеми оптимізації архітектур нейронних мереж для систем розпізнавання KWS, які можуть бути реалізовані на мікроконтролерах. Автори зосереджуються на підборі оптимальних параметрів мережі для забезпечення високої точності розпізнавання при обмежених обчислювальних ресурсах. Експерименти були проведені на наборі даних Google Speech Commands, призначених для побудови нейронних мереж, які мають класифікувати аудіо-

одані за категоріями, що відповідають одному з 10 ключових слів. Результати наведених досліджень показують, що глибокі роздільні згорткові нейронні мережі (DS-CNN) можуть забезпечити високу точність розпізнавання та ефективно використання обчислювальних ресурсів на мікроконтролерах. Вони також вимагають менше обчислювальних ресурсів та пам'яті в порівнянні з традиційними повнозв'язними згортковими мережами. Завдяки цьому, DS-CNN виявляються добре підходящими для застосування в системах KWS на пристроях з обмеженими ресурсами, такими як пристрої Інтернету речей. Крім DS-CNN розглянуто рекурентні нейронні мережі (RNN). Кожна з архітектур має свої переваги та недоліки:

1. Глибокі роздільні згорткові нейронні мережі (DS-CNN):

Переваги:

- Висока точність розпізнавання ключових слів.
- Ефективне використання обчислювальних ресурсів завдяки роздільним згорткам, що зменшують кількість параметрів моделі.
- Менш чутливі до змін у вхідних даних, що може поліпшити роботу на різних типах звукових сигналів.

Недоліки:

- Можуть мати складнішу архітектуру порівняно з RNN.
- Можуть вимагати більше часу на навчання та налаштування гіперпараметрів для досягнення оптимальної точності.

2. Рекурентні нейронні мережі (RNN):

Переваги:

- Здатні виявляти та використовувати властивості послідовності у вхідних даних, що може поліпшити розпізнавання ключових слів у звукових сигналах.
- Можуть мати менш складну архітектуру порівняно з DS-CNN, що сприяє меншому використанню обчислювальних ресурсів.

Недоліки:

- Точність розпізнавання ключових слів може бути нижчою порівняно з DS-CNN.
- Більш складні для оптимізації, що може призвести до проблем з обмеженими ресурсами мікроконтролерів.
- Можуть страждати від проблеми зникаючого градієнта, що ускладнює навчання моделі та досягнення оптимальної точності.

Враховуючи ці характеристики, дослідження підтверджує, що обидві архітектури мають свої сильні та слабкі сторони. Вибір оптимальної архітектури залежить від конкретних потреб та

обмежень, які ставляться перед моделлю розпізнавання ключових слів. Важливо враховувати фактори, такі як ресурси мікроконтролерів, точність розпізнавання та час навчання.

Ідеально, для досягнення найкращих результатів, рекомендується проводити експерименти з різними архітектурами та параметрами, щоб знайти оптимальне рішення для конкретної задачі. Застосування методів оптимізації, таких як прунінг та квантизація, може допомогти в подоланні обмежень ресурсів та покращити ефективність моделі без значного впливу на точність.

Дослідження підтверджує, що оптимізація архітектур нейронних мереж для систем розпізнавання ключових слів на мікроконтролерах є можливою та ефективною. Глибокі роздільні згорткові нейронні мережі (DS-CNN) виявилися підходящими для цієї задачі, забезпечуючи високу точність розпізнавання та ефективно використання обчислювальних ресурсів. Результати також вказують на можливість впровадження систем KWS на пристроях з обмеженими ресурсами, такими як пристрої Інтернету речей (IoT).

Слід враховувати, що моделі, оптимізовані для мікроконтролерів, можуть мати меншу точність, ніж більш потужні серверні моделі, а також можуть стикатися з технічними обмеженнями, такими як обмежена обчислювальна потужність та обсяг пам'яті.

Також була взята до уваги стаття [5], в якій представлено новий підхід до глибокого навчання для розпізнавання ключових слів (KWS – Keyword Spotting) з невизначеним словником, заснований на класифікації з темпоральним зв'язуванням (CTC – Connectionist Temporal Classification) з довготривалою короткотривалою пам'яттю (LSTM – Long Short-Term Memory). Запропонований метод навчає LSTM відрізняти фонему, використовуючи критерій CTC, та генерує решітку фонем, використовуючи інформацію про довготривалу залежність. Потім застосовується швидкий алгоритм пошуку підрядків, заснований на мінімальній редагуювальній відстані, для пошуку послідовності фонем ключового слова на решітці фонем. Підхід високоефективний і не залежить від словника.

Експериментальні результати на наборі даних WSJ0 показують, що запропонована система LSTM-CTC значно переважає традиційні підходи, засновані на прихованих марківських моделях (HMM – Hidden Markov Models), а також є більш ефективною, ніж базовий KWS, заснований на DNN-HMM (глибокі нейронні мережі – приховані марківські моделі). У майбутніх дослідженнях буде

розглянуто інтеграцію підходу LSTM-CTC з моделями Keyword-Filler для подальшого покращення роботи системи розпізнавання ключових слів.

Переваги запропонованого підходу до розпізнавання ключових слів (KWS) з використанням LSTM-CTC включають високу ефективність, незалежність від словника та можливість виявлення ключових слів на основі фонем. Також метод виявився більш успішним, ніж традиційні підходи, засновані на прихованих марківських моделях (HMM), і ефективніше, ніж базовий KWS, заснований на DNN-HMM. Однак, як і з будь-якою технікою, існують деякі недоліки. Зокрема, LSTM-CTC може вимагати значних обчислювальних ресурсів для навчання моделі, що може ускладнити швидке оновлення системи для додавання нових ключових слів. Крім того, інтеграція з моделями Keyword-Filler ще не була досліджена, тому можливий потенціал для подальшого покращення системи KWS залишається невідомим.

Також було розглянуто статтю [6], де автори розробляють системи розпізнавання ключових слів (KWS) на основі послідовності-допослідовності, використовуючи модель трансдукера рекурентної нейронної мережі (RNN-T). Ця модель дозволяє одночасно навчати акустичні та компоненти мовної моделі. У статті пропонується новий підхід до налаштування системи RNN-T для виявлення конкретного ключового слова за допомогою механізму уваги.

Експерименти показали, що система RNN-T, навчена з фонемними цілями, значно переважає над сильною базовою системою на основі CTC (класифікація за часовими зв'язками) з моделлю фонемного n-граму. Застосування запропонованої методики упередження дає додаткові вигоди порівняно з базовою моделлю RNN-T.

Автори використали рекурентну нейронну мережу для розробки ефективних систем розпізнавання ключових слів, які працюють в режимі потокової передачі. Вони довели, що модель RNN-T може значно покращити точність виявлення ключових слів порівняно з традиційними методами.

Переваги системи RNN-T для розпізнавання ключових слів:

1. Здатність одночасно навчати акустичні та мовні моделі, що спрощує архітектуру системи.
2. Використання механізму уваги для налаштування системи на конкретне ключове слово, що покращує точність виявлення ключових слів.
3. Вища точність порівняно з традиційними методами на основі CTC та HMM.

4. Здатність працювати в режимі потокової передачі.

Недоліки системи RNN-T для розпізнавання ключових слів:

1. Можливі проблеми з варіантами написання деяких ключових слів, що можуть призвести до погіршення точності моделі.
2. Необхідність додаткової роботи над впровадженням та оптимізацією алгоритмів для підвищення ефективності.
3. Складність для підтримки різних мов, які мають різні фонетичні та орфографічні властивості.
4. Більш високі обчислювальні вимоги для роботи з рекурентними нейронними мережами порівняно з деякими традиційними методами.

Проведений огляд моделей, методів та підходів до реалізації технології розпізнавання ключових слів дав змогу сформуванню базові характеристики зазначених засобів. В подальших дослідженнях, перелік сформованих характеристик на табл. 1 може бути розширений та модифікований згідно специфіки тієї чи іншої реалізації.

Таблиця 1

Перелік критеріїв ефективності засобів розпізнавання ключових слів в голосових сигналах

№	Опис критеріїв
N ₁	Обмеженість розмірів словника
N ₂	Використання складних обчислень
N ₃	Можливість обробки великих обсягів даних у реальному часі
N ₄	Стійкість до шуму та варіацій мовлення
N ₅	Можливість роботи на малоресурсних пристроях
N ₆	Висока точність
N ₇	Робота з довгими послідовностями
N ₈	Мінімізація енергоспоживання
N ₉	Можливість інтеграції з іншими методами розпізнавання

Сформовані характеристики дозволяють оцінити основні групи методів виявлення ключових слів у голосових сигналах та звернути увагу на недоліки тих чи інших методів. Оцінки, наведені в табл. 2, були сформовані за допомогою виставлення оцінок. Таким чином, ставилась 1, при наявності даної властивості, та 0 – при відсутності.

Як можна зрозуміти з табл. 2 вище, проаналізовані методи дозволяють реалізувати дану технологію в залежності від потреб, які ставляться – швидкість, точність, довжина ключових слів у словнику, обмеженість словника та ін.

Таблиця 2

Оцінка критеріїв ефективності засобів розпізнавання ключових слів в голосових сигналах

Засіб	Оцінка критерію								
	N ₁	N ₂	N ₃	N ₄	N ₅	N ₆	N ₇	N ₈	N ₉
Засоби на базі прихованих моделей Маркова	1	0	1	1	1	1	0	1	1
Засоби на базі згорткових нейронних мереж	0	1	0	1	0	1	1	0	1
Засоби на базі рекурентних нейронних мереж	0	1	0	1	0	1	1	0	1
Засоби на базі нейронних мереж з часовою затримкою	0	1	1	1	0	1	1	0	1

Варто відзначити, що мала кількість дослідників звертає увагу на класичні алгоритми, як приховані

Марківські моделі [7], які, судячи з табл. 2, гарно пристосовані для роботи в якості фонових процесу на малоресурсних комп'ютерних системах.

Сформований перелік характеристик дає змогу визначити наступні кроки дослідження – розробки методу виявлення ключових слів, який буде підходити для використання в системах з обмеженими ресурсами за аналогією з наявними рішеннями, як приховані Марківські моделі.

Висновки. У даному дослідженні було сформовано перелік критеріїв ефективності для технології розпізнавання ключових слів у голосових сигналах за допомогою огляду наявних рішень. Сформований перелік дає змогу провести оцінку засобів розпізнавання ключових слів у голосових сигналах, звертаючи увагу на вимоги, які ставляться перед ними.

Також було визначено перспективність подальших досліджень в даному напрямку, а саме покращення властивості роботи на малоресурсних пристроях, мінімізація енергоспоживання, особливо, за рахунок зменшення кількості складних обчислень.

Список літератури:

1. Михайленко В. М., Терейковська Л. О., Терейковський І. А., Ахметов, Б. Б. Нейромережеві моделі та методи розпізнавання фону в голосовому сигналі в системі дистанційного навчання. Київ: ЦП «Компринт», 2017. 252 с.
2. Терейковська Л. О. Методологія автоматизованого розпізнавання емоційного стану слухачів системи дистанційного навчання. Дисертація. Київ, 2002. 395 с.
3. Efficient keyword spotting using time delay neural networks. URL: <https://arxiv.org/pdf/1807.04353.pdf> (дата звернення: 23.05.2023)
4. Tutorial: Hello Edge: Keyword Spotting on Microcontrollers. URL: <https://arxiv.org/pdf/1711.07128.pdf> (дата звернення: 23.05.2023)
5. Unrestricted Vocabulary Keyword Spotting using LSTM-CTC. URL: https://www.isca-speech.org/archive_v0/Interspeech_2016/pdfs/0753.PDF (дата звернення: 23.05.2023)
6. STREAMING SMALL-FOOTPRINT KEYWORD SPOTTING USING SEQUENCE-TO-SEQUENCE MODELS. URL: <https://arxiv.org/pdf/1710.09617.pdf> (дата звернення: 23.05.2023)
7. Penagarikano, Mikel, & Bordel, German. (2004). Layered Markov models: A New architectural approach to automatic speech recognition. In Machine Learning for Signal Processing XIV – Proceedings of the 2004 IEEE Signal Processing Society Workshop (pp. 305-314). doi:10.1109/MLSP.2004.1422988. URL: https://www.researchgate.net/publication/4139645_Layered_Markov_models_A_New_architectural_approach_to_automatic_speech_recognition (дата звернення: 23.05.2023)
8. Грищук Т. В., Биков М. М. Моделювання процесу аналізу і класифікації голосових команд. Монографія. Вінниця: УНІВЕРСУМ-Вінниця, 2008. 146 с.

Dychka I.A., Tereikovskiy I.A., Didus A.V., Tereikovska L.O., Boyarinova Yu.Ye.

ASSESSMENT OF THE EFFICIENCY OF KEYWORD SPOTTING IN VOICE SIGNALS

The article analyzes the effectiveness of the process of recognizing keywords in a voice signal in order to increase efficiency and accuracy within the framework of limited computing resources. Different models of keyword recognition in voice signals using both classical and more modern approaches were considered, and the main efficiency criteria were also formed. Such criteria include such characteristics as limiting the size of the dictionary, applying complex computations, the ability to process data in real time, robustness to noise and language variations, suitability for low-resource devices, high accuracy, and processing of

long sequences. An evaluation of these performance criteria was carried out to compare means based on different approaches.

Hidden Markov models have been found to perform well for background processes in low-resource computing systems, while convolutional and recurrent neural networks have advantages when using larger computing resources.

The generated characteristics make it possible to evaluate different methods of detecting keywords in voice signals and indicate their limitations. Based on the obtained data, it is proposed to carry out further research in the field of keyword recognition in the direction of compliance with systems with limited computing resources, for example, taking into account the experience of tools that are created on the basis of hidden Markov models.

In addition, the article considers methods and approaches for recognizing keywords in voice signals in computer systems based on microcontrollers, which have limited characteristics compared to full-fledged computer systems.

Recent advances are reviewed, including the use of time-delayed neural networks, optimized neural network architectures on microcontrollers, deep learning techniques, and recurrent neural network systems. Existing research contributes to the development of more accurate and effective methods of identifying keywords, which helps to improve recognition systems and make them suitable for certain tasks.

Key words: *keyword recognition, voice signal, neural network methods, hidden Markov models, keyword spotting, convolutional neural networks, recurrent neural networks.*