

## РЕФЕРАТ

### на магістерську дисертацію

виконану на тему Метод розпізнавання слів у комп'ютерних системах  
студентом: Зіменков Дмитро Андрійович

**Актуальність теми.** Сучасний розвиток інформаційних технологій тісно пов'язаний із впровадженням інноваційних рішень у сфері обробки природної мови, зокрема автоматизацією розпізнавання мовлення в комп'ютерних системах. Це забезпечує підвищення ефективності взаємодії людини з комп'ютером, автоматизацію обробки аудіоданих і створення інтелектуальних систем, таких як голосові помічники, системи транскрипції та автоматичного перекладу. В умовах глобального зростання попиту на багатомовні та адаптивні системи зростає потреба в автоматизованих рішеннях для розпізнавання слів у реальному часі, які можуть працювати в різних умовах, включаючи шумні середовища та змішані мовні сценарії. Одним із перспективних підходів є застосування гібридних методів, що поєднують статистичні моделі, машинне навчання та глибоке навчання, для забезпечення високої точності й швидкості обробки аудіоданих.

Гібридні методи розпізнавання мовлення, які використовують сучасні алгоритми, відкривають нові можливості для універсального розпізнавання всіх мов, включаючи рідкісні, а також для обробки акцентів і діалектів у шумних умовах. Використання таких технологій дозволяє автоматизувати процес аналізу аудіо, зменшуючи залежність від людського фактору, підвищуючи швидкість і точність розпізнавання. Це особливо актуально для України, де розвиток інтелектуальних систем може сприяти підвищенню конкурентоспроможності на міжнародному ринку інформаційних технологій. Розробка методів автоматизованого розпізнавання слів у комп'ютерних системах є важливим науковим і практичним завданням, оскільки вона сприяє вдосконаленню інтерфейсів користувача та підвищенню ефективності обробки аудіоданих.

Об'єкт дослідження: процес розпізнавання слів у комп'ютерних системах.

Предмет дослідження: гібридні методи розпізнавання слів, що поєднують статистичні моделі, алгоритми машинного навчання та глибокі нейронні мережі, для обробки аудіоданих у комп'ютерних системах.

**Мета роботи:** Запропонувати ефективний гібридний метод обробки аудіоданих, адаптований до універсального розпізнавання всіх мов у реальному часі, включаючи шумні та багатомовні умови. Передбачено створення концепції програмного забезпечення, яке здатне автоматично розпізнавати слова, класифікувати мовні одиниці та оцінювати точність результатів.

**Наукова новизна полягає в наступному:**

- Розроблено гібридний метод розпізнавання слів, який інтегрує приховані марковські моделі (HMM), Support Vector Machines (SVM) і трансформери для забезпечення високої точності сегментації та класифікації аудіоданих за умов варіативного шуму та багатомовності.
- Запропоновано новий підхід до інтеграції компонентів гібридної моделі, що оптимізує конвеєр обробки аудіо, забезпечуючи низьку затримку (менше 100 мс) і високу точність (WER нижче 10%).
- Удосконалено методику попередньої обробки аудіоданих, яка включає адаптивне видалення шуму та витягнення ознак (MFCC і багатомовні embeddings), що підвищує стійкість моделі до шумних умов.
- Реалізовано концепцію програмного забезпечення, яке автоматизує процес розпізнавання слів, підтримуючи всі мови та адаптуючись до різних апаратних платформ, включаючи IoT-пристрої з обмеженими ресурсами.
- Запропоновано адаптивний підхід до тонкого налаштування моделі для рідкісних мов, який враховує обмеженість даних і забезпечує точність розпізнавання в реальних умовах.

Практична цінність полягає у можливості використання запропонованого гібридного методу для швидкого та точного розпізнавання слів у комп'ютерних системах. Застосування автоматизованої системи розпізнавання мовлення знижує витрати часу на обробку аудіоданих і зменшує залежність від суб'єктивного людського фактору, забезпечуючи об'єктивність і точність результатів. Запропонована концепція програмного забезпечення може бути використана в голосових помічниках, системах транскрипції, автоматичного перекладу та IoT-пристроях для обробки мовлення в реальному часі.

На відміну від існуючих методів, таких як традиційні приховані марковські моделі (HMM), глибокі нейронні мережі (DNN) чи сучасні трансформери (наприклад, Whisper від OpenAI), запропонований гібридний метод поєднує переваги трьох підходів – HMM, SVM і трансформерів – для досягнення вищої універсальності та ефективності. По-перше, інтеграція HMM забезпечує швидку сегментацію аудіосигналу (затримка <100 мс), що перевершує DNN і трансформери за швидкістю обробки на слабких платформах, де останні потребують >150–200 мс. По-друге, використання SVM із нелінійним ядром (RBF) підвищує стійкість до шуму (очікуваний WER <10% при SNR 0–10 дБ), тоді як окремі HMM чи DNN демонструють WER >15–20% у подібних умовах. По-третє, трансформери забезпечують точний контекстний аналіз, що дозволяє обробляти омоніми та рідкісні мови (наприклад, кримськотатарську, суахілі) ефективніше, ніж більшість комерційних систем, які обмежені поширеними мовами (WER >30% для рідкісних мов).

Крім того, оптимізаційні техніки, такі як дистилювання (зменшення моделі на 40%) і квантування (8-бітне представлення), знижують споживання пам'яті до <500 МБ, роблячи метод придатним для IoT-пристроїв (наприклад, Raspberry Pi), на відміну від ресурсоємних трансформерів, що потребують >2 ГБ. Це забезпечує енергоефективність і знижує витрати на апаратне забезпечення. Таким чином, запропонований метод перевершує існуючі рішення за комбінацією швидкості, точності, стійкості до шуму, підтримки всіх мов і адаптивності до різних платформ, сприяючи створенню конкурентоспроможних ІТ-рішень для автоматизації обробки аудіоданих.

Апробація роботи.

- ПРОБЛЕМАТИКА РОЗПІЗНАВАННЯ СЛІВ У КОМП'ЮТЕРНИХ СИСТЕМАХ Дмитро Зіменков, Ігор Терейковський. Матеріали 2-ї Міжнародної науково-практичної конференції «Інформаційні системи та технології: результати і перспективи» (IST 2025), 5 березня 2025 р. (Київ, Україна). К. : ФІТ КНУТШ, 2025 р. 425 с

матеріали магістерської дисертації використані при виконанні науково технічної роботи кафедри СПСКС Факультету Прикладної Математики Національного технічного університету України "Київський політехнічний інститут імені Ігоря Сікорського "Методи, моделі та засоби моніторингу закликів до тероризму у онлайн соціальних мережах" (Державний реєстраційний номер: 0124U003866, дата реєстрації: 05-09-2024)

Структура та обсяг роботи. Магістерська дисертація складається зі вступу, чотирьох розділів, висновків, списку використаних джерел та додатків. Загальний обсяг роботи становить 100 сторінок.

Ключові слова: розпізнавання слів, комп'ютерні системи, гібридний метод, обробка аудіоданих, машинне навчання, глибоке навчання, багатомовність.

## **ABSTRACT**

### **Relevance of the Topic**

The modern development of information technologies is closely linked to the implementation of innovative solutions in the field of natural language processing, particularly the automation of speech recognition in computer systems. This enhances the efficiency of human-computer interaction, automates the processing of audio data, and enables the creation of intelligent systems such as voice assistants, transcription systems, and automatic translation tools. Amid the global rise in demand for multilingual and adaptive systems, there is an increasing need for automated solutions for real-time word recognition capable of operating in diverse conditions, including noisy environments and mixed-language scenarios. One of the promising approaches is the application of hybrid methods that combine statistical models, machine learning, and deep learning to ensure high accuracy and speed in processing audio data.

Hybrid speech recognition methods utilizing modern algorithms open new opportunities for universal recognition of all languages, including rare ones, as well as for processing accents and dialects in noisy conditions. The use of such technologies enables the automation of audio analysis, reducing reliance on human intervention, and improving the speed and accuracy of recognition. This is particularly relevant for Ukraine, where the development of intelligent systems can enhance competitiveness in the global information technology market. The development of methods for automated word recognition in computer systems is a significant scientific and practical task, as it contributes to improving user interfaces and increasing the efficiency of audio data processing.

**Object of Research:** The process of word recognition in computer systems.

**Subject of Research:** Hybrid methods of word recognition that integrate statistical models, machine learning algorithms, and deep neural networks for processing audio data in computer systems.

**Objective of the Work:** To analyze modern methods of word recognition in computer systems and develop an effective hybrid method for processing audio data,

adapted for universal recognition of all languages in real time, including noisy and multilingual environments. The work aims to create a concept for software capable of automatically recognizing words, classifying linguistic units, and evaluating the accuracy of results.

Scientific Novelty:

- A hybrid word recognition method has been developed, integrating Hidden Markov Models (HMM), Support Vector Machines (SVM), and transformers to ensure high accuracy in audio data segmentation and classification under conditions of variable noise and multilingualism.
- A novel approach to integrating components of the hybrid model has been proposed, optimizing the audio processing pipeline to achieve low latency (less than 100 ms) and high accuracy (WER below 10%).
- An improved methodology for audio data preprocessing has been developed, incorporating adaptive noise reduction and feature extraction (MFCC), enhancing the model's robustness in noisy conditions.

Practical Significance: The practical significance lies in the potential application of the proposed hybrid method for fast and accurate word recognition in computer systems. The use of an automated speech recognition system reduces the time required for processing audio data and minimizes reliance on subjective human factors, ensuring objectivity and accuracy of results. The proposed software concept can be applied in voice assistants, transcription systems, automatic translation tools, and IoT devices for real-time speech processing.

Unlike existing methods, such as traditional Hidden Markov Models (HMM), Deep Neural Networks (DNN), or modern transformers (e.g., Whisper by OpenAI), the proposed hybrid method integrates the strengths of three approaches—HMM, SVM, and transformers—to achieve greater versatility and efficiency. Firstly, the incorporation of HMM enables rapid audio segmentation (latency <100 ms), surpassing DNN and transformers in processing speed on low-resource platforms, where the latter require >150–200 ms. Secondly, the use of Support Vector Machines (SVM) with a non-linear kernel (RBF) enhances robustness to noise (expected WER <10% at SNR 0–10 dB),

while standalone HMM or DNN exhibit WER >15–20% in similar conditions. Thirdly, transformers provide precise contextual analysis, enabling the processing of homonyms and rare languages (e.g., Crimean Tatar, Swahili) more effectively than most commercial systems, which are limited to widely spoken languages (WER >30% for rare languages).

Moreover, optimization techniques such as model distillation (reducing model size by 40%) and quantization (8-bit parameter representation) lower memory consumption to <500 MB, making the method suitable for IoT devices (e.g., Raspberry Pi), unlike resource-intensive transformers that require >2 GB. This ensures energy efficiency and reduces hardware costs. Thus, the proposed method outperforms existing solutions in terms of speed, accuracy, noise robustness, support for all languages, and adaptability to various platforms, contributing to the development of competitive IT solutions for automated audio data processing.

#### Work Validation:

- Zimenkov, D., Tereikovskiy, I. "Challenges of Word Recognition in Computer Systems." Proceedings of the 2nd International Scientific and Practical Conference "Information Systems and Technologies: Results and Perspectives" (IST 2025), March 5, 2025, Kyiv, Ukraine. Kyiv: FIT KNUTSH, 2025, 425 p.
- The materials of the master's thesis were used in the implementation of the scientific and technical project of the Department of System Programming and Specialized Computer Systems, Faculty of Applied Mathematics, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute," titled "Methods, Models, and Tools for Monitoring Calls to Terrorism in Online Social Networks" (State Registration Number: 0124U003866, Registration Date: 05-09-2024).

Structure and Scope of the Work: The master's thesis consists of an introduction, four chapters, conclusions, a list of references, and appendices. The total volume of the work is 100 pages.

Keywords: word recognition, computer systems, hybrid method, audio data processing, machine learning, deep learning, multilingualism.