

М. Е. БОНДАРЕНКОаспірант кафедри електронних обчислювальних машин
Харківський національний університет радіоелектроніки
ORCID: 0000-0002-2500-7626**Г. С. ІВАЩЕНКО**кандидат технічних наук, доцент,
доцент кафедри електронних обчислювальних машин
Харківський національний університет радіоелектроніки
ORCID: 0000-0003-1027-5262

АДАПТИВНИЙ РОЗПОДІЛ ОБЧИСЛЮВАЛЬНОГО НАВАНТАЖЕННЯ В СИСТЕМАХ ГОЛОСОВОЇ ІДЕНТИФІКАЦІЇ

У роботі запропоновано підхід до підвищення ефективності систем голосової ідентифікації шляхом розподілу обчислювального навантаження між центральним (CPU) та графічним (GPU) процесорами. Існуючі реалізації подібних систем ґрунтуються на фіксованих схемах розподілу, у межах яких окремі етапи обробки апіорі асоційовані з CPU або GPU. Така статична організація обчислень не враховує динамічної варіативності параметрів мовного сигналу, поточного стану обчислювальних ресурсів та істотних відмінностей у обчислювальній складності окремих операцій. За таких умов система позбавлена механізмів адаптивного перерозподілу навантаження, що зумовлює асиметричне використання CPU та GPU і, відповідно, призводить до зниження інтегральної продуктивності.

Запропоновано адаптивний алгоритм, який реалізує керування розподілом навантаження на основі аналізу комплексу характеристик мовних записів, що включає довжину кадру, ступінь перекриття, енергетичний вміст та спектральну насиченість сигналу. Використання цих параметрів дозволяє кількісно оцінювати обчислювальну складність поточного сегмента оброблюваного сигналу та динамічно визначати співвідношення обчислювальних операцій, виконуваних на CPU і GPU. Це забезпечує узгоджену взаємодію процесорів, мінімізацію простой і підвищення загальної швидкодії системи.

Проведений порівняльний аналіз показав, що використання адаптивного алгоритму дозволяє суттєво зменшити середній час обробки мовних фрагментів порівняно з підходами, у яких розподіл здійснюється лише за обсягом даних. Запропонований блок адаптивного керування та розподілу навантаження підвищує загальну продуктивність систем голосової ідентифікації, особливо під час обробки великих та структурно складних наборів сигналів, і може бути інтегрований у сучасні багатопроцесорні архітектури.

Ключові слова: системи голосової ідентифікації, MFCC, спектральне віднімання, вейвлет-фільтрація, CPU, GPU, адаптивний розподіл навантаження, керування навантаженням, розподіл задач, динамічне балансування ресурсів, продуктивність у реальному часі, багатопроцесорні системи.

М. Е. BONDARENKOPostgraduate Student at the Department of Electronic Computers
Kharkiv National University of Radio Electronics
ORCID: 0000-0002-2500-7626**H. S. IVASHCHENKO**Candidate of Technical Sciences, Associate Professor,
Associate Professor at the Department of Electronic Computers
Kharkiv National University of Radio Electronics
ORCID: 0000-0003-1027-5262

ADAPTIVE LOAD DISTRIBUTION IN VOICE IDENTIFICATION SYSTEMS

This paper examines an approach to improving the efficiency of voice identification systems by distributing the computational load between the central processing unit (CPU) and the graphics processing unit (GPU). Existing implementations of such systems are based on fixed distribution schemes, within which individual processing stages are a priori associated with either the CPU or the GPU. Such a static organisation of computations does not take into account

the dynamic variability of speech signal parameters, the current state of computing resources, and significant differences in the computational complexity of individual operations. Under such conditions, the system lacks mechanisms for adaptive load redistribution, resulting in asymmetric use of the CPU and GPU, and consequently, a decrease in overall performance.

The proposed approach involves an adaptive algorithm that implements load distribution control based on the analysis of a set of speech recording characteristics, including frame length, overlap degree, energy content, and spectral saturation of the signal. The use of these parameters enables a quantitative assessment of the computational complexity of the current segment of the processed signal and the dynamic determination of the ratio of computational operations performed on the CPU and GPU. This ensures coordinated interaction between processors, minimises downtime, and increases the overall performance of the system.

The comparative analysis revealed that the use of an adaptive algorithm significantly reduces the average processing time of speech fragments compared to approaches that rely solely on data volume. The proposed adaptive control and load distribution block increases the overall performance of voice identification systems, especially when processing large and structurally complex sets of signals, and can be integrated into modern multiprocessor architectures.

Key words: *voice identification systems, MFCC, spectral subtraction, wavelet filtering, CPU, GPU, adaptive load balancing, load management, task distribution, dynamic resource balancing, real-time performance, multiprocessor systems.*

Постановка проблеми

Голосова ідентифікація, як різновид біометрії, аналізує індивідуальні акустичні параметри мовного сигналу: частоту основного тону, формантну структуру, амплітудно-частотні характеристики та тембр, що відображають унікальні властивості мовного апарату людини: будову голосових зв'язок, форму резонаторних порожнин і особливості артикуляційних рухів. Голосова біометрія вже активно використовується у банківській сфері, службах підтримки, системах «розумного дому» та мобільних пристроях, забезпечуючи надійну та зручну перевірку особи [1-3].

Складність алгоритмів вилучення, аналізу та розпізнавання біометричних ознак потребує значних обчислювальних ресурсів [4]. Сучасні системи біометричної ідентифікації реалізуються на основі багатоядерних центральних процесорів (CPU), графічних процесорів (GPU) або їх комбінації у вигляді гібридних обчислювальних архітектур [5-6]. GPU забезпечує ефективну паралельну обробку великої кількості фреймів сигналу, що дозволяє скоротити час навчання та розпізнавання, тоді як CPU виконує керування потоками даних, попередню нормалізацію мовного сигналу та синхронізацію між етапами обчислень [7].

Одним із перспективних напрямів підвищення ефективності біометричних систем є керування розподілом обчислювального навантаження між CPU та GPU з урахуванням характеру операцій, ступіню їх паралельності, обсягів даних та завантаження апаратних ресурсів [8].

Аналіз останніх досліджень і публікацій

Розподілення навантаження між CPU та GPU з метою підвищення продуктивності гібридних обчислювальних систем висвітлюється у низці досліджень [9-11].

У [9] представлено підхід до спільної обробки аудіосигналів на центральному та графічному процесорах із використанням середовища OpenCL. Балансування навантаження здійснюється шляхом адаптивного розподілу фреймів або каналів сигналу між процесорами залежно від поточного завантаження CPU. Недоліком описаного підходу є ускладнена оцінка стану GPU, що знижує ефективність розподілу навантаження.

Дослідження [10] присвячено забезпеченню роботи в реальному часі при відеокодуванні H.264/AVC на гібридних системах, що поєднують багатоядерні CPU та GPU. CPU виконує керування, ініціалізацію та передачу даних через шину PCIe, тоді як GPU виконує частини алгоритмів у масовопаралельному режимі. Реалізація базується на технологіях CUDA, OpenMP та інструментах виробників апаратного забезпечення. Розподіл навантаження між CPU і GPU здійснюється адаптивно за допомогою ітераційного алгоритму балансування, який оцінює фактичний час обробки макроблоків, оновлює параметри продуктивності після обробки кожного кадру і корегує розподіл, враховуючи продуктивність пристроїв та пропускну здатність каналів. Основним недоліком є часові витрати на передавання даних через шину PCIe.

Дослідження [11] присвячене розробці методу адаптивного балансування обчислювального навантаження між CPU та GPU у системах розпізнавання мовлення. Обчислення виконуються паралельно: CPU обробляє етапи з низькою паралельністю або складною логікою (керування потоками, оцінка ймовірнісних моделей), а GPU прискорює масово паралельні обчислення, зокрема спектральні перетворення, фільтрацію та обчислення MFCC. Розподіл навантаження між процесорами здійснюється адаптивно, на основі аналізу часу виконання попередніх завдань та розміру оброблюваних блоків даних. Недоліком запропонованого підходу є підвищені витрати на обмін даними між процесорами, оскільки в процесі адаптивного балансування одні й ті самі фрейми можуть помилково передаватися між пристроями кілька разів, що спричиняє додаткові затримки та зменшує загальну ефективність обчислювального процесу.

Розглянуті дослідження підтверджують доцільність адаптивного керування розподілом навантаження між CPU та GPU в системах паралельної обробки даних, зокрема голосових сигналів.

Формулювання мети дослідження

Метою роботи є розробка та дослідження ефективності системи обробки голосових сигналів із алгоритмом адаптивного балансування навантаження між CPU та GPU. Підхід повинен передбачати динамічну оцінку складності кожного кадру сигналу на основі статистичних та спектральних характеристик і відповідне розподілення обчислювальних ресурсів при роботі з сигналами різної складності.

Викладення основного матеріалу дослідження

Обробка голосових сигналів у системах ідентифікації мовця включає підготовку сигналу (нормалізація, фільтрація, фреймування, кадрування за допомогою віконної функції), виділення ознак, класифікацію та прийняття рішення [7]. Перед обробкою кожен кадр сигналу проходить швидкий попередній аналіз у окремому потоці CPU. Це дозволяє оцінити характеристики сигналу та визначити його обчислювальну складність до застосування ресурсоемних алгоритмів.

Одним із параметрів, що визначаються при попередньому аналізі, є середньоквадратичне відхилення амплітуди кадру, що відображає розсіяння значень сигналу відносно середнього рівня та характеризує його коливальну активність:

$$\sigma = \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} (x[n] - \bar{x})^2}, \quad (1)$$

$$\bar{x} = \frac{1}{N} \sum_{n=0}^{N-1} x[n],$$

де $x[n]$ – відліки сигналу у кадрі, а N – кількість відліків у кадрі. Зростання σ вказує про високу амплітудну варіабельність і, відповідно, більшу обчислювальну складність операцій нормалізації та фільтрації.

Також враховується середня енергія кадру, яка відображає активність сигналу та використовується для визначення впливу амплітуди на обробку:

$$E = \frac{1}{N} \sum_{n=0}^{N-1} x^2[n] \quad (2)$$

Високі значення енергії вказують на наявність активних кадрів сигналу, що потребують інтенсивної обробки.

Для оцінки спектральної структури кадру розраховується коефіцієнт варіації спектра, який характеризує відносний розкид амплітуд спектральних компонентів та дозволяє визначити рівень деталізації сигналу:

$$CV_{\text{spec}} = \frac{\sigma_{\text{spec}}}{\mu_{\text{spec}}}, \quad (3)$$

$$\mu_{\text{spec}} = \frac{1}{N/2} \sum_{k=0}^{N/2-1} |X[k]|,$$

$$\sigma_{\text{spec}} = \sqrt{\frac{1}{N/2} \sum_{k=0}^{N/2-1} (|X[k]| - \mu_{\text{spec}})^2},$$

де $X[k]$ – дискретне Фур'є-перетворення кадру, μ_{spec} – середнє значення амплітуд спектра, σ_{spec} – середньоквадратичне відхилення амплітуд спектра. Значення CV_{spec} зростає зі збільшенням спектральної варіативності, що прямо впливає на обчислювальне навантаження при порівнянні з шаблонами.

Додатково розраховується спектральна ентропія, яка є мірою інформаційної насиченості та складності сигналу:

$$H = - \sum_{k=0}^{N/2-1} P[k] \log P[k], \quad (4)$$

$$P[k] = \frac{|X[k]|^2}{\sum_{j=0}^{N/2-1} |X[j]|^2},$$

де $P[k]$ – нормалізована спектральна потужність у точці частотного індексу k .

Високі значення спектральної ентропії вказують на наявність численних частотних компонентів, що ускладнює розпізнавання сигналу та підвищує потребу в обчислювальних ресурсах.

На основі розрахованих статистичних параметрів формується вектор ознак $F = [f_1, f_2, \dots, f_m]$, який передається модулю оцінювання складності. Попередньо всі параметри нормалізуються за виразом:

$$\rho_i^n = \frac{\rho_i - \rho_i^{\min}}{\rho_i^{\max} - \rho_i^{\min}}, \quad (5)$$

де ρ_i^{\min} та ρ_i^{\max} – мінімальне та максимальне значення параметра у поточній вибірці.

Модуль адаптивного керування розподілом навантаження обчислює зведений індекс складності кадру шляхом зваженого підсумовування всіх описаних нормалізованих параметрів із урахуванням вагових коефіцієнтів, що визначаються в процесі адаптації системи:

$$C_f = \sum_{i=1}^m \omega_i \rho_i^n, \quad (6)$$

де ω_i – вагові коефіцієнти, що характеризують відносну важливість кожного параметра у загальній оцінці. Вагові коефіцієнти визначаються таким чином, щоб сумарна вага всіх параметрів дорівнювала одиниці, що забезпечує нормування внеску кожного параметра у підсумкову оцінку C_f та робить індекс незалежним від кількості використовуваних характеристик:

$$\sum_{i=1}^m \omega_i = 1, \quad (7)$$

Оцінювання вагових коефіцієнтів відбувається динамічно: система безперервно відстежує статистику змін параметрів у часовому вікні, обчислюючи ковзні середні значення та дисперсії для кожного кадру. Це дозволяє адаптувати ваги в реальному часі до умов сигналу та підвищувати значущість енергетичного параметра при наявності шумів або спектрального при багатокomпонентній структурі мовлення.

Аналіз динаміки ознак голосового сигналу дає змогу системі реалізувати адаптивні стратегії розподілу навантаження між CPU та GPU. Для скорочення обчислень та підвищення ефективності роботи системи реалізовано механізм адаптивного пропускання кадрів, який приймає рішення на основі зваженої сумарної оцінки кожного кадру. Цей механізм дає змогу динамічно регулювати частоту виконання обчислень, уникаючи зайвого аналізу кадрів із низькою інформативністю.

Алгоритм прийняття рішення вибору виконавчого пристрою передбачає, що система обчислює вектор ознак для кожного кадру. Якщо виявлено суттєві зміни статистичних або спектральних характеристик (наприклад, різке зростання енергетичного рівня, зміна частотної структури або поява шумових компонентів), це свідчить про появу нового стану сигналу, що потребує оновлення параметрів для точної оцінки його складності. Якщо змін не виявлено, виконується розрахунок зваженої сумарної характеристики кадру, що дозволяє оцінити інтегральну складність без повторного повного аналізу:

$$S_i = \sum_{k=1}^m \omega_k f_{ik}, \quad (8)$$

де w_k – вагові коефіцієнти, визначені під час адаптації системи та відображають значущість окремих ознак.

Якщо обчислене значення S_i виявляється меншим за адаптивний поріг $S_{\text{пор}}$, кадр пропускається. Пропущений кадр вважається малозначущим для подальшої обробки, що дозволяє скоротити кількість операцій без втрати інформативності. Порогове значення $S_{\text{пор}}$ визначається статистично на основі середнього рівня складності попередніх кадрів із врахуванням їх дисперсії або ковзного стандартного відхилення:

$$T_{\text{адм}} = \mu_{\text{поперед}} + k + \sigma_{\text{поперед}}, \quad (9)$$

де $\mu_{\text{поперед}}$ – середнє значення складності попередніх кадрів, $\sigma_{\text{поперед}}$ – стандартне відхилення, k – коефіцієнт, що визначає чутливість алгоритму до змін сигналу.

У результаті система динамічно регулює кількість необхідних обчислень: при появі нових або складних фрагментів сигналу активується повний розрахунок параметрів, у стабільних фрагментах використовується лише оцінка складності, а в неінформативних інтервалах вихідного сигналу кадри пропускаються.

Система реалізує механізм повторного перенаправлення обчислювальних задач, у межах якого кожен вузол повідомляє блок керування про стан виконання кадру. Якщо час обробки перевищує встановлений поріг, кадр автоматично перенаправляється на процесор іншого типу. Такий підхід дозволяє адаптивно перерозподіляти навантаження між процесорами, уникаючи вузьких місць і підтримувати стабільну продуктивність навіть за значних коливань характеристик сигналів або завантаженості обчислювальних пристроїв, забезпечуючи ефективну роботу системи в режимі реального часу без простоїв та втрат швидкодії.

Для кожного кадру визначається поріг часу обробки $T_{\text{поріз}}$, перевищення якого або виявлення перевантаження вузла ініціює автоматичне перенаправлення кадру на інший вузол, що гарантує своєчасну обробку. Поріг часу обчислюється на основі середнього часу обробки останніх кадрів $T_{\text{сеп}}$ та стандартного відхилення σ_{\square} :

$$T_{\text{поріз}} = T_{\text{сеп}} + k \cdot \sigma_{\square}, \quad (10)$$

де $T_{сер}$ – середній час обробки попередніх кадрів, σ_T – стандартне відхилення часу обробки, k – емпіричний коефіцієнт, що визначає допустиме відхилення.

З урахуванням поточного завантаження вузла $L_{вузол}$ поріг коригується:

$$T_{поріг}^{кор} = T_{поріг} \cdot (1 - \alpha L_{вузол}), \tag{11}$$

де α – коефіцієнт адаптації порога до навантаження.

Для оцінки продуктивності системи голосової ідентифікації з адаптивним розподілом завдань між CPU та GPU метрикою ефективності обрано час виконання обробки кадру (у секундах) та ступінь завантаженості обчислювальних вузлів (у відсотках), що дозволяє оцінити, наскільки рівномірно розподіляються обчислювальні ресурси між CPU та GPU.

Експериментальна модель системи голосової ідентифікації була розгорнута у середовищі Google Colab, із підтримкою багатопотокових обчислень на CPU за допомогою бібліотек OpenMP та масово-паралельних обчислень на GPU. Такий підхід дозволяє оцінити ефективність адаптивного розподілу навантаження між CPU Intel Core i7-13700H та дискретним графічним прискорювачем NVIDIA архітектури Ada Lovelace (GTX 4060), забезпечуючи дослідження взаємодії між процесором і графічним модулем.

Для оцінки ефективності обробки голосових сигналів використано датасет SEWA, який містить понад 2000 хвилин аудіо-записів 398 учасників із шести культур. Емоційний стан учасників описується двома параметрами: позитивністю/негативністю та інтенсивністю прояву емоцій [12]. На основі цього датасету сформовано набори різного обсягу та складності, що включають короткі нейтральні фрагменти з низькою спектральною складністю і мінімальним шумом, середні за тривалістю сигнали з різною інтенсивністю емоцій та фоновим шумом, а також довгі діалогові сегменти з динамічними коливаннями енергетичної інтенсивності, паузами та періодичними шумовими перешкодами (Таблиця 1).

Таблиця 1

Тестові набори даних

№	Кількість файлів	Середня тривалість, с	Опис сигналу
1	100	0,8	Короткі фрази
2	200	1,5	Фрази середньої довжини
3	500	3,0	Запис із помірним шумом
4	1000	4,2	Фонові перешкоди
5	2000	5,0	Сильний шум середовища
6	5000	8,0	Розмовна мова у приміщенні
7	10000	12,0	Фонові голоси
8	20000	20,0	Уривки лекційних занять
9	50000	35,0	Складні акустичні умови
10	100000	60,0	Діалоги з високим фоновим шумом

Кадри організовано у багатовимірні тензори, де перша вісь відповідає номеру кадру, а наступні – часовим і частотним вимірам, що забезпечує паралельну обробку кількох фреймів і повніше використання ресурсів обчислювальних вузлів (CPU та GPU).

Для ускладнення роботи блоку адаптивного розподілу обчислювального навантаження в наборах тестових даних передбачено короткі імпульсні сегменти з високими піковими коливаннями амплітуди та фоновим шумом, а також чергування коротких, середніх та довгих кадрів у межах одного пакету, імітуючи реальні умови потокового мовлення з різкою зміною складності спектра та енергії.

Розглянуто чотири сценарії виконання обробки голосових сигналів, залежно від задіяних обчислювальних вузлів: лише на CPU, лише на GPU, спрощений розподіл за розміром сигналу (Size-based Scheduler) та повна адаптивна система (Full Adaptive Scheduler), яка аналізує характеристики сигналу для оптимального розподілу завдань.

Результати виконання методів попередньої обробки (спектральний аналіз, вейвлет-фільтрація), наведені в таблиці 2, свідчать, що час роботи адаптивної системи є найменшим серед розглянутих варіантів, що підтверджує її ефективність.

Розрахунок мел-частотних кепстральних коефіцієнтів MFCC в системі з застосуванням алгоритму адаптивного розподілу потребує найменший час (Таблиця 3), що підтверджує ефективність запропонованого підходу у задачах різного рівня навантаження. Водночас для окремих наборів даних спостерігається незначне зменшення часу виконання порівняно з CPU- та GPU-орієнтованими системами, що свідчить про коректність вибору обчислювального вузла відповідно до характеристик вхідних даних та їхньої обчислювальної складності.

Таблиця 2

Час виконання етапів попередньої обробки

№ тестового набору	CPU, с	GPU, с	Size-based Scheduler, с	Full Adaptive Scheduler, с
1	0,42	0,61	0,4	0,38
2	0,85	0,73	0,69	0,62
3	2,10	1,20	1,23	0,91
4	4,75	2,45	1,98	1,72
5	9,40	4,60	3,92	3,35
6	21,2	10,3	18,5	7,24
7	43,7	20,4	19,9	15,5
8	87,9	39,5	37,8	29,6
9	193	82,1	90,2	62,8
10	391	164	195	127

Таблиця 3

Час обчислення MFCC

№ тестового набору	CPU, с	GPU, с	Size-based Scheduler, с	Full Adaptive Scheduler, с
1	0,58	0,66	0,55	0,51
2	1,12	0,84	0,75	0,69
3	2,84	1,49	1,35	1,18
4	6,10	3,10	7,76	2,41
5	12,5	6,25	5,32	4,64
6	228,3	13,8	11,9	10,1
7	55,4	27,1	173,3	20,2
8	112	54,3	47,5	40,9
9	2238	113	2981,1	84,5
10	486	223	198	167

Проведене експериментальне дослідження дозволило оцінити навантаженість обчислювальних вузлів під час обробки різнорідних наборів мовних сигналів при використанні розподілу за обсягом та з урахуванням численних характеристик сигналу.

Таблиця 4

Результати навантаження вузлів системи під час роботи системи

№ тестового набору	Size-based Scheduler		Full Adaptive Scheduler	
	CPU, %	GPU, %	CPU, %	GPU, %
1	92	5	66	81
2	31	44	72	63
3	96	13	68	42
4	24	83	74	63
5	86	12	62	84
6	77	15	53	42
7	34	36	81	32
8	23	91	61	46
9	56	26	74	44
10	88	20	36	22

Як показано в таблиці 4, обсягоорієнтований підхід до розподілу навантаження (Size-based Scheduler) спричиняє дисбаланс: один вузол працює на межі обчислювальних можливостей, тоді як інший простоює, що знижує сумарну продуктивність і збільшує загальний час обробки. Тоді як запропонований адаптивний алгоритм забезпечує динамічний розподіл задач між CPU та GPU, враховуючи складність кадрів, поточне завантаження вузлів і варіативність характеристик сигналів.

Висновки

Запропоновано адаптивний розподіл обчислювального навантаження у системах обробки голосових сигналів, який використовує комплекс характеристик мовних записів, зокрема довжину кадру, рівень енергетичної активності та спектральну насиченість.

Експериментальні дослідження виконувалися на основі наборів даних різного обсягу та рівня складності сигналів, з урахуванням наявності викривлень, що дозволило всебічно перевірити ефективність адаптивного

механізму розподілу навантаження, оцінити його здатність визначати пріоритетність обробки кадрів та використовувати ресурси CPU та GPU незалежно від тривалості, спектральної складності, емоційного навантаження та наявності шумів у вхідних сигналах. Проведено порівняльний аналіз роботи системи без адаптивного розподілу, з його спрощеною версією (що враховує лише обсяг даних), та запропонованого рішення, яке дозволяє розподіляти обчислення між вузлами, підвищуючи ефективність використання ресурсів та загальну продуктивність системи.

Список використаної літератури

1. Samonte M. J. C., Callejo J. K., Lumbea D. C. N., Ocaya J. C. B. Mitigating Vishing in Digital Banking Through Caller Authentication and Verification Technologies. *2024 14th International Conference on Software Technology and Engineering ICSTE*, Macau, China. 2024. Pp. 102–108. DOI: 10.1109/ICSTE63875.2024.00025
2. Kambampati P., Rane S., Shoeb A., Dhannawat R. PAYV Payment Voice A Platform using Voice Recognition to Enable Payment Transactions. *2024 Asia Pacific Conference on Innovation in Technology APCIT*, Mysore, India. 2024. Pp. 1–6. DOI: 10.1109/APCIT62007.2024.10673442
3. Li Y., Gao X., Song Q., Wang Y., Lyu P., Zhang H. BoneAuth A Bone-Conduction-Based Voice Liveness Authentication for Voice Assistants. *IEEE Internet of Things Journal*. 2025. Vol. 12, no. 6. Pp. 6997–7009. DOI: 10.1109/JIOT.2024.3494024
4. Bao L., Zuo Y. Speaker Identification based on MFSC Voice Feature Extraction using Transformer. *2023 IEEE International Conference on Data Mining Workshops ICDMW*, Shanghai, China. 2023. Pp. 1–7. DOI: 10.1109/ICDMW60847.2023.00008
5. Chen Q., Gu Z., Lu L., Xu X., Ba Z., Lin F., Liu Z., Ren K. Conan's Bow Tie A Streaming Voice Conversion for Real-Time VTuber Livestreaming. *Proceedings of the 29th International Conference on Intelligent User Interfaces*. 2024. Pp. 35–50. DOI: 10.1145/3640543.3645146
6. Mykhailichenko I., Ivashchenko H., Barkovska O., Liashenko O. Application of Deep Neural Network for Real-Time Voice Command Recognition. *IEEE 3rd KhPI Week on Advanced Technology KhPIWeek*, Kharkiv, Ukraine. 2022. Pp. 1–4. DOI: 10.1109/KhPIWeek57572.2022.9916473
7. Бондаренко М. Е., Іващенко Г. С. Використання послідовності методів попередньої обробки в системах голосової ідентифікації. *Системи управління, навігації та зв'язку*. Полтава: ПНТУ. 2025. № 2 (80). С. 90–96. DOI: 10.26906/SUNZ.2025.2.090
8. Бондаренко М. Е., Іващенко Г. С. Організація паралельного виконання методів обробки голосових сигналів на багатоядерних CPU та GPU. *Системи управління, навігації та зв'язку*. Полтава: ПНТУ. 2025. № 4 (82). С. 39–44. DOI: 10.26906/SUNZ.2025.4.39-44
9. Gjermundsen A. CPU and GPU Co-processing for Sound. MA thesis. *Norwegian University of Science and Technology*. 2010. 173 p.
10. Momcilovic S., Ilic A., Roma N., Sousa L. Dynamic Load Balancing for Real-Time Video Encoding on Heterogeneous CPU+GPU Systems. *IEEE Transactions on Multimedia*. 2014. Vol. 16, no. 1. Pp. 108–121. DOI: 10.1109/TMM.2013.2284892
11. Kim J., Lane I. Accelerating Large Vocabulary Continuous Speech Recognition on Heterogeneous CPU-GPU Platforms. *2014 IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP*, Florence, Italy. 2014. Pp. 3291–3295. DOI: 10.1109/ICASSP.2014.6854209
12. Kossaifi J., Walecki R., Panagakis Y., Shen J., Schuller B., Pantic M. SEWA DB A Rich Database for Audio-Visual Emotion and Sentiment Research in the Wild. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2021. Vol. 43, no. 3. Pp. 1022–1040. DOI: 10.1109/TPAMI.2019.2944808

References

1. Samonte, M. J., C. Callejo, J. K., Lumbea, D. C. N., Ocaya, J. C. B. (2024). Mitigating Vishing in Digital Banking Through Caller Authentication and Verification Technologies. *2024 14th International Conference on Software Technology and Engineering (ICSTE)*. Macau, China, pp. 102–108. DOI: 10.1109/ICSTE63875.2024.00025
2. Kambampati, P., Rane, S., Shoeb, A., Dhannawat, R. (2024). PAYV – Payment Voice: A Platform using Voice Recognition to Enable Payment Transactions. *2024 Asia Pacific Conference on Innovation in Technology (APCIT)*. Mysore, India, pp. 1–6. DOI: 10.1109/APCIT62007.2024.10673442
3. Li, Y., Gao, X., Song, Q., Wang, Y., Lyu, P., Zhang, H. (2025). BoneAuth: A Bone-Conduction-Based Voice Liveness Authentication for Voice Assistants. *IEEE Internet of Things Journal*. Vol. 12, No. 6, pp. 6997–7009. DOI: 10.1109/JIOT.2024.3494024
4. Bao, L., Zuo, Y. (2023). Speaker Identification based on MFSC Voice Feature Extraction using Transformer. *2023 IEEE International Conference on Data Mining Workshops (ICDMW)*. Shanghai, China, pp. 1–7. DOI: 10.1109/ICDMW60847.2023.00008

5. Chen, Q., Gu, Z., Lu, L., Xu, X., Ba, Z., Lin, F., Liu, Z., Ren, K. (2024). Conan's Bow Tie: A Streaming Voice Conversion for Real-Time VTuber Livestreaming. *Proceedings of the 29th International Conference on Intelligent User Interfaces*, pp. 35–50. DOI: 10.1145/3640543.3645146
6. Mykhailichenko, I., Ivashchenko, H., Barkovska, O., Liashenko, O. (2022). Application of Deep Neural Network for Real-Time Voice Command Recognition. *IEEE 3rd KhPI Week on Advanced Technology (KhPIWeek)*. Kharkiv, Ukraine, pp. 1–4. DOI: 10.1109/KhPIWeek57572.2022.9916473
7. Bondarenko, M. E., Ivashchenko, H. S. (2025). Using a sequence of preprocessing methods in voice identification systems. *Control, navigation and communication systems*. Poltava: NUPP. Vol. 2 (80), pp. 90–96. DOI: 10.26906/SUNZ.2025.2.090
8. Bondarenko, M. E., Ivashchenko, H. S. (2025). Parallel implementation of voice signal processing methods on multicore CPU and GPU. *Control, navigation and communication systems*. Poltava: NUPP. Vol. 4 (82), pp. 39–44. DOI: 10.26906/SUNZ.2025.4.39-44
9. Gjermundsen, A. (2010). CPU and GPU Co-processing for Sound. MA thesis. *Department of Computer and Information Science, Norwegian University of Science and Technology*. Norwegian University of Science and Technology. 173 p.
10. Momcilovic, S., Ilic, A., Roma, N., Sousa, L. (2014). Dynamic Load Balancing for Real-Time Video Encoding on Heterogeneous CPU+GPU Systems. *IEEE Transactions on Multimedia*. Vol. 16, no. 1, pp. 108–121. DOI: 10.1109/TMM.2013.2284892
11. Kim, J., Lane, I. (2014). Accelerating Large Vocabulary Continuous Speech Recognition on Heterogeneous CPU-GPU Platforms. *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Florence, Italy, pp. 3291–3295. DOI: 10.1109/ICASSP.2014.6854209
12. Kossaifi, J., Walecki, R., Panagakis, Y., Shen, J., Schuller, B., Pantic, M. (2021). SEWA DB: A Rich Database for Audio-Visual Emotion and Sentiment Research in the Wild. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol. 43, no. 3, pp. 1022–1040. DOI: 10.1109/TPAMI.2019.2944808

Дата першого надходження рукопису до видання: 13.11.2025
Дата прийнятого до друку рукопису після рецензування: 10.12.2025
Дата публікації: 31.12.2025