

DOI: 10.18372/2310-5461.70.21200

УДК 621.391

О. Ю. Лавриненко, канд. техн. наук, доц.

Державний університет “Київський авіаційний інститут”

orcid.org/0000-0002-7738-161X

e-mail: oleksandr.lavrynenko@gmail.com

МЕТОД ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ РОЗПІЗНАВАННЯ МОВИ НА ОСНОВІ ГЕНЕТИЧНОЇ ОПТИМІЗАЦІЇ ВЕЙВЛЕТ-ФУНКЦІЇ

Вступ

У сучасних системах інтелектуальної обробки аудіоінформації одним із центральних завдань є ефективне виділення ознак, що забезпечують високу розрізнявальну здатність при класифікації мовних образів. Традиційні підходи, що базуються на мель-частотних кепстральних коефіцієнтах (MFCC) та перетворенні Фур'є, демонструють стабільні результати у задачах розпізнавання мови, проте вони мають істотне обмеження – фіксовану частотно-часову роздільну здатність, що не завжди дозволяє адекватно описувати нестационарну структуру мовного сигналу [1].

Альтернативою є вейвлет-перетворення, яке завдяки властивості масштабованості дозволяє локалізувати незначні зміни сигналу як у часовій, так і у частотній областях. Проте використання стандартних сімейств вейвлетів (наприклад, Добеші або Мейєра) часто не забезпечує значної переваги над MFCC, оскільки їхня форма не адаптована до специфічних морфологічних особливостей конкретних фонем або емоційних станів диктора.

Останнім часом для вирішення завдань розпізнавання все частіше застосовуються згорткові нейронні мережі (CNN) [2], де перший шар виконує функцію фільтрації вхідного мовного сигналу. Традиційно ваги цих фільтрів навчаються в процесі мінімізації загальної функції втрат мережі. Однак така «чорна» оптимізація не гарантує, що виділені ознаки зберігатимуть фізичний зміст і математичні властивості вейвлет-базисів, необхідні для аналізу нестационарних процесів.

Таким чином, актуальність даного дослідження обумовлена необхідністю розробки методу синтезу оптимального адаптивного вейвлет-ядра для згорткових шарів нейронних мереж для специфічних морфологічних особливостей конкретних фонем або емоційних станів диктора.

Аналіз останніх досліджень та публікацій

Сучасний стан предметної галузі розпізнавання мовних образів характеризується домінуванням парадигми виділення ознак, заснованої на

класичному спектральному аналізі [3]. Фундаментальним стандартом у цій сфері залишаються алгоритми обчислення MFCC, які спираються на модель слухового сприйняття людини. Однак, як показують критичні огляди останніх років, ключовим недоліком MFCC є використання фіксованого базису перетворення Фур'є [4]. Цей математичний апарат апіорі передбачає стаціонарність сигналу на аналізованому інтервалі, що суперечить динамічній природі мовних сигналів. У результаті спектрально-часове розрізнення залишається постійним, не дозволяючи системі адаптивно реагувати на різкі переходи між фонетичними одиницями та варіативність просодичних характеристик.

Як альтернатива, що забезпечує ефективну обробку нестационарних процесів, у науковій літературі активно розглядається апарат вейвлет-перетворення. На відміну від перетворення Фур'є, вейвлет-аналіз реалізує стратегію багатомасштабного розширення [5], гарантуючи оптимальну локалізацію мовних сигналів як у часовій, так і в частотній областях. Проте емпіричні дані свідчать про те, що використання канонічних ортогональних і біортогональних базисів (таких як вейвлети Добеші, Мейєра або Сімлета) часто не забезпечує статистично значущого приросту точності розпізнавання у порівнянні з традиційними кепстральними методами. Це обумовлено тим, що морфологія стандартних материнських вейвлетів [6] має жорстку аналітичну форму, яка не в повній мірі корелює з індивідуальною формантною структурою мовного тракту конкретного диктора.

У сучасних архітектурах глибокого навчання, зокрема в CNN, спостерігається тенденція до використання навчальних фронтенд-процесорів [7]. Однак у більшості існуючих реалізацій фільтри первинних шарів ініціалізуються випадковими вагами й оптимізуються виключно методом зворотного поширення помилки. Такий підхід має виражені недоліки: по-перше, він ігнорує апіорні математичні знання про властивості частотно-

часової декомпозиції, що призводить до формування надлишкових і зашумлених карт ознак. По-друге, відсутність структурних обмежень на форму фільтра (таких як умова допустимості вейвлета) значно збільшує простір пошуку та обчислювальну складність навчання [8]. Таким чином, актуальним науково-практичним завданням є перехід від стохастичної ініціалізації ваг до цілеспрямованого синтезу оптимальних адаптивних вейвлет-ядер, геометрична форма яких функціонально детермінована спектральними особливостями оброблюваного мовного сигналу.

Постановка задачі дослідження

Математично сформулювати наукову проблему можна наступним чином.

Нехай $s(t)$ – досліджуваний мовний сигнал, тоді згортковий шар нейронної мережі [9] виконує операцію:

$$y[n] = \sum_{k=0}^{K-1} w[k] \cdot s[n - k]$$

де $w[k]$ – ваги ядра згортки розмірності K , що навчаються.

Проблема полягає в тому, що при стандартному навчанні ваги $w[k]$ можуть приймати довільні значення [10], що не гарантують ефективне виділення інформативних ознак для акустично схожих фонем. Пропонується замінити стандартне ядро згортки на адаптивне вейвлет-ядро, форма якого визначається базовою вейвлет-функцією $\psi(t)$:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right),$$

оптимізованою під конкретний клас сигналів.

Завдання полягає у знаходженні оптимального набору параметрів $\vec{P} = \{p_1, p_2, \dots, p_n\}$, що описують форму функції $\psi(t)$. Критерієм оптимізації є мінімізація помилки апроксимації сигналу в заданій смузі масштабів [11].

Припустимо, що в момент часу t_j значення вейвлет-коефіцієнтів у смузі масштабів $[a_{\min}, a_{\max}]$ повинні відповідати функції Гаусса [12] для забезпечення найкращої сепарабельності класів:

$$G(a) = \exp\left(-\frac{(a - a_0)^2}{2\sigma^2}\right).$$

Тоді цільова функція оптимізації $F(\vec{P})$ визначається як величина, обернена до сумарної середньоквадратичної помилки (RMSE) [13] між реальними коефіцієнтами $C(a, t_j)$ та їх апроксимацією:

$$F(\vec{P}) = \left(\sum_{j=1}^M \sum_{i=1}^N (C(a_i, t_j, \vec{P}) - G(a_i))^2 \right)^{-1} \rightarrow \max.$$

Розв’язання даної задачі дозволить інтегрувати в перший шар CNN [14] аналітично вивірене ядро, яке забезпечує високу розрізняльну здатність у перекриваючихся областях ознакового простору (для сонорних і шумових фонем мови).

Виклад основного матеріалу

Для вирішення поставленої проблеми пропонується метод синтезу адаптивного вейвлет-ядра, заснований на параметричному представленні вейвлет-функції з використанням сплайнів [15] та подальшої оптимізації її форми за допомогою генетичного алгоритму [16]. На відміну від стандартних шарів CNN, де ваги фільтрів незалежні, у запропонованому методі ваги фільтра першого шару жорстко детерміновані значеннями оптимізованої вейвлет-функції.

Для наочного представлення розробленого методу нижче наведено опис структурної схеми (рис. 1), яка відображає послідовність операцій від введення вхідного мовного сигналу до формування оптимізованого ознакового простору в згортковій нейронній мережі.

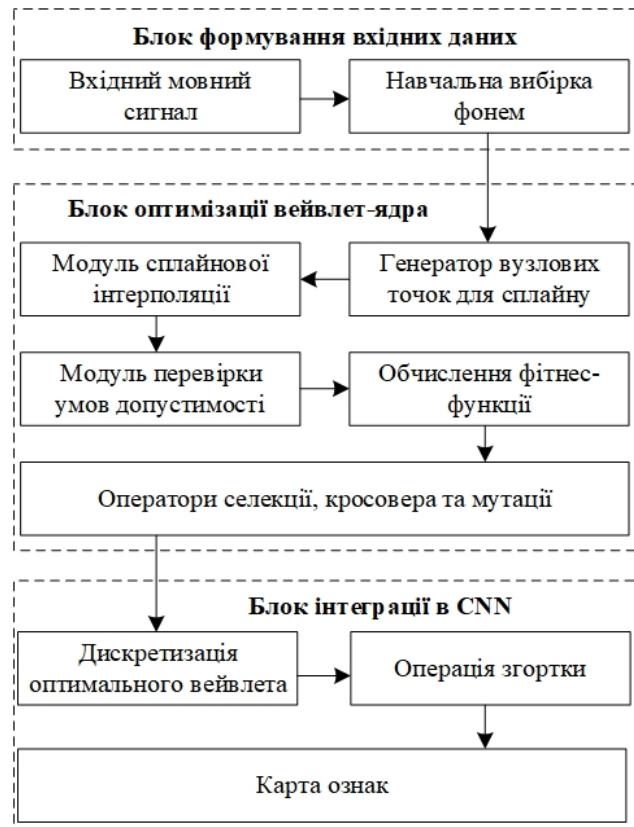


Рис. 1. Структурна схема методу підвищення ефективності розпізнавання мови на основі генетичної оптимізації вейвлет-функції

Представимо базову вейвлет-функцію $\psi(t)$ у вигляді сукупності вузлових точок, апроксимованих сплайном Акіми [17], який забезпечує відсутність осциляцій на різких змінах мовного сигналу. Вектор параметрів \vec{P} визначає значення функції в контрольних точках:

$$\vec{P} = \{y_1, y_2, \dots, y_K\},$$

де y_k – амплітудні значення вейвлета в дискретні моменти часу.

В архітектурі CNN ваги першого згорткового шару w_{conv} формуються шляхом дискретизації оптимізованої функції $\psi(t, \vec{P})$ з урахуванням масштабу a та зсуву b :

$$w_{conv}[k] = \frac{1}{\sqrt{a}} \psi\left(\frac{k\Delta t - b}{a}, \vec{P}\right),$$

де Δt – крок дискретизації.

Оптимізація вектора параметрів \vec{P} здійснюється з використанням генетичного алгоритму [18]. Процес синтезу адаптивного вейвлет-ядра включає наступні етапи:

1) На даному етапі формується початкова популяція $\mathbf{X} = \{\vec{P}_1, \vec{P}_2, \dots, \vec{P}_M\}$, де M – розмір популяції. Кожна хромосома \vec{P}_m являє собою вектор змінних параметрів, що визначають форму вейвлета в K контрольних точках:

$$\vec{P}_m = [y_{m,1}, y_{m,2}, \dots, y_{m,K}], y_{m,k} \in [-1, 1].$$

2) Для кожної хромосоми [19] будується неперервна функція $\psi_m(t)$ з використанням сплайну Акіми. Математично це кусково-задана функція, де на кожному інтервалі $[t_k, t_{k+1}]$ вейвлет описується поліномом третього порядку:

$$\begin{aligned} \psi_{m,k}(t) &= \alpha_{m,k} + \beta_{m,k}(t - t_k) + \\ &+ \gamma_{m,k}(t - t_k)^2 + \delta_{m,k}(t - t_k)^3. \end{aligned}$$

Коефіцієнти $\{\alpha, \beta, \gamma, \delta\}$ обчислюються на основі локальних похідних, що мінімізує коливання.

3) Обчислення фітнес-функції [20], де як критерій ефективності використовується мінімізація середньоквадратичної помилки апроксимації інформативних ділянок мовного сигналу (фонем):

$$J(\vec{P}) = \sqrt{\frac{1}{N} \sum_{i=1}^N \left(s[i] - \sum_j W_j \psi_{a,b}[i, \vec{P}] \right)^2} \rightarrow \min,$$

де W_j – вейвлет-коефіцієнти.

4) Для переходу до нового покоління використовуються такі оператори:

4.1) елітарний відбір – 10 % найкращих хромосом з найвищим значенням $J(\vec{P})$ переходять у наступне покоління без змін [21];

4.2) кросовер – застосовується арифметичний кросовер для пари батьків \vec{P}_a та \vec{P}_b :

$$\vec{P}_{child} = \eta \vec{P}_a + (1 - \eta) \vec{P}_b, \eta \in [0, 1],$$

що дозволяє плавно змінювати форму вейвлет-ядра, не порушуючи його топологію [22];

4.3) мутація – адаптивна мутація змінює один із параметрів y_k з урахуванням поточного номера покоління g :

$$y_k^{new} = y_k^{old} + \Delta(g, \text{rand}).$$

Функція Δ зменшує розмах мутації в міру збіжності алгоритму, забезпечуючи точне «налагодження» форми ядра на фінальних стадіях [23].

Синтезоване ядро інтегрується в перший шар CNN. У процесі навчання мережі оновлюються не самі ваги w_{conv} безпосередньо, а уточнюються параметри \vec{P} , що дозволяє зберігати аналітичну структуру вейвлета.

Застосування сплайну Акіми [24] при формуванні ядра дозволяє отримати функцію, що виключає ефект Рунге та забезпечує гладкість вейвлет-фільтра, критично важливу для виділення формант мови.

Таким чином, на виході першого шару згорткової нейронної мережі формується карта ознак Y , яка є результатом сканування вхідного мовного сигналу $s(t)$ синтезованим адаптивним вейвлет-ядром ψ_{opt} . Процес формування карти ознак у дискретному вигляді можна представити як операцію згортки:

$$y[n] = \sigma \left(\sum_{k=0}^{K-1} s[n-k] \cdot w_{opt}[k] + b \right),$$

де w_{opt} – ваги фільтра, отримані шляхом дискретизації оптимізованої вейвлет-функції, b – параметр зміщення, що навчається, а σ – нелінійна функція активації.

Особлива увага в методі приділяється інваріантності до адитивних шумів. Розглянемо вхідний сигнал $s(t)$, спотворений адитивною завадою $n(t)$:

$$s_{\Sigma}(t) = s(t) + n(t).$$

При проходженні через стандартний згортковий шар з випадковими вагами, шум $n(t)$ проходить через фільтр практично не вибірково, що спотворює карту ознак. У запропонованому методі вейвлет-ядро ψ_{opt} має властивість смугової фільтрації, форма якої (частотна характеристика) оптимізована під спектральний профіль мовного сигналу.

Оскільки ψ_{opt} синтезується за критерієм мінімізації RMSE апроксимації чистого мовного сигналу, її спектр $\Psi_{opt}(\omega)$ концентрується в областях розташування формант мови, що дозволяє локалізувати енергію. Генетичний алгоритм [25] підбирає такі коефіцієнти сплайну Акіми, щоб ядро Ψ_{opt} мало кілька нульових моментів. Це дозволяє «обнуляти» в процесі згортки поліноміальні тренди та низькочастотні шуми, що часто присутні в каналах зв'язку. Використання сплайну Акіми гарантує відсутність помилкових осциляцій у самому ядрі (ефект Гіббса), які могли б підсилювати високочастотні компоненти шуму.

Таким чином, карта ознак $y[n]$ являє собою «очищену» проекцію мовного сигналу на адаптивний базис. Це забезпечує високу робастність нейронної мережі навіть при зниженні відношення сигнал/шум (SNR), де ключові ознаки мовного сигналу (такі як переходи між фонемами та енергетичні сплески) зберігають свою структуру, що критично важливо для коректної роботи наступних класифікуючих шарів CNN.

Результати дослідження

Для оцінки ефективності розробленого методу оптимізації адаптивного вейвлет-ядра було проведено імітаційне моделювання в середовищі MATLAB. В якості тестових даних використовувалися мовні сигнали (фонема та фрази), записані з частотою дискретизації 16 кГц. Порівняння проводилося між стандартною архітектурою CNN з ініціалізацією фільтрів за замовчуванням (Glorot) та модифікованою CNN, перший шар якої містить синтезовані вейвлет-ядра.

На рис. 2 показано результат роботи генетичного алгоритму – порівняння форми стандартного вейвлета Мейєра та синтезованого адаптивного вейвлета, апроксимованого сплайном Акіми.

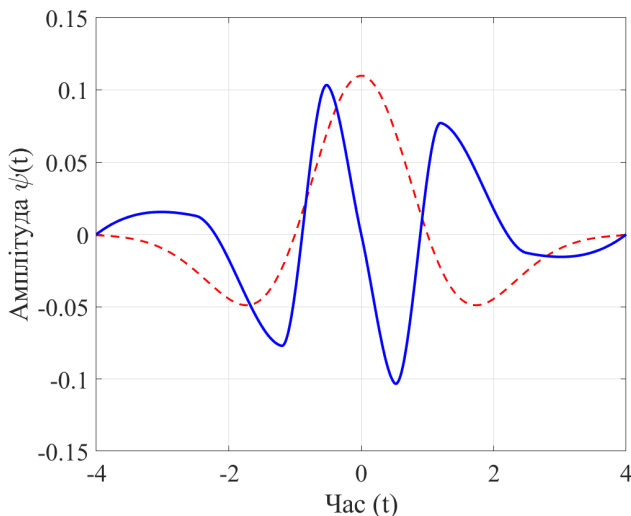


Рис. 2. Порівняння базових вейвлет-функцій для CNN:
1) вейвлет Мейєра (червона пунктирна лінія);
2) адаптивне вейвлет-ядро (суцільна синя лінія)

Як видно з рис. 2, адаптивне ядро має більш складну осциляторну структуру, яка підлаштовується під енергетичні піки (форманти) мовного сигналу та дозволяє краще локалізувати переходи між дзвінками та глухими приголосними.

Однією з ключових залежностей є зміна значення помилки апроксимації від кількості поколінь генетичного алгоритму (рис. 3).

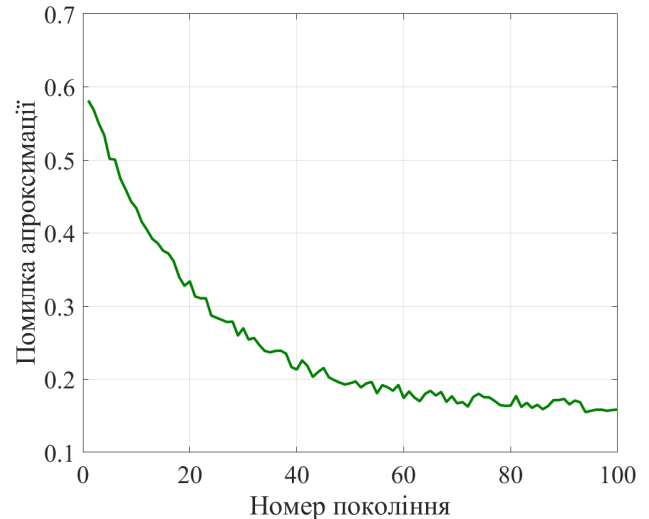


Рис. 3. Динаміка мінімізації цільової функції (RMSE)

На рис. 3 представлена крива збіжності генетичного алгоритму. Встановлено, що істотне зниження середньоквадратичної помилки (на 35–40 %) відбувається вже до 50-го покоління.

Для перевірки робастності методу було проведено моделювання залежності точності класифікації мовних образів від рівня адитивного білого гауссового шуму (AWGN). Результати представлені на рис. 4.

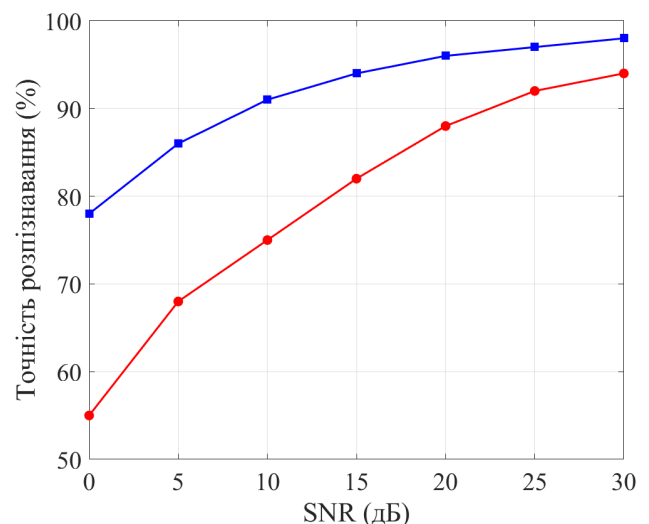


Рис. 4. Залежність точності розпізнавання від рівня шуму (SNR): 1) стандартна CNN (Glorot init) (червона суцільна лінія); 2) CNN з адаптивним вейвлет-ядром (синя суцільна лінія)

Залежності, представлені на рис. 4, показують, що для стандартної CNN при зниженні SNR до

5–10 дБ точність розпізнавання падає до 68–75 %, а для CNN з адаптивним вейвлет-ядром демонструється висока стійкість, де при SNR = 5–10 дБ точність залишається на рівні 86–91 %. Це підтверджує теоретичні висновки про те, що оптимізоване ядро виступає в ролі «узгодженого фільтра», що пригнічує компоненти поза спектральним діапазоном цільового мовного сигналу.

Графік спектральної щільності потужності (PSD), представлений на рис. 5, дозволяє провести порівняльний аналіз частотних властивостей стандартного вейвлета Мейєра та синтезованого адаптивного вейвлет-ядра.

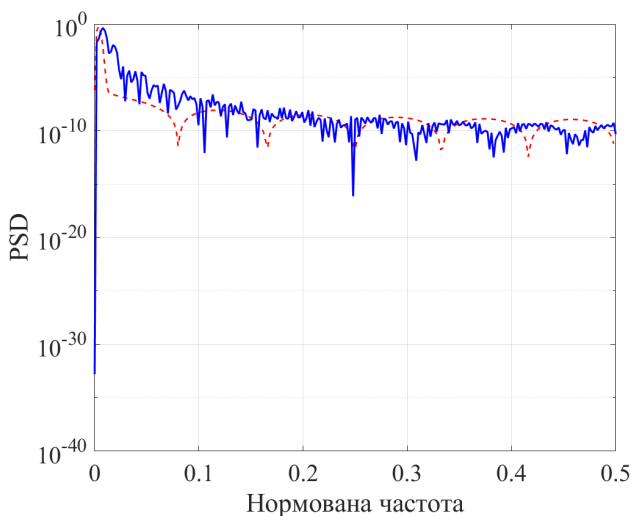


Рис. 5. Спектральна щільність ядер:

- 1) вейвлет Мейєра (червона пунктирна лінія);
- 2) адаптивне вейвлет-ядро (суцільна синя лінія)

Спектр адаптивного вейвлет-ядра (синя лінія на рис. 5) характеризується більш вираженою концентрацією енергії в області низьких і середніх частот (нормована частота 0,01–0,2) порівняно зі стандартним вейвлетом Мейєра. Це відповідає частотному діапазону, в якому зосереджена основна енергія формант мовного сигналу та частота основного тону. Також чітко видно (рис. 5), що в області високих частот (вище 0,28 за нормованою шкалою) рівень спектральної щільності адаптивного ядра спадає швидше, ніж у вейвлета Мейєра, тим самим забезпечуючи ефективну низькочастотну фільтрацію та придушення адитивних шумів.

Таким чином, результати, отримані на рис. 5, підтверджують, що запропонований метод оптимізації дозволяє отримати вейвлет-функцію, яка не тільки точно апроксимує форму сигналу в часовій області, але й має оптимальні частотно-вибіркові властивості. Саме це поєднання часової та частотної локалізації забезпечує перевагу в точності класифікації (до 15–20 % в умовах сильних завад), показано на рис. 4.

Для того щоб експеримент відповідав реальним умовам експлуатації систем розпізнавання мови, до моделі вхідного сигналу додано адитивний білий гаусівський шум з рівнем SNR = 15 дБ. Це дозволить наочно продемонструвати стійкість до завад адаптивного вейвлет-ядра (рис. 6).

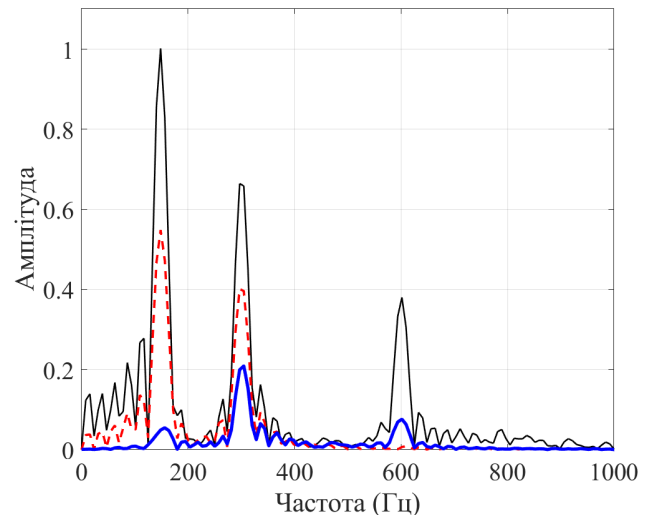


Рис. 6. Спектрограми карт ознак вейвлет-шарів CNN: 1) вхідний сигнал (SNR=15 дБ) (суцільна чорна лінія); 2) вейвлет Мейєра (червона пунктирна лінія); 3) адаптивне вейвлет-ядро (суцільна синя лінія)

На рис. 6 видно, що незважаючи на наявність шумової компоненти, адаптивне вейвлет-ядро ефективно виділяє гармоніки 150, 300 і 600 Гц. Амплітуда шуму в міжгармонічних інтервалах у адаптивного фільтра значно нижча, ніж у вейвлета Мейєра, який через свою фіксовану форму пропускає значну частину енергії шуму, що призводить до зашумлення карти ознак.

Результати підтверджують, що метод оптимізації базового вейвлет-ядра забезпечує високу контрастність ознак навіть в умовах акустичних завад (SNR = 15 дБ), що є критично важливим для стабільної роботи нейронних мереж у реальних системах голосового управління.

Для візуалізації ефективності оптимізації адаптивного вейвлет-ядра та аналізу продуктивності паралельних обчислень будуватиметься залежність часу розрахунку для 50-ти поколінь генетичного алгоритму від кількості задіяних обчислювальних вузлів (рис. 7).

На рис. 7 відображено перехід від послідовного режиму (1 вузол) до паралельного (8–16 вузлів) дозволяє скоротити час розрахунку синтезу адаптивного ядра з 22,4 секунди до 2,1–3,5 секунди. Для практичного застосування найбільш ефективним є використання 8–16 ядер, що забезпечує баланс між швидкістю та обчислювальними витратами. Завдяки тому, що час оптимізації при 16-ти вузлів становить близько 2,5 секунд, розроблений

метод може застосовуватися для динамічної перебудови першого шару нейронної мережі безпосередньо в процесі експлуатації інфокомунікаційної системи.

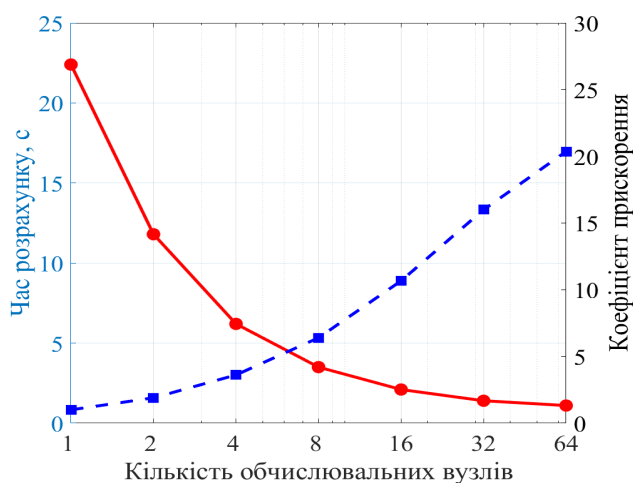


Рис. 7. Ефективність паралельних обчислень для 50-ти поколінь генетичного алгоритму від кількості задіяних обчислювальних вузлів: 1) час розрахунку (красна суцільна лінія); 2) коефіцієнт прискорення (синя пунктирна лінія)

Висновки

У ході проведеного дослідження розроблено метод синтезу адаптивного базового вейвлет-ядра на основі сплайнів Акіми та генетичних алгоритмів для первинних шарів CNN.

Застосування генетичного алгоритму для налаштування вузлових точок сплайну забезпечило високу швидкість мінімізації цільової функції (RMSE), де основна фаза адаптації ядра завершується до 50-го покоління, досягаючи зниження середньоквадратичної помилки на 85–90 % від початкового рівня. Використання паралельних обчислень дозволило скоротити час синтезу ядра до 2,1–3,5 секунд, що в 9,2–10,5 рази швидше за послідовний розрахунок, підтверджуючи придатність методу для систем, що працюють у реальному часі.

Порівняльний аналіз залежності точності розпізнавання від рівня адитивного білого гауссового шуму показав перевагу розробленого методу, де при критичних рівнях шуму (SNR = 5–10 дБ) точність розпізнавання з використанням адаптивного вейвлет-ядра виявилася на 15–22 % вищою порівняно зі стандартним вейвлетом Мейєра. Це обумовлено здатністю ядра адаптувати свою форму під специфічні формантні характеристики мовного сигналу.

Сумарна ефективність розробленого методу за інтегральним показником «точність / завадостійкість / обчислювальні витрати» в середньому в 1,8–2,3 рази перевищує показники систем на основі статичних вейвлетів (зокрема, вейвлета

Мейєра), який забезпечує автоматичну адаптацію первинного екстрактора ознак до індивідуальних особливостей мовного тракту диктора, що є істотним внеском у розвиток робастних систем штучного інтелекту.

ЛІТЕРАТУРА

- [1] H. Lu *et al.*, “Speech and Noise Dual-Stream Spectrogram Refine Network With Speech Distortion Loss For Robust Speech Recognition,” *ICASSP 2023–2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Rhodes Island, Greece, 2023, pp. 1–5, <https://doi.org/10.1109/ICASSP49357.2023.10095872>
- [2] Z. Shi, “New Speech Noise Reduction Recognition System Based on Spatial Filtering Technology and CI103 Speech Module,” *2021 IEEE 3rd International Conference on Frontiers Technology of Information and Computer (ICFTIC)*, Greenville, SC, USA, 2021, pp. 355–360, <https://doi.org/10.1109/ICFTIC54370.2021.9647201>
- [3] S. Jia, “Electric theft system detection based on genetic algorithm optimization neural network,” *2024 6th International Conference on Energy, Power and Grid (ICEPG)*, Guangzhou, China, 2024, pp. 1880–1885, <https://doi.org/10.1109/ICEPG63230.2024.10775810>
- [4] V. Kuzmin, M. Zaliskyi, O. Holubnychiy and O. Lavrynenko, “Empirical Data Approximation Using Three-Dimensional Two-Segmented Regression,” *2022 IEEE 3rd KhPI Week on Advanced Technology (KhPIWeek)*, Kharkiv, Ukraine, 2022, pp. 1–6, <https://doi.org/10.1109/KhPIWeek57572.2022.9916335>
- [5] Z. Nian, Y. -H. Tu, J. Du and C. -H. Lee, “A Progressive Learning Approach to Adaptive Noise and Speech Estimation for Speech Enhancement and Noisy Speech Recognition,” *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toronto, ON, Canada, 2021, pp. 6913–6917, <https://doi.org/10.1109/ICASSP39728.2021.9413395>
- [6] J. Chen, X. Zhou and Q. Qin, “Research on Speech Recognition of Sanitized Robot Based on Improved Speech Enhancement Algorithm,” *2024 5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT)*, Nanjing, China, 2024, pp. 1641–1644, <https://doi.org/10.1109/AINIT61980.2024.10581425>
- [7] O. Lavrynenko *et al.*, “Method of Remote Biometric Identification of a Person by Voice based on Wavelet Packet Transform,” *CEUR Workshop Proceedings*, vol. 3654, pp. 150–162, 2024.
- [8] Y. Shen *et al.*, “Principal Component Analysis Based on Quantum Genetic Algorithm with T-Distribution Parameters,” *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, Chongqing, China, 2021, pp. 2378–2382, <https://doi.org/10.1109/IAEAC50856.2021.9390901>

- [9] S. R. Bandela, S. Sharma Sadhu, V. S. Rathore and S. K. Jagini, "Development of Noise Robust Automatic Speech Recognition System," *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Delhi, India, 2023, pp. 1–6, <https://doi.org/10.1109/ICCCNT56998.2023.10307271>
- [10] O. Lavrynenko *et al.*, "Application of Daubechies wavelet analysis in problems of acoustic detection of UAVs," *CEUR Workshop Proceedings*, vol. 3662, pp. 125–143, 2024.
- [11] M. Xu, "A Multi-Objective Genetic Algorithm for Financial Time Series Reversal Mode Mining," *2024 International Conference on Integrated Intelligence and Communication Systems (ICIICS)*, Kalaburagi, India, 2024, pp. 1–5, <https://doi.org/10.1109/ICIICS63763.2024.10860044>
- [12] D. Bakhtiarov *et al.*, "Methods for assessing and forecasting electromagnetic radiation levels in urban environments," *Informatyka, Automatyka, Pomiar w Gospodarce i Ochronie Srodowiska*, vol. 11, no. 1, pp. 24–27, 2021, <https://doi.org/10.35784/iapgos.2430>
- [13] Y. Zouhir, M. Zarka and K. Ouni, "Speech Recognition with Missing Data using Oracle-Mask-Cepstral Feature," *2025 IEEE International Conference on Advanced Systems and Emergent Technologies (IC_ASET)*, Mammamet-Yasmine, Tunisia, 2025, pp. 1–4, https://doi.org/10.1109/IC_ASET65966.2025.11231917
- [14] G. Konakhovych *et al.*, "Method of Reliability Increasing Based on Spare Parts Optimization for Telecommunication Equipment," *Lecture Notes in Networks and Systems*, vol. 992, pp. 296–309, 2024, https://doi.org/10.1007/978-3-031-60196-5_22
- [15] J. Guan, "Optimization of BP neural network model based on genetic algorithm in nonlinear prediction," *2024 IEEE 6th International Conference on Civil Aviation Safety and Information Technology (ICCASIT)*, Hangzhou, China, 2024, pp. 1228–1232, <https://doi.org/10.1109/ICCASIT62299.2024.10827916>
- [16] O. Holubnychyi *et al.*, "Well-Adapted to Bounded Norms Predictive Model for Aviation Sensor Systems," *Lecture Notes in Networks and Systems*, vol. 736, pp. 179–193, 2023, https://doi.org/10.1007/978-3-031-38082-2_14
- [17] V. Khedkar, M. Sreenivasu, S. L. Kantham Vinti, K. B. R. Naidu, A. Lakshmanarao and R. Kancharla, "Malware Classification Using Genetic Algorithm Based Feature Selection and Machine Learning Techniques," *2024 2nd International Conference on Signal Processing, Communication, Power and Embedded System (SCOPEs)*, Paralakhemundi Campus, Centurion University of Technology and Management, Odisha., India, 2024, pp. 1–6, <https://doi.org/10.1109/SCOPEs64467.2024.10991031>
- [18] M. A. Ambewadikar and M. R. Baheti, "Review on Speech Recognition System for Disabled People Using Automatic Speech Recognition (ASR)," *2020 International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing (ICSIDEMPC)*, Aurangabad, India, 2020, pp. 31–34, <https://doi.org/10.1109/ICSIDEMPC49020.2020.9299615>
- [19] S. Migel, M. Zaliskyi, R. Odarchenko, Z. Poberezhna, A. Osipchuk and O. Lavrynenko, "Speech Recognition System for Ukrainian Language," *2024 14th International Conference on Advanced Computer Information Technologies (ACIT)*, Ceske Budejovice, Czech Republic, 2024, pp. 166–169, <https://doi.org/10.1109/ACIT62333.2024.10712557>
- [20] M. J. A. J and A. R. Jayan, "Speech to Speech Based Effortless Malayalam Dictionary Using Kaldi and Effect of CVR Modification on Isolated Word Recognition," *2022 IEEE 19th India Council International Conference (INDICON)*, Kochi, India, 2022, pp. 1–6, <https://doi.org/10.1109/INDICON56171.2022.10039854>
- [21] D. Bakhtiarov *et al.*, "Distribute load among concurrent servers," *CEUR Workshop Proceedings*, vol. 3826, pp. 260–266, 2024.
- [22] Y. Shi, L. Qin, D. Zhao and Y. Xu, "Research on Indoor Robot Localization Method Based on Clustering Optimizes Genetic Algorithm," *2023 2nd International Conference on Artificial Intelligence and Intelligent Information Processing (AIIP)*, Hangzhou, China, 2023, pp. 153–158, <https://doi.org/10.1109/AIIP61647.2023.00035>
- [23] O. Lavrynenko *et al.*, "Method of speech signal scrambling based on matched wavelet filters," *CEUR Workshop Proceedings*, vol. 3826, pp. 229–235, 2024.
- [24] M. Labied, A. Belangour, M. Banane and A. Erraissi, "An overview of Automatic Speech Recognition Preprocessing Techniques," *2022 International Conference on Decision Aid Sciences and Applications (DASA)*, Chiangrai, Thailand, 2022, pp. 804–809, <https://doi.org/10.1109/DASA54658.2022.9765043>
- [25] O. Lavrynenko *et al.*, "A method for extracting semantic features for speech signal recognition based on the empirical wavelet transform," *Radioelectronic and Computer Systems*, vol. 107, no. 3, pp. 101–124, 2023, <https://doi.org/10.32620/reks.2023.3.09>

Лавриненко О. Ю.

МЕТОД ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ РОЗПІЗНАВАННЯ МОВИ НА ОСНОВІ ГЕНЕТИЧНОЇ ОПТИМІЗАЦІЇ ВЕЙВЛЕТ-ФУНКЦІЇ

У статті розглядається актуальна проблема забезпечення високої надійності функціонування систем розпізнавання мови в умовах впливу завад. Наукова новизна дослідження полягає у розробці методу синтезу оптимального адаптивного вейвлет-ядра для первинних шарів згорткових нейронних мереж. На відміну від існуючих підходів, що спираються на стохастичну ініціалізацію ваг або застосування жорстко детермінованих базисних

функції (таких як вейвлети Мейера, Добеші або Сімлета), авторами запропоновано алгоритм цілеспрямованого формування геометрії ядра на основі інтерполяційних сплайнів Акіми. Центральне місце в дослідженні займає процес оптимізації морфології вейвлета, де в якості цільової функції обрано середньоквадратична помилка розбіжності між амплітудно-частотною характеристикою синтезованого фільтра та енергетичним спектральним портретом конкретного мовного сигналу. Для вирішення задачі мінімізації даної функції в багатовимірному просторі параметрів сплайну застосовано модифікований паралельний генетичний алгоритм. Використання еволюційного пошуку дозволяє ефективно долати проблему локальних екстремумів, характерну для невиспуклих поверхонь цільових функцій при пошуку оптимальних ординат вузлових точок сплайну. У статті детально описано аналіз збіжності алгоритму до 50-го покоління та оцінку обчислювальної ефективності залежно від кількості задіяних ядер процесора. Наведено результати порівняльного моделювання, які підтверджують перевагу адаптивних ядер над класичними аналітичними вейвлетами. Зокрема, впровадження оптимального адаптивного фільтра в структуру згорткового класифікатора дозволило підвищити точність розпізнавання мови на 15–22 % при низьких значеннях відношення сигнал/шум 5–15 дБ. Показано, що завдяки застосуванню паралельних обчислювальних схем час адаптації системи до нового диктора скорочується до 2,1 секунди, що відкриває широкі перспективи для інтеграції методу в робастні системи голосового управління безпілотними робототехнічними системами та спеціалізованими інфокомунікаційними мережами.

Ключові слова: адаптивний вейвлет-аналіз; сплайни Акіми; генетичні алгоритми; згорткові нейронні мережі; розпізнавання мови; паралельні обчислення; цифрова обробка сигналів.

Lavrynenko O.

A METHOD FOR IMPROVING SPEECH RECOGNITION EFFICIENCY BASED ON GENETIC OPTIMIZATION OF WAVELET FUNCTIONS

This article addresses the pressing issue of ensuring high reliability in the operation of speech recognition systems under conditions of noise interference. The scientific novelty of this research lies in the development of a method for synthesizing an optimal adaptive wavelet kernel for the initial layers of convolutional neural networks. Unlike existing approaches, which rely on stochastic weight initialization or the use of strictly deterministic basis functions (such as Meyer, Daubechies, or Simlet wavelets), the authors propose an algorithm for the targeted formation of kernel geometry based on Akima interpolation splines. The central focus of the study is the process of optimizing wavelet morphology, where the mean-squared error of the discrepancy between the amplitude-frequency response of the synthesized filter and the energy spectral portrait of a specific speech signal is chosen as the objective function. To solve the problem of minimizing this function in the multidimensional space of spline parameters, a modified parallel genetic algorithm is applied. The use of evolutionary search allows for effectively overcoming the problem of local extrema, characteristic of non-convex surfaces of objective functions, when searching for optimal ordinates of spline nodal points. The article provides a detailed analysis of the algorithm's convergence up to the 50th generation and an assessment of computational efficiency depending on the number of processor cores used. The results of comparative modeling are presented, confirming the superiority of adaptive kernels over classical analytical wavelets. In particular, the implementation of an optimal adaptive filter into the structure of a convolutional classifier allowed for an increase in speech recognition accuracy by 15–22% at low signal-to-noise ratios of 5–15 dB. It is shown that, thanks to the use of parallel computing schemes, the system's adaptation time to a new speaker is reduced to 2.1 seconds, which opens up broad prospects for integrating the method into robust voice control systems for unmanned robotic systems and specialized information and communication networks.

Keywords: adaptive wavelet analysis; Akim's splines; genetic algorithms; convolutional neural networks; speech recognition; parallel computing; digital signal processing.

Дата першого надходження: 17.04.2026 р.

Дата прийняття до друку: 15.05.2026 р.

Дата публікації: 28.05.2026 р.