УДК 004.89 **doi: 10.26906/SUNZ.2023.2.90-96**

М. Е. Бондаренко, Г. С. Іващенко

Харківський національний університет радіоелектроніки, Харків, Україна

ВИКОРИСТАННЯ ПОСЛІДОВНОСТІ МЕТОДІВ ПОПЕРЕДНЬОЇ ОБРОБКИ В СИСТЕМАХ ГОЛОСОВОЇ ІДЕНТИФІКАЦІЇ

Анотація. Актуальність. На сьогоднішній день, голосова ідентифікація є найпоширенішим біометричним методом, однак все ще існують певні технічні обмеження, пов'язані з впливом навколишнього середовища та варіаціями голосу, через що актуальні подальші дослідження. З розвитком технологій створення підроблених аудіозаписів виникає потреба вдосконалення існуючих методів голосової біометрії для забезпечення надійної ідентифікації, зокрема шляхом впровадження методів попередньої обробки голосових даних. **Метою даної роботи** є дослідження методів попередньої обробки в системах голосової ідентифікації людини. **Об'єктом дослідження** є модуль фільтрації голосових сигналів у системі голосової ідентифікації користувача. **Предметом дослідження** є методи попередньої обробки голосових сигналів. **Результати.** У даній роботі досліджується ефективність застосування послідовності методів попередньої обробки в системах голосової ідентифікації для зменшення впливу фонових шумів та інших викривлень на якість розпізнавання. Проведено порівняльний аналіз ефективності методів видалення сталих і динамічних шумів. **Висновок.** Запропоновано підхід організації процесу попередньої обробки голосових сигналів, що передбачає вибір послідовності методів з урахуванням таких параметрів, як якість запису, обчислювальні ресурси та необхідний рівень точності розпізнавання. Отримані результати демонструють, що послідовне використання методів шумозаглушення та нормалізації сигналу підвищує точність голосової ідентифікації.

Ключові слова: методи попередньої обробки, послідовне застосування методів, голосові сигнали, системи голосової ідентифікації, фільтрація, нормалізація, виділення характеристик, шум, спектральний аналіз

Вступ

Інформаційні технології змінили сучасне суспільство, забезпечивши глобальний доступ до хмарних сховищ, онлайн-банкінгу та відеоконференцій. Поширення використання мобільних пристроїв і персональних комп'ютерів сприяло розвитку цифрових технологій, але водночас підвищило небезпеку кіберзагроз [1]. Захист конфіденційних даних у хмарних сховищах та онлайн-сервісах став одним із головних завдань інформаційної безпеки.

Використання паролю залишається основним методом захисту конфіденційної інформації, проте зростання кількості онлайн-сервісів змушує користувачів запам'ятовувати дедалі більше унікальних комбінацій. Це часто призводить до створення слабких або повторюваних паролів, що підвищує ризик компрометації облікових записів і знижує загальний рівень безпеки даних користувача.

Використання біометричних методів суттєво спрощує процес ідентифікації [2]. Висока швидкість сканування біологічних характеристик та відсутність необхідності запам'ятовувати складні паролі роблять цю технологію найпоширенішим вибором для систем з високою пропускною здатністю, таких як задіяні у аеропортах та громадських установах [3, 4].

Серед широкого спектру біометричних методів ідентифікації, які базуються на унікальних фізіологічних та поведінкових характеристиках людини, голосова ідентифікація є одним з перспективних напрямків, що швидко розвивається [5]. Цей метод використовує аналіз звукових хвиль, через що має особливості та переваги, які роблять доцільним його застосування у різних сферах людської діяльності [6].

Створення універсальних систем голосової ідентифікації, що здатні адаптуватися до різних варіацій мовлення, вимагає детального аналізу голосових сигналів та їх акустичного середовища. Це включає дослідження фонового шуму, реверберації та інших викривлень. Шум є одним із основних чинників, що знижують точність таких систем, і може мати різне походження, наприклад, наявність навколишніх звуків у приміщенні або перешкод, які виникають під час передачі сигналу [7].

З метою підвищення точності аналізу аудіосигналів, що використовуються в системах голосової ідентифікації, застосовується попередня обробка. Мета даного етапу полягає у мінімізації впливу акустичних шумів та викривлень, які можуть негативно впливати на якість розпізнавання мовлення. У контексті голосової ідентифікації шум визначається як будь-який небажаний сигнал, що викривляє мовленнєвий сигнал та ускладнює процес автоматичного розпізнавання. Джерела шуму можуть бути як зовнішніми (зумовлені акустичним середовищем), так і внутрішніми (спричинені обмеженнями апаратного забезпечення) [8]. Проблема організації попередньої обробки й вибору методів для неї широко представлена у сучасних наукових дослідженнях. Методи попередньої обробки поділяються на статичні та динамічні, які характеризуються специфічними особливостями та застосовуються з урахуванням конкретних завдань та властивостей наданого аудіосигналу.

Статичні методи (методи кепстрального та спектрального аналізу, аналіз нульових переходів та енергії), засновані на аналізі спектральних характеристик голосового сигналу, є ефективними для виділення статичних характеристик мовця, але потребують додаткової обробки для врахування динамічних аспектів мовлення [9]. Вони дозволяють ефективно розділити аудіозапис на окремі мовні відрізки та зменшити вплив шумів. Статичні методи забезпечують достатню точність для задач, таких як сегментація чистого мовлення, але можуть бути неефективними при видаленні складних шумів, таких як музичний шум або шум, корельований з сигналом [10].

Динамічні методи (засновані на прихованих Марківських ланцюгах, методи засновані на штучних нейронних мережах, на авторегресивних моделях та на хвильових перетвореннях) враховують часову структуру сигналу, що дозволяє аналізувати складні нелінійні процеси, які відбуваються при мовленні.

Дослідження [11] описує розробку програмного забезпечення, що реалізує етап попередньої обробки голосових сигналів. Цей етап здійснюється через модулі наповнення, виявлення голосової активності (VAD), нормалізації, шумозаглушення, фільтрації та квантування. Застосування VAD дозволяє ідентифікувати мовленнєві фрагменти сигналу та виключати інтервали тиші або фонового шуму, що підвищує ефективність подальшої обробки. Нормалізація використовується для вирівнювання рівня гучності між різними записами та їх окремими частинами. Фільтрація здійснюється для виділення частотних компонентів, що містять найбільш значущу інформацію для розпізнавання мовлення. Методи шумозаглушення базуються на застосуванні Швидкого Перетворення Фур'є (FFT). Водночас, для досягнення оптимальних результатів часто необхідна комбінація FFT з іншими підходами до аналізу та обробки голосових сигналів, а також ретельне налаштування параметрів системи [12].

У дослідженні [13] описується застосування статистичних критеріїв для голосової ідентифікації та сегментації мовних сигналів з метою виділення мовних та немовних фрагментів. Рішення базується на методі оцінки часових характеристик (STE) – це метод у обробці мовлення, спрямований на визначення часових характеристик звукових сигналів, таких як початок і кінець звуків, тривалість пауз тощо. Метод дозволяє не тільки знайти мовні фрагменти мовленнєвого запису, а й нормалізувати швидкість записаного мовлення до заданого значення та розбивати запис на окремі фонеми.

У дослідженні [14] для зменшення переривчатості мовного сигналу на початку і в кінці кожного кадру пропонується використовувати віконний метод обробки кожного кадру сигналу для збільшення кореляції мел-частотних кепстральних коефіцієнтів (MFCC).

В ході аналізу результатів існуючих досліджень, зроблено висновок, що результативність описаних підходів наразі є недостатньою, та існують ще не задіяні напрямки вдосконалення поширених методів попередньої обробки, що потребують окремого детального експериментального дослідження.

Метою цієї роботи є дослідження послідовного застосування методів попередньої обробки мовних записів для підвищення точності роботи систем голосової ідентифікації. Вибір методів попередньої обробки потребує аналізу впливу різних типів шумових завад та інших факторів викривлення на якість записів та на їх подальшу обробку.

Постановка задачі

Голосовий сигнал можна представити як часовий ряд, де кожне окреме значення відповідає амплітуді сигналу в певний момент часу (рис. 1), $x(t) = [x_1, x_2, \dots, x_N]$, де $x(t)$ – вектор, що представляє голосовий сигнал, x_i – значення амплітуди сигналу в момент часу i, N– загальна кількість відліків сигналу.

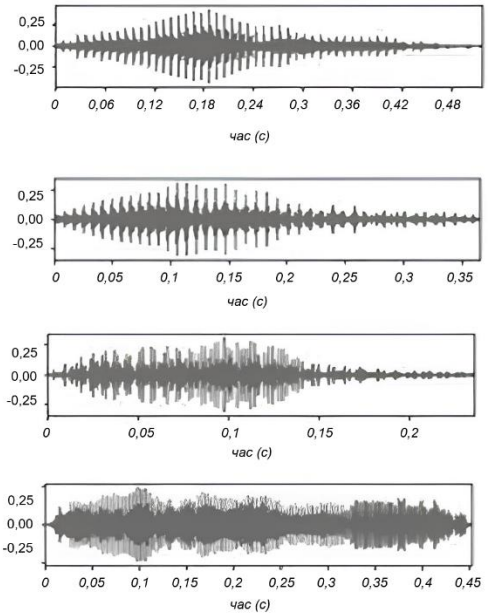

Рис. 1. Спектральний вигляд аудіозаписів чотирьох мовців, котрі промовляють однаковий текст з додаванням шумових ефектів

Дослідження передбачає використання методів попередньої обробки, таких як видалення шуму, нормалізація та фільтрація. Необхідно проаналізувати ефективність різних методів попередньої обробки та їх поєднання у системах голосової ідентифікації.

З метою дослідження стійкості системи голосової ідентифікації до різних типів викривлень було сформовано спеціалізовану базу даних, яка включала 4900 аудіозаписів [15, 16]. Дана база містить записи мовлення 10 різних дикторів, які вимовляли одне й те саме слово з варіаціями швидкості мовлення в діапазоні від 0,5 до 2,0.

Крім того, аудіозаписи були виконані в умовах різних рівнів акустичних завад для моделювання впливу шумового середовища.

Для оцінки ефективності системи в умовах відсутності мовленнєвого сигналу або під час обробки мовлення, що належить іншому диктору, до бази даних також були додані відповідні типи записів. Такий підхід дозволив всебічно проаналізувати стійкість системи до різних сценаріїв реального використання та оцінити її здатність коректно функціонувати за умов змінної акустичної обстановки.

Рішення задачі попередньої обробки голосових сигналів

У процесі попередньої обробки голосового сигналу застосовуються методи зниження шуму, нормалізації та фільтрації. Зокрема, використовуються підходи спектрального віднімання та фільтра Вінера, а також динамічні методи, такі як LMS-алгоритм, Гаар-вейвлет і Морлет-вейвлет, що дозволяють ефективно покращити якість сигналу перед його подальшим аналізом.

У залежності від характеру шумів та викривлень зазначені методи можуть застосовуватися як окремо, так і в комбінації, забезпечуючи багатоступеневу обробку сигналу. Мультимодальний підхід, що передбачає комбінацію різних методів попередньої обробки, дозволяє адаптувати систему до широкого спектру умов запису та підвищити її стійкість до різноманітних типів шумів (рис. 2).

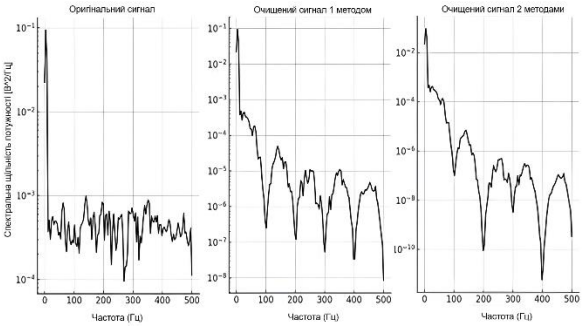

Рис. 2. Спектральний вигляд голосового фрагменту до обробки й після обробки одним та двома методами

Першим етапом є пригнічення сталих шумів за допомогою спектрального віднімання та фільтру Вінера, що забезпечує очищення сигналу від небажаних компонентів.

Спектральне віднімання базується на припущенні, що шум має постійний спектр, який можна оцінити під час пауз у мовленні:

$$S(f) = X(f) - N(f), \quad (1)$$

де $S(f)$ – спектр очищеного сигналу, $X(f)$ – спектр вихідного сигналу, $N(f)$ – оцінка спектра шуму.

Фільтр Вінера здійснює фільтрацію на основі статистичних властивостей як корисного сигналу, так і шуму. Його робота вимагає оцінки відповідних характеристик, зокрема розрахунку автокореляційних функцій та спектрів потужності. Використовуючи відомі статистичні характеристики, розраховується оптимальний фільтр, який мінімізує середньоквадратичну похибку. Формула для фільтра Вінера в частотній області має вигляд:

$$H(f) = \frac{S_{xx}(f)}{S_{xx}(f)+S_{nn}(f)}, \quad (2)$$

де $H(f)$ – передаточна функція фільтра Вінера, $S_{xx}(f)$ – спектр потужності корисного сигналу, $S_{nn}(f)$ – спектр потужності шуму.

Наступним етапом є нормалізація, яка вирівнює амплітудні та частотні характеристики сигналу. Амплітудна нормалізація здійснюється шляхом ділення значень сигналу на його максимальну амплітуду, що дозволяє уникнути перевищення динамічного діапазону. Частотна нормалізація виконується з використанням технік фільтрації, таких як високочастотні та низькочастотні фільтри, що дозволяє видалити небажані частоти.

Фільтрація, що є наступним кроком, видаляє специфічні небажані частотні компоненти сигналу. На цьому етапі застосовуються смугові фільтри, які дозволяють виділити найбільш інформативні частотні компоненти для задачі розпізнавання:

$$H(f) = \begin{cases} 1, f_1 \le f \le f_2 \\ 0, інакше \end{cases}, \quad (3)$$

де f_1 та f_2 – межі смуги пропускання.

Фреймування є необхідним етапом для підготовки сигналу до спектрального аналізу. Сигнал розбивається на короткі сегменти, або фрейми, тривалістю від 20 до 40 мс, що дозволяє врахувати зміну властивостей сигналу у часі. Застосування віконних функцій, таких як функція Хеннінга або Хеммінга, допомагає згладити краї фреймів, що мінімізує спектральні витікання. Віконна функція $\omega(n)$ може бути описана як:

$$\omega(n) = 0{,}54 - 0{,}46\, cos\left(\frac{2\pi n}{N-1}\right), \quad (4)$$

де N – довжина вікна, коефіцієнт 0,54 обирається таким чином, щоб забезпечити оптимальне співвідношення згладжування та збереження корисного сигналу (відповідає за зсув середнього значення амплітуди, забезпечуючи підйом центральної частини вікна), коефіцієнт 0,46 використовується для корекції і згладжування країв вікна (забезпечує зменшення амплітуди косинусної функції, що сприяє мінімізації спектрального витікання).

Перетворення сигналу з часової області в частотну (рис. 3) здійснюється за допомогою швидкого перетворення Фур'є (FFT). Перетворення Фур'є є одним із найпоширеніших методів спектрального аналізу, що дозволяє розкласти складний мовний сигнал на простіші частотні компоненти, забезпечуючи детальне представлення його спектральної структури. Аналізуючи ці компоненти, можна виявити індивідуальні особливості мовлення, такі як інтонація, ритм, тембр та артикуляційні характеристики, які відіграють ключову роль у розпізнаванні мови та ідентифікації мовця. Таке перетворення дозволяє отримати спектральне представлення сигналу, яке є більш зручним для подальшого аналізу. Дискретне перетворення Фур'є виконується згідно:

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-j2\pi k/N}, \quad (5)$$

де $X(k)$ – спектральні компоненти, $X(n)$ – часові відліки сигналу, N – кількість відліків.

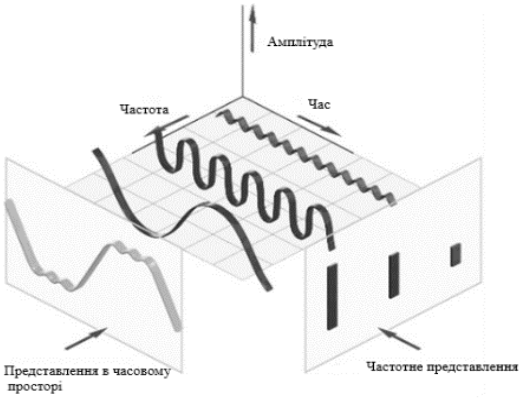

Рис. 3. Перехід від часової області до частотної області сигналу

Подальша обробка включає перетворення спектра в мел-шкалу з використанням мел-фільтробанку.

Мел-шкала більш точно відображає особливості сприйняття частот людським слуховим апаратом. Спектральні компоненти агрегуються в мел-частотні кепстральні коефіцієнти (MFCC) за допомогою дискретного косинусного перетворення (DCT). Розрахунок MFCC виконується відповідно до:

$$c_j = \sum_{k=1}^{K} log(S(k)) \cos\left(\frac{\pi i}{K}(k-0{,}5)\right), \quad (6)$$

де c_i – кепстральний коефіцієнт, $S(k)$ – відповідний енергетичний спектр, K – кількість фільтрів у мел-фільтробанку.

Логарифмування енергетичного спектру зменшує динамічний діапазон сигналу, що забезпечує меншу чутливість до варіацій у гучності. Це досягається шляхом застосування логарифмічної функції $log(E)$, де E – енергія спектрального компонента.

Для завершального етапу (декореляції ознак значень вхідних даних) виконується дискретне косинусне перетворення (DCT), що дозволяє зменшити взаємозв'язок між ознаками та робить ознаки (кепстральні, спектральні коефіцієнти та форманти) більш незалежними одна від одної.

Декореляція спрощує подальший аналіз та обробку даних і дозволяє відфільтрувати високочастотні компоненти, які в більшості випадках містять в собі шум. Отримано мел-кепстральні коефіцієнти:

$$C_n = \sum_{m=0}^{M-1} log\, M_m \cdot cos\left(\frac{\pi n(2m+1)}{2M}\right), \quad (7)$$

де M – кількість мел-каналів.

Після отримання мел-кепстральних коефіцієнтів можна застосувати алгоритми пригнічення динамічного шуму, які потребують точної оцінки шумового порогу. Для цього використовуються алгоритми адаптивних фільтрів, таких як LMS (Least Mean Squares) та RLS (Recursive Least Squares), які передбачають постійну корекцію ваг фільтра на основі поточної помилки, що дозволяє їм адаптуватися до змінних умов шляхом корегування свої ваги $\omega(n)$ на кожному кроці на основі поточної помилки $e(n)$:

$$\omega(n+1) = \omega(n) + \mu e(n) x(n), \quad (8)$$

де $\omega(n)$ – вектор ваг фільтра на n-му кроці, μ – крок навчання, що визначає швидкість адаптації, $e(n) = d(n) - y(n)$ – поточна помилка, де $d(n)$ – бажаний сигнал, а $y(n) = \omega(n)^T x(n)$ – вихід адаптивного фільтра, $x(n)$ – вхідний сигнал на n-му кроці. Алгоритм RLS використовує вдосконалену рекурсивну схему для мінімізації середньоквадратичної помилки:

$$\omega(n) = \omega(n-1) + k(n)[d(n) - \omega(n-1)^T x(n)], \quad (9)$$

де $k(n)$ – коефіцієнт підсилення.

Також можуть бути використані методи вейвлет-фільтрації та маскування спектрального середовища.

Вейвлет-фільтрація заснована на розкладі сигналу на різні частотні підполоси за допомогою вейвлет-перетворення. Це дозволяє аналізувати сигнал на різних масштабах і виявляти компоненти, пов'язані з шумом.

$$W(j,k) = \sum_{n=0}^{N-1} x[n] \cdot \varphi_{j,k}[n], \quad (10)$$

де $W(j,k)$ – коефіцієнти вейвлет-перетворення.

Далі, виконується порогова модифікація коефіцієнтів вейвлет-перетворення, при якій коефіцієнти, що відповідають шуму, зменшуються або обнуляються, тоді як значення, що відповідають корисному сигналу, залишаються незмінними:

$$\hat{W}(j,k) = \begin{cases} W(j,k) - \lambda, якщо\ W(j,k) > \lambda, \\ W(j,k) + \lambda, якщо\ W(j,k) < -\lambda, \\ 0, якщо\ |W(j,k)| \le \lambda, \end{cases} \quad (11)$$

Метод маскування спектрального середовища заснований на оцінці спектрального відношення сигнал-шум (SNR) в кожній частотній смузі. На основі отриманого SNR будується маска, яка використовується для ослаблення або посилення різних частотних компонентів сигналу:

$$H(k) = \begin{cases} 0, якщо\ SNR(k) \le \gamma, \\ 1, якщо\ SNR(k) > \gamma, \end{cases} \quad (12)$$

де $H(k)$ – функція маскування на k-й частоті, SNR – відношення сигнал/шум на k-й частоті, γ – порогове значення для SNR.

Результати порівняльного аналізу

У ході дослідження створено проект на основі засобів Pandas та Teras (бібліотеки для обробки аудіо, яка розроблена для спрощення аналізу та перетворення аудіо сигналів), підготовлені тестові дані в форматі WAV-файлів (Waveform Audio File Format) з записами промови чисел різними мовцями.

В процесі експериментів порівняно результати застосування різних методів пригнічення шумів у системі. Проведено окремі експерименти з фільтрацією лише статичних шумів: за допомогою спектрального віднімання та фільтра Вінера та досліджено ефективність фільтрації лише динамічних шумів: за допомогою алгоритму LMS та вейвлет-фільтрації – Гааровий вейвлет та Морлет вейвлет.

В експериментах аналізувалися голосові сигнали, зокрема оригінальний та зашумлений, представлені у спектральному вигляді. Для оцінки змін у спектрі було обрано п'ять контрольних точок, у яких визначалися відповідні коефіцієнти згідно з графічними даними. Подальша обробка сигналів здійснювалася із застосуванням різних методів шумопригнічення, зокрема спектрального віднімання, фільтра Вінера, LMS-фільтра та вейвлет-фільтрації. Після фільтрації для кожного методу було повторно визначено спектральні коефіцієнти у тих самих контрольних точках.

Основним критерієм оцінки якості фільтрації є ступінь наближення коефіцієнтів фільтрованого сигналу до коефіцієнтів оригіналу. Перевищення цих значень після обробки вказує на залишкові шуми, тоді як занижені коефіцієнти свідчать про втрату корисної інформації, що погіршує відновлення сигналу.

Таким чином, мета експерименту полягала у визначенні ефективності різних методів шумопригнічення шляхом аналізу впливу змін коефіцієнтів у спектрі після обробки сигналів.

Отримані результати дають можливість оцінити ступінь збереження корисної інформації при застосуванні кожного з розглянутих методів фільтрації та обробки (табл. 1).

Таблиця 1 – **Результати обробки сигналу обраними методами фільтрації шумів**

	Вхідні сигнали				
Вид сигналу	**1**	**2**	**3**	**4**	**5**
Оригінальний сигнал	0,50	0,32	0,26	0,68	0,44
Зашумлений сигнал	0,71	0,70	0,40	0,89	0,87
Результати обробки оригінального сигналу					
Метод обробки	**1**	**2**	**3**	**4**	**5**
Спектральне віднімання	0,45	0,48	0,35	0,62	0,31
Фільтр Вінера	0,52	0,33	0,32	0,63	0,39
LMS	0,53	0,32	0,31	0,62	0,46
Гааровий вейвлет	0,49	0,32	0,29	0,65	0,43
Морлет вейвлет	0,46	0,34	0,25	0,64	0,41
Результати обробки зашумленого сигналу					
Зашумлений сигнал	0,71	0,70	0,40	0,89	0,87
Спектральне віднімання	0,64	0,39	0,42	0,54	0,43
Фільтр Вінера	0,58	0,34	0,32	0,60	0,46
LMS	0,51	0,38	0,25	0,66	0,52
Гааровий вейвлет	0,48	0,37	0,24	0,67	0,47
Морлет вейвлет	0,43	0,36	0,23	0,69	0,44

Повне видалення шуму є складною задачею у умовах динамічного рівня шуму та може бути вирішено шляхом організації двоступеневої обробки голосового сигналу, яка передбачає поєднання методів попередньої обробки та глибшого аналізу акустичних характеристик. Такий підхід дозволяє поєднати переваги різних методів, забезпечуючи ефективне відфільтровування шумів, нормалізацію сигналів і покращення якості голосового сигналу перед подальшим розпізнаванням. Вибір методів фільтрації для двоступеневої обробки голосового сигналу визначається характеристиками шуму, співвідношенням сигнал-шум та бажаною якістю обробленого сигналу. Ефективність фільтрації оцінюється шляхом експериментального дослідження використання різних методів, таких як лінійні, нелінійні фільтри та нейронні мережі. Для порівняння ефективності статичних методів (спектрального віднімання (1) та фільтра Вінера (2)), а також динамічних підходів (LMS-алгоритму (3), Гаар-вейвлета (4) і Морлет-вейвлета (5)), було реалізовано двоступеневу систему пригнічення шумів, яка передбачає послідовне застосування фільтрації як статичних, так і динамічних шумів Для цього поєднано: спектральне віднімання та LMS алгоритм (6), Фільтр Вінера та Гаар вейвлет (7), Фільтр Вінера та Морлет вейвлет (8), а також реалізовано повторну обробку мовного запису за допомогою статичного методу – спектрального віднімання (9).

Для оцінки точності алгоритмів фільтрації шуму шляхом порівняння очищеного сигналу з оригінальним сигналом використовується середня абсолютна помилка, яка розраховується згідно:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|x_i - \hat{x}_i|, \qquad (13)$$

де N – кількість елементів у сигналу, x_i – значення оригінального (еталонного) сигналу в момент часу, $\hat{x}_i$ – значення очищеного сигналу в момент часу.

Результати, наведені у табл. 2, демонструють, що більшість застосованих методів фільтрації знижують рівень шуму в сигналах, що підтверджується зменшенням середньої абсолютної помилки (рис. 4). Комбіновані методи фільтрації, такі як поєднання фільтра Вінера та вейвлет-фільтрації (7, 8) є більш стійкими до різнотипових шумів та викривлень.

Таблиця 2 – **Результати обробки сигналів методами фільтрації та успішність ідентифікації користувача**

№ методу	Оригінальний сигнал		Зашумлений сигнал	
	Успішність ідентифікації (%)	**MAE**	**Успішність ідентифікації (%)**	**MAE**
Методи фільтрації статичних шумів				
1	57,23	9,34	48,32	11,44
2	53,61	10,53	54,68	12,04
Методи фільтрації динамічних шумів				
3	89,12	3,24	82,49	6,48
4	85,81	4,92	83,26	8,52
5	83,62	7,12	83,73	7,57
Двоступенева система пригнічення шумів				
6	91,12	2,13	87,34	5,36
7	94,23	1,58	92,14	2,16
8	98,85	0,22	97,39	1,14
9	47,32	11,64	44,99	12,11

Організація етапу фільтрації дозволяє покращити якість голосового сигналу, що, у свою чергу, позитивно впливає на успішність ідентифікації користувача. Згідно з отриманими експериментальними результатами, комбінація фільтра Вінера та Морлет вейвлета (8) продемонструвала найвищу ефективність серед усіх розглянутих методів, забезпечивши 97,39% успішності ідентифікації навіть у зашумлених умовах, що свідчить про здатність цього підходу ефективно адаптуватися до динамічних шумів (рис. 5). Поєднання фільтра Вінера та Гаар вейвлета (7) також показало значне покращення результатів порівняно з окремим використанням цих методів.

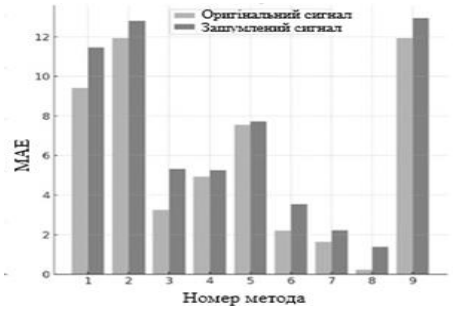

Рис. 4. Діаграма порівняння показника MAE досліджуваних методів фільтрації шумів

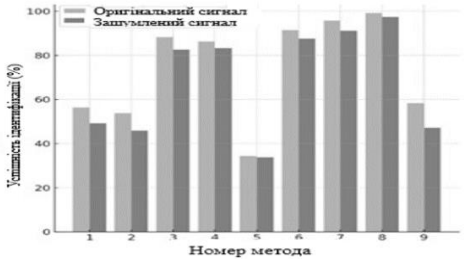

Рис. 5. Діаграма порівняння показника успішності ідентифікації користувача після в залежності від метода попередньої обробки

Застосування повторної обробки мовного запису із використанням статичного методу спектрального віднімання [9] виявило тенденцію до зниження загальної якості обробленого сигналу, що зумовлено перетворенням частини корисної мовленнєвої інформації на шумові артефакти. Аналіз отриманих результатів показує, що при повторному застосуванні алгоритму спостерігається його схильність до некоректного інтерпретування залишкових низькоенергетичних компонентів мовного сигналу як небажаних шумів, що призводить до їхнього видалення. Це може знижувати розбірливість корисної інформаційної складової мовленнєвого сигналу та призводити до втрати важливих акустичних характеристик, критичних для точного розпізнавання та ідентифікації мовця. Цей ефект особливо помітний при низькій інтенсивності вхідного сигналу або значній варіативності амплітудного спектра. Повторне спектральне віднімання без адаптивного налаштування може негативно впливати на якість фільтрованого сигналу.

У табл. 3 представлено середній час обробки голосового сигналу для різних методів фільтрації шумів, що дозволяє здійснити порівняльний аналіз їхньої обчислювальної складності.

Таблиця 3 – **Час виконання методів очищення сигналу**

№ методу	Час виконання, с	
	Оригінальний сигнал	Защумлений сигнал
1	41,1	50,3
2	70,1	73,5
3	22,3	26,5
4	73,3	72,9
5	145,1	144,1
6	67,5	71,5
7	124,3	132,6
8	227,7	241,4
9	91,2	98,4

Отримані результати свідчать про те, що послідовне застосування двох методів попередньої обробки очікувано призводить до збільшення загального часу виконання, що зумовлено накопиченням обчислювальних витрат на кожному етапі фільтрації. Дане явище пояснюється необхідністю проведення додаткових операцій аналізу, фреймування, оцінки характеристик сигналу та їхньої корекції, що, у свою чергу, підвищує вимоги до обчислювальних ресурсів.

Зростання обчислювальної складності особливо актуальне в системах, що працюють у режимі реального часу, де збільшення часу обробки може суттєво впливати на продуктивність та швидкість реагування. Вибір методів шумозаглушення має враховувати не лише покращення якості сигналу, а й обчислювальні витрати, важливі для застосування в умовах обмежених ресурсів.

Результати проведених досліджень підтверджують високу ефективність двоступеневої системи шумозаглушення, яка забезпечує значне покращення якості обробки голосового сигналу шляхом комбінованої фільтрації статичних і динамічних шумів. Комплексний підхід до пригнічення акустичних перешкод демонструє суттєві переваги порівняно з методами, що застосовують лише однокомпонентну фільтрацію, оскільки дозволяє більш точно виділяти корисну мовленнєву інформацію. Крім того, результати експериментальних досліджень вказують на особливу важливість попередньої обробки голосових записів для підвищення точності ідентифікації користувача, що особливо актуально в умовах наявності високого рівня фонових шумів або змін у характеристиках мовлення протягом запису.

Висновки

В роботі досліджено методи попередньої обробки голосових сигналів та їх вплив на кінцеву точність роботи системи голосової ідентифікації. Проведено порівняння отриманих результатів методів пригнічення шумів, що дозволило визначити найбільш ефективні підходи для покращення якості голосового сигналу та підвищення точності ідентифікації користувача. Запропонований двоступеневий підхід пригнічення шуму (сталого та динамічного) є сучасним рішенням поліпшення якості мовних сигналів в системах голосової ідентифікації та потребує подальших досліджень, спрямованих на розробку більш ефективних і робастних методів.

При послідовному застосуванні двох методів кожен з них може вносити власні артефакти або залишкові шуми, відсутні на початковому етапі обробки. Несумісність методів шумозаглушення може призводити до зміни структури шуму першим методом, що знижує ефективність другого на етапі його усунення. Важливим етапом конфігурації двоступеневих методів є точне налаштування параметрів кожного методу та оцінка їх сумісності. У середньому похибка у двоступеневих системах шумозаглушення на 4-6% менша, ніж у динамічних методів, і на 8-9% нижча порівняно зі статичними.

Аналіз отриманих даних свідчить про те, що послідовне застосування методів для видалення статичних та динамічних шумів дозволяє досягти значно кращих результатів у відновленні чистого сигналу.

Список літератури

1. 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, pp. 1-4, doi: https://doi.org/10.1109/KhPIWeek57572.2022.9916473
2. Barnwal S.K., Gupta P. (2022), “Evaluation of AI System’s Voice Recognition Performance in Social Conversation”, 5th International Conference on Contemporary Computing and Informatics (IC3I), Uttar Pradesh, India, pp. 804-808, doi: https://doi.org/10.1109/IC3I56241.2022.10073242
3. Cornacchia M., Papa F., Sapio B. (2020), “User acceptance of voice biometrics in managing the physical access to a secure area of an international airport”, Technology Analysis & Strategic Management, Vol. 32(10), pp. 1236-1250, doi: https://doi.org/10.1080/09537325.2020.1758655

4. Li Z.L., Sajat M.S., Yusof Y., Fazea Y., Santana Purba H. (2021), "The Design and User Acceptance of IoT-based Access and Entrance Control System Using Voice Recognition", Knowledge Management International Conference (KMICe), pp. 315-321, doi: https://doi.org/10.13140/RG.2.2.12611.73762
5. Kinkiri S., Keates S. (2020), "Speaker Identification: Variations of a Human voice", 2020 International Conference on Advances in Computing and Communication Engineering (ICACCE), Las Vegas, NV, USA, 2020, pp. 1-4, doi: https://doi.org/10.1109/ICACCE49060.2020.9154998
6. Kambampati P., Rane S., Shoeb A., Dhannawat R. (2024), "PAYV - Payment Voice: A Platform using Voice Recognition to Enable Payment Transactions", Asia Pacific Conference on Innovation in Technology (APCIT), MYSORE, India, 2024, pp. 1-6, doi: https://doi.org/10.1109/APCIT62007.2024.10673442
7. Papadopoulos P., Tsiartas A., Gibson J., Narayanan S. (2014), "A supervised signal-to-noise ratio estimation of speech signals", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, pp. 8237-8241, doi: https://doi.org/10.1109/ICASSP.2014.6855207
8. Mishra P., Singh S., Singh S.K., Dixit A. (2018), "Pre-Processing and Partition of Voice for Semi-Voice Authentication", Fourth International Conference on Advances in Computing, Communication & Automation (ICACCA), Subang Jaya, Malaysia, pp. 1-5, doi: https://doi.org/10.1109/ICACCAF.2018.8776849
9. Awan S.N., Giovinco A., Owens J. (2012), "Effects of Vocal Intensity and Vowel Type on Cepstral Analysis of Voice", Journal of voice, Vol. 26(5), pp. 15-20, doi: https://doi.org/10.1016/j.jvoice.2011.12.001
10. Lecomte I., Lever M., Boudy J., Tassy A. (1989), "Car noise processing for speech input", International Conference on Acoustics, Speech, and Signal Processing, Glasgow, vol. 1, pp. 512-515, doi: https://doi.org/10.1109/ICASSP.1989.266476
11. Mamatov N., Niyozmatova N., Samijonov A. (2021), "Software for preprocessing voice signals", Tashkent University of information technologies, Vol. 18, pp. 1-8, doi: https://doi.org/10.6703/IJASE.202103_18(1).006
12. Slimane A.B.; Zaid A.O. (2021), "Real-Time Fast Fourier Transform-Based Notch Filter for Single-Frequency Noise Cancellation Application to Electrocardiogram Signal Denoising", Journal of Medical Signals & Sensors, Vol. 11(1), pp. 52-61, doi: https://doi.org/10.4103/jmss.JMSS_3_20
13. Zaw T.H., War N. (2017), "The combination of spectral entropy, zero crossing rate, short time energy and linear prediction error for voice activity detection", 20th International Conference of Computer and Information Technology (ICCIT), Dhaka, Bangladesh, pp. 1-5, doi: https://doi.org/10.1109/ICCITECHN.2017.8281794
14. Singh S., Rajan E. (2011), "MFCC VQ based speaker identification and its accuracy affecting factors", International Journal of Computer Applications, Vol. 21(6), pp. 1-6, doi: https://doi.org/10.5120/2519-3423
15. Ahmed A., Khondkar M.J.A., Herrick A., Schuckers S., Imtiaz M.H. (2024), "Descriptor: voice pre-processing and quality assessment dataset", IEEE Data Descriptions, Vol. 1, pp. 146-153, DOI: https://doi.org/10.1109/IEEEDATA.2024.3493798
16. Nagrani A., Chung J.S., Ziccerman A. (2017), "VoxCeleb: a large-scale apeaker identification dataset", Sound, p. 6, doi: https://doi.org/10.21437/Interspeech.2017-950

Received (Надійшла) 15.01.2025
Accepted for publication (Прийнята до друку) 26.03.2025

ВІДОМОСТІ ПРО АВТОРІВ / ABOUT THE AUTHORS

Іващенко Георгій Станіславович – кандидат технічних наук, доцент, доцент кафедри електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна;
Heorhii Ivashchenko – Candidate of Technical Sciences, Associate Professor at the Department of Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine;
e-mail: heorhii.ivashchenko@nure.ua; ORCID Author ID: http://orcid.org/0000-0003-1027-5262;
Scopus Author ID: https://www.scopus.com/authid/detail.uri?authorId=57217030807.

Бондаренко Максим Едуардович – аспірант кафедри електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна;
Maksym Bondarenko – PhD student at the Department of Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine;
e-mail: maksym.bondarenko@nure.ua; ORCID Author ID: http://orcid.org/0000-0002-2500-7626;
Scopus Author ID: https://www.scopus.com/authid/detail.uri?authorId=57216159508.

Using a sequence of preprocessing methods in voice identification systems

Maksym Bondarenko, Heorhii Ivashchenko

Abstract. Topicality. To date, voice identification is the most common method, but some technical limitations associated with environmental influences and voice variations require further research. With the development of technologies for creating fake audio recordings, there is a need for additional improvements to ensure reliable identification using existing voice biometrics methods, mainly by introducing voice data pre-processing methods. **The goal of this work** is to study the methods of pre-processing in human voice identification systems. **The object of research** is the voice signal filtering module in the user's voice identification system. **The subject of research** is the methods of pre-processing voice signals. **Results**. This article studies the effectiveness of applying a sequence of pre-processing methods in voice identification systems to reduce the impact of background noise and other distortions on recognition quality. A comparative analysis of the effectiveness of removing steady and dynamic noise was performed. **Conclusion**. The proposed approach to organizing the procedure of pre-processing voice signals involves the selection of an optimal sequence of methods, considering such parameters as recording quality, computing resources, and the required level of accuracy. The obtained results demonstrate that the sequential use of noise reduction and signal normalization methods increases the accuracy of voice identification.

Keywords: preprocessing methods, sequential application of methods, voice signals, voice identification systems, filtering, normalization, feature extraction, noise, spectral analysis.