Одеський національний університет імені І.І. Мечникова
Факультет математики, фізики та інформаційних технологій
Кафедра оптимального керування і економічної кібернетики

Кваліфікаційна робота

на здобуття ступеня вищої освіти «бакалавр»

«Методи та алгоритми розпізнавання звукових образів»

«Methods and algorithms for recognizing sound images»

Виконав: здобувач денної форми навчання
спеціальності 113 Прикладна математика
Освітня програма «Прикладна математика»
Кулик Данііл Вячеславович
Керівник: канд. техн. наук, доц. Мороз В.В. _____
Рецензент: канд. техн. наук, доц. Мазурок І.Є.

Рекомендовано до захисту:
Протокол засідання кафедри
№ ___ від _____2023 р.

Завідувач кафедри

_____ _____
(підпис) (прізвище, ініціали)

Захищено на засіданні ЕК №_____
протокол № ____ від _____ 2023 р.
Оцінка_____/___ ___/_____
(за національною шкалою, шкалою ECTS, бали)

Голова ЕК

_____ _____
(підпис) (прізвище, ініціали)

Одеса – 2023

ЗМІСТ

ВСТУП

На відміну від задачі розпізнавання музичних сигналів та мови [1-2], відомі алгоритми не дають прийнятних результатів при розпізнаванні звичайних звукових образів. Тому ціль даного дослідження спрямована на розробку методу розпізнавання звукових образів, який може бути застосований у різних сферах, включаючи мілітарні застосунки, IoT, системи відеоспостереження та інше. Один з ключових елементів запропонованого підходу полягає у застосуванні дискретного вейвлет-перетворення (DWT) [3] для знаходження особливостей певних класів аудіосигналів [4]. Готових продуктів чи алгоритмів у відкритому доступі майже не існує, що дає свободу вибору в створенні власних методів.

Однією з проблем при роботі з аудіосигналами є наявність шуму, який суттєво впливає на розпізнавання звукових образів та знижує точність класифікації. Тому важливо виконати якісну фільтрацію до етапу виявлення особливостей.

Актуальність кваліфікаційної роботи полягає у розробці ефективного алгоритму розпізнавання звукових образів з використанням дискретного вейвлет-перетворення (DWT).

Мета роботи полягає в створені та дослідженні методів та алгоритмів розпізнавання звукових образів для подальшого використання у різних сферах.

Об'єктами дослідження даної роботи є записані звукові образи крилатих ракет, реактивних бойових літаків, гелікоптерів та побутові звуки для їх подальшої класифікації.

Предметом дослідження роботи є розробка, опрацювання та модифікація методів та алгоритмів, що використовуються для розпізнавання звукових образів.

Результати роботи були апробовані на міжнародній науково-технічній конференції "Інформатика, управління та штучний інтелект (ІУШІ-2023)".

РОЗДІЛ 1.
ОГЛЯД ПРОБЛЕМИ РОЗПІЗНАВАННЯ ЗВУКОВИХ ОБРАЗІВ

1.1 Проблеми із розпізнавання звукових образів

Написання методу не єдина проблема в розпізнаванні звукових образів. Людство ще не створило такий комп'ютер, який міг би пародіювати нейронні клітини голови людини. Через це обробка інформації займає багато ресурсу та часу. І це не кажучи про аналіз великої кількості аудіо-даних в реальному часі, де запис аудіосигналу може починатися задовго до потрібного моменту, збільшуючи кількість непотрібних даних, які все ще потрібно проаналізувати.

Припускаючи, що певні алгоритми та моделі можуть витягти лише потрібні частини сигналу, або скоротити поступаючі зразки, все ще залишається проблема з шумом в аудіосигналі. Шум – це коливання частинок навколишнього середовища, що сприймається органами слуху людини як небажані сигнали [5]. Для аудіосигналу - нестійкі або випадкові акустичні коливання, що характеризуються випадковою зміною амплітуди і частоти [6]. Якщо відтворити цифровий запис аудіосигналу, знятого з непрофесійного пристрою, можна помітити спотворення та тремтіння. Це відбувається через те, що при дискретизації сигналу відбувається вибірка з відхиленням від рівномірного проміжку часу, тобто відліки беруться трохи раніше або пізніше за ідеальний часовий інтервал. Як приклад, при частоті дискретизації в 44100 Гц відліки беруться не точно 1/44100 секунди, як були б правильно, а з відхиленням.

Варіабельність звуків також є складною задачею для розпізнавання образів: подібно до розпізнавання мови та музики, звуки можуть значно варіюватись. Наприклад, звук реактивного літака може звучати по-різному залежно від моделі, швидкості та відстані до мікрофона. Ці варіації можуть ускладнити розробку універсального алгоритму розпізнавання звуку.

Розглядаючи алгоритми машинного навчання також з'являються складнощі. По-перше – це потреба значного обсягу інформації для навчання розпізнавання

звукових образів та підвищення їх точності. По-друге – потреба в надзвичайно потужних обчислювальних можливостях.

Також наразі виникає нова проблема, пов'язана з конфіденційністю. Можливість використання технології розпізнавання звуку для спостереження може викликати занепокоєння щодо приватності. Деяким людям може бути неприємно, що їхні розмови чи звуки записуються та аналізуються машинами.

1.2 Відомі алгоритми розпізнавання звукових образів

Переходячи до створення чи покращення алгоритмів щодо розпізнавання звукових образів важливо подивитися на готові продукти та методи. Напевно однією з найвідоміших програм з розпізнавання пісень є «Shazam». Це додаток, який за лічені секунди здатен впізнати метадані пісні за коротким наданим аудіозаписом з мікрофона. Іншими прикладами з цієї сфери є:

- SoundHound / Midomi
- Chromaprint
- Echoprint
- Musixmatch

Якщо розглядати розпізнавання мовлення, що є дуже актуальним в сьогоденні, то тут існує ще більше готових рішень. Одні з найвідоміших – голосові помічники «Siri» від компанії «Apple Inc», або «Alexa» від «Amazon». Також існують програми для відтворення мови у текст. Яскравим прикладом буде «Windows Speech Recognition». Але цікавить нас саме те, як вони працюють.

1.3 Алгоритм пошуку звуку промислової потужності

В 2003 році була опублікована стаття Ейвері Лі-Чун Вана під назвою «An Industrial-Strength Audio Search Algorithm» [1]. У статті розповідається алгоритм для знаходження особливостей пісень. Ці особливості були названі «audio

fingerprints», і кожна пісня може мати сотні-тисяч таких відбитків. Ці «відбитки пальців» знаходяться за допомогою використання комбінаторно хешованого частотно-часового аналізу «сузір’я» аудіо.

Якщо коротко – робимо віконне Фур’є перетворення сигналу, з якого створюємо спектрограму. В створеній спектрограмі виділяємо пікові значення амплітуди. Визначаємо пік як пару (час, частота), що відповідає значенню амплітуди, яке є найбільшим у локальній "околиці" навколо нього.

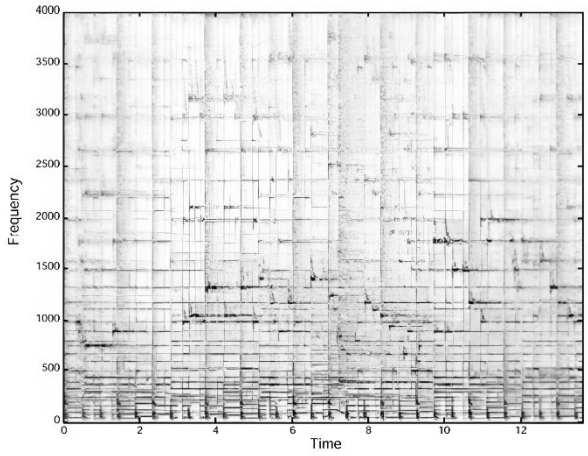

Рис. 1.1 A – Спектрограма

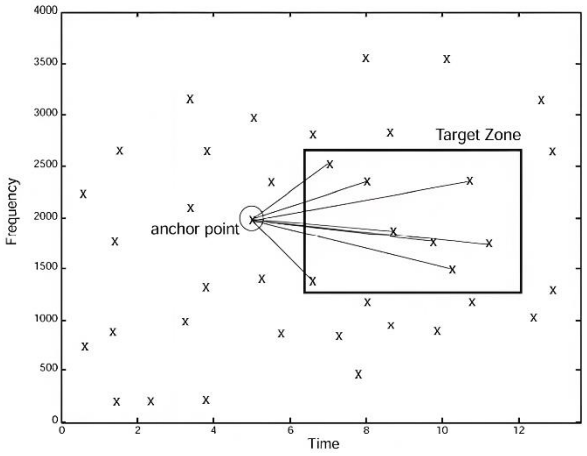

Рис. 1.1 C – Комбінаторна генерація хешу

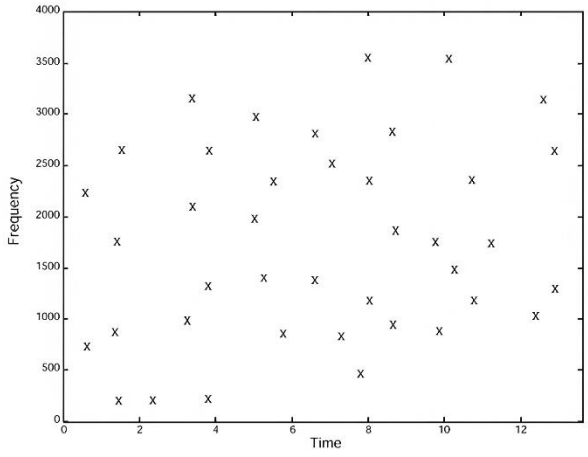

Рис. 1.1 B – Карта сузір'їв

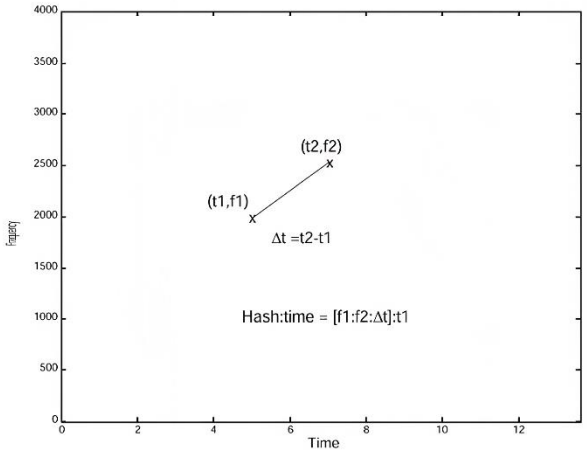

Рис. 1.1 D – Деталі хешу

У статті автор показує на прикладі, як складна спектрограма (на рис. 1.1 A) може бути зведена до розрідженого набору координат (як показано на рис. 1.1 B).

Хеші відбитків формуються з карти сузір'їв, де точки комбінаторно пов'язані між собою. Кожна опорна точка (сузір’я) асоціюється з цільовою зоною.

Послідовно об’єднуємо опорні точки в пари з точками в межах цільової зони, де кожна пара дає дві частотні складові плюс час різниця в часі між точками (Рисунок 1.1 C і 1.1 D). Таким чином отримуємо «відбиток пальця» для поточної точки і повторюємо дії для всіх точок. Далі отримані відбитки порівнюються з відбитками в базі даних. Де більше всього співпадінь по «відбиткам пальців», ту пісню і повертаємо як розпізнану.

РОЗДІЛ 2.
МАТЕМАТИЧНІ ІНСТРУМЕНТИ ДЛЯ АНАЛІЗУ АУДІОСИГНАЛІВ

Звукові образи вимагають спеціальних методів аналізу для розуміння їхньої структури та властивостей. Розглянемо різні математичні інструменти, які використовуються для аналізу звукових образів у даній роботі.

2.1 Перетворення Фур’є

Перетворення Фур'є (Fourier transform, FT) є одним з фундаментальних математичних інструментів, що використовуються для аналізу сигналів та даних [7]. Воно назване на честь французького математика Жана Батіста Жозефа Фур'є.

Перетворення Фур'є дозволяє розкласти сигнал або функцію на складові частоти, що дозволяє розглядати його в частотному домені. В основі цього перетворення лежить ідея, що будь-яку складену хвилю або сигнал можна представити як суму простих синусоїдальних хвиль різних амплітуд і фаз [8].

Формально, перетворення Фур'є зводиться до розкладання функції у вигляді інтегралу за допомогою комплексних експонент:

$$F(x) = \int_{-\infty}^{\infty} f(x) \cdot e^{-2\pi i k x} dx, \tag{2.1}$$

та обернене перетворення Фур'є (Inverse Fourier transform, IFT) має вигляд:

$$f(x) = \int_{-\infty}^{\infty} F(k) \cdot e^{2\pi i k x} dk, \tag{2.2}$$

де $f(x)$ - вихідна функція, $F(k)$ - її перетворення Фур'є, k - частота (в частотному домені), x - час або просторова координата (у вихідному домені).

Результатом перетворення Фур'є є спектр сигналу, який представляє себе як функцію частоти. Він вказує на наявність або відсутність певних частот у сигналі та їхні амплітуди. Перетворення Фур'є знаходить широке застосування в

сигнальній обробці, зображеннях, телекомунікаціях, оптиці, криптографії та інших галузях науки та технологій.

Перетворення Фур'є має свої обернене перетворення, яке дозволяє відновити вихідну функцію з її спектра. Це дозволяє переміщуватися між часовим та частотним представленнями сигналів, що є надзвичайно корисним у багатьох областях.

2.1.1 Дискретне перетворення Фур’є

Для розгляду подальших методів потрібно ознайомитися з дискретним Фур’є перетворенням (Discrete Fourier Transform, DFT). Дискретне Фур'є-перетворення (DFT) є математичним інструментом, який використовується для аналізу і обробки дискретних сигналів [9]. Воно є дискретним відповідником неперервного Фур'є-перетворення і дозволяє розкласти сигнал, представлений у вигляді послідовності чисел, на набір комплексних амплітуд та фазових компонент.

Процес DFT включає поділ послідовності дискретних даних на різні частини, які розглядаються як синусоїдальні хвилі різних частот. Кожну з цих частот розглядають як комплексну функцію, що описує амплітуду та фазу хвилі. Застосування DFT до сигналу дозволяє отримати спектральне представлення сигналу, яке може бути використане для подальшого аналізу та обробки сигналу [10].

$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N} kn} = \\ = \sum_{n=0}^{N-1} x_n \left(\cos\left(\frac{2\pi kn}{N} \right) - i \cdot \sin\left(\frac{2\pi kn}{N} \right) \right), \quad k = 0,\ldots,N-1, \tag{2.3}$$

де:

- N : кількість зразків;
- n : поточна вибірка;
- k : поточна частота, $k \in [0, N-1]$;

- x_n : значення синусоїди на відліку n;
- X_k : масив комплексних чисел після DFT, який включає інформацію про амплітуду та фазу;

Дискретне Фур'є-перетворення може бути оберненим (Inverse Discrete Fourier Transform, IDFT), що означає, що воно також може використовуватись для відновлення сигналу з його спектрального представлення [11].

$$x_n = \frac{1}{N}\sum_{n=0}^{N-1} X_k e^{\frac{2\pi i}{N}kn} = \\ = \frac{1}{N}\sum_{n=0}^{N-1} X_k \left(\cos\left(\frac{2\pi kn}{N}\right) + i\cdot\sin\left(\frac{2\pi kn}{N}\right)\right), \quad \mathrm{n} = 0,\ldots,N-1. \tag{2.4}$$

Це дозволяє здійснювати фільтрацію сигналу, вилучати шуми та виконувати інші операції над сигналами у частотному домені.

Як було зазначено у статті Омара Алкуса [12], після застосування дискретного перетворення Фур'є (DFT) до вхідного сигналу отримуємо масив комплексних чисел, які містять інформацію про частоти, амплітуди та фази синусоїд, що складаються з вхідного сигналу. Масив DFT $\left(X_k\right)$ поділяється на дві половини: перша половина містить додатні частотні складові, а друга половина містить від'ємні частотні складові. Якщо вхідний сигнал є лише дійсним, то перша половина є спряженою з другою половиною частотних складових, і спектр стає симетричним. Тому ми зосереджуємося лише на першій половині (додатні частотні складові) у випадку дійсних сигналів [13]. На наведеному нижче рисунку показані додатні та від'ємні частотні складові для випадку, коли кількість вхідних відліків $\left(N\right)$ є непарною або парною.

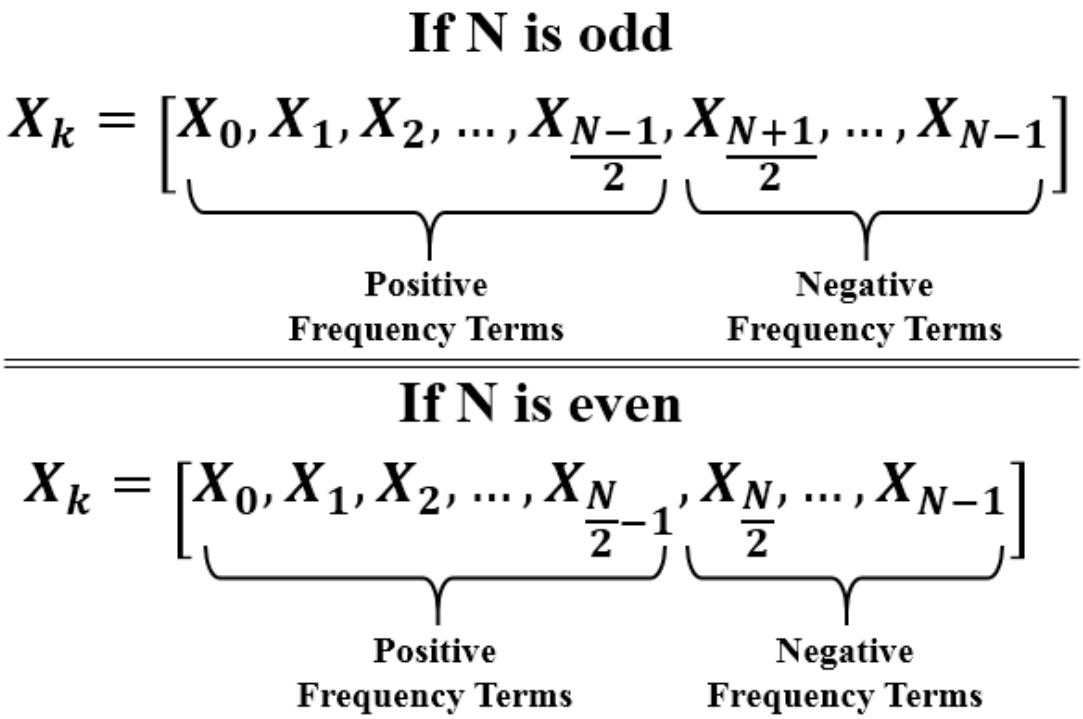

Рис. 2.1 Додатні та від'ємні частотні члени.

За допомогою комплексного масиву (X_k) можна визначити амплітуду та фазу кожної синусоїди, яка додається для формування сигналу. Ці значення можна обчислити з використанням уявної (Im) та дійсної (Re) частин комплексних чисел, які складають масив.

$$\text{Амплітуда} = \frac{|X_k|}{N} = \frac{1}{N}\sqrt{\text{Re}^2(X_k) + \text{Im}^2(X_k)}, \tag{2.5}$$

$$\text{фаза} = \arg(X_k) = \arctan\left(\text{Im}(X_k), \text{Re}(X_k)\right) = -i \cdot \ln\left(\frac{X_k}{|X_k|}\right). \tag{2.6}$$

2.1.2 Дискретно-часове перетворення Фур’є

Дискретно-часове перетворення Фур'є (Discrete-time Fourier transform, DTFT) - це математичний інструмент, який використовується для аналізу дискретних сигналів у частотній області. Воно перетворює послідовність дискретних відліків часового сигналу у послідовність комплексних чисел, які представляють амплітуду та фазу відповідних частот [9-10].

За допомогою рівномірно розташованих відліків DTFT формує функцію частоти, яка є періодичним сумуванням неперервного перетворення Фур'є (FT) вихідної неперервної функції. За певних теоретичних умов, описаних теоремою про дискретизацію, початкова неперервна функція може бути повністю відновлена з дискретно-часового перетворення Фур'є, а отже, з вихідних дискретних відліків. Саме DTFT є неперервною функцією частоти, але її дискретні відліки можна легко обчислити за допомогою дискретного перетворення Фур'є (DFT) [14].

$$X(\omega) = \sum_{n=-\infty}^{\infty} x(n) e^{-j\omega n}. \tag{2.7}$$

Дискретне перетворення Фур'є $X(\omega)$ дискретної послідовності $x(n)$ можна описати як представлення частотного змісту послідовності $x(n)$. За допомогою перетворення Фур'є, дискретна послідовність розкладається на складові залежно від їх частот. Тому Дискретно-часове перетворення Фур'є $X(\omega)$ також називається спектром сигналу, оскільки він відображає частотний склад послідовності.

2.2 Швидке перетворення Фур’є

Швидке перетворення Фур'є (Fast Fourier Transform, FFT) - це алгоритм, що використовується для ефективного обчислення дискретного перетворення Фур'є (DFT). Традиційний алгоритм обчислення DFT має складність $O(N^2)$, де N - це кількість вхідних точок [15]. Це означає, що обчислення дискретного перетворення Фур’є може бути дуже часомістким для великих наборів даних. FFT є ефективнішим методом обчислення дискретного перетворення Фур’є і має складність $O(N \log_2 N)$, що дозволяє значно прискорити процес обчислення [16].

Основна ідея швидкого перетворення Фур'є полягає в ефективному використанні симетрій і періодичності вхідних даних. Використовуючи рекурсивний підхід і поділяючи вхідний сигнал на менші частини, FFT здійснює швидке обчислення ДПФ, зменшуючи кількість операцій.

$$x[n]\cdot y[n]=\sum_{k=-\infty}^{\infty}x[k]y[n-k]\xleftrightarrow{DTFT}X\left(e^{j\omega}\right)Y\left(e^{j\omega}\right). \tag{2.8}$$

Вищезазначене рівняння вказує на те, що згортка двох сигналів може бути представлена як перемноження їх перетворень Фур'є. З цього випливає, що при перетворенні вхідного сигналу в частотний простір, згортка може бути замінена поелементним множенням. Іншими словами, вхідні дані для згорткового шару і ядра можуть бути перетворені в частотну область за допомогою перетворення Фур'є, помножені разом, а потім зворотно перетворені за допомогою зворотного перетворення Фур'є. Хоча це вимагає додаткових обчислювальних витрат на перетворення вхідних даних у частотну область та зворотне перетворення Фур'є для отримання результатів в просторовій області, це компенсується прискоренням, отриманим завдяки заміні множення ядра з різними частинами зображення на одне множення.

2.3 Віконне Фур’є перетворення

Віконне Фур’є перетворення (Short-Time Fourier Transform, STFT) - це метод аналізу сигналів у часово-частотній області. Він використовується для розкладання сигналу на його спектральні компоненти на коротких проміжках часу. STFT широко використовується в сигнальній обробці і акустичному аналізі для виявлення та вивчення змін у часі та частотному складі сигналу [17-18].

Віконне Фур’є перетворення засновано на ідеї застосування перетворення Фур'є до коротких відрізків сигналу. Вхідний сигнал розбивається на невеликі відрізки, які називаються вікнами, і для кожного вікна обчислюється перетворення Фур'є. Це дозволяє отримати інформацію про частотний склад сигналу на кожному відрізку часу.

Одна з основних переваг STFT полягає у тому, що вона дозволяє аналізувати спектр сигналу у часовому вимірі. Це корисно для виявлення змін в частотному складі сигналу від часу до часу, таких як зміна тональності або поява нових частотних складових.

Однак, STFT має обмеження, пов'язані з вибором розміру вікна. Велике вікно надає кращу частотну роздільну здатність, але погіршує часову роздільну здатність, тоді як маленьке вікно забезпечує кращу часову роздільну здатність, але погіршує частотну роздільну здатність.

2.3.1 Віконне Фур’є перетворення в неперервному часі

У випадку неперервного часу, процес перетворення функції можна пояснити наступним чином: спочатку вхідна функція помножується на віконну функцію, яка має значення, відмінне від нуля, лише протягом короткого проміжку часу. Потім застосовується перетворення Фур'є до результату, що представляє собою одновимірну функцію. Цей процес повторюється, зсуваючи вікно вздовж осі часу до кінця сигналу. В результаті отримується двовимірне представлення сигналу. У математичному вигляді процес записується наступним чином:

$$STFT\{x(t)\}(\tau,\omega) \equiv X(\tau,\omega) = \int_{-\infty}^{\infty} x(t)w(t-\tau)e^{-i\omega t}dt, \qquad (2.9)$$

де $w(\tau)$ - віконна функція (вікно Ганна, вікно Геммінга, вікно Гауса з центром навколо нуля, прямокутне вікно, вікно Кайзера та інші), $x(t)$ - сигнал, що підлягає перетворенню, а ω - частота. У сутності, це означає, що ми маємо справу з перетворенням Фур'є $X(\tau,\omega)$, де комплексна функція представляє фазу і амплітуду сигналу в залежності від часу та частоти.

2.3.2 Віконне Фур’є перетворення в дискретному часі

У випадку дискретного часу, дані, які мають бути піддані перетворенню, можна розділити на фрагменти або кадри, які, як правило, перекриваються для зменшення артефактів на межах. Кожен фрагмент проходить перетворення Фур'є, де комплексний результат додається до матриці, в якій зберігається амплітуда і фаза для кожної точки часу і частоти. Цей процес можна виразити таким чином:

$$STFT\{x[n]\}(m,\omega) \equiv X(m,\omega) = \sum_{n=-\infty}^{\infty} x[n]w[n-m]e^{-i\omega n}, \qquad (2.10)$$

аналогічно з сигналом x[n] і вікном w[n]. У цьому випадку змінна m є дискретною, тоді як змінна ω є неперервною. Проте, у більшості типових застосувань, обчислення віконного перетворення Фур'є (STFT) зазвичай виконуються на комп'ютері за допомогою алгоритмів швидкого перетворення Фур'є (FFT). Це означає, що обидві змінні, m і ω, фактично є дискретними і піддаються квантуванню.

Квадрат амплітуди віконного Фур’є перетворення дає спектрограмне представлення спектральної щільності потужності функції:

$$\text{spectrogram}\{x(t)\}(\tau,\omega) \equiv |X(\tau,\omega)|^2. \qquad (2.11)$$

2.4 Вейвлет-перетворення

Для побудови спектрограми використовується віконне Фур’є перетворення (short-time Fourier transform, STFT). STFT передбачає, що сигнал стаціонарний на короткому часовому інтервалі. Це, наприклад, є пісні (але важливо підкреслити, що не всі пісні можуть бути стаціонарними). Натомість при використанні нестаціонарних сигналів виникають проблеми зі зміною спектральних характеристик. Нестаціонарні сигнали можуть мати спектральні характеристики, що змінюються, як-от амплітуда, частота або фаза, у різні

моменти часу. Це може ускладнити точне визначення спектральних особливостей сигналу з використанням методів, призначених для стаціонарних сигналів. Також проблеми можуть бути зі складністю часового аналізу. Нестаціонарні сигнали вимагають складнішого і гнучкішого часового аналізу, щоб врахувати зміни в часі. Традиційні методи, такі як FFT (STFT), не завжди здатні точно представити еволюцію сигналу в часі, особливо якщо сигнал швидко змінюється.

Тому для нестаціонарних сигналів можна спробувати використати Вейвлет-перетворення. Вейвлет - це хвилеподібне коливання, локалізоване в часі [19]. Вейвлети мають дві основні властивості: масштаб і розташування. Масштаб (або розширення) визначає, наскільки «розтягнутий» або «здавлений» вейвлет. Ця властивість пов’язана з частотою, визначеною для хвиль. Розташування визначає, де вейвлет розташований у часі (або просторі). Гарний приклад з масштабом та розташуванням наведено в статті Шавхіна Талебі [20].

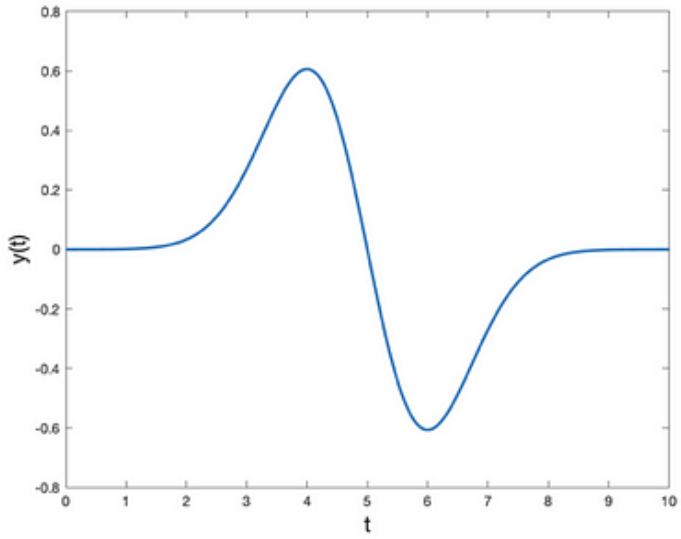

Рис. 2.2 Перша похідна функції Гауса

Масштаб вейвлету вищезазначеного виразу визначається параметром "a". При зменшенні значення цього параметра, вейвлет стає більш сплющеним, що спричиняє захоплення високочастотної інформації. Навпаки, збільшення

значення "a" розтягує вейвлет і дозволяє захоплювати низькочастотну інформацію.

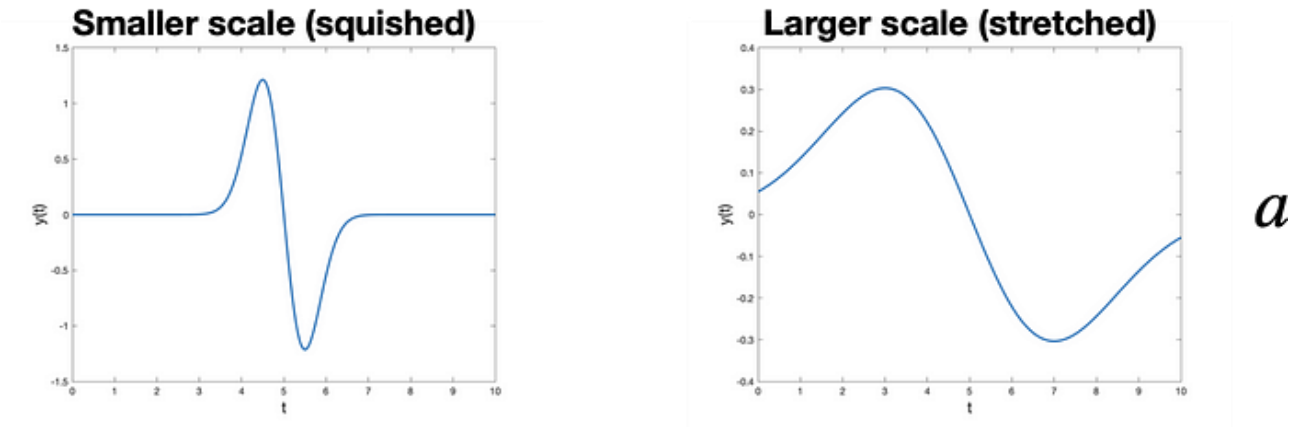

Рис. 2.3 Ліворуч: приклад вейвлету зі зменшеним масштабом. Праворуч: приклад вейвлету зі збільшеним масштабом

Місцезнаходження вейвлету визначається параметром "b". Якщо зменшити значення "b", вейвлет буде зсунутий вліво, а збільшення "b" зсуне його вправо. Розташування вейвлету має велике значення, оскільки, на відміну від хвиль, вейвлети діють лише на обмеженому проміжку. Крім того, при аналізі сигналу нас цікавить не тільки його коливання, а й місце, де ці коливання відбуваються.

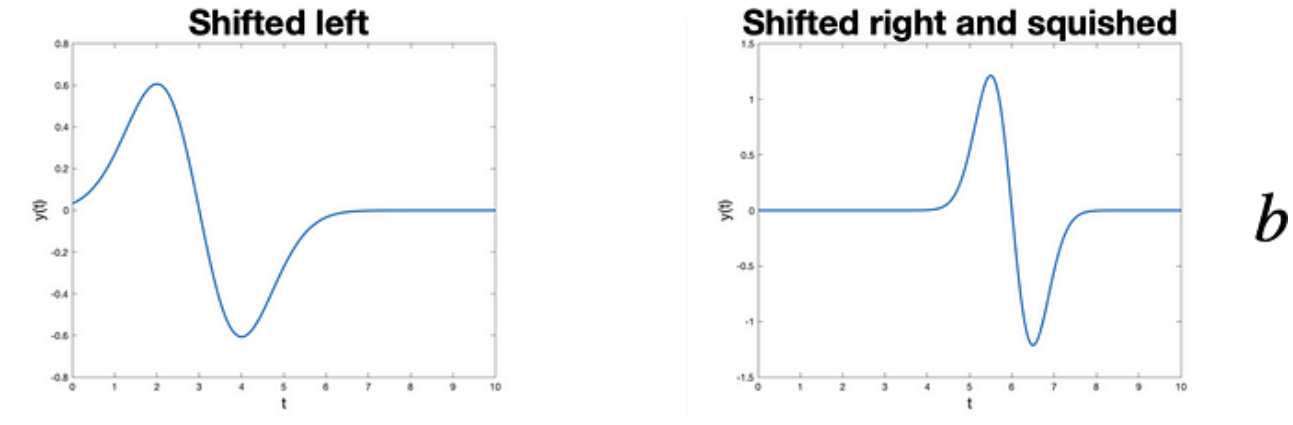

Рис. 2.4 Ліворуч: Приклад вейвлету зі зменшеним розташуванням. Праворуч: Приклад вейвлету зі збільшеним розташуванням та зменшеним масштабом

Основна ідея полягає у визначенні кількості вейвлетів у сигналі для певного масштабу і розташування. Це означає, що спочатку обирається вейвлет певного масштабу, після чого цей вейвлет зміщується по всьому сигналу, змінюючи його розташування. На кожному кроці часу проводиться множення вейвлету на сигнал, і отримується коефіцієнт для даного масштабу вейвлету. Після цього масштаб вейвлету збільшується, і процес повторюється.

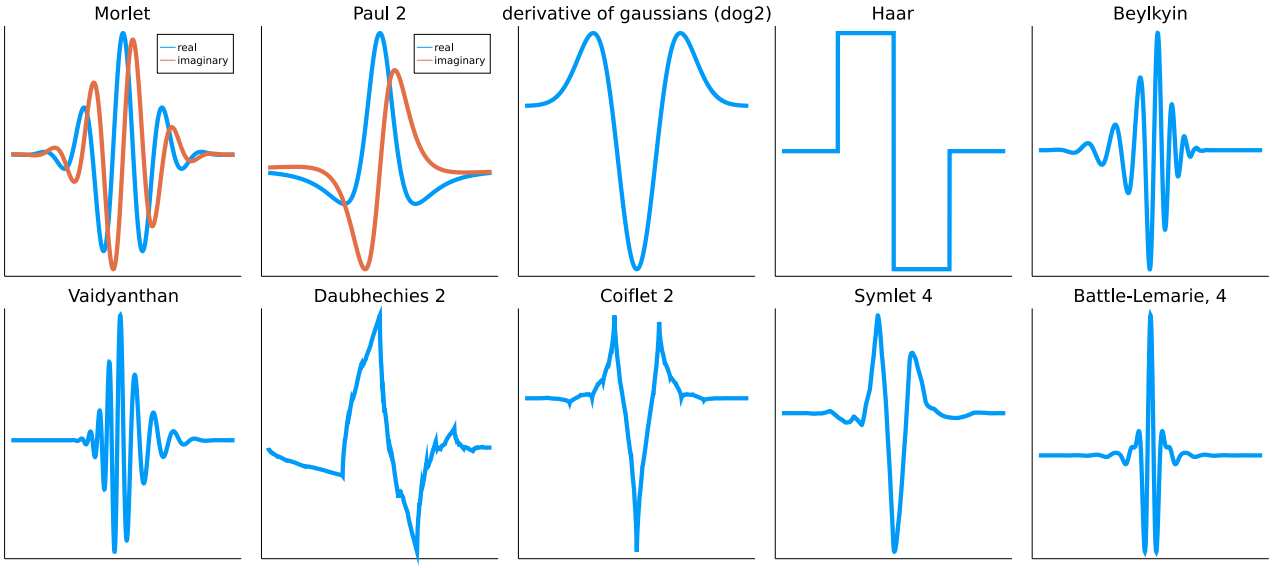

Рис. 2.5 Сімейства вейвлетів

Існує два типи вейвлет-перетворень: безперервне (Continuous Wavelet Transform, CWT) та дискретне (Discrete Wavelet Transform, DWT) [21-22].

2.4.1 Неперервне вейвлет-перетворення

$$T(a,b)=\frac{1}{\sqrt{a}}\int_{-\infty}^{\infty}x(t)\psi^{*}\frac{(t-b)}{a}dt \qquad (2.12)$$

Неперервне вейвлет-перетворення - це часово-частотне перетворення, цілком придатне для аналізу нестаціонарних сигналів, яке представляє собою часово-частотне перетворення. Нестаціонарність сигналу вказує на зміну його представлення в частотній області протягом часу. Це перетворення схоже на

віконне перетворення Фур'є (STFT), однак воно використовує вікна змінної ширини для розбиття частотно-часової площини. Розмір вікна збільшується з часом, що дозволяє аналізувати низькочастотні явища, і зменшується для високочастотних явищ. Безперервне вейвлет-перетворення можна успішно використовувати для аналізу перехідних процесів, швидких змін частот та повільних змін в поведінці.

2.4.2 Дискретне вейвлет-перетворення

$$T_{m,n} = \int_{-\infty}^{\infty} x(t)\psi_{m,n}(t)dt. \tag{2.13}$$

Основна відмінність між цими двома типами полягає в тому, що безперервне вейвлет-перетворення (CWT) використовує нескінченну кількість вейвлетів у всьому діапазоні масштабів і розташувань. Дискретне вейвлет-перетворення (DWT) використовує лише обмежений набір вейвлетів, які визначені на певному наборі масштабів і розташувань. За допомогою DWT масштаби дискретизуються менш деталізовано, ніж за допомогою CWT. Це робить DWT корисним для стиснення та зменшення шуму сигналів, зберігаючи важливі функції. Крім того, DWT є швидшим за CWT, що має значення для наших досліджень.

2.4.3 Порівняння Вейвлет перетворення з Фур’є перетворенням

На рисунку 2.6 [23] можна побачити порівняння часових рядів, Фур’є перетворення, віконне Фур’є перетворення та Вейвлет-перетворення. Часові ряди мають фіксовану роздільну здатністю за часом і не мають роздільної здатності за частотою. У перетворенні Фур'є ситуація навпаки: з фіксованою частотною роздільною здатністю і без часової роздільної здатності.

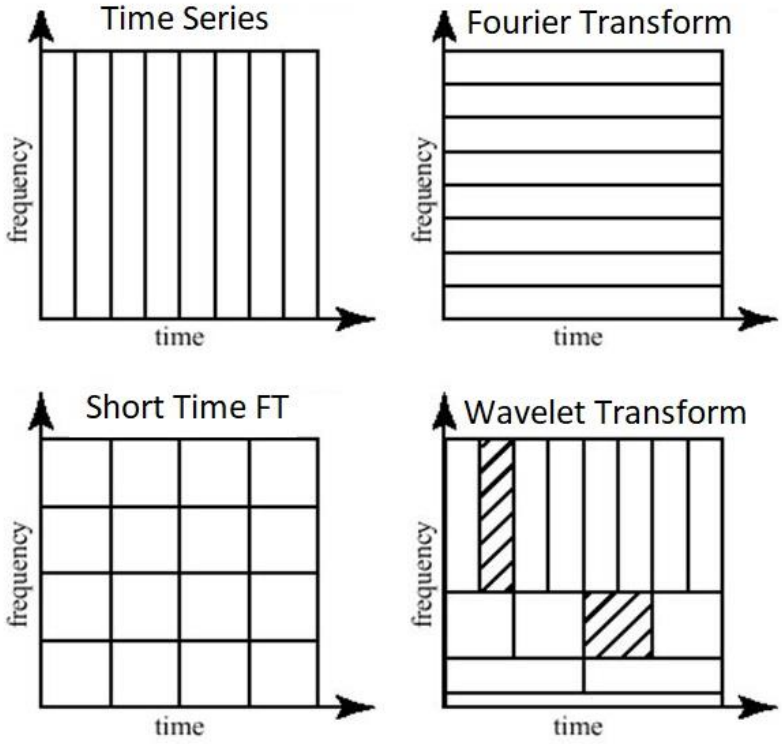

Рис. 2.6 Порівняння FFT та Wavelet

Натомість маємо приклад із віконним перетворенням Фур'є, де час розбивається на різні вікна. Тепер ми можемо мати як фіксовану часову, так і частотну роздільну здатність у кожному вікні. Слід зазначити:

- Вузьке вікно - це хороша роздільна здатність за часом, але погана роздільна здатність за частотою.
- Широке вікно - погана роздільна здатність за часом, хороша роздільна здатність за частотою.
- Низькочастотні компоненти часто тривають довгий період часу, тому потрібна висока частотна роздільна здатність.
- Високочастотні компоненти часто з'являються у вигляді коротких сплесків, що вимагає вищої роздільної здатності за часом.

При цьому вейвлет-перетворенні ми можемо адаптувати ширину вейвлету так, щоб його часова роздільна здатність відповідала частоті, яка нас цікавить. Отже маємо:

- Хорошу роздільну здатність за часом і погану роздільну здатність за частотою на високих частотах.
- Хорошу частотну роздільну здатність і погану часову роздільну здатність на низьких частотах.

Таким чином, ми отримуємо кращий компроміс між роздільною здатністю та частотою [24].

Принцип невизначеності Гейзенберга стверджує, що неможливо точно виміряти одночасно роздільну здатність (просторову або часову) і частотну роздільність сигналу. Коли ми аналізуємо нестаціонарні сигнали, як от сигнали зі змінною частотою або змінною амплітудою, такі методи, як віконне перетворення Фур'є (STFT), можуть мати обмежену ефективність через проблему невизначеності Гейзенберга.

Однак, вейвлет-перетворення є альтернативним методом аналізу нестаціонарних сигналів, який враховує принцип невизначеності Гейзенберга. Це дозволяє досягти більш точного аналізу нестаціонарних сигналів, оскільки вейвлети забезпечують кращу локалізацію в часі і частотному просторі, ніж стандартні базисні функції, що використовуються в STFT, такі як синусоїди. Виходячи з цих властивостей для аналізу та класифікації звукових образів було обрано дискретне вейвлет-перетворення.

РОЗДІЛ 3.
СТВОРЕННЯ ТА ЗАСТОСУВАННЯ НОВИХ МЕТОДІВ КЛАСИФІКАЦІЇ

3.1 Звукові образи для розпізнавання

Для розгляду та створенню алгоритмів і методів розпізнавання звукових образів було вирішено використати звукові сигнали крилатих ракет та бойових реактивних літаків для їх подальшого розпізнавання. Також, для дослідження точності розпізнавання використовуються звукові образи запуску двигуна автомобіля, стукіт пальців по столу та звуки гвинтокрилів. Деякі аудіосигнали були записані на мікрофон мобільного телефона, інші були вирізані з відео, яке також було записано на телефон. Іншими словами – всі звукові образи були записані на непрофесійні пристрої, через що в аудіо присутній шум, а іноді й інші звукові образим окрім дослідженого, що створює додаткову складність при розпізнаванні.

Ціль дослідження полягає в тому, щоб звуки крилатих ракет було можливо точно відрізнити від звуку двигуна реактивного літака, який має велику подібність до звуків двигуна ракети. Інші звуки повинні відразу відсіюватися. Аналогічна ситуація для літаків – потрібно зробити алгоритм, який дозволить розпізнавати модель при різній швидкості реактивного літака на допустимій відстані об’єкту до мікрофона.

3.2 Розпізнавання звукових образів за допомогою спектрограми

Після ознайомлення з алгоритмом "Shazam" з'явилася ідея дослідити методи розпізнавання звукових сигналів на основі ідентифікації пісень. Для початку розглянемо спектрограму композиції «Blurred Lines», якої була узята зі статті Уїлла Древо [25], та звуку ракети.

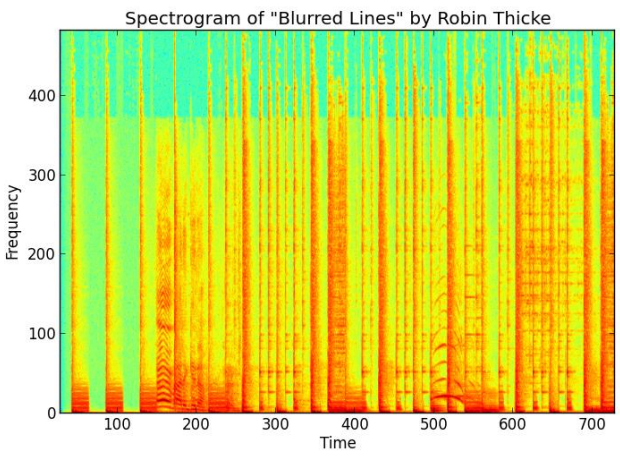

Рис. 3.1 – Спектрограма пісні

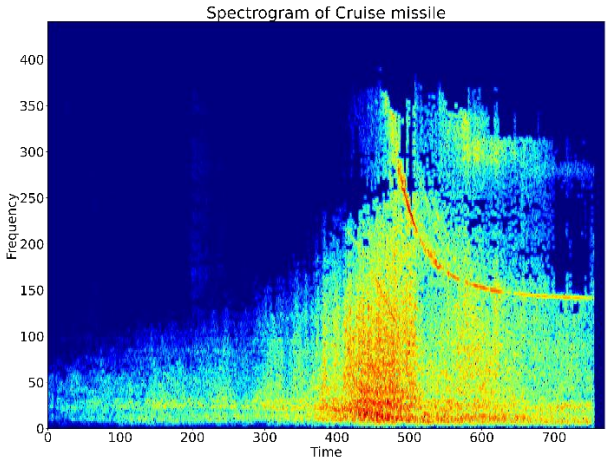

Рис. 3.2 – Спектрограма ракети

Обидві спектрограми звукових образів майже на 7 секунд. І з цих прикладів видно, що стаціонарний сигнал пісні краще структурований, ніж нестаціонарний сигнал ракети. Повторюючи дії алгоритму сузір'я не дадуть прийнятного результату. Також слід зазначити, що пісня, зазвичай, має тривалість більше хвилини, у той час як звук ракети, літака та навіть спів пташок приблизно до хвилини, чи навіть декількох секунд. Тому і відбитків пальців буде набагато менше, що означає гіршу розпізнаваність. Отже подальші дослідження в цій області, при поточних методах, не є ефективними.

3.3 Розпізнавання звукових образів за допомогою скейлограми

Альтернативою спектрограми для Вейвлет-перетворення є скейлограма. Вейвлет-скейлограма представляє двовимірне відображення одновимірних даних. У скейлограмі горизонтальна вісь відображається час, а вертикальна вісь - масштаб або частоту. Колір або яскравість кожного пікселя вказує на інтенсивність або амплітуду сигналу в заданий час і на заданій масштабній частоті.

Для початку були побудовані скейлограми для звуку крилатої ракети як для CWT перетворення (рис. 3.3), так і для DWT перетворення (для 1го та 5го рівнів декомпозиції)(рис. 3.4 та рис. 3.5 відповідно).

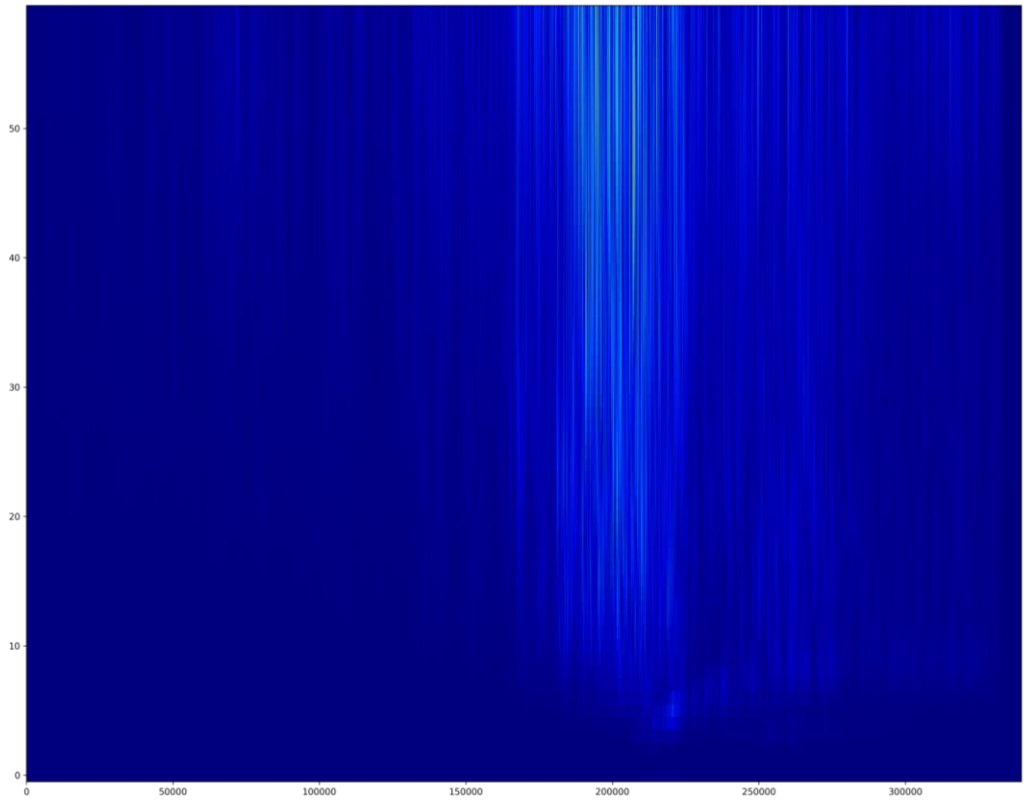

Рис. 3.3 Скейлограма CWT крилатої ракети

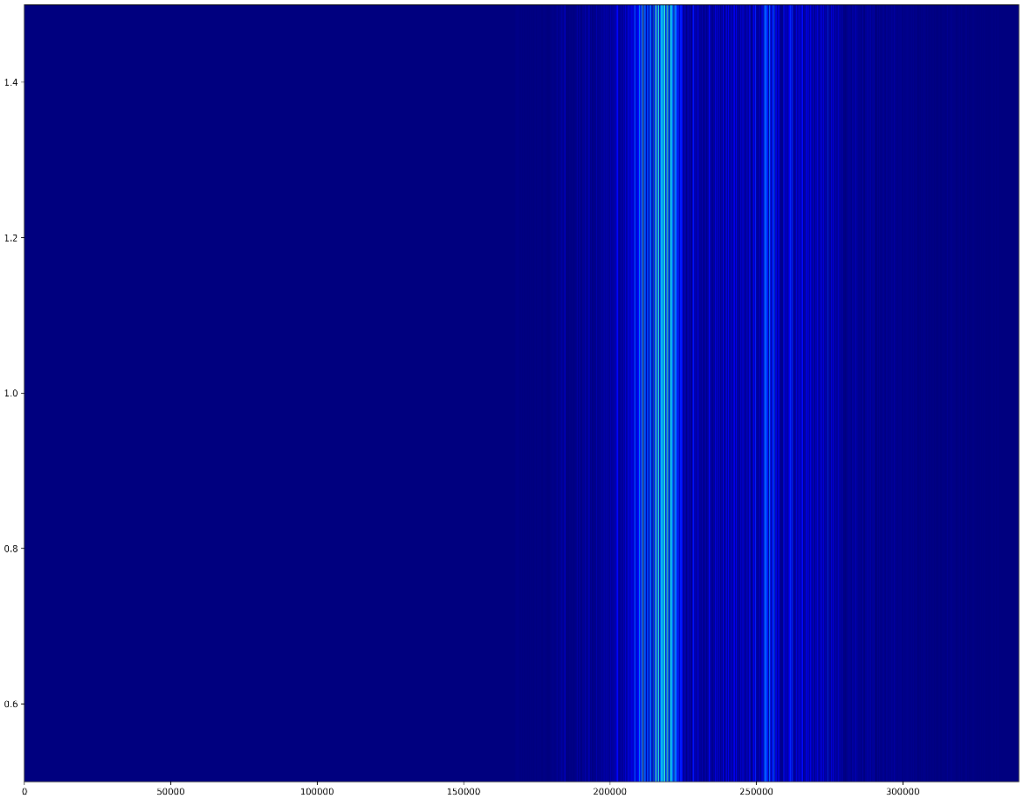

Рис. 3.4 Скейлограма DWT з 1 рівнем декомпозиції крилатої ракети

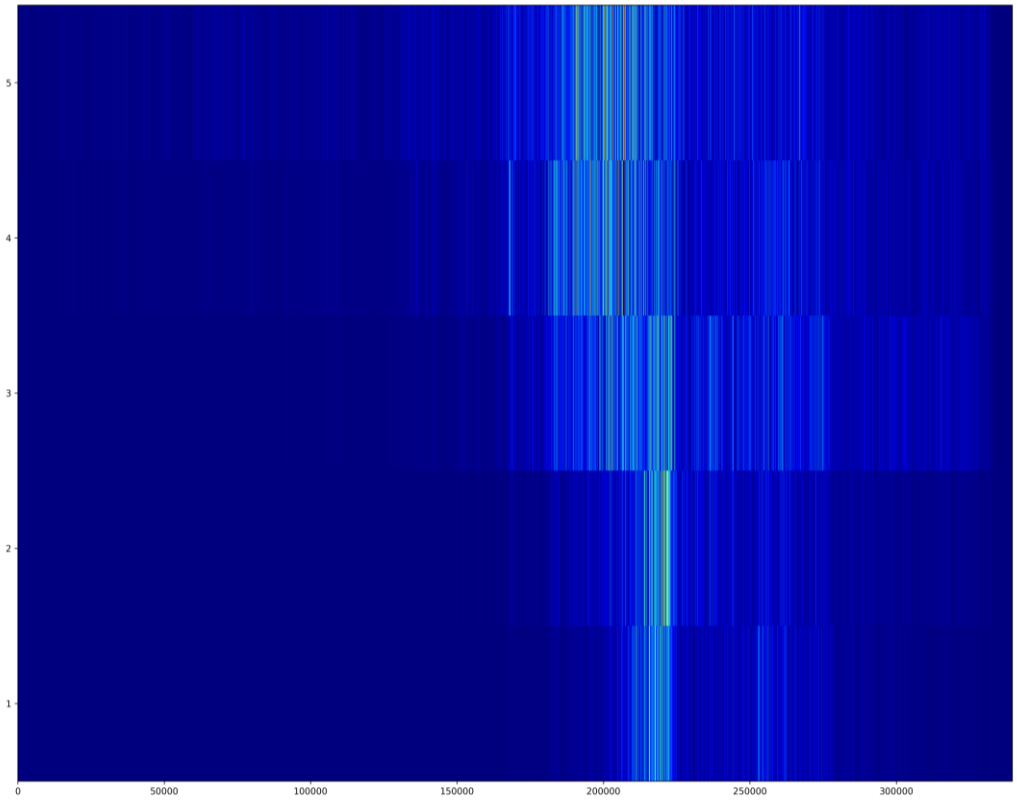

Рис. 3.5 Скейлограма DWT з 5 рівнями декомпозиції крилатої ракети

З отриманих скейлограм видно, що містять вони достатньо мало корисної інформації. Припускаючи, що це може бути проблема із записом аудіофайлу, розглянемо скейлограму неперервного вейвлет-перетворення та дискретного вейвлет-перетворення 5го рівня декомпозиції реактивного літака МіГ-29.

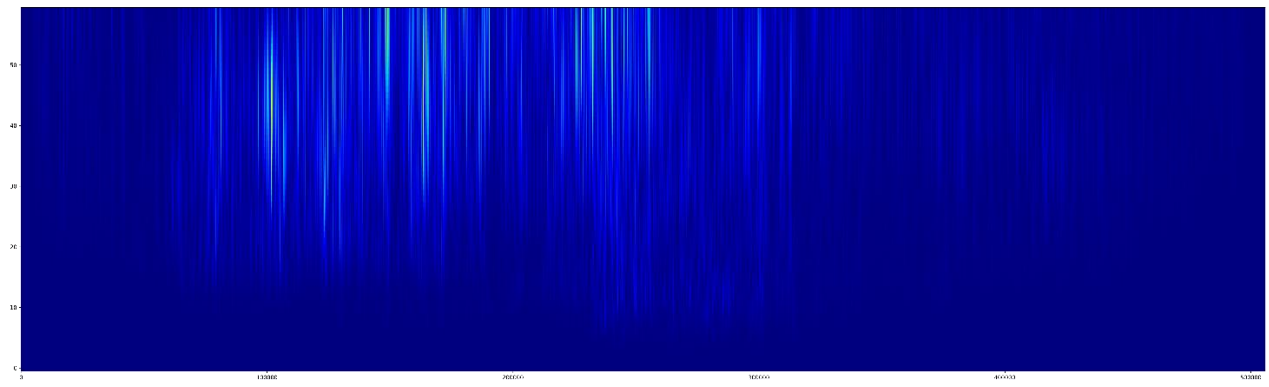

Рис. 3.6 Скейлограма CWT літака МіГ-29

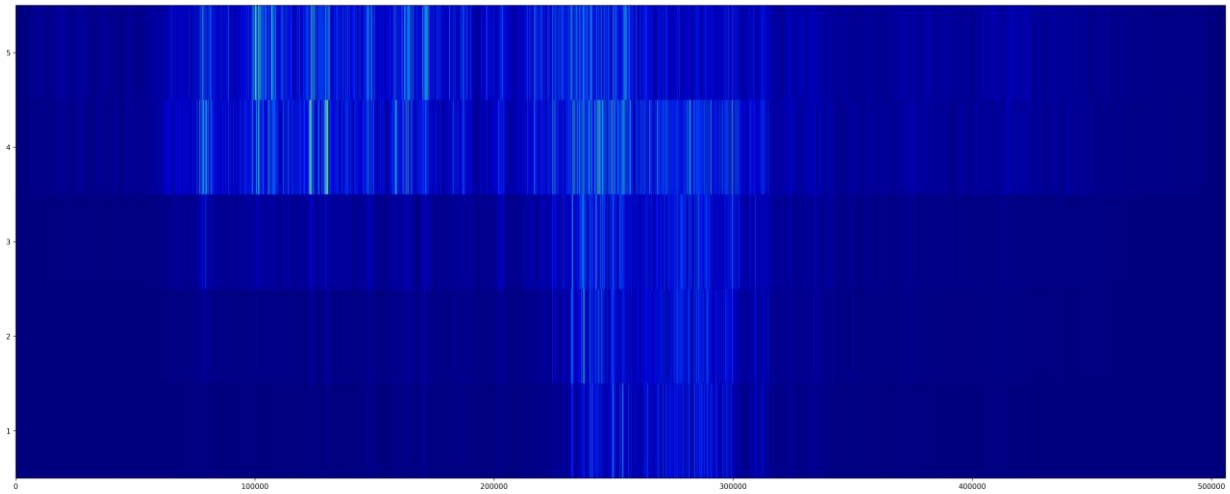

Рис. 3.7 Скейлограма DWT з 5 рівнями декомпозиції літака МіГ-29

Отримали майже аналогічні скейлограми. Проводячи експерименти із порівняння цих зображень, разом із іншими звуковими образами крилатих ракет та літаків у пошуку певних шаблонів зображення, прийшли до висновку, що порівняння поточних скейлограм не є ефективним методом порівняння. Алгоритми порівняння зображень частіше за все намагаються знайти певні особливості чи максимальні/мінімальні значення у зображенні. Отже при порівнянні скейлограм будемо отримувати точки у верхній чи ніжній частині рівнів декомпозиції, що не наддасть ніякої інформації для їх порівняння та класифікації.

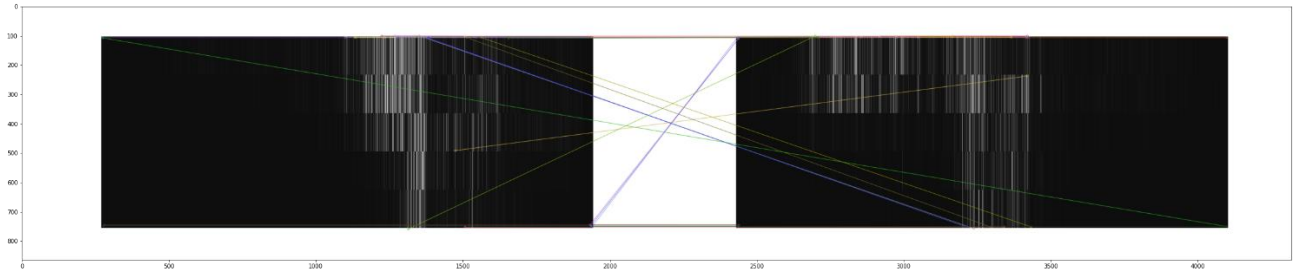

Рис. 3.8 Приклад порівняння скейлограм DWT ракети (зліва) та літака (справа) за допомогою відповідності локальних особливостей AKAZE

3.4 Пошук особливостей в звукових образах

Інша ідея полягає у виділенні особливостей аудіосигналу після дискретного Вейвлет перетворення. Перш за все – при використанні DWT можна видалити шум із аудіо-файлу, щоб краще класифікувати звук. Це актуальна проблема, бо більшість звуків записуються на непрофесійні записуючі пристрої та не у знешумленному приміщені. Як приклад, можна виділити наступні особливості для пошуку унікальних характеристик звукового образу:

1. Crest factor
2. Середнє значення піків сигналу
3. MFCC
4. Спектральні особливості (спектральна енергія, спектральна ширина смуги, спектральний центроїд, спектральна плотність і т. п.)
5. Стандартне відхилення
6. Середньоквадратичне значення
7. Коефіцієнт кореляції
8. Часові та частотні статистики

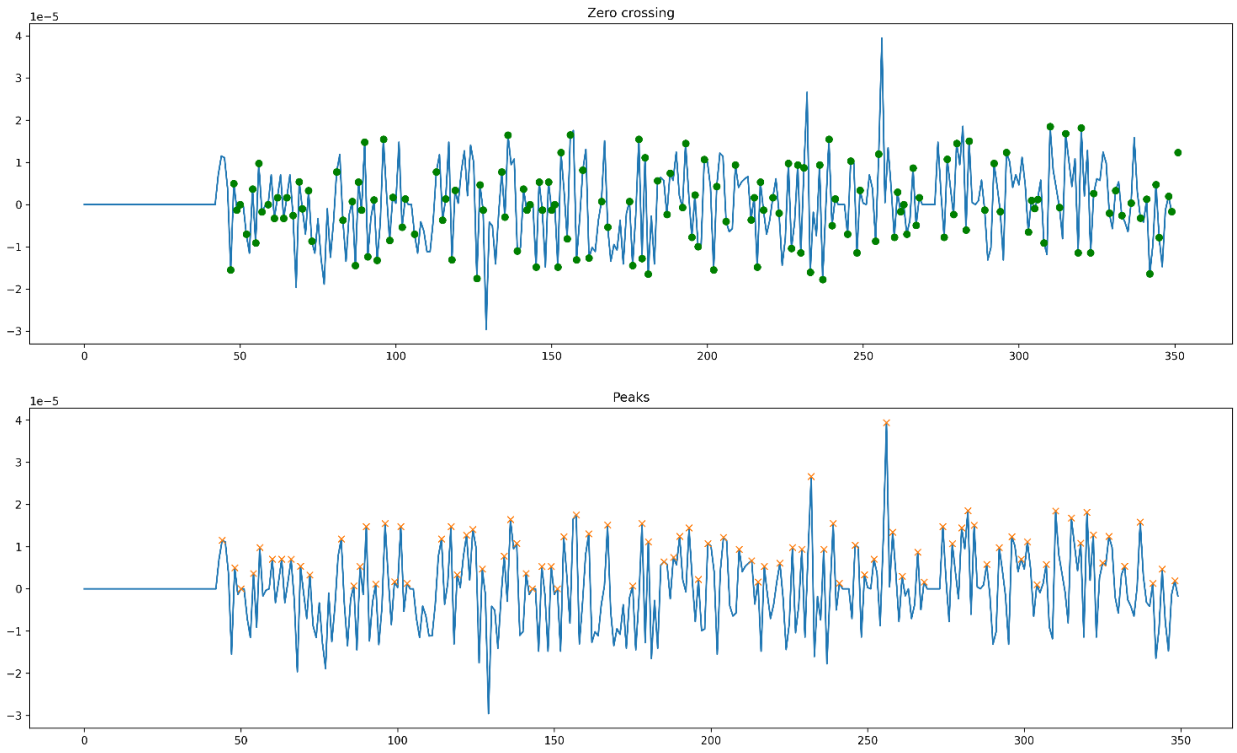

Рис. 3.9 Перехід через нуль та пікові значення сигналу

Для порівняння звукових образів за особливостями були зроблені наступні кроки:

- Крок 1. Попередня обробка у вигляді нормалізації за допомогою RMS (Root Mean Square) амплітуди сигналу від «-1» до «1».
- Крок 2. Нормалізований аудіосигнал подається на обробку з використанням дискретного вейвлет-перетворення (DWT) для визначення областей високої потужності в сигналі.
- Крок 3. За допомогою розрахунку медіанного абсолютного відхилення (MAD) від квадратів коефіцієнтів, обчислюються порогові значення.
- Крок 4. Визначаються області високої потужності шляхом порівняння квадратів коефіцієнтів з пороговим значенням. Якщо квадрат коефіцієнта перевищує поріг, він залишається, в іншому випадку встановлюється в нуль.
- Крок 5. Для виявлення особливостей в аудіосигналі використовується потужність коефіцієнтів апроксимації. Коефіцієнти деталізації нас не цікавлять, оскільки в них попадає весь низькочастотний шум під час декомпозиції сигналу.
- Крок 6. Конвертування фіч у вигляді чисел одинарної точності float32 у двійкові числа. Така дія значно зменшить розмір бази даних та прискорить порівняння.
- Крок 7. Об’єднання двійкових чисел в один рядок.
- Крок 8. Порівняння рядків різних звукових образів за допомогою відстані Геммінга [26].

Процес нормалізації амплітуди зазвичай включає в себе масштабування значень амплітуд сигналу таким чином, щоб максимальна амплітуда стала рівною 1 або іншому обраному максимальному значенню. В результаті всі значення амплітуд сигналу будуть знаходитися в діапазоні від -1 до 1. Це дозволить краще порівнювати аудіо сигнали, адже кожен сигнал був записаний

із різною відстанню до об’єкту та на різні пристрої. Процес нормалізації відбувався за допомогою RMS (Root Mean Square). Нормалізація за допомогою RMS передбачає обчислення RMS-значення амплітуди сигналу, а потім поділ всіх амплітудних значень на це значення для досягнення бажаного рівня гучності. Цей метод дозволяє зберегти відносні рівні гучності різних елементів всередині аудіо сигналу.

Важливо зауважити, що отримані фічі в більшості випадків є десятковим дробом. Тому конвертація таких чисел у двійковий формат дасть погані результати при порівнянні. Як приклад – двійкові числа можна порівняти за допомогою відстані Геммінга. І при порівнянні чисел 2 і 15 відстань між їх двійковими значеннями буде 4. А між 2 та 2.3 буде 11, що є критичним для порівняння особливостей аудіосигналу. Також значення отриманих фіч можуть бути числами з низькою величиною. Тому для таких випадків підбирався певний коефіцієнт для подання у тисячних значеннях. Після отримання та приведення фіч з малими величинами до тисячних значень (у протилежному випадку залишаємо як є) особливості округляємо до цілих значень для точнішого порівняння двійкових значень.

3.4.1 Результати першого порівняння звукових образів за особливостями

Для першого порівняння аудіосигналів було вирішено взяти crest factor та середньоквадратичне відхилення.

Виходячи з отриманих результатів табл. 3.1 можна зробити висновок, що особливості звукових образів дають значно кращий результат при порівнянні, ніж попередні методи. Було прийнято рішення, що сигнали вважатимуться подібними, якщо їх результат схожості перевищує 90%.

Таблиця 3.1

Результати першого порівняння звукових образів за особливостями

Відсоток схожості (crest factor та середньоквадратичне відхилення)		
	Ракета 1	МіГ-29 (1)
Ракета 1	100	84.38
Ракета 2	98.44	85.94
Ракета 3	79.69	82.81
Ракета 4	90.62	84.38
Ракета 5	89.06	89.06
МіГ-29 (1)	84.38	100
МіГ-29 (2)	84.38	87.5
Су-25 (1)	82.81	89.06
Су-25 (2)	84.38	84.38
Гвинтокрил	90.62	87.5
Двигун машини	85.94	82.81
Стук пальцями по столу	85.94	92.19

Досить непогані результати розпізнавання отримали при порівнянні звукових образів за допомогою crest factor та середньоквадратичного відхилення. 2 з 4 ракет розпізнаються доволі точно, четверта ракета має майже 90% схожості з «еталонним» звуком першої ракети, а от третя ракета, яка досить погано записана, має схожість лише 79.69%. Також отримано, що гвинтокрил схожий на першу ракету на 90.62%, що є доволі дивним. Інші звукові образи, такі як літаки, двигун автомобіля та стукіт пальці мають схожість не вище 86%.

Інша ситуація з результатами порівняння літака та звукових образів. МіГ-29 ні на що не схожий, окрім стуканню пальців по столу. Другий МіГ-29 було погано розпізнано. Результат порівняння співпадає з гвинтокрилом. І на 89.06%

образ сигналу першого МіГ-29 схожий на п’яту ракету. Варто зазначити, що наявні звукові образи Су-25 записані обрізано, та у житті їх двигуни відрізняються по звуку із «МіГом». Щодо ситуації з стуканням – використанні особливості не змогли краще відрізнити ці два звукові образи.

3.4.2 Результати другого порівняння звукових образів за особливостями

Для другого порівняння аудіосигналів було вирішено взяти crest factor, середньоквадратичне відхилення, spectral flatness та мелчастотні кепстральні коефіцієнти (Mel-frequency cepstral coefficients, MFCCs).

Таблиця 3.2

Результати другого порівняння звукових образів за особливостями

Відсоток схожості (crest factor, середньоквадратичне відхилення, spectral flatness, мелчастотні кепстральні коефіцієнти)		
	Ракета 1	МіГ-29 (1)
Ракета 1	100	82.03
Ракета 2	92.19	82.03
Ракета 3	81.25	83.59
Ракета 4	84.38	78.91
Ракета 5	83.59	85.94
МіГ-29 (1)	82.03	100
МіГ-29 (2)	85.94	86.72
Су-25 (1)	85.16	87.5
Су-25 (2)	85.16	84.38
Гвинтокрил	88.28	85.94
Двигун машини	88.28	82.81
Стук пальцями по столу	85.94	86.72

При додаванні spectral flatness та мелчастотні кепстральні коефіцієнти результат стає іншим. В цьому випадку, перша та друга ракети, які були добре записані, сходяться на 92.19%. Усі інші ракети мають збіжність менше 85%. Натомість гвинтокрил схожий менше на першу ракету – його результат 88.28, ніж 90.62% при першому порівнянні. У всіх інших відсоток схожості трошки став вищим, але все ще меншим 90%.

Ситуація з МіГ-29 стала краща. Результат порівняння другого МіГ-29 із першим має майже найвищий результат серед інших образів – 87.5%. Такий самий відсоток має стукіт пальців по столу, але при цьому його відсоток знизився в порівнянні із 92.19% першого розпізнавання. Серед інших звукових образів немає жодних подібностей.

3.5 Розпізнавання звукових образів за допомогою матриці особливостей

Наступним кроком в реалізації алгоритму розпізнаванні звукових образів було створення матриці особливостей аудіосигналу для детальнішого порівняння фіч. Сама матриця будується на основі n-ї кількості особливостей, які розташовуються по стовпцях. По рядкам тепер розглядається рівень декомпозиції, щоб поглянути, як особливості аудіосигналу змінюються з рівнем розкладання.

В практичному випадку було вирішено розглядати п’ять рівнів декомпозиції дискретного вейвлет-перетворення, щоб зберегти характеристику звукового образу та не видалити більше, ніж потрібно. Особливості знаходимо аналогічним чином, як у попередній реалізації з пошуком особливостей, окрім самого запису чисел. Тепер, щоб не втрачати інформацію через множення на коефіцієнт, округлення та переводячи числа в двійкове представлення, фічі записуються у звичайному вигляді десяткового дробу (float 32).

Для побудови матриці особливостей користуємося наступним алгоритмом:

- Крок 1. Нормалізація звукового образу за допомогою RMS.
- Крок 2. Нормалізований аудіосигнал подається на виконання DWT з 5 рівнями декомпозиції, записуючи в вихідний масив лише масив коефіцієнтів апроксимації. Коефіцієнти деталізації також не цікавлять в даній реалізації.
- Крок 3. Побудова матриці особливостей з використання масиву коефіцієнтів апроксимації та бажаних особливостей.
- Крок 4. Отримані матриці особливостей нормалізуємо по кожному стовпцю для отримання матриці у відтінках сірого, що містить значення в діапазоні від 0 (чорний) до 1 (білий).
- Крок 5. Порівняння двох зображень у відтінках сірого відбувається за допомогою метрики пікового співвідношення сигналу до шуму (англ. Peak signal-to-noise ratio (PSNR)).

Зазначимо, якщо результат при порівняння менший за 20 децибел – то два звукових образи сильно відрізняються. У випадку результату більше 35 децибел – звукові образи є схожими.

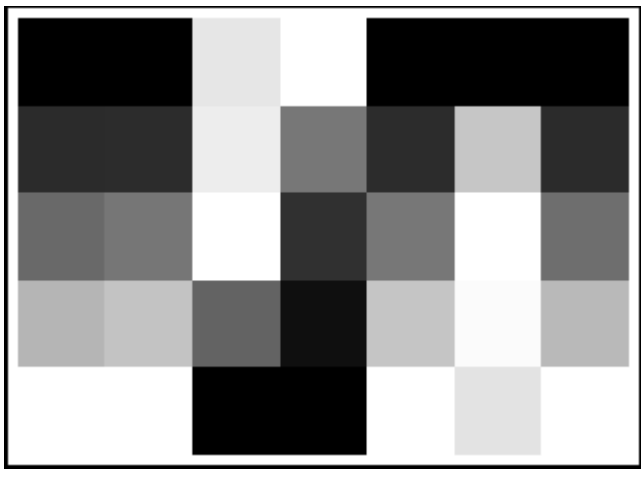

Рис. 3.10 Приклад матриці особливостей у відтінках сірого із 5 рівнями декомпозиції дискретного вейвлет-перетворення та 7 фічами

3.5.1 Результати першого порівняння матриць особливостей

Для першого порівняння скористаємося фічами crest factor та середньоквадратичним відхиленням, як у варіанті розпізнавання звукових образів за особливостями, оскільки вони дали непогані результати.

Рис. 3.11 Матриця особливостей для першої ракети

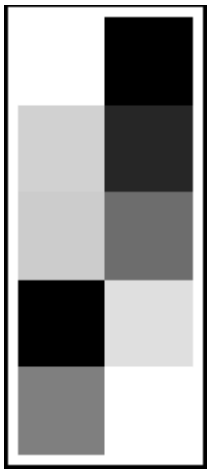

Рис. 3.12 Матриця особливостей для літака МіГ-29 (1)

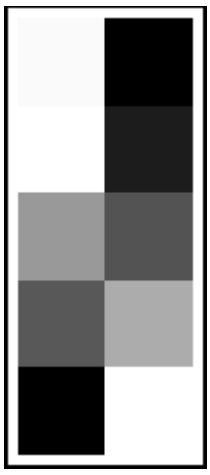

Рис. 3.13 Матриця особливостей для другої ракети

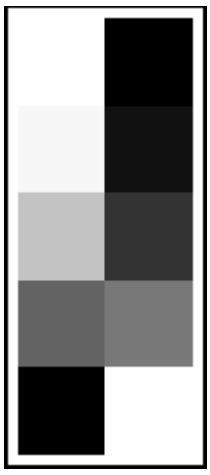

Рис. 3.14 Матриця особливостей для стукоту пальців по столу

З рисунків 3.11-3.14 можна побачити невелику відмінність у відтінках сірого, що дозволяє алгоритму саме за незначною зміною кольору зробити порівняння зображень.

Таблиця 3.3

Результати першого порівняння матриць особливостей

Значення PSNR у децибелах (crest factor та середньоквадратичне відхилення)		
	Ракета 1	МіГ-29 (1)
Ракета 1	∞	14.20
Ракета 2	17.75	14.02
Ракета 3	8.43	12.67
Ракета 4	19.14	12.17
Ракета 5	9.15	14.02
МіГ-29 (1)	14.20	∞
МіГ-29 (2)	9.66	14.70
Су-25 (1)	6.07	6.02
Су-25 (2)	6.09	7.27
Гвинтокрил	12.89	9.27
Двигун машини	9.43	7.18
Стук пальцями по столу	17.08	12.84

Результати вийшли всі менше 20 дБ, алгоритм не знайшов схожих особливостей у матрицях.

3.5.2 Результати другого порівняння матриць особливостей

Для другого порівняння скористаємося наступними фічами:

- середнє квадратичне значення,
- crest factor,
- середньоквадратичне відхилення,
- мелчастотні кепстральні коефіцієнти,

- нелінійна різницева частотна область,
- середнє значення піків аудіосигналу,
- спектральний центроїд.

В теорії ці особливості повинні краще виділяти характеристики аудіо.

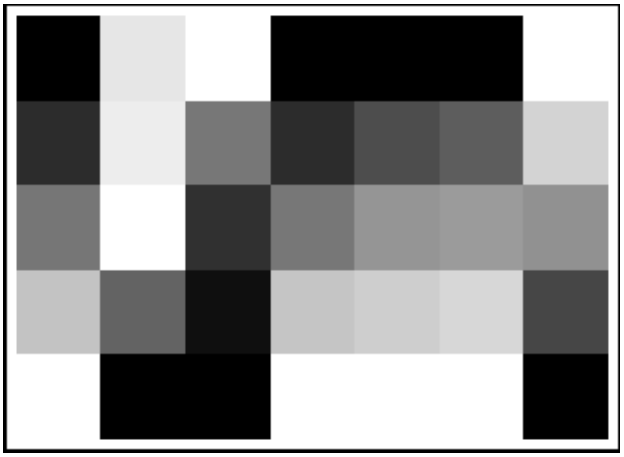

Рис. 3.15 Матриця особливостей для першої ракети

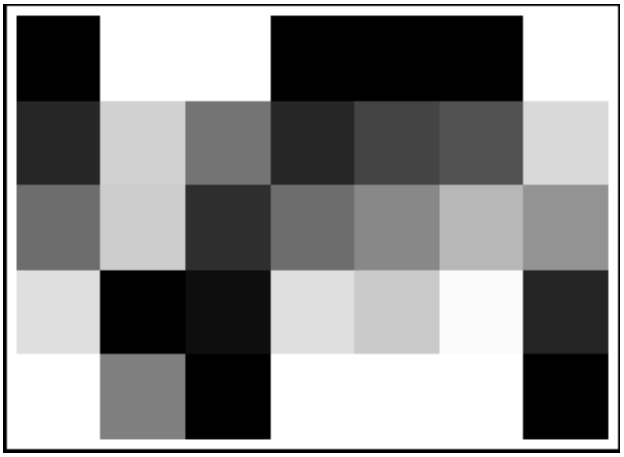

Рис. 3.16 Матриця особливостей для літака МіГ-29 (1)

Слід зауважити, що збільшення особливостей може погіршити результат розпізнавання, оскільки фічі, які погано знаходять особливості звукових образів, будуть виступати шумом для порівняння в метриці PSNR.

Хоча результатів порівняння у таблиці 3.4 більше 21 дБ немає, можна побачити значне покращення розпізнавання звукових образів. Порівнюючи з результатами звукових образів за особливостями (crest factor, середньоквадратичне відхилення) бачимо, що з 4 ракет 2 були майже розпізнанні. Більше немає збіжностей із гвинтокрилом чи іншими сигналами. З МіГ-29 ситуація також стала кращою. У попередніх результатах Мали схожість більше 89.06% (майже 90%, поріг, який був встановлений) з п’ятою ракетою та Су-25 (1). І 92.19% було у порівнянні з стуканням пальців по столу. Зараз же звуковий образ МіГ-29 значно відрізняється від інших. Але, на жаль, звуковий образ із іншим МіГ-29 не був розпізнаний. Результат всього 16.58 дБ. Іншим цікавим фактом є те, що Су-25 між собою теж мають 16.58 дБ.

Таблиця 3.4

Результати другого порівняння матриць особливостей

Значення PSNR у децибелах (середнє квадратичне значення, crest factor, середньоквадратичне відхилення, мелчастотні кепстральні коефіцієнти, нелінійна різницева частотна область, середнє значення піків аудіосигналу, спектральний центроїд)		
	Ракета 1	МіГ-29 (1)
Ракета 1	∞	18.49
Ракета 2	20.57	17.24
Ракета 3	13.01	16.31
Ракета 4	20.75	15.54
Ракета 5	13.65	17.11
МіГ-29 (1)	18.49	∞
МіГ-29 (2)	13.81	16.58
Су-25 (1)	9.35	9.34
Су-25 (2)	8.94	9.61
Гвинтокрил	16.16	12.99
Двигун машини	11.38	9.53
Стук пальцями по столу	16.18	13.69

Слід помітити, що при нормалізації матриці особливостей по стовпцям для отримання зображення у сірих відтінках трохи губиться унікальність особливостей. Це відбувається через те, що діапазон для нормування вибирається між мінімальним та максимальним значеннями у стовпці, а потім на основі цих значень нормалізується решта фіч. Тому дослідити різницю між двома однаковими особливостями стає складніше, бо два різні максимальні числа можуть інтерпретуватися як 1 (білий колір), і від цього залежить колір

інших фіч у стовпці. Натомість, без нормалізації по стовпцям будемо мати лише чорні стовпці (значення 0), оскільки особливості мають широкий діапазон значень і великі значення просто «перекривають» малі, та білі стопці (значення 1), де знаходяться максимальні чи наближені до них значення. Логарифмування значень трошки покращує зображення. В такому випадку стовпці мають «градієнт», який також не дає можливостей для кращого розпізнавання з збереженням унікальності значень особливостей.

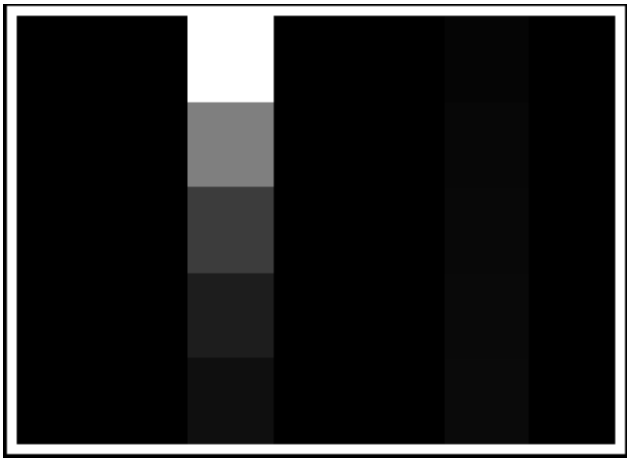

Рис. 3.17 Матриця особливостей без нормалізації по стовпцям

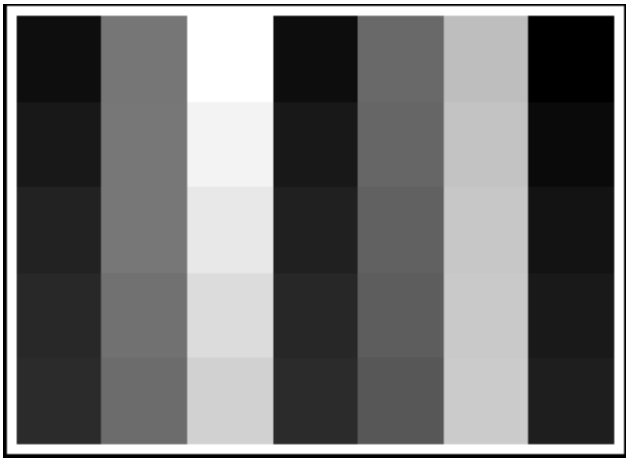

Рис. 3.18 Матриця особливостей з логарифмічною нормалізацією

3.6 Розпізнавання звукових образів за допомогою коваріаційної матриці

Під час порівняння матриць особливостей з’явилася ідея порівняння коваріаційних матриць. Ідея полягає в тому, щоб зробити дві коваріаційні матриці для порівнювальних аудіосигналів і вже за допомогою матриць у відтінках сірого та метрики PSNR визначити їх схожість. При цьому аудіосигнали, які порівнюються, зазвичай мають різну довжину. Для пришвидшення роботи було вирішено більший звуковий образ приводити до розміру меншого.

Для побудови коваріаційної матриці був реалізован наступний метод:

- Крок 1. Порівнюємо два вхідних аудіосигнали для визначення більшого і його подальшого зменшення.
- Крок 2. Нормалізація звукових образів за допомогою RMS.
- Крок 3. За допомогою дискретного вейвлет-перетворення вибираємо лише коефіцієнти апроксимації на 6-му рівні декомпозиції.
- Крок 4. Будуємо коваріаційну матрицю для обох сигналів.
- Крок 5. Використовуємо сингулярний розклад матриці (Singular Value Decomposition, SVD) для знаходження сингулярних значень більшої матриці.
- Крок 6. Отримані сингулярні значення більшого сигналу сортуються за спаданням. Чим менше значення – тим менше корисної інформації він містить.
- Крок 7. Залишаються найбільші значення такої кількісті, щоб відповідало розмірам меншого сигналу.
- Крок 8. З отриманих коваріаційних матриць одного розміру створюємо зображення у відтінках сірого. Виключенням є те, що тепер нормалізація по кожному стовпцю не потрібна.
- Крок 9. Порівнюємо отримані зображення за допомогою метрики PSNR для класифікації звукових образів.

Нормалізація амплітуди аудіосигналу відбувається як і у попередніх методах. Виключенням є те, що тепер не рахуються області високої потужності в сигналі за допомогою медіанного абсолютного значення (MAD). Щодо 6-го рівня декомпозиції DWT – то такий рівень було обрано для швидшого розрахунку коваріаційної матриці, методом зменшення даних вхідного сигналу без критичної втрати інформації.

Таблиця 3.5

Результати розпізнавання звукових образів за допомогою коваріаційної матриці

Значення PSNR у децибелах		
	Ракета 1	МіГ-29 (1)
Ракета 1	∞	48.66
Ракета 2	25.71	28.68
Ракета 3	82.85	48.50
Ракета 4	21.79	42.61
Ракета 5	21.79	21.39
МіГ-29 (1)	48.66	∞
МіГ-29 (2)	20.01	39.14
Су-25 (1)	18.81	23.55
Су-25 (2)	23.06	37.18
Гвинтокрил	33.07	34.63
Двигун машини	24.02	26.37
Стук пальцями по столу	30.57	30.56

Дивлячись на отримані результати можна зробити висновок, що розпізнавання звукових образів за допомогою коваріаційної матриці є поганим методом. Результати у таблиці свідчать про те, що ніякі звуки не були нормально класифіковані, а результати PSNR вийшли майже випадковими. Іншим недоліком виявився час, необхідний для порівняння двох звукових образів. Для прикладу час обробки та розпізнавання для ракети 1 та МіГ-29 (1) у середньому дорівнює 2 хвилини 51 секунди. Розпізнавання ракети 1 та ракети 2 відбувається у середньому за 55 секунд. Для порівняння час розпізнавання звукових образів при використанні матриці особливостей для ракети 1 та МіГ-29 (1) у середньому дорівнює 450 мілісекунд, для ракети 1 та ракети 2 – 300 мілісекунд.

У цілому метод із коваріаційною матрицею непогано працює для зменшення розміру зображень, оскільки в такому випадку розмір коваріаційної матриці буде значно меншим, тож і час розрахунків сингулярних значень буде швидшим. Іншим фактором є те, що високий рівень декомпозиції також може впливати на отримані результати. Хоч критичної втрати інформації не повинно бути, але при високому рівні декомпозиції відбувається зменшення розміру даних звукового образу. Це може видалити певні особливості, які могли б допомогти при аналізі сигналів. Тому подальше дослідження цього методу без глобальної модифікації є недоцільним.

РОЗДІЛ 4.
МОЖЛИВІ МОДИФІКАЦІЇ ТА ПОДАЛЬШІ КРОКИ НА ОСНОВІ ОТРИМАНИХ РЕЗУЛЬТАТІВ

Підбиваючи підсумки, можна впевнено сказати, що метод знаходження особливостей звукового образу за допомогою коефіцієнтів апроксимації дискретного вейвлет-перетворення є досить ефективним для своїх цілей. Ці методи із фічами показали непогану ефективність у розпізнаванні ракет, відкидаючи усі інші звукові образи.

Найефективнішим методом, опираючись на результати, є побудова матриці особливостей на основі декількох рівнів декомпозиції. Він працює краще, ніж звичайний метод з знаходженням фіч та їх подальшою конвертацією у двійкове число, що, як було описано в методі, може призвести до втрати певних характеристик самої особливості. Натомість, побудова матриці також впливає на особливості при нормалізації по стовпцям. При подальших дослідженнях слід знайти метод для створення матриці з можливістю зберігання унікальності кожної фічі на заданому рівні декомпозиції. Це повинно значно краще дозволити розпізнавати звукові образи. Іншою можливою модифікацією методу є знаходження нових особливостей. Непоганий результат у всіх дослідженнях показав crest factor та спектральні властивості сигналу. Можливо, що подальше додавання фіч, які можуть краще знаходити часово-частотні характеристики звукових образів.

Якщо розглядати побудову скейлограми, або спектрограми, то тут важливо зазначити, що при нестаціонарних звукових сигналах потрібно змінювати подання зображення. Необхідно отримати таке зображення, яке б дало змогу виділити певні особливості, чи навіть шаблон рисунку, а не вертикальні лінії, які взагалі неможливо ніяк оцінити.

Метод із побудовою коваріаційної матриці, на мою думку, не має подальшої актуальності у дослідженні. Одна з найголовніших причин – це час, яких необхідний для знаходження сингулярних значень да подальшої модифікації

матриці. Найдовшим звуковим образом, що використовувався в роботі, був образ МіГ-29 (1). Його тривалість трошки більше 7 секунд, але навіть при цьому алгоритм працює приблизно 3 хвилини. Такий результат є дуже критичним для розпізнавання звукових образів, адже подальші дослідження можуть включати в себе напряму запис образів на мікрофон і їх розпізнавання у реальному часі. Може бути ситуація, коли заздалегідь невідомо час потрапляння потрібного звукового образу на мікрофон. Навіть при допоміжному алгоритмі, який буде відсіювати шум, на запис може потрапити інший звук, який почне оброблятися за невідомий час. Можливо, що у майбутньому з’явиться прискорений алгоритм для знаходження сингулярних значень значно швидше, але поки робота з каваріаційною матрицею не надає прийнятних результатів.

Складніша ситуація з реактивними літаками. Їх звук надзвичайно сильно відрізняється, залежно від швидкості літака та допустимій відстані до мікрофона. З розглянутих методів у розділі 3 жоден не зміг добре класифікувати звукові образи літаків. Через це можна стверджувати, що при створенні алгоритму для розпізнавання образів літаків слід звернути увагу на інші характеристики сигналів, чи створити власний метод для отримання фіч. Алгоритм, що зможе якісно розпізнавати модель літака (а можливо й навіть швидкість та напрямок) буде здатен розпізнавати й інші звукові образи без труднощів. Тому є актуальність у дослідженні саме звукових образів реактивних літаків.

Отже, для подальшого дослідження одним з варіантів розвитку та модифікації є розгляд матриці особливостей із декількома рівнями декомпозиції. Також слід визначити кращі методи для знаходження особливостей звукових образів.

ВИСНОВОК

В даній роботі були досліджені та розроблені прототипи методів розпізнавання звукових образів, які можуть використовуватися у різних сферах, включаючи мілітарні застосунки, IoT, системи відеоспостереження та інше.

Дискретне вейвлет-перетворення (DWT) показало непогану ефективність у видаленні шуму, а також знаходження особливостей певних класів аудіосигналів. Доповнюючи DWT попередньою обробкою можна якісно відфільтрувати сигнал для подальшої класифікації.

З отриманих результатів можна зробити висновок, що використання методів для розпізнавання пісень не підходять для розпізнавання нетривалих, нестаціонарних звукових образів. Використання скейлограми на основі DWT, як при використанні спектрограми віконного Фур'є перетворення, не дає можливості отримати прийнятні результати. Натомість, використання дискретного вейвлет-перетворення для подальшого знаходження особливостей аудіосигналу із коефіцієнтів апроксимації дають значно кращі результати. При використанні таких фіч як crest factor та середньоквадратичне відхилення алгоритм дає змогу непогано розпізнати звуки ракет від інших звукових образів. А використання матриці особливостей, побудованій на основі коефіцієнтів апроксимації із декількома рівнями декомпозиції DWT, для порівняння і розпізнавання значно покращують результат класифікації аудіосигналів. Що не можна сказати про коваріаційну матрицю, яка дала досить випадкові результати.

На підставі аналізу результатів можна припустити, що знаходження особливостей та побудова матриці особливостей із декількома рівнями декомпозиції мають потенціал для подальшого дослідження у цій області. Щодо порівняння звукових образів за допомогою спектрограми та скейлограми було зроблено висновок, що у поточній реалізації немає актуальності їх використання. Можливо модернізація цих методів та їх побудова зображень зможуть дати більш прийнятні результати. Стосовно коваріаційної матриці робимо майже аналогічний висновок. Метод непогано працює для зображень, але, без значної

модернізації чи покращення, ніяк не для розпізнавання звукових образів, що мають значно більші розміри матриці та працюють набагато довше алгоритмів, що досліджувалися в даній роботі.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Avery Li-Chun Wang, Shazam Entertainment, Ltd. An Industrial-Strength Audio Search Algorithm, 2003.

2. Arvind Kumar, Sandeep Singh Solanki, Mahesh Chandra. Hilbert Spectrum Based Features for Speech/Music Classification, Vol 19, No.2, 2022, p. 239-259.

3. Andrey Osadchiy; Aleksandr Kamenev; Vladimir Saharov; Sergei Chernyi, Signal Processing Algorithm Based on Discrete Wavelet Transform. Designs. 2021. URL: https://www.researchgate.net/publication/353223804_Signal_Processing_Algorithm_Based_on_Discrete_Wavelet_Transform (дата звернення: 12.04.2023).

4. Garima Sharma, Kartikeyan Umapathy, Sridhar Krishnan. Trends in audio signal feature extraction methods. 2020.

5. ДСТУ 2325-93. Шум. Терміни та визначення.

6. ДСТУ 3515-97 Акустика й електроакустика. Терміни та визначення.

7. Boualem Boashash. Time-Frequency Signal Analysis and Processing: A Comprehensive Reference, 2003.

8. Chris Chatfield. The Analysis of Time Series: An Introduction, Fifth Edition (2016).

9. Julius O. Smith III, “Mathematics of the Discrete Fourier Transform (DFT), with Audio Applications --- Second Edition”, 2007. 247 p.

10. G. E. P. Box, L. R. Connor, W. R. Cousins, O. L. Davies (Ed.), F. R. Hirnsworth & G. P. Silitto, The Design and Analysis of Industrial Experiments, Oliver & Boyd, Edinburgh, 1954. I. J. Good, "The interaction algorithm and practical Fourier series," J. Roy. Statist. Soc. Ser. B., v. 20, 1958, p. 361–372; Addendum, v. 22, 1960, pp. 372–375.

11. Thakur, B., & Mehra, R. Discrete Fourier Transform Analysis with Different Window Techniques Algorithm. International Journal of Computer Applications (0975-8887), Volume 125-No.12, 2016.

12. Omar Alkousa, Learn Discrete Fourier Transform (DFT). URL: https://towardsdatascience.com/learn-discrete-fourier-transform-dft-9f7a2df4bfe9 (дата звернення: 27.05.2023).

13. Kong, Q., Siauw, T., & Bayen, A. Fourier Transform. Python programming and numerical methods: A guide for engineers and scientists, Academic Press, 2020.

14. Alan V. Oppenheim; Ronald W. Schafer; John R. Buck. Discrete-time signal processing (2nd ed., international edition), 1999. 893 p.

15. Ulrich Oberst, The Fast Fourier Transform, 2007. SIAM Journal on Control and Optimization 46(2): pp. 496-540. URL: https://www.researchgate.net/publication/220259110_The_Fast_Fourier_Transform (дата звернення: 11.03.2023).

16. Paul Heckbert, Fourier Transforms and the Fast Fourier Transform (FFT) Algorithm, Feb. 1995 (Revised 27 Jan. 1998).

17. Jeon, Hohyub & Jung, Yongchul & Lee, Seongjoo & Jung, Yunho. (2020). Area-Efficient Short-Time Fourier Transform Processor for Time–Frequency Analysis of Non-Stationary Signals. Applied Sciences. URL: https://www.researchgate.net/publication/346243843_Area-Efficient_Short-Time_Fourier_Transform_Processor_for_Time-Frequency_Analysis_of_Non-Stationary_Signals (дата звернення: 16.03.2023).

18. Sejdic, Ervin & Djurovic, Igor & Jiang, Jin. (2009). Time–frequency feature representation using energy concentration: An overview of recent advances. Digital Signal Processing. 2019. URL: https://www.researchgate.net/publication/223900360_Time-frequency_feature_representation_using_energy_concentration_An_overview_of_recent_advances (дата звернення: 23.05.2023).

19. Debnath, Lokenath & Antoine, Jean-Pierre. Wavelet Transforms and Their Applications. Physics Today - PHYS TODAY, 2003. URL: https://www.researchgate.net/publication/238958371_Wavelet_Transforms_and_Their_Applications (дата звернення: 02.06.2023).

20. Shawhin Talebi, The Wavelet Transform. An Introduction and Example. 2020. URL: https://towardsdatascience.com/the-wavelet-transform-e9cfa85d7b34 (дата звернення: 01.06.2023).

21. Heil, Christopher E., and David F. Walnut. “Continuous and Discrete Wavelet Transforms.” SIAM Review, vol. 31, no. 4, 1989, pp. 628–66.

22. Florian Bömers, Wavelets in real time digital audio processing: Analysis and sample implementations, 2000, pp. 36-38.

23. Mostafa, Abeer & Barghash, Toka & Assaf, Asmaa & Gomaa, Walid. Multi-sensor Gait Analysis for Gender Recognition. 2020, pp. 629-636. URL: https://www.researchgate.net/publication/342995592_Multi-sensor_Gait_Analysis_for_Gender_Recognition (дата звернення: 19.04.2023)

24. S. A. A. Karim, M. H. Kamarudin, B. A. Karim, M. K. Hasan and J. Sulaiman, "Wavelet Transform and Fast Fourier Transform for signal compression: A comparative study," 2011 International Conference on Electronic Devices, Systems and Applications (ICEDSA), Kuala Lumpur, Malaysia, 2011, pp. 280-285.

25. Will Drevo, Audio Fingerprinting with Python and Numpy, 2013. URL: https://willdrevo.com/fingerprinting-and-audio-recognition-with-python/ (дата звернення: 18.02.2023).

26. Richard W. Hamming, Error detecting and error correcting codes, 1950. Bell System Technical Journal 29 (2): pp. 147–160.