

Вінницький національний технічний університет
Факультет інтелектуальних інформаційних технологій та автоматизації
Кафедра автоматизації та інтелектуальних інформаційних технологій

Бакалаврська кваліфікаційна робота на тему:

«РОЗРОБКА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ АВТОМАТИЗОВАНОЇ
СИСТЕМИ ІДЕНТИФІКАЦІЇ ТА ЗАХИСТУ АУДІОКОНТЕНТУ»

Виконав: студент IV курсу, групи ІАКІТ-216
спеціальності 151 – Автоматизація та
комп'ютерно-інтегровані технології

(шифр і назва спеціальності)

Антон ГНАТЕНКО

(прізвище та ініціали)

Керівник: доцент кафедри АІТ

Володимир ГАРМАШ

(прізвище та ініціали)

« 16 » 06 2025 р.

Рецензент: к.т.н., проф. канд. р.с.ч.

Микола БІСІКАЛО

« 18 » 06 2025 р.

Допущено до захисту
завідувач кафедри АІТ
Олег БІСІКАЛО

« 19 » 06 2025 р.

Вінниця ВНТУ – 2025 рік

Вінницький національний технічний університет
(повне найменування вищого навчального закладу)

Факультет інтелектуальних інформаційних технологій та автоматизації
Кафедра автоматизації та інтелектуальних інформаційних технологій
Рівень вищої освіти перший (бакалаврський)
Галузь знань 15 - Автоматизація та приладобудування
Спеціальність Автоматизація та комп'ютерно-інтегровані технології
Освітньо-професійна програма Інтелектуальні комп'ютерні системи управління

ЗАТВЕРДЖУЮ

завідувач кафедри АІТ

д.т.н., проф. Олег БІСІКАЛО

«19» 06 2025 року

ЗАВДАННЯ

НА БАКАЛАВРСЬКУ КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТУ

Гнатенко Антону Валерійовичу

(прізвище, ім'я, по батькові)

1. Тема роботи «Розробка програмного забезпечення для автоматизованої системи ідентифікації та захисту аудіоконтенту»

керівник роботи Гармаш Володимир Володимирович, к.т.н., доцент

(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом вищого навчального закладу від «20» березня 2025 року №97

2. Термін подання студентом роботи 10.06.2025 р.

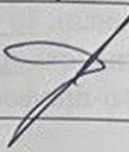
3. Вихідні дані до роботи:

- тип аудіоконтенту – монофонічний та стереофонічний;
- формат аудіоданих – WAV, MP3, AAC;
- мінімальна частота дискретизації – 16 кГц;
- максимальна частота дискретизації – 48 кГц;
- точність ідентифікації аудіоконтенту – від 90% до 99% залежно від рівня шуму та якості запису.

4. Зміст текстової частини (перелік питань, які потрібно розробити) Вступ. Загальні відомості про системи захисту та ідентифікації аудіоконтенту. Розробка системи ідентифікації та захисту аудіоконтенту. Розробка програмного забезпечення. Висновки.

5. Перелік ілюстративного матеріалу (з точним зазначенням обов'язкових креслень) схема запропонованого алгоритму отримання вектора на підставі ознак спектра модуляції, головне вікно робочої програми, блок представлення цифрових аудіовідбитків, Залежність коефіцієнта бітових відмінностей від зсуву при порівнянні музичного відбитка з вихідною послідовністю, в присутності шуму SNR = 0db (Б) і масштабування в часі на 15%.

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	виконання прийняв
1-3	к.т.н. доцент каф. АПТ Володимир ГАРМАШ		

7. Дата видачі завдання 20.03.2025 р.

КАЛЕНДАРНИЙ ПЛАН

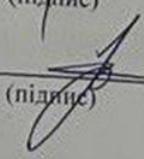
№ з/п	Назва етапів бакалаврської кваліфікаційної роботи	Термін виконання		При-мітка
1	Вибір, узгодження та затвердження теми БКР	03.10.2024	09.10.2024	<i>вик.</i>
2	Аналіз предметної області та опрацювання літературних джерел.	12.03.2025	29.03.2025	<i>вик.</i>
3	Постановка задач для вирішення в БКР	01.04.2025	26.04.2025	<i>вик.</i>
4	Вирішення поставлених задач	29.04.2025	10.05.2025	<i>вик.</i>
5	Підтвердження достовірності отриманих результатів	13.05.2025	24.05.2025	<i>вик.</i>
7	Оформлення пояснювальної записки та графічної частини	27.05.2025	31.05.2025	<i>вик.</i>
8	Перевірка ПЗ на відповідність вимогам	03.06.2025	07.06.2025	<i>вик.</i>
9	Попередній захист, доопрацювання, та рецензування БКР	10.06.2025	15.06.2025	<i>вик.</i>
10	Захист БКР	16.06.2025	23.06.2025	<i>вик.</i>

Студент

Керівник роботи


(підпис)

АНТОН ГНАТЕНКО
(ім'я та прізвище)


(підпис)

ВОЛОДИМИР ГАРМАШ
(ім'я та прізвище)

АНОТАЦІЯ

Бакалаврська кваліфікаційна робота складається з 90 сторінок формату А4, включаючи додатки, що містить 19 рисунків. Список використаних джерел містить 29 найменувань.

Робота присвячена розробці програмного забезпечення для автоматизованої системи ідентифікації та захисту аудіоконтенту в комп'ютерно-інтегрованих системах на основі формування цифрових аудіовідбитків. Метою є створення ефективного інструменту для розпізнавання аудіоданих та їх захисту від несанкціонованого використання з мінімальними обчислювальними витратами. Проведено аналіз існуючих методів ідентифікації аудіоконтенту, досліджено можливості вдосконалення алгоритмів спектрального аналізу та їх застосування в системах захисту. Запропоновано модифікований алгоритм формування цифрових відбитків, який забезпечує точність ідентифікації до 99% при збереженні низької трудомісткості. Розроблено програмний модуль, який реалізує запропонований метод, та проведено експерименти, що підтверджують його ефективність у підвищенні продуктивності обробки аудіоданих у комп'ютерно-інтегрованих системах.

Ключові слова: аудіоконтент, цифровий відбиток; система захисту, ідентифікація аудіоконтенту, програмне забезпечення.

ABSTRACT

The bachelor's qualification work consists of 90 pages of A4 format, including appendices, containing 19 figures. The list of used sources contains 29 items.

The work is devoted to the development of software for an automated system for identifying and protecting audio content in computer-integrated systems based on the formation of digital audio fingerprints. The goal is to create an effective tool for recognizing audio data and protecting it from unauthorized use with minimal computational costs. An analysis of existing methods for identifying audio content has been conducted, the possibilities of improving spectral analysis algorithms and their application in protection systems have been investigated. A modified algorithm for forming digital fingerprints has been proposed, which provides identification accuracy up to 99% while maintaining low labor intensity. A software module has been developed that implements the proposed method, and experiments have been conducted to confirm its effectiveness in increasing the productivity of audio data processing in computer-integrated systems.

Keywords: audio content, digital fingerprint; protection system, audio content identification, software.

ЗМІСТ

ВСТУП.....	4
1 ЗАГАЛЬНІ ВІДОМОСТІ ПРО СИСТЕМИ ЗАХИСТУ ТА ІДЕНТИФІКАЦІЇ АУДІОКОНТЕНТУ.....	7
1.1 Аналіз об'єкту автоматизації	7
1.2 Вимоги до системи захисту та ідентифікації аудіоконтенту.....	10
1.3 Аналіз складових комплексної системи захисту аудіовізуального контенту від незаконного копіювання та розповсюдження	14
1.4 Основні принципи отримання та використання цифрових відбитків в системах ідентифікації аудіоконтенту	18
1.5 Висновок до розділу 1	29
2 РОЗРОБКА СИСТЕМИ ІДЕНТИФІКАЦІЇ ТА ЗАХИСТУ АУДІОКОНТЕНТУ.....	31
2.1 Експериментальне порівняння і вибір дескрипторів аудіоматеріалів.	31
2.2 Аналіз стійкості дескрипторів на основі моделей сприйняття акустичних сигналів людиною	36
2.3 Метод цифрових аудіо відбитків.....	43
2.4 Висновок до розділу 2	52
3 РОЗРОБКА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ.....	53
3.1 Обґрунтування вибору мови програмування	53
3.2 Реалізація додатку для автоматизованої системи ідентифікації та захисту аудіоконтенту	54
3.3 Тестування розробленої програми	59
3.4 Характеристики підсистеми пошуку та розпізнавання аудіоматеріалів за цифровими відбитками	65
2.4 Висновок до розділу 3	72
ВИСНОВКИ	73
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	74

ДОДАТКИ	78
Додаток А (обов'язковий) – Технічне завдання	79
Додаток Б (обов'язковий) – Ілюстративна частина	83
Додаток В (обов'язковий) Лістинг модуля програми	87
Додаток Г (обов'язковий) Протокол перевірки кваліфікаційної роботи..	90

ВСТУП

Актуальність. У сучасному цифровому світі аудіоконтент відіграє ключову роль у різних сферах: від розважальної індустрії до освітніх платформ і корпоративних комунікацій. Зростання популярності стрімінгових сервісів, подкастів, аудіокниг та інших форм цифрового аудіо супроводжується стрімким збільшенням обсягів контенту, що створюється та поширюється. Проте це також породжує низку викликів, пов'язаних із захистом авторських прав, запобіганням піратству та забезпеченням безпеки аудіоданих. У цьому контексті розробка програмного забезпечення для автоматизованої системи ідентифікації та захисту аудіоконтенту набуває особливої актуальності.

Однією з ключових проблем є незаконне використання аудіоконтенту, що завдає значних збитків правовласникам. За даними Міжнародної федерації фонографічної індустрії (IFPI), у 2023 році втрати від піратства в музичній індустрії сягнули мільярдів доларів. Автоматизовані системи ідентифікації, такі як цифрові відбитки (fingerprinting) або водяні знаки (watermarking), дозволяють ефективно виявляти та відстежувати використання аудіофайлів у мережі, що сприяє захисту інтелектуальної власності. Такі технології дають змогу не лише виявляти порушення, а й автоматизувати процеси моніторингу, зменшуючи витрати часу та ресурсів.

Крім того, з появою штучного інтелекту та генеративних технологій з'явилися нові виклики, пов'язані з створенням глибоких фейків (deepfakes) та маніпуляціями з аудіоконтентом. Це створює загрозу для безпеки інформації, зокрема в медіа, судових процесах чи політичній сфері. Автоматизовані системи, які здатні ідентифікувати оригінальність аудіо та виявляти підробки, стають важливим інструментом для забезпечення довіри до цифрового контенту.

Розробка такого програмного забезпечення також відповідає сучасним тенденціям цифровізації та потребам автоматизації. Інтеграція алгоритмів машинного навчання та обробки великих даних дозволяє створювати високоточні системи, які можуть працювати в реальному часі, аналізуючи величезні обсяги аудіоконтенту. Це особливо актуально для платформ, які обробляють мільйони аудіофайлів щодня, таких як Spotify, YouTube чи SoundCloud [1-4].

Таким чином, створення програмного забезпечення для автоматизованої системи ідентифікації та захисту аудіоконтенту є актуальним завданням, яке вирішує проблеми захисту авторських прав, запобігання піратству та забезпечення безпеки цифрового середовища. Така розробка сприятиме розвитку медіаіндустрії, підвищенню довіри до цифрових платформ і створенню нових можливостей для правовласників та споживачів.

Для ефективного захисту мультимедійного контенту необхідна розробка нових методів ідентифікації та контролю за його поширенням. Одним із перспективних рішень є технологія «цифрових відбитків», яка аналогічна ідентифікації людини за відбитками пальців. Останнім часом ця технологія, зокрема метод аудіовідбитків, широко застосовується для моніторингу P2P-мереж, контролю трансляцій на радіо й телебаченні, перевірки рекламного контенту та виявлення піратських копій ліцензійних матеріалів. Аудіовідбитки стали основою міжнародних стандартів, зокрема MPEG-7, і викликають значний інтерес у компаній, що працюють у сфері доставки, пошуку та захисту мультимедійного контенту та інтелектуальної власності в Інтернеті [5-6].

Водночас існуючі методи цифрових відбитків і пошуку в базах даних досі не забезпечують повноцінної ідентифікації значної частини мультимедійного контенту, поширеного в Інтернеті, що й визначає актуальність подальших досліджень у цій сфері.

Метою даної роботи є підвищення ефективності автоматизованої системи ідентифікації, захисту та пошуку аудіоконтенту шляхом створення програмного забезпечення, яке використовує вдосконалений метод цифрового відбитка для забезпечення стійкості до спотворень, надійного захисту від несанкціонованого розповсюдження та швидкого доступу до аудіоданих.

Об'єктом дослідження даної бакалаврської роботи є процес ідентифікації, захисту та пошуку аудіоконтенту на основі методу цифрового відбитка.

Предметом дослідження даної бакалаврської роботи є методи захисту аудіоматеріалів від несанкціонованого копіювання.

Для досягнення мети в роботі поставлені наступні основні **задачі**:

- проаналізувати сучасний стан вирішення проблеми захисту мультимедійного контенту;
- розробити вдосконалений метод створення та порівняння цифрових відбитків, що підходить для широкого діапазону типів аудіосигналів;
- виконати тестування й оптимізацію запропонованих методів для забезпечення високої надійності захисту аудіофайлів від несанкціонованого поширення.

Практична цінність роботи полягає в розробці програмного забезпечення для автоматизованої системи ідентифікації та захисту аудіоконтенту на основі методу цифрових відбитків. Розроблений метод може бути використаний для ідентифікації та пошуку широкого класу акустичних сигналів, включаючи мову, музику, структурований і неструктурований шум, а також для виявлення матеріалів, що містять аудіоканал, зокрема телевізійних передач, фільмів і музичних кліпів. Система також придатна для моніторингу об'єктів з унікальним акустичним профілем і може використовуватись для контролю над поширенням захищених аудіоматеріалів.

1 ЗАГАЛЬНІ ВІДОМОСТІ ПРО СИСТЕМИ ЗАХИСТУ ТА ІДЕНТИФІКАЦІЇ АУДІОКОНТЕНТУ

1.1 Аналіз об'єкту автоматизації

Система ідентифікації та захисту аудіоконтенту є спеціалізованим програмним комплексом, призначеним для автоматизації процесів виявлення, аналізу, класифікації та захисту аудіоданих від несанкціонованого використання, копіювання чи модифікації. Об'єктом автоматизації в даному контексті виступає сукупність процесів, пов'язаних із обробкою аудіоконтенту, включаючи його ідентифікацію за унікальними характеристиками, моніторинг використання в різних середовищах (онлайн-платформи, радіо, телебачення, соціальні мережі тощо) та забезпечення захисту авторських прав. Система спрямована на підвищення ефективності управління аудіоконтентом, зниження ризиків піратства та забезпечення дотримання законодавчих норм у сфері інтелектуальної власності [7].

Об'єкт автоматизації включає кілька ключових компонентів, які взаємодіють для досягнення поставлених цілей:

1. Модуль ідентифікації аудіоконтенту.

Цей компонент відповідає за аналіз аудіоданих та їх унікальну ідентифікацію. Ідентифікація здійснюється шляхом створення цифрових "відбитків" (audio fingerprints), які є унікальними сигнатурами звукових файлів. Ці відбитки базуються на акустичних характеристиках, таких як спектральні особливості, темп, ритм, частотний діапазон та інші параметри. Модуль використовує алгоритми машинного навчання та обробки сигналів для порівняння аудіофрагментів із базою даних відомого контенту. Наприклад, система може розпізнавати пісню чи звуковий фрагмент навіть у разі спотворення, шумів чи зміни формату.

2. Модуль моніторингу та пошуку

Цей модуль забезпечує постійний моніторинг аудіоконтенту в різних джерелах, таких як стрімінгові платформи (YouTube, Spotify), соціальні мережі, радіостанції чи телевізійні трансляції. Він використовує технології веб-сканування та API-інтеграцію для автоматичного виявлення використання аудіофайлів. У разі виявлення контенту система зіставляє його з базою даних для визначення правовласника та законності використання.

3. Модуль захисту аудіоконтенту

Захист аудіоконтенту реалізується через кілька механізмів:

- Вбудовування цифрових водяних знаків: Цифрові водяні знаки (watermarks) додаються до аудіофайлів для підтвердження їх походження та авторства. Вони є стійкими до модифікацій, таких як стиснення чи перекодування.
- Шифрування: Для захисту від несанкціонованого доступу аудіофайли можуть шифруватися за допомогою сучасних криптографічних алгоритмів.
- Управління доступом: Система забезпечує контроль доступу до контенту, дозволяючи лише авторизованим користувачам відтворювати чи модифікувати файли.
- Автоматизоване сповіщення: У разі виявлення несанкціонованого використання система автоматично надсилає повідомлення правовласникам або адміністраторам платформ для вжиття заходів.

4. Модуль аналітики та звітності

Цей компонент відповідає за обробку даних, отриманих у процесі ідентифікації та моніторингу, та формування звітів. Звіти можуть включати статистику використання контенту, географічний розподіл, виявлені порушення та рекомендації щодо захисту. Аналітика допомагає правовласникам приймати обґрунтовані рішення щодо ліцензування чи судового переслідування [2-6].

Об'єкт автоматизації охоплює кілька основних процесів, які виконуються системою:

1. Збір та обробка аудіоданих

Система автоматично збирає аудіофайли з різних джерел, включаючи локальні сховища, хмарні платформи чи потокові сервіси. Обробка включає нормалізацію аудіоданих, видалення шумів та підготовку до аналізу.

2. Ідентифікація та класифікація

На основі цифрових відбитків система визначає, чи належить аудіофайл до відомого контенту в базі даних. Класифікація дозволяє визначити тип контенту (музика, звукові ефекти, голосові записи) та його правовий статус.

3. Моніторинг використання

Система відстежує використання аудіоконтенту в реальному часі, аналізуючи цифрові платформи та медіа. Це дозволяє виявляти як легальне, так і нелегальне використання.

4. Захист та реагування

У разі виявлення порушень система автоматично застосовує захисні механізми, такі як блокування доступу до контенту, сповіщення правовласників або ініціація юридичних процедур через інтеграцію з відповідними платформами [5].

Для реалізації системи ідентифікації та захисту аудіоконтенту використовуються сучасні технології:

- Алгоритми обробки сигналів: Для створення цифрових відбитків та аналізу аудіоданих.
- Машинне навчання: Для підвищення точності ідентифікації та класифікації контенту.
- Хмарні технології: Для зберігання великих обсягів даних та забезпечення масштабованості системи.
- Блокчейн: Для забезпечення прозорості та безпеки управління авторськими правами.

- API-інтеграція: Для взаємодії з платформами стрімінгу, соціальними мережами та іншими сервісами.

Автоматизація процесів ідентифікації та захисту аудіоконтенту забезпечує:

- Ефективність: Значне скорочення часу на обробку та аналіз аудіоданих порівняно з ручними методами.
- Точність: Висока точність ідентифікації завдяки використанню передових алгоритмів.
- Безпека: Захист авторських прав та зниження ризиків піратства.
- Масштабованість: Можливість обробки великих обсягів даних у реальному часі.

Система ідентифікації та захисту аудіоконтенту є складним об'єктом автоматизації, який поєднує передові технології обробки даних, машинного навчання та захисту інформації. Вона дозволяє ефективно управляти аудіоконтентом, захищати права авторів та забезпечувати дотримання законодавства у сфері інтелектуальної власності. Впровадження такої системи має значний потенціал для медіаіндустрії, дозволяючи автоматизувати рутинні процеси та зосередитися на створенні якісного контенту [3-7].

1.2 Вимоги до системи захисту та ідентифікації аудіоконтенту

Система захисту та ідентифікації аудіоконтенту призначена для забезпечення безпеки цифрового аудіоконтенту, його ідентифікації та захисту від несанкціонованого використання. Вона має вирішувати завдання автоматичного розпізнавання аудіофайлів, забезпечення їхньої цілісності, автентичності та захисту авторських прав. У сучасному цифровому середовищі, де аудіоконтент широко розповсюджується через Інтернет, потокові платформи та соціальні мережі, виникає потреба у створенні

надійних механізмів для запобігання піратству, несанкціонованому копіюванню та модифікації аудіофайлів. Цей документ описує основні вимоги до такої системи, включаючи функціональні, нефункціональні та технічні аспекти.

Система повинна забезпечувати точну ідентифікацію аудіофайлів незалежно від їхнього формату, якості чи ступеня модифікації. Основні вимоги до функції ідентифікації включають [3-7]:

- Розпізнавання аудіо за цифровими відбитками: Система має генерувати унікальні цифрові відбитки (fingerprinting) для кожного аудіофайлу на основі його акустичних характеристик, таких як спектральні особливості, темп, тональність тощо. Ці відбитки повинні бути стійкими до змін, таких як стиснення, шум, перекодування чи часткове обрізання.
- Порівняння з базою даних: Система повинна мати доступ до централізованої або розподіленої бази даних, що містить цифрові відбитки аудіофайлів. Вона має порівнювати отримані відбитки з базою для визначення відповідності.
- Обробка великих обсягів даних: Система повинна ефективно обробляти великі масиви аудіоданих, включаючи потокові трансляції, для ідентифікації контенту в реальному часі.

Система має забезпечувати захист аудіоконтенту від несанкціонованого доступу, копіювання та модифікації. До ключових вимог належать:

- Вбудовування водяних знаків: Система повинна підтримувати технологію цифрових водяних знаків (watermarking), які вбудовуються в аудіофайл без помітного впливу на якість звуку. Водяні знаки мають бути стійкими до спроб їх видалення чи зміни.
- Шифрування аудіоданих: Для захисту від несанкціонованого доступу система повинна використовувати сучасні алгоритми шифрування (наприклад, AES-256) для кодування аудіофайлів під час передачі та зберігання.

- **Контроль доступу:** Система має надавати можливість налаштування прав доступу до аудіоконтенту, включаючи обмеження за географічним регіоном, типом користувача чи пристроєм.

Система повинна мати функції для моніторингу використання аудіоконтенту та виявлення порушень:

- **Автоматичне сканування платформ:** Система має сканувати популярні платформи (YouTube, SoundCloud, соціальні мережі тощо) для виявлення несанкціонованого використання аудіоконтенту.
- **Журналювання подій:** Усі дії, пов'язані з ідентифікацією та захистом контенту, мають фіксуватися в журналі подій для подальшого аналізу та аудиту.
- **Сповіщення про порушення:** У разі виявлення несанкціонованого використання система повинна автоматично повідомляти власника контенту або адміністратора системи.
- **Швидкість обробки:** Система повинна забезпечувати швидке виконання операцій ідентифікації (менше 1 секунди для файлів тривалістю до 5 хвилин) та обробки водяних знаків.
- **Масштабованість:** Система має бути здатною обробляти тисячі запитів одночасно, підтримуючи роботу з великими обсягами даних у хмарному або локальному середовищі.
- **Надійність:** Система повинна мати рівень доступності не нижче 99.9% (SLA) для забезпечення безперебійної роботи.
- **Захист даних користувачів:** Усі дані, що обробляються системою, включаючи цифрові відбитки та журнали подій, мають бути захищені від несанкціонованого доступу відповідно до стандартів GDPR та інших регуляторних вимог.
- **Стійкість до атак:** Система повинна бути стійкою до атак типу "man-in-the-middle", спроб злому бази даних чи видалення водяних знаків.

- Зручність використання: Система має надавати інтуїтивно зрозумілий інтерфейс для користувачів, включаючи веб-панель та API для інтеграції з іншими платформами.
- Багатомовна підтримка: Інтерфейс повинен підтримувати кілька мов, включаючи українську, англійську та інші за потребою.

Система повинна підтримувати широкий спектр аудіоформатів, включаючи, але не обмежуючись:

- MP3, WAV, FLAC, AAC, OGG.
- Різні бітрейти (від 32 кбіт/с до 320 кбіт/с) та частоти дискретизації (від 8 кГц до 48 кГц).
- API: Система має надавати RESTful або GraphQL API для інтеграції з зовнішніми платформами, такими як потокові сервіси, соціальні мережі чи системи управління контентом.
- Сумісність: Система повинна бути сумісною з популярними операційними системами (Windows, macOS, Linux) та хмарними платформами (AWS, Google Cloud, Azure).
- Алгоритми ідентифікації: Використання машинного навчання (наприклад, нейронні мережі для аналізу аудіосигналів) та алгоритмів цифрових відбитків (наприклад, Chromaprint).
- База даних: Використання реляційних (PostgreSQL) або нереляційних (MongoDB) баз даних для зберігання цифрових відбитків та метаданих.
- Шифрування: Використання стандартів AES-256 для шифрування та RSA для обміну ключами.

Система захисту та ідентифікації аудіоконтенту повинна поєднувати передові технології ідентифікації, захисту та моніторингу для забезпечення безпеки цифрового аудіоконтенту. Вона має бути гнучкою, масштабованою та зручною для користувачів, забезпечуючи високу точність ідентифікації та надійний захист від несанкціонованого використання. Реалізація зазначених

вимог дозволить створити ефективне рішення для захисту авторських прав у цифровому середовищі [9].

1.3 Аналіз складових комплексної системи захисту аудіовізуального контенту від незаконного копіювання та розповсюдження

У сучасному цифровому світі аудіовізуальний контент є одним із найцінніших активів медіаіндустрії. Фільми, музика, телевізійні програми та інші форми медіа стають об'єктами незаконного копіювання та розповсюдження, що завдає значних збитків правовласникам. Для протидії цим загрозам розробляються комплексні системи захисту, які поєднують технологічні, юридичні та організаційні заходи. Цей документ присвячено аналізу ключових складових таких систем, їх функціональних особливостей та викликів, пов'язаних із їхньою реалізацією.

Цифрове водяне маркування є однією з ключових технологій для захисту аудіовізуального контенту. Воно передбачає вбудовування унікальних ідентифікаторів у аудіо- чи відеофайли, які дозволяють відстежувати джерело контенту та виявляти факти його незаконного використання. Водяні знаки можуть бути видимими або невидимими, залежно від мети їх застосування. Наприклад, невидимі водяні знаки використовуються для відстеження контенту в піратських мережах, тоді як видимі знаки можуть позначати право власності на контент.

Переваги цифрового водяного маркування включають:

- **Стійкість до модифікацій:** сучасні алгоритми дозволяють водяним знакам залишатися читабельними навіть після стиснення чи перекодування файлу.
- **Відстеження джерела:** унікальний ідентифікатор дозволяє визначити, звідки саме було скопійовано контент.

- Інтеграція з іншими системами: водяні знаки можуть працювати в поєднанні з системами DRM (Digital Rights Management).

Однак є й виклики:

- Обчислювальна складність: створення стійких водяних знаків вимагає значних ресурсів.
- Потенційна деградація якості: вбудовування знаків може впливати на якість аудіо чи відео, якщо алгоритми не оптимізовані.

Системи управління цифровими правами (DRM)

DRM-системи є основою захисту аудіовізуального контенту від несанкціонованого доступу. Вони включають набір інструментів і протоколів, які обмежують копіювання, відтворення та розповсюдження контенту. Наприклад, DRM може блокувати відтворення контенту на неавторизованих пристроях або вимагати ліцензію для доступу.

Ключові елементи DRM-систем:

- Шифрування контенту: використання криптографічних методів для захисту файлів від несанкціонованого доступу.
- Аутентифікація користувачів: перевірка прав доступу через ліцензійні ключі чи облікові записи.
- Обмеження функціональності: заборона копіювання, запису на зовнішні носії чи знімання екрана.

Проблеми DRM включають:

- Складність для користувачів: надмірні обмеження можуть погіршувати користувацький досвід.
- Уразливості: хакери постійно знаходять способи обходу DRM, що вимагає регулярного оновлення систем.

Технології ідентифікації, такі як Content ID від YouTube, дозволяють автоматично виявляти захищений контент на платформах. Вони базуються на аналізі аудіо- та відеоданих для створення унікальних "цифрових відбитків" (fingerprints). Ці відбитки порівнюються з базою даних правовласників, щоб

виявити порушення.

Переваги:

- Автоматизація: швидке виявлення піратського контенту без ручного втручання.
- Масштабованість: можливість обробки великих обсягів даних на платформах із мільйонами файлів.

Недоліки:

- Помилкові спрацювання: системи можуть помилково позначати легальний контент як піратський.
- Високі витрати на інфраструктуру: створення та підтримка бази відбитків вимагає значних ресурсів.

Ефективна система захисту аудіовізуального контенту неможлива без міцної юридичної основи. Міжнародні угоди, такі як Бернська конвенція чи DMCA (Digital Millennium Copyright Act), встановлюють стандарти захисту авторських прав. У багатьох країнах діють закони, які криміналізують піратство та передбачають штрафи чи інші санкції.

Ключові аспекти:

- Міжнародна співпраця: гармонізація законодавства між країнами для боротьби з транскордонним піратством.
- Швидке реагування: механізми для оперативного видалення піратського контенту з платформ.

Виклики:

- Різниця в законодавстві: різні країни мають різні підходи до захисту авторських прав.
- Складність судового переслідування: відстеження порушників у цифровому середовищі є складним і дорогим.

Організаційні заходи включають створення внутрішніх політик компаній, навчання персоналу та співпрацю з платформами. Наприклад, медіакомпанії можуть:

- Впроваджувати програми навчання для творців контенту щодо захисту їхніх робіт.
- Співпрацювати з платформами для швидкого видалення піратського контенту.
- Створювати спеціалізовані підрозділи для моніторингу та реагування на порушення.

Сучасні пірати використовують дедалі складніші методи обходу захисту, такі як VPN, анонімні мережі чи модифікація файлів для видалення водяних знаків. Це вимагає постійного вдосконалення технологій захисту.

Надмірний захист може відштовхувати легальних користувачів. Наприклад, складні DRM-системи можуть ускладнювати доступ до контенту, що спонукає користувачів шукати піратські альтернативи.

Перспективи розвитку включають:

- Штучний інтелект: використання AI для аналізу великих обсягів контенту та виявлення порушень у реальному часі.
- Блокчейн: створення децентралізованих систем для управління правами та відстеження контенту.
- Хмарні технології: забезпечення безпечного зберігання та потокового відтворення контенту.

Комплексна система захисту аудіовізуального контенту від незаконного копіювання та розповсюдження вимагає інтеграції технологічних, юридичних та організаційних складових. Цифрове водяне маркування, DRM та технології ідентифікації контенту є основними інструментами, але їх ефективність залежить від постійного вдосконалення та адаптації до нових загроз. Юридична база та організаційні заходи доповнюють технології, створюючи цілісний підхід до захисту. У майбутньому розвиток штучного інтелекту та блокчейн-технологій може значно посилити захист, але важливо знайти баланс між безпекою та зручністю для користувачів.

1.4 Основні принципи отримання та використання цифрових відбитків в системах ідентифікації аудіоконтенту

Цифрові відбитки (або цифрові водяні знаки) є ключовим елементом сучасних систем ідентифікації та захисту аудіоконтенту. Вони дозволяють ідентифікувати аудіофайли, відстежувати їх використання, а також захищати авторські права. Технологія цифрових відбитків базується на вбудовуванні унікальних ідентифікаторів в аудіодані, які залишаються непомітними для слухача, але можуть бути виявлені спеціальними алгоритмами. У цій статті розглядаються основні принципи створення, вбудовування та використання цифрових відбитків у системах ідентифікації аудіоконтенту, їх переваги та виклики.

Цифровий відбиток — це унікальний набір даних, який генерується на основі аудіосигналу та відображає його унікальні характеристики. Процес створення цифрових відбитків включає кілька етапів:

1. Аналіз аудіосигналу: На першому етапі аудіофайл аналізується для виділення його унікальних акустичних особливостей. Це можуть бути спектральні характеристики, амплітудні піки, частотні компоненти або часові параметри. Наприклад, алгоритми можуть аналізувати мел-частотні кепстральні коефіцієнти (MFCC), які відображають спектральну форму сигналу.

2. Генерація унікального ідентифікатора: На основі отриманих характеристик створюється компактний цифровий відбиток, який є унікальним для конкретного аудіофайлу. Цей відбиток зазвичай представлений у вигляді хеш-коду або бінарної послідовності, що дозволяє швидко порівнювати його з іншими відбитками в базі даних.

3. Стійкість до модифікацій: Важливою вимогою до цифрових відбитків є їхня стійкість до змін аудіосигналу, таких як стиснення, перекодування, додавання шуму чи обрізання. Для цього використовуються

алгоритми, які фокусуються на інваріантних характеристиках сигналу, що зберігаються навіть після обробки.

Вбудовування цифрових відбитків у аудіоконтент є складним процесом, який вимагає балансу між якістю звуку та надійністю захисту. Існує два основних підходи до вбудовування:

1. Непомітне вбудовування: Цифровий відбиток додається до аудіосигналу таким чином, щоб зміни були непомітними для людського слуху. Це досягається за допомогою психоакустичних моделей, які визначають, які частини аудіосигналу можна модифікувати без впливу на сприйняття. Наприклад, відбиток може бути вбудовано у високочастотні компоненти, які менш чутливі для слуху.

2. Стійкість до атак: Вбудовані відбитки повинні бути стійкими до спроб видалення або зміни, таких як компресія у формати MP3 чи AAC. Для цього використовуються методи, які розподіляють відбиток по всьому аудіосигналу, наприклад, за допомогою розсіювання спектру (spread spectrum) або модуляції низькочастотних компонентів.

3. Додавання метаданих: У деяких випадках цифровий відбиток може містити додаткову інформацію, таку як ідентифікатор власника, дата створення чи ліцензійні дані. Це дозволяє не лише ідентифікувати контент, але й відстежувати його використання.

Цифрові відбитки знаходять широке застосування в системах ідентифікації та захисту аудіоконтенту. Основні сценарії використання включають:

1. Ідентифікація контенту: Цифрові відбитки дозволяють швидко порівнювати аудіофайли з базою даних для визначення їх походження. Наприклад, сервіси як Shazam використовують цифрові відбитки для ідентифікації пісень за короткими уривками.

2. Захист авторських прав: Відбитки допомагають відстежувати незаконне використання аудіоконтенту в інтернеті. Системи автоматичного

моніторингу, такі як Content ID від YouTube, сканують завантажені файли та порівнюють їх відбитки з базою даних, щоб виявити порушення.

3. Аутентифікація та верифікація: Цифрові відбитки можуть використовуватися для підтвердження автентичності аудіозаписів, наприклад, у судових справах або при перевірці оригінальності контенту.

4. Персоналізація та аналітика: Відбитки дозволяють відстежувати, як і де використовується аудіоконтент, що корисно для маркетингових досліджень або аналізу популярності треків.

Незважаючи на численні переваги, технологія цифрових відбитків стикається з певними викликами:

1. Якість звуку: Неправильне вбудовування відбитків може призвести до погіршення якості аудіо, що неприпустимо для професійного контенту.

2. Обчислювальна складність: Створення та порівняння відбитків вимагає значних обчислювальних ресурсів, особливо при обробці великих баз даних.

3. Безпека: Зловмисники можуть намагатися видалити або змінити цифрові відбитки, що вимагає розробки більш стійких алгоритмів.

4. Етичні аспекти: Використання цифрових відбитків для моніторингу може викликати занепокоєння щодо приватності, особливо якщо інформація про користувачів збирається без їхньої згоди.

Цифрові відбитки є потужним інструментом для ідентифікації та захисту аудіоконтенту. Вони дозволяють ефективно вирішувати завдання ідентифікації, захисту авторських прав і відстеження використання контенту. Однак успішне впровадження цієї технології вимагає ретельного балансу між якістю звуку, стійкістю до модифікацій і обчислювальною ефективністю. Подальший розвиток алгоритмів і методів вбудовування цифрових відбитків сприятиме підвищенню їхньої надійності та розширенню сфер застосування, включаючи нові галузі, такі як аналіз потокового аудіо та захист мультимедійного контенту в реальному часі.

Система обробки аудіо відбитків складається з двох основних елементів: модуля генерації відбитків і підсистеми пошуку. Схему запропонованої системи ідентифікації цифрових відбитків можна побачити на рисунку 1.1.

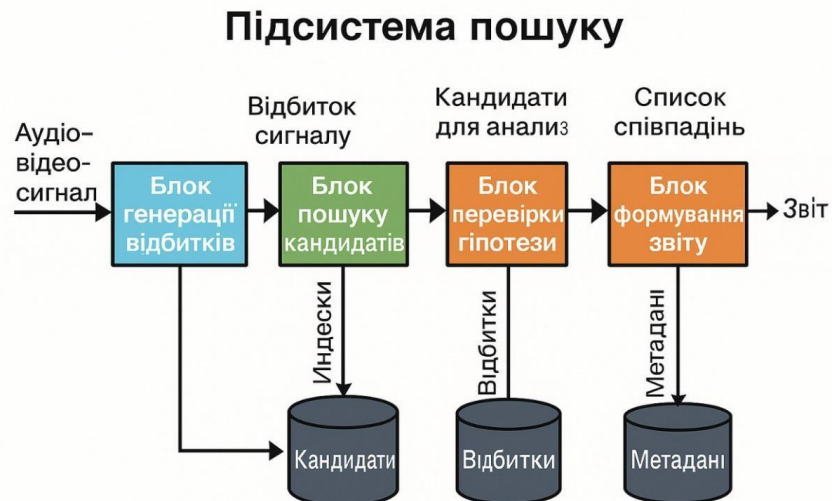


Рисунок 1.1 – Схема системи ідентифікації матеріалів з використання цифрових відбитків

Система складається з таких компонентів:

1. Блок генерації цифрового відбитку вхідного аудіовізуального сигналу;
2. Блок індексації цифрових відбитків та пошуку кандидатів;
3. Блок перевірки гіпотез та порівняння відбитків на співпадіння;
4. Блок пошуку метаданих та формування відповідних відбитків.

Система ідентифікації матеріалів може бути розподіленою. У цьому випадку обчислення відбитків може виконуватись на портативних пристроях під час запису сигналу в реальному часі, наприклад, на мобільних телефонах. Створений відбиток, який має компактний розмір, передається від клієнтського пристрою на сервер, де відбувається пошук відповідних метаданих, які потім надсилаються назад клієнту.

Під час пошуку аудіо відбиток порівнюється з даними в базі даних, де розраховується відстань між заданим відбитком та найближчим відбитком у

базі. Оскільки кількість порівнянь може бути дуже великою, а розрахунок відстані потребує значних обчислювальних ресурсів, процес пошуку поділяється на кілька етапів. Спочатку здійснюється швидкий пошук кандидатів за допомогою індексації або спрощених методів порівняння відбитків. Потім на наступному етапі проводиться більш детальний аналіз вибраних кандидатів.

Основними вимогами до системи пошуку та ідентифікації аудіовідбитків є:

- Швидкість. Система повинна забезпечувати оперативний пошук кандидатів у базах даних, що містять мільйони записів;
- Точність. Підсистема порівняння повинна мати низький рівень помилок, зокрема мінімізувати випадки, коли матеріали не ідентифікуються (false negative rate, FNR) або неправильно ідентифікуються (false positive rate, FPR);
- Мінімальні вимоги до пам'яті. Більшість систем використовують таблиці індексів для первинного пошуку кандидатів, що займають весь доступний об'єм оперативної пам'яті. Обсяг пам'яті, що використовується, суттєво впливає на ефективність пошуку та максимальний розмір бази даних.

Існуючі методи отримання аудіовідбитків можна поділити на три основні категорії [10]:

- Локальні характеристики сигналу. Вектори ознак обчислюються через рівні проміжки часу, що дає можливість сформувати траєкторії в евклідовому просторі. Прикладом може бути спектральна рівномірність, що використовується в стандарті MPEG-7. Для обчислення відстані між елементами відбитків часто застосовується відстань Хеммінга. Пошук і порівняння послідовностей елементів відбитків зазвичай виконуються за допомогою методу динамічного програмування (Dynamic Time Warping, DTW).

- Довгострокові характеристики аудіосигналу. Це, наприклад, середнє значення і дисперсія параметрів, таких як ритм або частота домінантних тонів, обчислені для всього музичного твору. Такі характеристики не потребують точного позиціювання при порівнянні, але мають низьку стійкість до шуму та спотворень [7]. Вони також не дозволяють порівнювати окремі фрагменти або матеріали, що передаються як неперервний потік.
- Приховані Марківські моделі. Незважаючи на високу ефективність та поширеність у розпізнаванні мовлення [8] і використанні для ідентифікації випадкових аудіосигналів [5][14], на сьогодні не існує ефективних систем цифрових відбитків, заснованих на цих моделях.

Аналіз цих методів показав, що найкращі результати для вирішення задачі ідентифікації аудіовізуальних матеріалів дають методи, засновані на аналізі локальних параметрів сигналу, які визначають структуру процесу обчислення набору ознак аудіосигналу. Схема цього процесу зображена на рисунку 1.2.

Первинна обробка. На першому етапі акустичні сигнали перетворюються в цифровий формат і приводяться до основного формату для подальшої обробки (зазвичай 16-бітний моно сигнал з частотою дискретизації від 5 до 44,1 КГц). Інколи аудіо піддається специфічним перетворенням, щоб моделювати або компенсувати спотворення каналу передачі. Інші типи перетворень включають GSM-кодування в телефонних системах, амплітудну нормалізацію та фільтрацію [7].

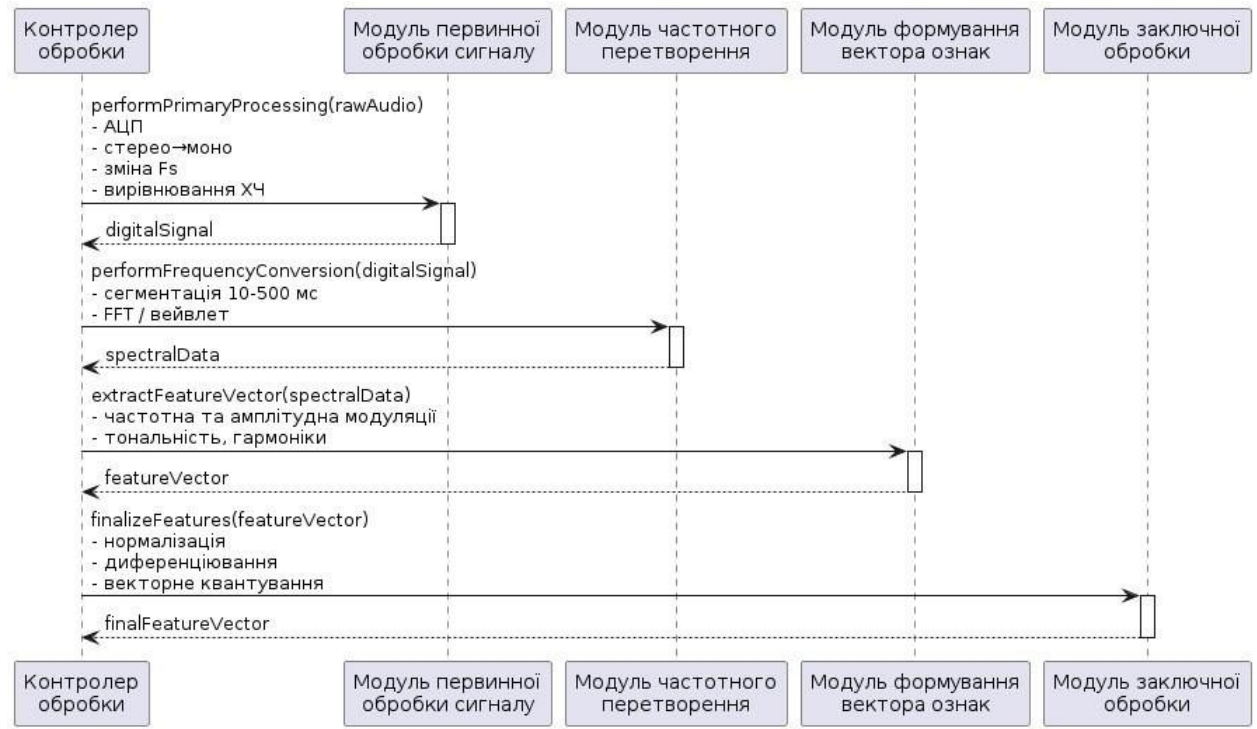


Рисунок 1.2 – Діаграма обробки сигналу в блоці отримання цифрових аудіовідбитків

Частотне перетворення сигналу. Перетворення використовуються для виділення характерних параметрів акустичних сигналів і зменшення надлишковості отриманих даних. Існують перетворення, які мають оптимальні властивості для стиснення і декореляції, такі як перетворення Карунена-Лоева (Karhunen Loeve Transform, KLT) та метод головних компонентів (Principal Component Analysis, PCA) [6]. Ці методи дозволяють значно зменшити розмірність векторів ознак, але мають істотні недоліки. Основним недоліком є те, що параметри перетворень обчислюються динамічно і залежать від структури вхідного сигналу, тому їх неможливо відтворити після того, як сигнал зазнав пошкоджень від спотворень чи модифікацій. Підвищений рівень шуму може призвести до значної втрати корисного сигналу. З цієї причини більшість перетворень, що використовуються, мають фіксовані ортогональні базові функції і здійснюють перетворення даних в частотний домен. Методи, засновані на аналізі змін

потужності сигналу [15][16], можна розглядати як спрощену форму частотно-часового розподілу з однією частотною смугою. Найбільш поширеним перетворенням є швидке перетворення Фур'є (Fast Fourier Transform, FFT). Також використовуються дискретне косинусне перетворення (Discrete Cosine Transform, DCT), перетворення Хаара (Haar Transform) та модельоване комплексне перетворення (Modulated Complex Transform, MCLT).

При цьому вважається, що сигнал X^t є стаціонарним на невеликому інтервалі часу, і для аналізу він розбивається на фрагменти тривалістю $T=10-500$ мс. Амплітудно-частотний спектр $F_i = f_i(m)$, $m \in (1, M)$ для i -го фрагменту x^i сигналу $X^t = \{x_1, x_2, \dots\}$ зазвичай розбивається на N діапазонів, заданих в інтервалі $[a_n, b_n]$, $a_n < b_n$, $a_n \in M$, $b_n \in M$, де M є кількістю спектральних коефіцієнтів.

Вектор ознак розраховується на основі характерних властивостей сигналу в цих частотних діапазонах.

Отримання вектора ознак. Після перетворення сигналу в частотне представлення, він проходить додаткову обробку для виділення ключових акустичних параметрів [17]. Головною метою цього етапу є зменшення розмірності вектора ознак при збереженні його стійкості до спотворень. Для цього часто застосовуються алгоритми, які імітують слухову систему людини для виявлення перцепційно важливих характеристик, зазвичай із використанням банку смугових фільтрів, що моделюють критичні смуги слухової системи людини. Найпоширенішими є такі методи:

Ентропія сигналу (Spectrum Entropy, SE):

$$SE_{i,n} = \sum_{m=a_n}^{b_n} |f_i(m)| \log_2 |f_i(m)| \quad (1.1)$$

Енергія спектральних діапазонів (Spectral Band Energy, SBE):

$$SBE_{i,n} = \frac{\sum_{m=a_n}^{b_n} |f_i(m)|^2}{\sum_{m=1}^M |f_i(m)|^2} \quad (1.2)$$

Спектральний центроїд (Spectral Centroid, SC). Визначає центр маси спектру і виступає мірою «яскравості» звучання аудіо сигналу

$$SC_{i,n} = \frac{\sum_{m=a_n}^{b_n} m |f_i(m)|^2}{\sum_{m=a_n}^{b_n} |f_i(m)|^2} \quad (1.3)$$

Спектральна широта (Spectral Bandwidth, SB). Являє собою зважену середню відстань між спектральними компонентами і спектральним центроїдом SC:

$$SC_{i,n} = \frac{\sum_{m=a_n}^{b_n} (m - SC_{i,n})^2 |f_i(m)|^2}{\sum_{m=a_n}^{b_n} |f_i(m)|^2} \quad (1.4)$$

Міра рівномірності спектру (Spectral Flatness Measure, SFM). SFM виступає мірою тональності чи шумності сигналу в даному частотному діапазоні:

$$SC_{i,n} = \frac{\sum_{m=a_n}^{b_n} (m - SC_{i,n})^2 |f_i(m)|^2}{\sum_{m=a_n}^{b_n} |f_i(m)|^2} \quad (1.5)$$

Зміна енергії в частотних діапазонах (Spectral Band Difference, SBD):

$$SBD_{i,n} = (SBE_{i,n} - SBE_{i,n+1}) - (SBE_{i+1,n} - SBE_{i+1,n+1}) \quad (1.6)$$

Коефіцієнт переходів рівня сигналу через нуль (Zero Crossing Rate, ZSR):

$$ZCR_i = \frac{1}{T} \sum_{t=0}^{T-1} \text{sgn}(x_{i,t} - x_{i,t-1} < 0) \quad (1.7)$$

Мел-частотні кепстральні коефіцієнти (MFCC, Mel Frequency Cepstrum Coefficients) [18]. Для отримання MFCC амплітудно-частотний спектр Фур'є згладжується і розподіляється на частотні діапазони, що відповідають так званій Mel-шкалі. Mel-шкала розроблена для моделювання здатності людського вуха розрізняти різні тональності: вона є лінійною для частот менше 1000 Гц і логарифмічною для вищих частот. На наступному етапі здійснюється обчислення амплітуди сигналу в кожному діапазоні з подальшим частотним перетворенням отриманих значень. Зазвичай застосовують дискретне косинусне перетворення (DCT), що дозволяє концентрувати спектральні дані на коефіцієнтах з низькими частотами.

Для моделювання музичних характеристик також використовуються параметри, такі як фундаментальна частота, гармоніки, гучність і ритм, а також методи класифікації типів аудіосигналів [19].

Заклучна обробка і кодування. Вищезгадані характеристики акустичних сигналів є абсолютними величинами. Для більш повного та стійкого представлення сигналів, що змінюються по амплітуді (амплітудна модуляція) та частоті (частотна модуляція), використовуються похідні цих параметрів. Наприклад, в роботі [14] для отримання векторів ознак застосовуються MFCC та їх похідні. Використання похідних параметрів (швидкість і прискорення) може збільшувати шум, але разом з тим зменшує спотворення в каналі передачі. Для цього ж може застосовуватись нормалізація спектру модуляції [5] та кепстральних коефіцієнтів (Cepstrum Mean Normalization, CMN) [20], що дозволяє зменшити вплив повільно змінюваних лінійних спотворень каналів

передачі. На останньому етапі проводиться квантування значень векторів ознак з низьким розширенням, що підвищує стійкість до шуму і зменшує розмір отриманих відбитків.

Аналіз цих методів дозволяє зробити висновок, що для відновлення векторів ознак доцільно використовувати алгоритм, заснований на аналізі змін енергії в спектральних діапазонах і їх похідні, а для компенсації спотворень каналу передачі варто застосовувати нормалізацію значень отриманого вектора ознак на основі середніх параметрів.

Розділ стандарту, що описує аудіосигнали, включає не тільки метадані, але й специфічні параметри, що безпосередньо характеризують акустичний сигнал [9]. Стандарт MPEG-7 описує структури, які можна поділити на два основних класи параметрів: основні характеристики акустичних сигналів і методи, специфічні для конкретних додатків. Останні включають інструменти для виконання таких функцій:

- Розпізнавання окремих звуків серед суміші сигналів.
- Анотація, ідентифікація і пошук мовних матеріалів.
- Класифікація музичних інструментів, включаючи ідентифікацію та пошук за зразком інструменту.
- Опис мелодії: пошук і ідентифікація музичних творів, пошук за зразком мелодії.
- Ідентифікація довільних аудіосигналів за їх відбитками.

Стандарт MPEG-7 надає різноманітні структурні елементи для побудови аудіовідбитків. Один із них, AudioSpectrumFlatness Descriptor, містить набір низькорівневих описувачів сигналу (Low Level Descriptors, LLD) і є універсальним для даних типів додатків. Конкретні характеристики низькорівневих описувачів визначаються через схему, яка описує, зокрема, часові та частотні параметри відбитка. За стандартом, розмір вікна для спектральних обчислень становить 30 мс, а частотний діапазон від 240 Гц до 4 кГц поділяється на 16 діапазонів шириною $1/4$ октави, для кожного з яких обчислюється середнє значення та дисперсія спектральних характеристик. Як основний дескриптор використовується спектральна рівномірність SFM.

Обчислювальна ефективність алгоритму є дуже високою — більше ніж 100 разів швидше за швидкість програвання. За даними розробників [7], система стійка до типових спотворень сигналу і здатна ідентифікувати фрагменти незалежно від їх зміщення в тестовому сигналі. Для сигналів, що піддаються невеликим викривленням, точність розпізнавання зазвичай становить 98-99% [7].

Попри велику кількість публікацій, присвячених технології MPEG-7, на сьогодні не існує ефективної реалізації цього методу, що забезпечує надійну ідентифікацію та прийнятну швидкість пошуку для практичного використання. Високий рівень помилкових ідентифікацій, низька швидкість пошуку і обмеження на розмір бази даних роблять цей метод недостатньо ефективним для ідентифікації та пошуку аудіоматеріалів в Інтернеті. Хоча стандарт MPEG-7 був створений як універсальна мова для опису параметрів аудіовізуальних сигналів, його застосування обмежене через недостатню гнучкість для конкретних завдань ідентифікації матеріалів.

1.5 Висновок до розділу 1

Існуючі системи захисту аудіовізуальної інформації спрямовані на запобігання копіюванню матеріалів у цифровому форматі, однак не дозволяють захистити матеріали після їх перетворення в аналоговий формат під час перегляду користувачем. Користувач може легко створювати аналогові копії зображення і звуку, а також стискати та поширювати їх через Інтернет.

Для контролю за поширенням матеріалів пропонується використовувати пошук та ідентифікацію несанкціонованих копій за допомогою цифрових відбитків.

Цифрові відбитки дозволяють знаходити схожі аудіовізуальні матеріали за малими взірцями, причому надійна ідентифікація можлива незалежно від

формату представлення даних, рівня стиснення, наявності шуму та різних типів спотворень.

Аналіз літературних даних показав, що всі існуючі методи аудіо відбитків мають достатньо високий рівень похибок у системах ідентифікації аудіоматеріалів.

Таким чином, на сьогоднішній день немає систем, стійких до всіх типів спотворень та шуму, особливо до їх комбінацій. Для забезпечення необхідної ефективності та точності пошукових і ідентифікаційних систем розроблені нові методи цифрових відбитків, які будуть розглянуті в наступному розділі.

2 РОЗРОБКА СИСТЕМИ ІДЕНТИФІКАЦІЇ ТА ЗАХИСТУ АУДІОКОНТЕНТУ

Однією з ключових проблем при створенні систем для пошуку та розпізнавання аудіо- та відеоматеріалів є необхідність обробки великих обсягів даних, а також обмеження в ефективності та точності наявних підходів [8][16]. У цьому розділі проведено експериментальне дослідження стійкості різних видів дескрипторів акустичних сигналів до впливу спотворень з метою вибору оптимальних параметрів для швидкого та надійного порівняння цифрових «відбитків». Розглянуто нові підходи до пошуку та ідентифікації на основі цифрових слідів аудіо- і відеофрагментів, які ґрунтуються на локальних характеристиках сигналу. Запропоновано метод узгодження потоків аудіо- та відеовідбитків для їх спільного використання в межах єдиної системи ідентифікації мультимедійного контенту.

2.1 Експериментальне порівняння і вибір дескрипторів аудіоматеріалів

У цьому розділі наведено результати експериментів і підсумки тестування дескрипторів, описаних у підрозділах 1.1–1.7. Основна мета — визначити найбільш ефективний варіант представлення ознакового вектора. Проведено реалізацію та аналіз наступних підходів:

1. **Muscle Fish.** Це одна з перших розробок, орієнтованих на розпізнавання аудіофрагментів за їх внутрішніми характеристиками [5]. Метод фокусується на виділенні ключових перцептивних властивостей, таких як рівень гучності, частотна структура, тембральна палітра, динаміка ритму та зміни темпу. Для кожного показника обчислюються такі статистичні параметри, як середнє значення, дисперсія і автокореляція.

Порівняння здійснюється шляхом обчислення відстані між ознаковими векторами в евклідовому просторі [6]. Подальшим розвитком цієї ідеї стала технологія, реалізована компанією Audible Magic [9], яка, крім вищезазначених характеристик, також включає у розрахунки коефіцієнти кепструму на основі Mel-шкали (MFCC).

2. AudioDNA. Методика формування унікального аудіопрофілю передбачає аналіз спектру сигналу із застосуванням MFCC [14].
3. RARE (Robust Audio Recognition Engine). Цей підхід був створений у дослідницькій лабораторії компанії Microsoft [12][18]. Його структурна реалізація наведена на рисунку 2.2. Для представлення сигналу у частотно-часовій площині застосовується модульоване комплексне перетворення (MCLT), що дає змогу зберігати як амплітудну, так і фазову інформацію при обробці.

Попередній етап обробки спектру сигналу дозволяє усунути спотворення в частотному діапазоні. Це досягається шляхом обчислення різниці між коефіцієнтами MCLT і згладженим спектром. Після цього застосовується психоакустична модель (Psycho Acoustic Model, PAM), що дозволяє відфільтрувати складові сигналу, які не сприймаються людським слухом. Для зменшення кількості параметрів, що описують аудіовідбиток, використовується вдосконалений варіант методу головних компонент — Oriented Principle Component Analysis (OPCA). Цей підхід реалізується у два послідовні кроки: спочатку OPCA обробляє спектрально-амплітудне представлення сигналу в накладених часових вікнах тривалістю 372 мс, де для навчання алгоритму використовується кореляційна матриця, побудована між незмінним і змінним сигналом. Далі об'єднуються результати 32 таких перетворень, що охоплюють часовий відрізок тривалістю 6,1 секунди, після чого здійснюється додаткове стискання й формується 64-вимірний вектор ознак.

4. Philips. Один із підходів, який характеризується низькими обчислювальними вимогами, був розроблений J. A. Naitsma у компанії Philips [3]. Для кожного трисекундного фрагмента звуку формується набір із 256 32-бітних субвідбитків, які обчислюються через кожні 11,7 мс. Щоб сформувати субвідбитки, аудіосигнал поділяється на перекривні відрізки довжиною 0,37 с з перекриттям 31/32. Далі ці фрагменти перетворюються у спектрограму за допомогою швидкого перетворення Фур'є (FFT). Частотна область між 300 і 2000 Гц розділяється на 33 ізольовані смуги із логарифмічним розподілом. Для кожної з них визначається знак зміни енергії в часово-частотному контексті — ця ознака є характерною і стійкою до змін.

Порівняння відбитків здійснюється за критерієм, який враховує кількість бітів, що відрізняються (Bit Error Rate, BER). Якщо кількість різних бітів менша за певний поріг, який балансує між вірними та хибними збігами, то відбитки вважаються однаковими. Експериментально було встановлено, що оптимальним значенням BER є 0.35. Цей підхід через простоту й стабільність результатів набув широкого розповсюдження й фактично став еталонним у тестуванні ефективності аудіоідентифікаційних алгоритмів.

5. Philips (модифікований). Щоб компенсувати зміну масштабу у часово-частотному представленні сигналу, спричинену розтягуванням або стисканням у часі, було запропоновано вдосконалення попереднього методу, що базується на використанні перетворення Фур'є-Мелліна (Fourier-Mellin Transform, FMT) [13]. Зокрема, спектр потужності логарифмічно масштабується по частоті, після чого виконується перетворення Фур'є. За результатами тестів, такий підхід у п'ять разів менш вразливий до варіацій швидкості відтворення звуку, ніж класичний метод. Подібний ефект був зафіксований також при застосуванні автокореляційної функції замість Фур'є-перетворення [7].

6. MusicDNA. Методика, розроблена компанією Cantametrix і згодом придбана Gracenote [12], базується на обчисленні середніх значень енергії та її варіацій у 15 барк-смугах частот, аналізуючи 12 перекривних трисекундних фрагментів. Ці ознаки нормалізуються, утворюючи набір із 30 компонентів, що формують аудіовідбиток.

7. Google. У дослідженні [12] було використано алгоритми, типові для розпізнавання зображень, зокрема людських облич. Унікальною особливістю є трактування спектрограми як графічного об'єкта. Втіленням такого підходу став удосконалений варіант методики Haitsma [3], де замість ядра, заданого формулою 1.6, використовуються комбінації Хаар-фільтрів. У процесі навчання було обрано набір фільтрів, які одночасно забезпечують ефективну компресію даних та високу точність при ідентифікації спотворених звуків. Подальший розвиток цієї ідеї привів до впровадження вейвлет-перетворень для спектрограм — це рішення було реалізоване фахівцями компанії Google [13].

Виявлені недоліки методів:

- MFCC-підходи (1 і 2) вразливі до шуму (якщо співвідношення сигнал/шум менше ніж 6 дБ) та до втрат при цифровому стисканні, зокрема MP3 (32 kbps);
- Метод головних компонент (3) має обмежену стійкість, яка залежить від умов навчання і не гарантує надійність при комбінованих викривленнях;
- Методи, що базуються на енергетичних змінах у частотних смугах (4 і 5), погано працюють при реверберації понад 30 мс або при використанні мовних кодеків (типу GSM);
- Використання прихованих марковських моделей не вирішує проблему ефективного індексування і підходить лише для перевірки попередньо обраних варіантів;
- Жодна з методик не має стійкості до масштабування аудіо у часі понад 5% або змін швидкості понад 2%.

Варто зауважити, що здатність до розпізнавання при часовому масштабуванні є критично важливою для контролю ефірних трансляцій, де ступінь зміни тривалості сигналу може досягати 10–15%. На рисунку 2.1 відображено пороги коректного розпізнавання для початкових та стислих аудіо (MP3, GSM). Навіть удосконалений алгоритм Philips, попри заявлену масштабну інваріантність, не гарантує стабільного розпізнавання: на практиці цей метод демонструє ефективність лише при варіаціях швидкості в межах $\pm 5\%$.

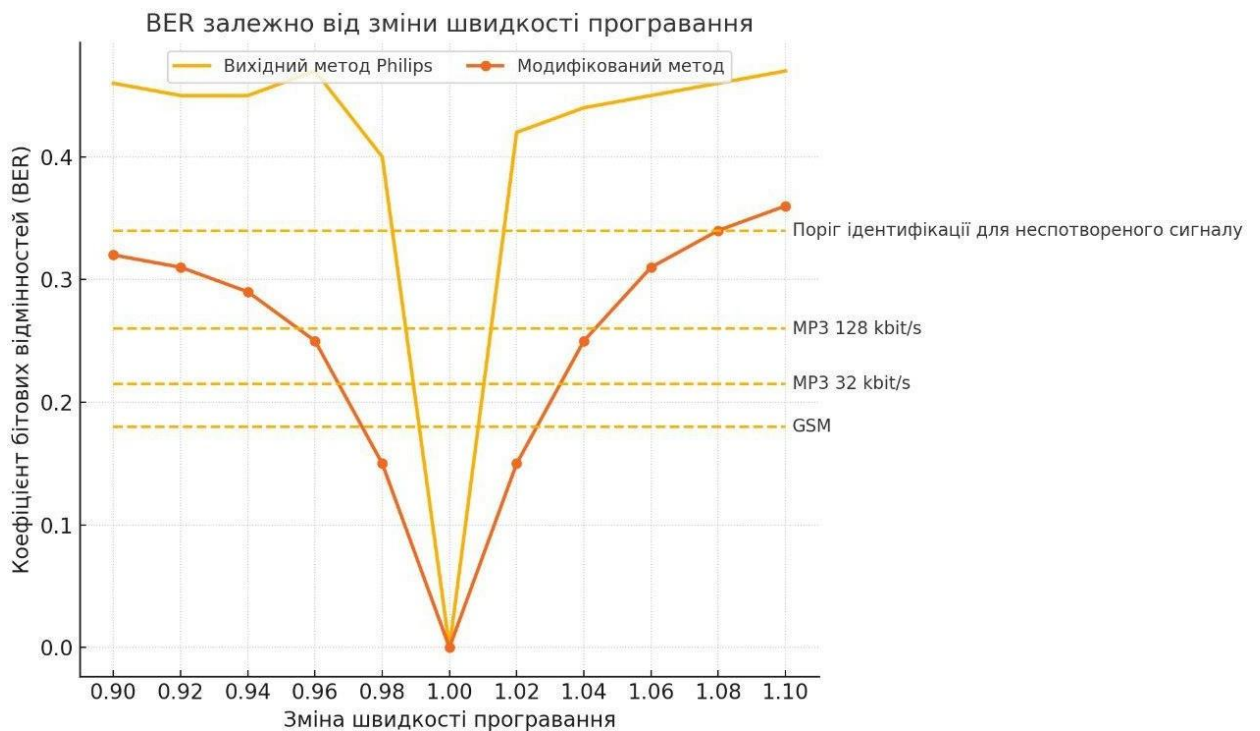


Рисунок 2.1 – Залежність коефіцієнта бітових відмінностей від швидкості програвання аудіосигналу для методу цифрових відбитків фірми Philips і його модифікації на основі перетворення Фур'є-Мелліна

Суттєвою вадою усіх проаналізованих підходів є те, що під час їх розробки не брали до уваги специфіку людського слухового сприйняття та змін, які виникають при стисканні звукового сигналу цифровими методами. Варто підкреслити, що велика частина аудіо- та відеоконтенту поширюється у

форматі, стислому за допомогою алгоритмів, які опираються на принципи психоакустичного сприйняття та видаляють фрагменти, непомітні для людського слуху. У зв'язку з цим пропонується створювати алгоритми ідентифікації цифрових мультимедійних даних з урахуванням психофізіологічних характеристик сприйняття звуку, адаптуючи дескриптори до психоакустичних моделей, що лежать в основі сучасних методів стискання.

Як підтвердили експерименти, наявні технології неспроможні гарантувати точне виявлення й пошук аудіоінформації, що використовується, наприклад, для моніторингу її поширення. У наступному розділі буде розглянуто дескриптори, здатні витримувати спотворення, а також проведено оцінку їх ефективності з урахуванням особливостей слухового аналізатора людини.

2.2 Аналіз стійкості дескрипторів на основі моделей сприйняття акустичних сигналів людиною

Людина може розуміти мову навіть за умов значного шуму, реверберації чи складних акустичних змін. Існує припущення, що в ході еволюції покращення здатності до мовного розпізнавання відбувалося завдяки не лише розвитку слухових механізмів, але й формуванню ефективного способу зберігання та передачі фонетичної інформації. Аналіз функціонування людської слухової системи є актуальним для виявлення мовних характеристик, які мають найбільшу інформативність. Застосування таких характеристик допомагає ідентифікувати риси звуку, що залишаються стабільними при трансформаціях, включаючи фільтрацію та стискання. Відомо, що типові аудіокодеки, які використовуються в телефонному зв'язку, наприклад GSM чи CELP, ґрунтуються на моделюванні людського вокального апарату і слуху, проте можуть сильно спотворювати нефонетичний контент,

зокрема музику або шум. Зважаючи на те, що більшість звукових файлів зберігаються та передаються у стисненому вигляді, здатність аудіовідбитків зберігати свої властивості після кодування є критично важливою при формуванні параметрів дескрипторів.

Первинне перетворення звуку у частотно-часовий формат у слуховій системі людини здійснюється за допомогою базальної мембрани, яка виконує функцію фільтрації, подібної до набору смугових фільтрів із шириною смуги, що змінюється відповідно до частоти. Слух людини має надзвичайно високу чутливість до змін у частоті — менш ніж 0,2% у випадку окремих тональних компонентів. Для характеристики сприйняття висоти тону застосовується шкала мел, де значення 1000 мел відповідає 1000 Гц на рівні 40 дБ. Ще однією важливою властивістю частотного аналізу є наявність критичних смуг, які визначають, коли два сигнали обробляються незалежно один від одного. Наприклад, при спектральному згладжуванні мовні сигнали залишаються зрозумілими, якщо розмір фільтрації не виходить за межі відповідної критичної смуги.

Також існує взаємозв'язок між можливостями частотного аналізу та музичним сприйняттям, зокрема — між критичними смугами і приємністю консонансних інтервалів. Консонанс визначається як гармонійне поєднання тонів, яке здається приємним. Як показано у [34], два чистих тони сприймаються як гармонійні, коли їх частотна відстань трохи більша за ширину критичної смуги.

Згідно з дослідженням [17], структура європейської тональної музики тісно пов'язана з цими критичними інтервалами: для кожного діапазону передбачається наявність лише одного домінуючого тону, що забезпечує стабільність і гармонійність звучання.

Поняття критичних діапазонів є основою барк-шкали, яка застосовується при симуляції слухової системи людини через набір фільтрів певної смуги, а також у багатьох форматах аудіокодування (наприклад, MPEG-1, MPEG-2). Через тісний зв'язок цих частотних зон із механізмами сприйняття мовлення і музичних структур, а також беручи до уваги специфіку архітектур алгоритмів стискання, що здійснюють обробку сигналів саме у межах цих критичних діапазонів, можна зробити висновок, що для ефективної ідентифікації звукових даних достатньо використання фільтрів, побудованих із частотним розміщенням, відповідним до критичних смуг. Їхня ширина змінюється залежно від частоти, і для методу аудіовідбитків вона може бути з достатньою точністю змодельована шляхом використання фільтрів із шириною $1/3$ октави.

У зв'язку з цим у нашій роботі для побудови банку фільтрів у системі створення відбитків пропонується застосовувати накладені смугові фільтри, які мають логарифмічне розташування по осі частот і однакову смугу пропускання, що дорівнює третині октави.

Однією з ключових рис людської мови є те, що більшість лінгвістичної інформації передається через відносно повільні зміни структури мовного сигналу. Це підтверджують експерименти з каналним вокодером, який був розроблений у 1928 році в лабораторії Bell Labs для моделювання та стиснення мови. Вокодер працює шляхом кодування мови через витягування параметрів мовного сигналу, після чого здійснюється синтез сигналу на основі цих параметрів. Вокодер включає 10 аналогових смугових фільтрів, які покривають частотний діапазон 250-3000 Гц. На етапі аналізу кожен з цих фільтрів виділяє основну частоту та обвідну для кожного частотного діапазону.

Під час синтезу обвідна кожного сигналу накладається на відповідну несучу частоту, і результати всіх діапазонів складаються в єдиний сигнал. Під час досліджень було встановлено, що розбірлива мова може бути відтворена

за умови, що огинаюча сигналу виділяється за допомогою фільтра низьких частот з пропускнуою здатністю 0-25 Гц, при цьому основна частина мовного сигналу складається з частот 10 Гц і нижче. У подальших роботах було доведено, що частоти понад 16 Гц не впливають на розбірливість мови.

Часові зміни спектральних характеристик мови можуть бути виражені через спектр модуляції. Для обчислення спектра модуляції проводиться спектральний аналіз огинаючої сигналу, після чого значення нормалізуються.

$$|m(\omega)| = \frac{1}{\langle s(t) \rangle} \left| \int_{-\infty}^{\infty} s(t) e^{-j\omega t} dt \right| \quad (2.1)$$

де, $s(t)$ являє собою енергію кривої сигналу і $\langle s(t) \rangle$ є середнім значенням $s(t)$.

Огинаюча сигналу отримується шляхом фільтрації сигналу за допомогою фільтра низьких частот. Для складних сигналів, таких як мова, різні частоти можуть мати слабку кореляцію і відрізнятися за параметрами спектра модуляції. Тому для таких сигналів доцільно проводити аналіз спектра модуляції окремо для кожного діапазону. Модуляція мови має дві основні причини. Перша – це зміни загального рівня енергії сигналу, пов'язані зі зміною фоном, які однаково впливають на всі діапазони. Друга причина – зміни в спектральних розподілах формант для кожного частотного діапазону.

Ступінь спотворень сигналу у каналі передачі критично важливий для розпізнавання мови. Якісною характеристикою каналу є функція передачі модуляції (MTF), яка є відношенням спектра модуляції прийнятого сигналу до спектра модуляції вихідного сигналу та враховує різні акустичні спотворення каналу передачі. У каналі з шумом модуляційна передавальна функція є сталою, визначаючи ступінь ослаблення коефіцієнтів модуляції, що не залежить від частоти модуляції [15].

$$H_m(\omega) = \frac{E_s}{E_s + E_n} \quad (2.2)$$

де E_s є середньою енергією сигналу, а E_n - середня енергія шуму. Канал з імпульсною характеристикою $h(t)$ має передавальну модуляційну функцію MTF такого вигляду:

$$H_m(\omega) = \left| \frac{\int h^2(t)e^{-j\omega t} dt}{\int h^2(t) dt} \right| \quad (2.3)$$

Відлуння в приміщеннях спричиняє помітний підйом MTF на низьких модуляційних частотах. Для поліпшення сприйняття мови в приміщеннях з великим часом реверберації пропонується фільтрація сигналів у модуляційно-частотному домені за допомогою смугових фільтрів, які мають підсилення в діапазоні 10-15 dB на частотах 4-6 Гц.

Функцію передачі модуляції можна використовувати для оцінки якості мови, зокрема для визначення індексу передачі мови (STI). STI є чисельним показником MTF каналу для частот 125-8000 Гц та модуляційних частот 0.625-12.5 Гц. Було показано, що STI добре корелює з експериментальними результатами сприйняття мови в умовах реверберації, шуму та нелінійних спотворень. Використання STI в ряді стандартів якості сприйняття мови підтверджує важливість повільних змін спектральних характеристик для лінгвістичної інформації.

Пізніші дослідження продемонстрували, що для ефективного розпізнавання мови в умовах реверберації достатньо модуляційних частот в діапазоні 8 Гц, з особливим акцентом на частоти близько 4 Гц. Для вимірювання якості сприйняття мови був розроблений індекс частотно-часової модуляції (STMI), який враховує зміни сигналу як в часі, так і по частоті, включаючи амплітудну (AM) і частотну (FM) модуляцію.

Існує гіпотеза, що в слуховій системі є канали, налаштовані на сприйняття модуляційних частот так само, як сприйняття спектральних частот

у критичних діапазонах. Особливості сприйняття тимчасової динаміки звукових сигналів можна пояснити фільтрацією в модуляційно-частотних каналах слухової системи.

Аналізуючи особливості сприйняття аудіоінформації людиною, доцільно використовувати дескриптори, що базуються на аналізі амплітудної і частотної модуляції акустичних сигналів у частотних діапазонах з логарифмічним розподілом частот. Стійкість запропонованого методу до цифрового стиснення порівняно з іншими типами дескрипторів буде розглянута в наступному розділі.

Більшість аудіовізуальних матеріалів поширюються в цифровому форматі. Алгоритм стиснення та його ступінь істотно впливають на якість цифрових сигналів. Для мультимедійних застосунків використовуються алгоритми, які забезпечують високу якість звуку, що досягається за допомогою стиснення з високим бітрейтом. Ідентифікація таких сигналів зазвичай не становить труднощів для більшості сучасних методів цифрових аудіовідбитків. Однак у деяких програмах, зокрема в IP і стільниковій телефонії, застосовуються спеціалізовані алгоритми стиснення, оптимізовані для відтворення певних типів аудіосигналів і забезпечують максимальний рівень стиснення. Наприклад, стандарт GSM, широко використовуваний у телефонії, оптимізований для стиснення мови і погано підходить для кодування складних багатотональних сигналів, таких як музика.

Стійкість алгоритмів цифрових відбитків до стиснення в стандарті GSM має важливе значення для додатків, що займаються ідентифікацією аудіоматеріалів, переданих через цифрові канали, включаючи IP і стільникову телефонію. У роботі [12] було досліджено, як спотворення, що виникають при стисненні через GSM, впливають на різні типи аудіосигналів. Результати показали, що повільні зміни сигналу, зокрема з частотою 4 Гц, значно меншою мірою зазнають спотворень під час стиснення за допомогою GSM, ніж дескриптори, які ґрунтуються на спектральних характеристиках сигналу,

таких як MFCC, енергія сигналу SBE (вираз 1.2), спектральний центроїд SC (1.3), спектральна рівномірність SFM (1.5), коефіцієнт переходу рівня сигналу через нуль ZCR (1.7) та їх похідні для різних частотних діапазонів.

Рівень шуму SNR (Signal to Noise Ratio) для кожного типу дескриптора визначався за допомогою такої формули:

$$SNR = 10 \log_{10} \frac{\sum (F_m)^2}{\sum (F_m - F_M^c)^2} \quad (2.4)$$

де F_m – значення дескриптора в вихідному сигналі і значення цього дескриптора після GSM кодування/декодування. Результати експерименту наведені в таблиці 2.1.

Таблиця 2.1 –Величини співвідношення сигнал/шум (db) для різних типів дескрипторів сигналів при їх передачі по GSM каналу.

Тип сигналу	Мова		Музичні твори різних жанрів			
	Мова	Спів	Піаніно	Віолончель	Класична	Танцювальна
SBE	3.45	2.62	2.65	2.22	1.29	0,57
SFM	9.13	6.33	6.65	4.22	4.48	2.78
SC	5.06	6.06	-3.18	-8.00	-2.53	-3.41
MFCC	-2.50	-3.07	-2.10	1.00	0,38	0.09
ZCR	5.93	4.79	8.20	14.24	7.18	8.62
4 Гц	26.73	24.64	23.92	23.77	19.63	19.85

Згідно з даними, представленими в таблиці 2.1, співвідношення сигнал/шум для дескриптора, заснованого на спектральному аналізі з частотою 4 Гц у каналі GSM, в середньому на 19 dB перевищує відповідні значення для спектральних параметрів. Ці результати підтверджують, що повільні зміни амплітуди акустичних сигналів мають значно більшу стійкість до цифрового стиснення GSM, порівняно з акустичними спектральними

параметрами. Тому для ідентифікації аудіосигналів пропонується використовувати дескриптори, засновані на аналізі спектра модуляції акустичних сигналів.

Сприйняття мовлення та музики людиною базується на комбінованому часовому, частотному та модуляційному аналізі акустичних сигналів. Частотний аналіз здійснюється в критичних смугах, а спектральний аналіз модуляції проводиться в модуляційних каналах, найбільш важливі з яких знаходяться в діапазоні 0-8 Гц. Повільні зміни огинаючої акустичного сигналу є критичними для розпізнавання мови в складних умовах, таких як шум, спотворення та реверберація, і також демонструють високу стійкість до цифрового стиснення з низьким бітрейтом. Окрім цього, для коректного розпізнавання мови важливими є не тільки амплітудні коливання (АМ), а й зміни основної частоти, спричинені частотною модуляцією (FM) в голосовому тракті. З урахуванням особливостей аудиторної системи та існуючих технологій, запропоновано новий метод ідентифікації, що базується на аналізі спектру АМ/FM модуляції акустичних сигналів у діапазоні 0,5-8 Гц. Реалізація цього методу наведена в наступному розділі.

2.3 Метод цифрових аудіо відбитків

Спираючись на характеристики повільних коливань сигналу, які були розглянуті в попередньому розділі, зокрема на їх стійкість до шумів, спотворень і цифрового стиснення, було розроблено новий метод ідентифікації аудіофрагментів через аналіз спектра модуляції [14]. Цей метод заснований на понятті спектрограми амплітуди модуляції, яка демонструє частотний розподіл амплітуди повільних змін сигналу в часі. Спектрограма амплітуди модуляції формується на основі стандартної спектрограми сигналу та показує залежність спектральної енергії від часу і частоти. У цьому розділі

досліджуються властивості частотно-часових уявлень, специфіка їх використання, а також схема створення цифрових відбитків за допомогою спектра модуляції.

При обробці сигналів часто вважається, що сигнал є стаціонарним. Однак це припущення рідко є правильним для акустичних сигналів. Частота та амплітуда музики та мовних сигналів можуть змінюватися протягом часу. Для аналізу змінних сигналів використовуються методи, такі як локальне перетворення Фур'є (STFT) [9] та частотно-тимчасовий розподіл (TFD) [6].

Локальне перетворення Фур'є полягає в поділі сигналу на маленькі перекриваючі частини, кожна з яких повинна бути достатньо короткою, щоб у межах фрагмента сигнал можна було вважати стаціонарним. Для зменшення спотворень, що виникають через розриви на межах інтеграційної області, на кожен фрагмент застосовується віконна функція.

$$STFT(t, \omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x(\tau) w(\tau - t) e^{-j\omega\tau} d\tau \quad (2.5)$$

Розмір фрагмента, що визначається характеристиками віконної функції w , впливає на тимчасове та частотне розширення. Іншим методом отримання частотно-часових уявлень є частотно-часовий розподіл, який теоретично забезпечує більш високе розширення, ніж локальне перетворення Фур'є. Одним із найбільш простих варіантів є розподіл Вігнера-Віллі (Wigner-Ville Distribution, WVD).

$$WVD(t, \nu) = \int_{-\infty}^{\infty} x(t + \tau/2) x^*(t - \tau/2) e^{-j2\pi\nu\tau} d\tau \quad (2.6)$$

Основним недоліком частотно-часових розподілів є значні артефакти, спричинені взаємодією численних частотних компонентів. Інші типи таких розподілів, наприклад, Чоя-Вільямса (Choi-Williams), дозволяють зменшити ці

артефакти, але їх розширення все одно обмежується принципом невизначеності. Для задач ідентифікації та порівняння аудіосигналів не потрібно високе частотне розширення. Водночас інтеркомпонентні артефакти, характерні для TFD, є дуже небажаними і можуть значно знизити реальну роздільну здатність алгоритму, особливо в умовах сильних перешкод і спотворень. Тому в цій роботі запропоновано використовувати локальне перетворення Фур'є, де для досягнення необхідного частотного розширення спектра модуляції 8 Гц розмір вікна становить 62.5 мс. У такому випадку частотне розширення складає 32 Гц, що є меншим за ширину критичних смуг і, як показано в параграфі 2.2.1, достатнім для ідентифікації аудіосигналів. Крім того, локальне перетворення Фур'є є простим у реалізації та дає передбачувані результати.

Таким чином, в запропонованому методі спектрограма $S(t, \omega)$ акустичного сигналу $x(t)$ отримується за допомогою локального перетворення:

$$S(t, \omega) = \frac{1}{2\pi} \left| \int_{-\infty}^{\infty} x(\tau) w_s(\tau - t) e^{-j\omega t} d\tau \right|^2 \quad (2.7)$$

де $w_s(t)$ є віконної функцією.

Так як, віконна функція $w(t)$ має ненульові значення в досить вузькому часовому інтервалі, локальне перетворення Фур'є відображає частотні властивості сигналу в даний відрізок часу. Спектр модуляції $M(t, \omega, v)$ обчислюється за допомогою другого перетворення Фур'є для кожного частотного діапазону:

$$M(t, \omega, v) = \frac{1}{r} \int_{-\infty}^{\infty} S(t, \omega) e^{-jvt} dt \quad (2.8)$$

В даному випадку фактор r використовується для нормалізації значень в кожному діапазоні. Відповідно спектрограма амплітуди модуляції P обчислюється як

$$P(t, \omega, v) = |M(t, \omega, v)|^2 \quad (2.9)$$

Дослідним шляхом було виявлено, що побудована спектрограма зберігає свою стійкість до різноманітних спотворень і навіть після суттєвого квантування може слугувати основою для порівняння та ідентифікації різних аудіосигналів [3].

Як зазначалося раніше, аналіз спектра модуляцій у звукових сигналах демонструє високу ефективність у задачах автоматичного розпізнавання мови [9]. У ньому вхідний сигнал поділяється на частотні смуги за допомогою набору фільтрів з незначним перекриттям. Після детектування сигнал у кожній смузі проходить фільтрацію, нормалізацію, а потім спектр модуляції визначається через перетворення Фур'є.

Подібний принцип покладено в основу методу цифрових відбитків, що базується на спектрі модуляції та був запропонований Sukittanon і Atlas [17]. Головна відмінність цієї технології від згаданого раніше підходу до розпізнавання мовлення [4] полягає у способі вилучення параметрів модуляції: замість повторного застосування перетворення Фур'є тут використовується вейвлет-аналіз, який забезпечує кращу деталізацію в області низьких частот. У реалізації, побудованій за алгоритмом із [7], дескриптор формується шляхом оцінювання геометричного середнього модуляційної спектрограми аудіосигналу тривалістю 12 секунд. Сигнал при цьому розбивається на перекривані вікна по 4 секунди з частотним розкладом на 19 барків.

Результати дослідження продемонстрували, що метод, незважаючи на свої особливості, поступається за загальними характеристиками рішенню компанії Philips [3]. Зокрема, алгоритм виявився дуже вразливим до

спектральних спотворень: застосування еквалайзерної корекції призводить до зниження рівня повноти розпізнавання до 90%, що є критично неприйнятним для практичного застосування в системах ідентифікації та пошуку аудіовізуального контенту. Крім того, метод надто чутливий до ефектів, що виникають при компресії сигналів. Наприклад, для MP3-кодування з бітрейтом 96 кбіт/с, яке зазвичай забезпечує прийнятну якість для монофонічного звучання, повнота ідентифікації склала лише 96,5%. Для порівняння, алгоритм Philips демонструє високу стійкість навіть при стисканні до 32 кбіт/с.

Варто підкреслити, що витривалість до впливу стиснення має велике прикладне значення, оскільки безпосередньо визначає ефективність роботи пошукових систем, що оперують сигналами, переданими через цифрові канали. Слід також зазначити, що всі вищенаведені експериментальні дані отримані для інтервалу інтегрування тривалістю 10 секунд. При скороченні цього періоду до 5 секунд (тобто при зменшенні довжини фрагмента для аналізу) середня повнота розпізнавання знижується до 80%. У практичних умовах слід враховувати, що аудіосигнали рідко піддаються лише одному типу спотворення — зазвичай це комбінація впливів, включаючи шум, спектральні зміщення та цифрове стискання.

Запропонований нами новий алгоритм формування цифрових відбитків, схема якого наведена на рисунку 2.2, позбавлений вищезазначених недоліків.

У його основі лежить спектрограма сигналу, що будується з використанням локального перетворення Фур'є та ділиться на $N = 13$ частотних смуг із логарифмічним розподілом частот. Для кожної з цих смуг обчислюється спектр модуляцій за допомогою повторного застосування перетворення Фур'є. Далі результати проходять процедуру нормалізації, а також стискання через диференціальне кодування із подальшим квантуванням. Після надходження нового аудіофрагмента тривалістю 50 мс,

виконується повторне обчислення, внаслідок чого формується послідовність цифрових відбитків із продуктивністю 240 байт на секунду.

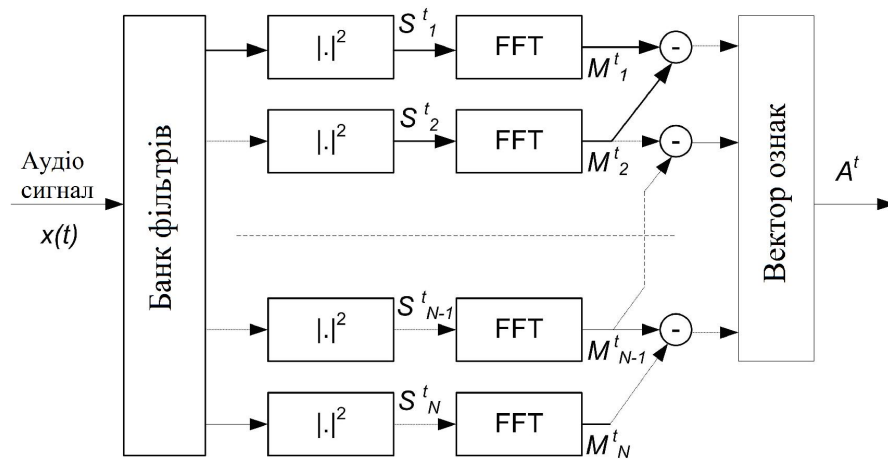


Рисунок 2.2 – Схема запропонованого алгоритму отримання вектора на підставі ознак спектра модуляції

Розділення за частотними смугами виконується з використанням банку фільтрів, сформованого з набору кусочно-безперервних смугових фільтрів із перекриттям, які мають лінійну характеристику спадання та охоплюють частотний діапазон від 200 до 3200 Гц .

Лінійна характеристика похилих фільтрів дозволяє звести до мінімуму спотворення під час демодуляції FM-сигналів, а також сприяє ефективному розділенню амплітудно- та частотно-модульованих компонентів у спектрі модуляції. Враховуючи розширення частотного діапазону слухової системи людини, а також особливості структури європейської музичної шкали, можна зробити припущення, що для більшості реальних аудіосигналів у межах кожної частотної смуги, ширина якої не перевищує критичне значення, присутній лише один переважаючий тональний компонент. Для перевірки цього припущення досліджена модель, за якої акустичний сигнал $x(t)$

представлений у вигляді несучої частоти ω_c , модульованої по частоті сигналом $f_m(t)$, а по амплітуді сигналом $a_m(t)$:

$$x(t) = [1 + \alpha a_m(t)] \cdot A_c \cdot \sin(\omega_c t + \varphi_o + \beta f_m(t)) \quad (2.10)$$

де α і β є коефіцієнтами АМ і FM модуляції.

У разі модуляції синусоїдальними сигналами $f_m(t)$ і $a_m(t)$ без урахування фази сигналів і амплітуди несучої частоти цей вираз може бути представлено в наступному вигляді:

$$x(t) = [1 + \alpha \sin(vat)] \cdot \sin(\omega_c t + \beta \sin(\vartheta_f t)) \quad (2.11)$$

Сумісне виділення амплітудно- та частотно-модульованих компонентів сигналу здійснювалося за допомогою дискримінатора, реалізованого у вигляді смугового фільтра, налаштованого на частоту, дещо вищу за основну частоту вхідного сигналу. Після проходження через фільтр виконувалося детектування та екстракція огибаючої.

Після детектування спектр модуляції для кожного частотного діапазону обчислюється за допомогою перетворення Фур'є.

Квантування спектра модуляції, що складається з K коефіцієнтів, здійснюється за наступною формулою:

$$A^t(i, k) = \begin{cases} 1, \text{ якщо } M^t(i, k) - M^t(i + 1, k) > 0 \\ 0 \text{ в іншому випадку} \end{cases} \quad (2.11)$$

де $k, 1 \leq k \leq K$ - індекс коефіцієнта спектра модуляції;

i - є індексом частотного діапазону в банку фільтрів.

Результати експериментів засвідчили, що незважаючи на відносну простоту реалізації, розроблений алгоритм демонструє ефективність, співставну з існуючими підходами, як за точністю, так і за повнотою розпізнавання аудіосигналів, зокрема в умовах наявності шумів та різного роду спотворень. Водночас, як і більшість методів, розглянутих у попередніх розділах, цей підхід має певну вразливість до змін тривалості сигналу. Алгоритм здатен коректно обробляти звукові фрагменти, змінені в часі в межах $\pm 5\%$. Однак на практиці, наприклад у радіо- або телевізійній трансляції, зокрема в рекламних вставках, аудіоінформація може навмисно стискатись більш ніж на 10% з метою оптимального заповнення ефірного часу, що створює додаткові складнощі для надійної ідентифікації.

Одним з важливих властивостей перетворення Фур'є $F(\omega)$ є інваріантність спектра до зрушення. Перетворення Фур'є сигналу $f(t)$ і його зрушено в часі версії $f_t(t + t_0)$ відрізняються зрушенням фаз:

$$F_t(\omega) = e^{j\omega t_0} F(\omega) \quad (2.13)$$

Схожу здатність до інваріантності при масштабних змінах сигналу, таких як його стискання чи розтягування, має і Меллін-перетворення.

Перетворення Мелліна D_f для функції $f(t)$ визначається наступним виразом:

$$D_f(s) = \int_0^\infty f(t)t^{s-1} dt \quad (2.14)$$

У контексті цифрової обробки сигналів перетворення Мелліна частіше відоме як масштабне перетворення (Scale Transform) [100], [46], що фактично є його окремим випадком при $\beta = 1/2$

$$D_f(c) = \frac{1}{\sqrt{2\pi}} \int_0^\infty f(t) e^{(-jc+\beta-1) \ln t} dt = \frac{1}{\sqrt{2\pi}} \int_0^\infty f(t) \frac{e^{-jc \ln t}}{\sqrt{t}} dt, t \geq 0 \quad (2.15)$$

Перетворення Мелліна може бути виражено через перетворення Фур'є F функції $f_k(t) = f(e^t) e^{\beta t}$, $\beta \in R$, яка представляє собою варіант вихідного сигналу, представленого в логарифмічній шкалі часу

$$F_k(c) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty f_k(t) e^{-jct} dt = \frac{1}{\sqrt{2\pi}} \int_0^\infty f(\tau) e^{(-jc+\beta-1) \ln \tau} d\tau \quad (2.16)$$

Подібно до локального перетворення Фур'є, локальне Меллін-перетворення може бути записане у такій формі:

$$D_t(c) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty f(\tau) w_m(\tau - t) e^{(-jc+\beta-1) \ln \tau} d\tau \quad (2.17)$$

У цій роботі швидке Меллін-перетворення реалізовано шляхом застосування швидкого перетворення Фур'є до сигналу, масштабованого в часі за експоненційним законом у межах інтервалу, де віконна функція $h(\tau)$ набуває ненульових значень. Щоб здійснити масштабування у часі, використовувалась інтерполяція значень локального сигналу, розташованих з експоненційним кроком.

Ключовою властивістю масштабного перетворення є його інваріантність до змін масштабу сигналу $f(t)$. Під масштабуванням зазвичай мають на увазі стискання або розтягнення сигналу в часовій області без зміни його енергії. В результаті, масштабований варіант сигналу $g(t)$ можна побудувати таким чином: $g(t) = \sqrt{\alpha} f(\alpha t)$ де $\alpha \in R^+$ представляє масштаб перетворення.

Відношення між масштабним перетворенням початкового сигналу $f(t)$ та його часово стисненою або розтягнутою версією виражається наступним чином:

$$D_g(c) = \alpha^{jc} D_f(c) = e^{jc \ln \alpha} D_f(c) \quad (2.18)$$

Слід зауважити, що хоча перетворення Мелліна є масштабно-інваріантним, воно не зберігає інваріантності до зсуву. Тому в практичних застосуваннях для зменшення чутливості цифрового відбитка до зсуву та послаблення вимог до синхронізації сигналів доцільно використовувати більший ступінь перекриття вікон під час обчислення спектрограми за допомогою локального перетворення Фур'є.

2.4 Висновок до розділу 2

Експериментальні дослідження показали, що наявні методи не забезпечують надійної ідентифікації акустичних сигналів, змінених у часовому масштабі, і тому є непридатними для моніторингу радіо- та телевізійного контенту.

Проведено аналіз властивостей різних типів дескрипторів акустичних сигналів. На основі експериментальних результатів встановлено, що найвищу стійкість до спотворень і цифрового стиснення демонструє метод, який ґрунтується на довготривалому аналізі змін енергії в окремих частотних діапазонах.

Удосконалено метод ідентифікації акустичних сигналів шляхом використання аналізу спектра модуляції у частотних діапазонах. Для підвищення стійкості алгоритму до змін часової шкали сигналу запропоновано застосування перетворення Мелліна та його модифікованої версії, основаної на дискретному косинусному перетворенні. Розроблений алгоритм демонструє високу стійкість до часових деформацій аудіосигналів і забезпечує надійну ідентифікацію аудіовізуального контенту з аудіосупроводом.

3 РОЗРОБКА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

3.1 Обґрунтування вибору мови програмування

Розробка додатку здійснюється з використанням мови програмування C++, яка є універсальною алгоритмічною мовою, широко застосовуваною для створення як системного, так і складного прикладного програмного забезпечення. C++ дозволяє оперувати не лише змінними, а й їхніми адресами, розміщувати дані як у пам'яті, так і в регістрах, застосовувати непряму адресацію, а також автоматично змінювати адреси.

З розвитком інформаційних технологій виникла потреба у створенні інструментів, які забезпечують швидку розробку програм для найпоширенішої операційної системи — Windows. До таких інструментів належать Microsoft Visual Basic, Borland Delphi, а також більш сучасна система — Borland C++ Builder. Ці середовища програмування реалізують принципи візуального проектування інтерфейсів та подієвого програмування, що значно спрощує процес створення додатків. Програмування в таких середовищах здебільшого зводиться до компонування інтерфейсу з готових елементів та написання обробників подій. Велика частина коду генерується автоматично, що дає змогу розробникам зосередитися на логіці програми.

Мова C++ є мовою високого рівня, на основі якої створено низку популярних середовищ програмування, зокрема Borland C++, Visual C++ та Borland C++ Builder. Найбільш широке застосування серед них отримав саме C++ Builder — потужне середовище візуального об'єктно-орієнтованого програмування, яке дає змогу як новачкам, так і досвідченим розробникам створювати користувацькі інтерфейси для прикладних програм різної складності.

C++ Builder, розроблений компанією CodeGear (нині належить Embarcadero), поєднує Бібліотеку візуальних компонентів (VCL) та

інтегроване середовище розробки (IDE), що використовує компілятор C++. Архітектура C++ Builder схожа з Delphi, однак має низку вдосконалень. Більшість компонентів, створених у Delphi, можуть використовуватись у C++ Builder без модифікації, що забезпечує гнучкість розробки. Завдяки підтримці візуального проектування за принципом drag-and-drop, середовище дозволяє швидко створювати інтерфейси з використанням редактора WYSIWYG.

Під час розробки додатку було використано компоненти з вкладок Standard, Additional, System, Dialog, Samples, зокрема:

1. Label — для виведення текстових написів на формі;
2. Edit — для створення полів вводу текстових рядків;
3. Memo — для редагування та виведення великих текстових блоків;
4. Button — стандартні кнопки з підписами для запуску функцій;
5. DrawGrid — компонент для створення таблиць з можливістю графічного відображення;
6. Timer — для реалізації функціоналу, що залежить від часу;
7. OpenDialog і SaveDialog — стандартні діалоги відкриття та збереження файлів;
8. PerformanceGraph — для побудови графіків залежно від обраних семплів.

3.2 Реалізація додатку для автоматизованої системи ідентифікації та захисту аудіоконтенту

Для створення програмного забезпечення, що реалізує формування цифрових аудіовідбитків, окрім середовища C++ Builder, було додатково використано бібліотеку bass.dll. Ця бібліотека надає широкий набір функцій для обробки аудіосигналів у реальному часі, включаючи роботу з потоками, відтворенням та аналізом аудіоданих.

Основні етапи алгоритму

Процес формування цифрового відбитка базується на попередньо описаному алгоритмі і включає такі ключові етапи:

Аудіофайл розбивається на послідовні частини (кадри), кожна з яких є окремою вибіркою, що використовується для аналізу. Як було зазначено у підрозділі 2.1, для ідентифікації композиції достатньо обробити фрагмент тривалістю 3 секунди.

Перед використанням будь-якого аудіопотоку необхідно викликати функцію ініціалізації.

Вона задає пристрій, частоту дискретизації (44.1 кГц) і параметри відтворення.

На рисунку 3.1 наведена діаграма етапу ініціалізації та створення потоку

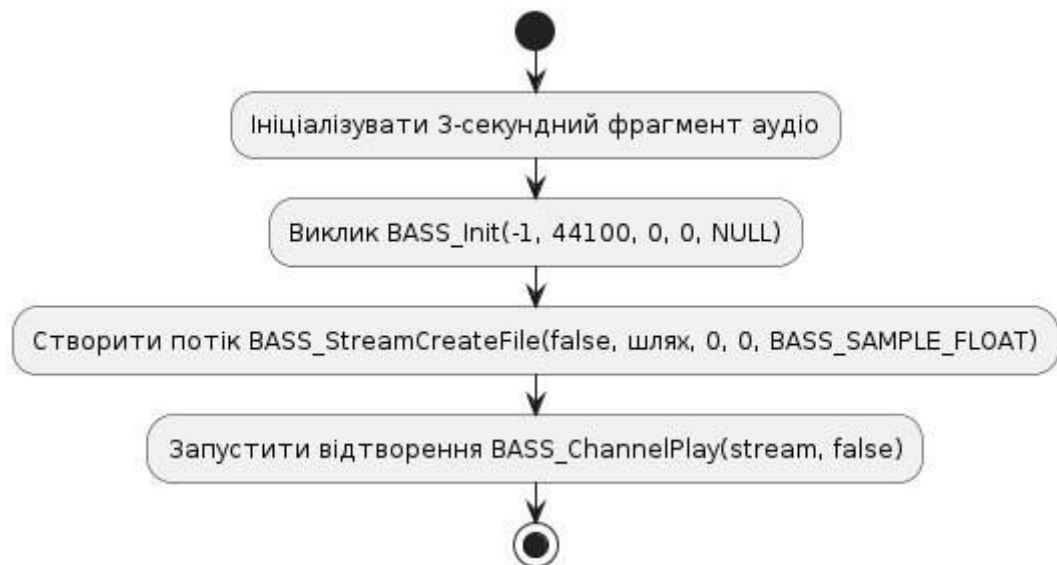


Рисунок 3.1 – Діаграма етапу ініціалізації та створення потоку

Цей виклик дозволяє встановити позицію потоку, з якої буде зчитано дані. На рисунку 3.2 наведена діаграма фреймування, обчислення енергії та створення бітової карти

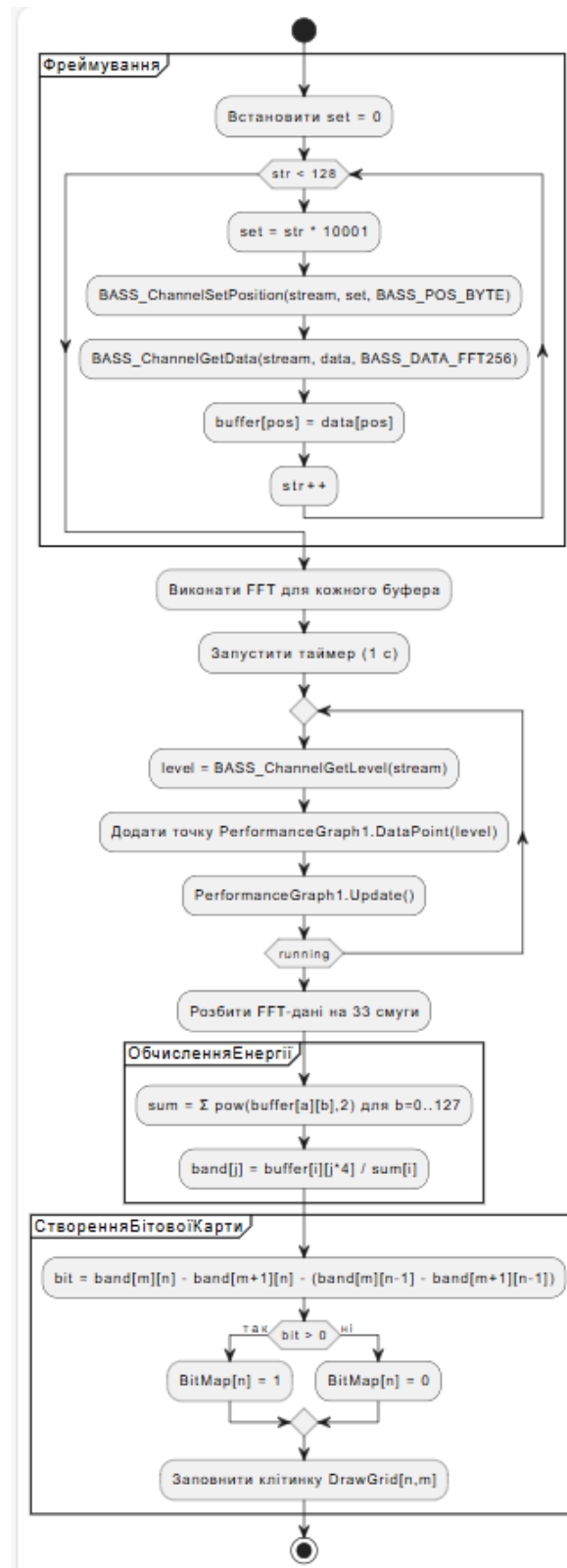


Рисунок 3.2 – Діаграма фреймування, обчислення енергії та створення бітової карти

Функція для отримання вибірки дозволяє отримати набір частотних компонент, необхідних для подальшого спектрального аналізу.

Для аналізу спектра застосовується функція FFT

Під час виклику `BASS_ChannelGetData` із параметром `BASS_DATA_FFT256` обробляється 256 семплів, результатом чого є масив із 128 значень (оскільки половина спектра є дзеркальною, з протилежними знаками).

Для візуалізації результатів обробки використовується таймер та компонент `PerformanceGraph` (вкладка *Samples*), який оновлює графік значень з певною періодичністю.

Таким чином, за допомогою поєднання засобів середовища `C++ Builder`, бібліотеки `bass.dll`, алгоритмів спектрального аналізу та візуальних компонентів було реалізовано додаток, що дозволяє формувати та відображати цифрові аудіовідбитки для подальшого використання у системах захисту та ідентифікації аудіоконтенту.

Для створення цифрового відбитку аудіофайлу дані ділять на смуги за допомогою циклу.

Отримується 33 неперекриваючі смуги. Далі обчислюють енергію сигналу за формулою 1.2. Спочатку сумують рядки, які є константою для субвідбитків, що прискорює програму

Потім розраховують чисельник формули 1.2 для кожного субвідбитку.

Значення субвідбитку ділять на квадрат суми всіх значень 32-бітного субвідбитку

У масиві `band` зберігається енергія кожного біта субвідбитку, необхідна для створення бітової карти. Для визначення значень бітів застосовують формулу 1.6: якщо результат ≤ 0 , біт білий, якщо > 0 — чорний.

Для відображення відбитку використовують компонент `DrawGrid` із закладки `Additional`, фарбуючи клітинки в чорний або білий колір.

На рисунку 3.3 наведена діаграма відображення відбитку відбитку та його збереження.

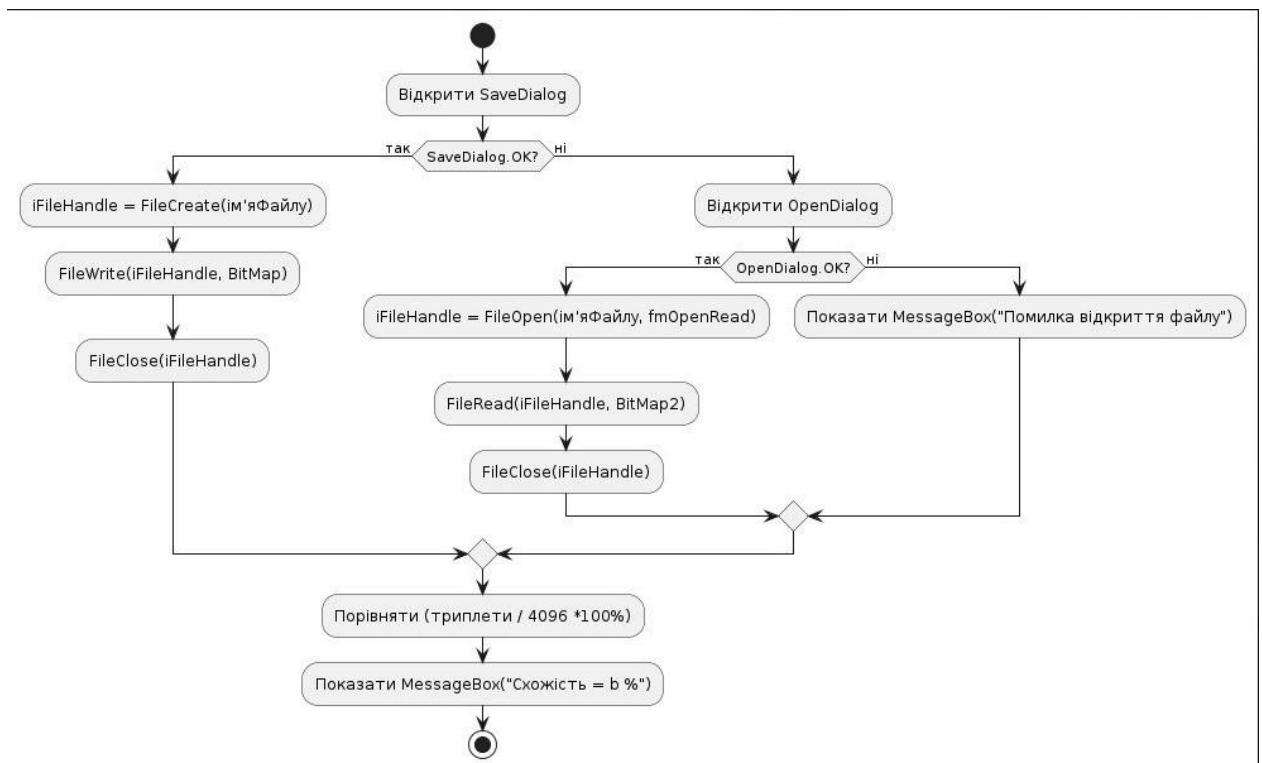


Рисунок 3.3 – Діаграма відображення відбитку та його збереження

Для збереження відбитків створено функцію з використанням FileWrite та компонента SaveDialog (зкладка Dialogs). Змінна iFileHandle вказує шлях або ім'я файлу. У разі помилки програма видає повідомлення через MessageBox.

Для завантаження відбитків використовують функцію з компонентом OpenFileDialog та командою FileRead. У разі помилки також видається повідомлення.

Для порівняння відбитків створено функцію, яка аналізує три біти поспіль. Якщо вони збігаються, лічильник збільшується. Відсоток схожості розраховують діленням на 4096 і множенням на 100%.

Додатково програма має функції для завантаження аудіофайлів, керування відтворенням і зупинкою. Повний код наведено в додатку В.

3.3 Тестування розробленої програми

Результатом бакалаврської роботи стала програма для створення цифрових відбитків, що захищають аудіофайли будь-якого формату. На основі вибірок даних додаток формує двовимірну матрицю, яка відображає унікальний відбиток кожної композиції.

При запуску відкривається головне вікно програми, зображене на рисунку 3.1.

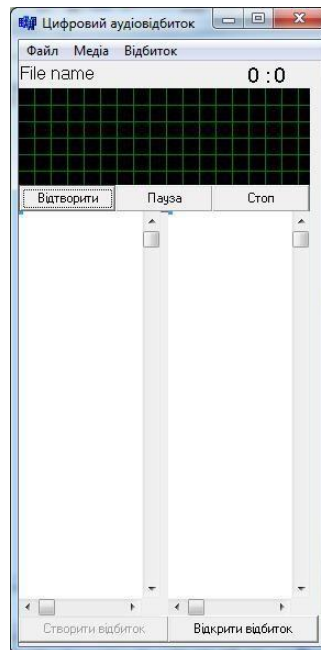


Рисунок 3.1 – Головне вікно робочої програми

Інтерфейс інтуїтивно зрозумілий, подібний до стандартних програм. У верхній частині — кнопки керування: згортання, розгортання на весь екран, відновлення розміру або закриття програми. Інтерфейс поділено на блоки: головне меню, поле з назвою аудіофайлу та секундоміром, що показує час відтворення. Нижче — блок осцилограми, який відображає графік залежності фізичних величин звуку. Ще нижче — кнопки для керування відтворенням: запуск, пауза, зупинка.

Два найбільші блоки інтерфейсу призначені для відображення цифрових відбитків. У першому показується відбиток композиції, що відтворюється, у другому — завантажується раніше створений відбиток для порівняння. Під цими блоками розташовані кнопки для основних дій з відбитками: створення нового та завантаження збереженого. На цьому інтерфейс головного вікна завершується.

Головне меню програми, зображене на рисунку 3.2, містить групи команд: «Файл», «Медіа» та «Відбиток». Пункт «Файл» дозволяє завантажити аудіофайл для створення цифрового відбитку або порівняння з існуючими, а також завершити роботу програми.

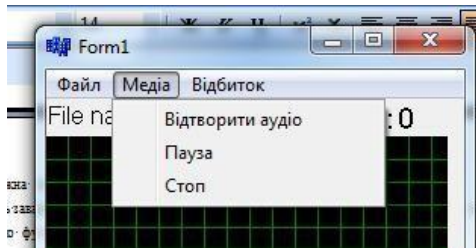


Рисунок 3.2 – Головне меню програми

Для зручного вибору аудіофайлу або цифрового відбитку використовується діалогове вікно, яке допомагає швидко знайти та обрати потрібний файл.

Пункт меню «Медіа» забезпечує керування відтворенням композиції. Як показано на рисунку 3.2, він включає команди «Відтворити аудіо», «Пауза» та «Стоп», які також дублюються на головному вікні програми.

Пункт меню «Відбиток» призначений для роботи з цифровими відбитками. Його перша команда — «Зберегти відбиток» — дозволяє зберегти сформований відбиток у файл формату txt. Це необхідно для подальшого порівняння відбитків, виявлення неліцензійних копій та захисту авторських прав.

Наступна команда меню «Відбиток» — «Відкрити відбиток» — дозволяє завантажити раніше створені цифрові аудіо відбитки для порівняння та визначення унікальності композиції.

Третя команда, «Порівняти відбиток», запускає процес порівняння створеного відбитку з раніше збереженими. Обидва відбитки завантажуються у відповідні блоки інтерфейсу для роботи.

Під головним меню розташований «Інформаційний блок» (рисунок 3.3), який відображає назву поточної композиції та таймер, що показує момент відтворення аудіофайлу.

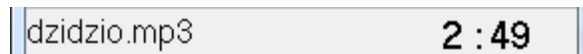


Рисунок 3.3 – Інформаційний блок

Наступний елемент програми відображає осцилограму, яка характеризує аудіофайл, що відтворюється в певний момент. Цей блок візуалізує вхідний сигнал у графічному вигляді.

При цифровому звукозаписі аудіодоріжка представлена осцилограмою, що показує форму звукової хвилі, тобто залежність амплітуди від часу. Осцилограма дозволяє спостерігати ключові події відтворення, такі як зміни гучності, паузи між фрагментами композиції чи навіть окремі ноти в сольних інструментальних записах.

На основі цих даних створюється унікальний опис аудіофайлу, необхідний для формування цифрового аудіовідбитку (рисунок 3.4).

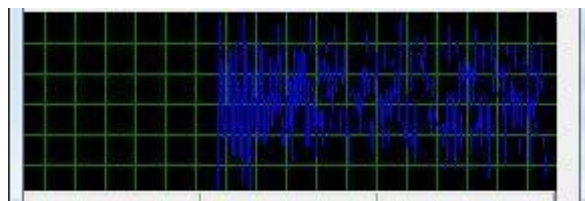


Рисунок 3.4 – Блок відображення осцилограми

Наступний блок, розташований після графічної частини для відображення осцилограми, призначений для керування відтворенням композиції (рисунок 3.5). Він містить три кнопки: «Відтворити» запускає аудіофайл, «Пауза» призупиняє програвання, а «Стоп» повністю зупиняє відтворення.

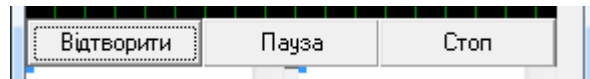


Рисунок 3.5 – Блок керування програвання аудіо файлу

На рисунку 3.6 показано блок відображення цифрових відбитків. Ця частина інтерфейсу призначена для показу цифрових відбитків, створених програмою або попередньо сформованих. Кожен відбиток складається з 32 бітів у рядку, загалом 256 рядків.

Як зазначалося раніше, звуковий відбиток (сигнатура) — це компактний опис аудіофайлу, що відображає його звукові характеристики. Ці характеристики мають бути стійкими до модифікацій файлу, таких як стиск, аналого-цифрове чи цифро-аналогове перетворення, фільтрація тощо, і водночас пов’язані зі сприйняттям звуку.



Рисунок 3.6 – Блок представлення цифрових аудіовідбитків

Цифровий відбиток представлено у вигляді хеш-таблиці, де ключем є значення частоти піків інтенсивності та їхніх опорних точок, а відповідним значенням — час у секундах від початку треку.

Для створення відбитків використовуються стійкі до спотворень звукові характеристики, зокрема коефіцієнти Фур'є, коефіцієнти косинусного перетворення Фур'є (MFFC) та нерівномірність спектра (spectral flatness).

На рисунку 3.8 зображено дві кнопки, що спрощують роботу з цифровими аудіовідбитками. Кнопка «Створити відбиток» запускає генерацію відбитка для аудіофайлу, що відтворюється. Під час відтворення програма обробляє набір даних (вибірки) для кожного моменту часу, формуючи на їх основі бітову карту цифрового відбитка. Після натискання кнопки відбиток відображається в лівій частині блоку, зображеного на рисунку 3.7.

Кнопка «Відкрити відбиток» дозволяє завантажити раніше створений відбиток у праву частину цього блоку. Це дає змогу порівнювати відбитки для перевірки їхньої унікальності, забезпечуючи захист авторських прав і запобігаючи нелегальному використанню чи поширенню аудіокомпозицій.



Рисунок 3.7 – Кнопки для керування аудіо відбитками

Щоб користувач міг опублікувати аудіокомпозицію в інтернеті, вона повинна пройти перевірку на авторські права. Якщо відбиток аудіофайлу виявиться подібним до наявного, розповсюдження буде заборонено.

Програма видасть повідомлення з результатом схожості (рисунок 3.8).

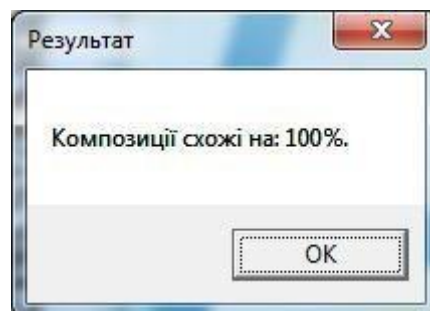


Рисунок 3.8 – Інформаційне повідомлення

Головна мета цифрових аудіовідбитків – захист композицій від піратства та нелегального поширення. Розроблена програма створює відбитки на основі даних, що характеризують кожну аудіокомпозицію.

За підсумками дипломної роботи, запропонований метод захисту аудіофайлів від несанкціонованого копіювання є ефективним для вирішення завдання. Однак вбудувати ідентифікаційну мітку в усі копії неможливо. Цифрові відбитки дозволяють відстежувати навіть значно змінені версії оригінального файлу. Особливість методу полягає в імовірнісній оцінці

схожості аудіофайлів, що супроводжується двома типами помилок: помилкою першого роду (нерозпізнавання наявного файлу) та помилкою другого роду (хибне визначення іншого файлу). У методі спектрограм помилка другого роду в 2,5 раза вища, ніж у покадровому обчисленні відбитків. Для підвищення точності розпізнавання пропонується додати надмірність, що може збільшити тривалість пошуку.

3.4 Характеристики підсистеми пошуку та розпізнавання аудіоматеріалів за цифровими відбитками

Для порівняння аудіофрагментів обчислювали коефіцієнт бітових розбіжностей BER між відповідними відбитками. Фрагменти вважаються ідентичними, якщо BER нижчий за встановлений поріг. На рисунку 3.9 показано залежність BER від зсуву при порівнянні відбитка музичного твору з оригінальним відбитком (рис. 3.10А), з відбитком сигналу з шумом (SNR = 0 дБ, рис. 3.11Б) та з відбитком сигналу, стисненого в часі зі збереженням тональності (рис. 3.10В). Для незбіжних відбитків BER становить 0,45–0,55, тоді як при збігу відбитків значення нижче 0,40.

Експерименти показали, що для точної ідентифікації послідовностей відбитків поріг BER має бути 0,37, що забезпечує відсутність помилок навіть для коротких послідовностей. У практичній реалізації застосовується динамічний поріг BER, який залежить від довжини послідовності. Це дозволяє алгоритму розпізнавати слабко спотворені 5-секундні фрагменти з порогом $BER = 0,37$.

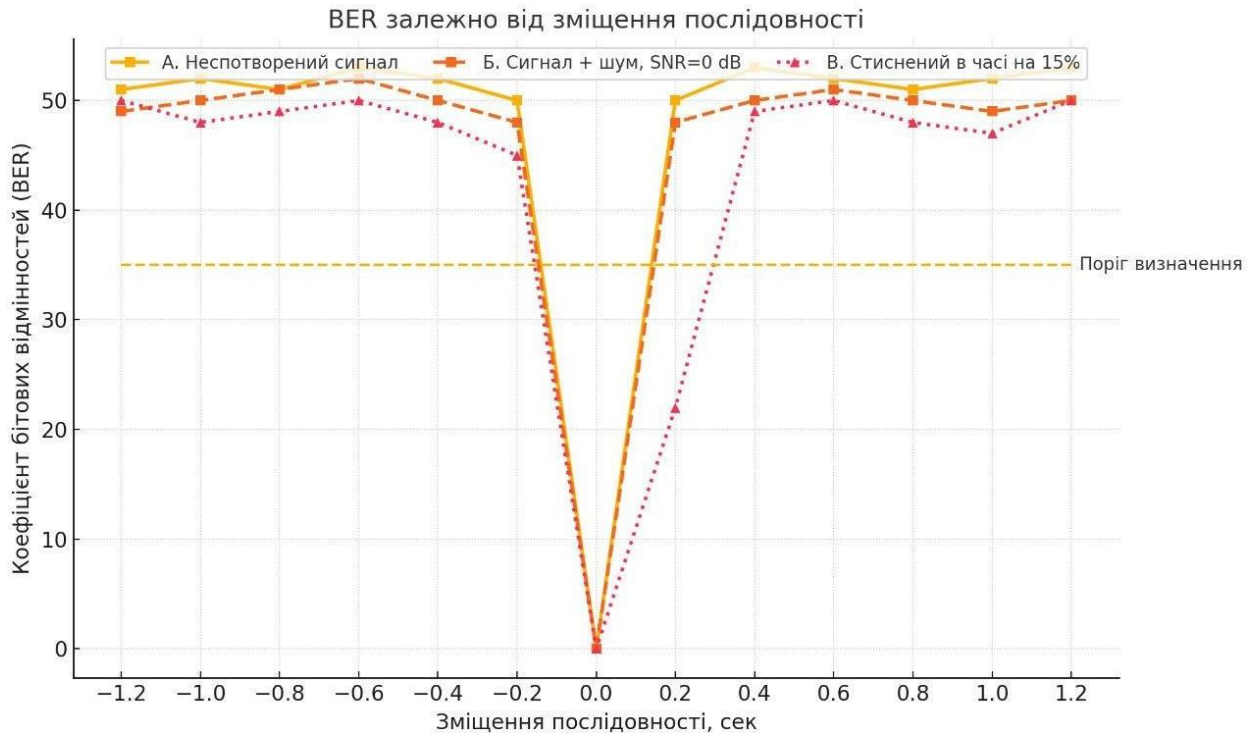


Рисунок 3.9 – Залежність коефіцієнта бітових відмінностей від зсуву при порівнянні музичного відбитка з вихідною послідовністю (А), в присутності шуму $SNR = 0\text{ dB}$ (Б) і масштабування в часі на 15% (В)

Для довгих збіжних послідовностей поріг ідентифікації підвищується до $BER = 0,40$, що забезпечує розпізнавання сигналів за значного шуму без втрати точності методу.

Чутливість сучасних методів цифрових відбитків до змін швидкості відтворення є значним обмеженням для ідентифікації аудіоматеріалів, що транслюються радіоканалами. Компанії, які надають контент через кабельні мережі чи радіо, часто навмисно змінюють швидкість програвання, щоб оптимізувати ефірний час і дотримуватися розкладу. Такі зміни також виникають при конвертації аудіо між цифровими форматами, зокрема при переході від стандарту PAL до NTSC, коли частота кадрів знижується з 25 до 23,976 кадрів/с, а група з 4 кадрів перетворюється на 5 (3:2 pulldown). Це призводить до сповільнення відтворення на 4%, що непомітно для слухача, але ускладнює роботу алгоритмів ідентифікації.

На рисунку 3.10 показано залежність коефіцієнта бітових розбіжностей запропонованого методу від ступеня стиснення/розтягування сигналу в часі. Порівняно з базовим методом, застосування перетворення Мелліна підвищило стійкість алгоритму до часового масштабування в 2,5 раза. При порозі $BER = 0,37$ це відповідає максимальному прискоренню/уповільненню на 40%.

У реальних умовах масштабування рідко перевищує 15%, але для надійного розпізнавання матеріалів із комплексними спотвореннями потрібен значний запас продуктивності.

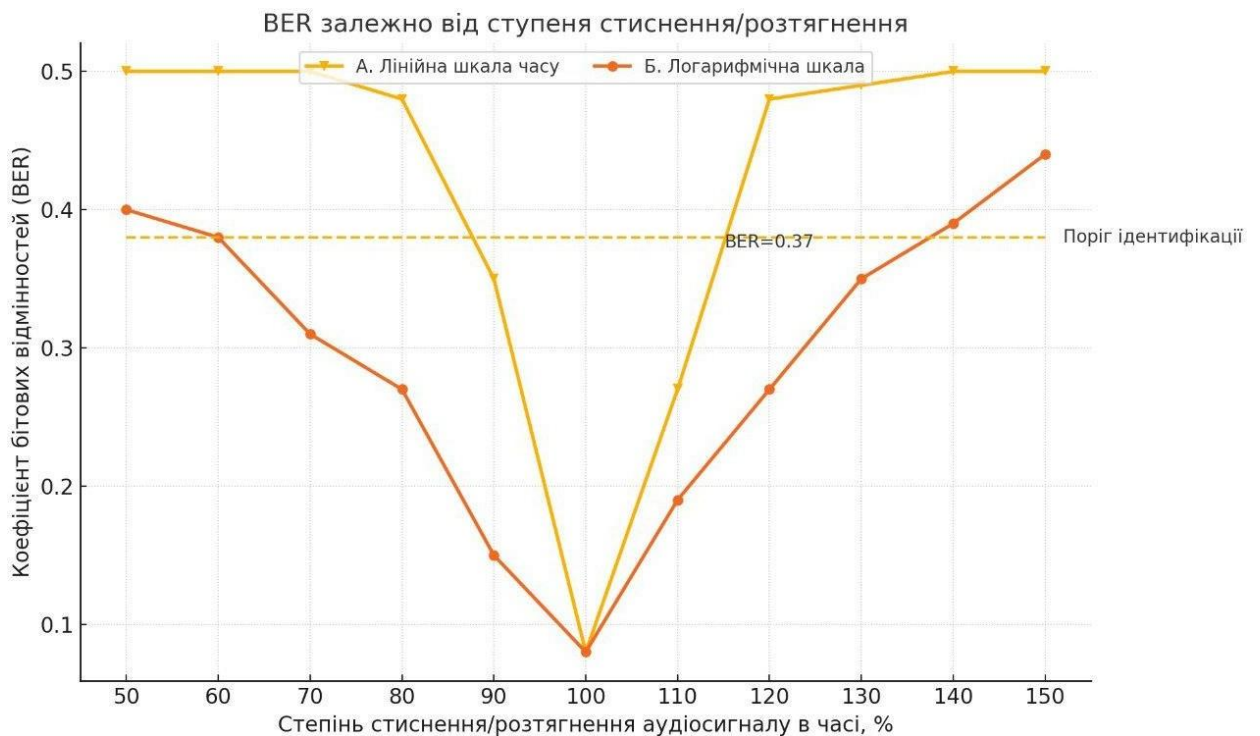


Рисунок 3.10 – Залежність коефіцієнта бітових відмінностей від ступеня масштабування сигналу в часі для відбитків, отриманих на підставі аналізу спектра модуляції в лінійній і логарифмічній шкалі

Варто відзначити, що навмисне введення часових спотворень є одним із методів атаки на системи ідентифікації аудіосигналів [17]. Підвищення стійкості до таких атак має важливе значення для систем пошуку та виявлення

несанкціонованих копій аудіо- та відеоматеріалів в Інтернеті та за його межами.

Функціональні характеристики приймача (Receiver Operating Characteristics, ROC) для двох розглянутих методів за умов помірного шуму (співвідношення сигнал/шум 12 дБ) зображено на рисунку 3.11. ROC відображає залежність частки правильних ідентифікацій сигналу (True Positive Rate, TPR) від частки помилкових спрацьовувань (False Positive Rate, FPR) при різних порогових значеннях коефіцієнта бітових розбіжностей BER.

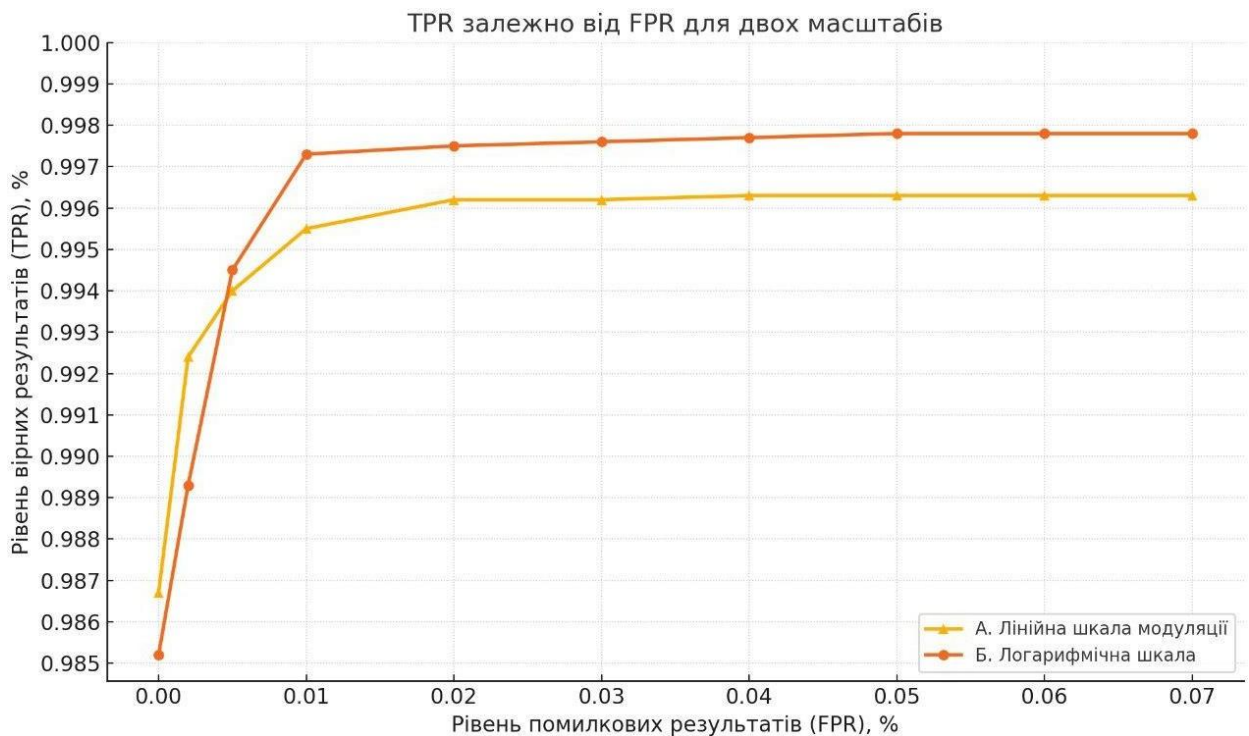


Рисунок 3.11 – Функціональні характеристики приймача аудіосигналу (SNR = 12db) для відбитків, отриманих на підставі аналізу спектра модуляції в лінійній і логарифмічній шкалі

Аудіовідбитки, створені шляхом аналізу спектру модуляції в логарифмічній шкалі, мають удвічі нижчий рівень помилок (0,0025%) порівняно з відбитками, отриманими в лінійній шкалі (0,0045%).

Стійкість алгоритму до зсувів значною мірою залежить від рівня перекриття вікон локального перетворення Фур'є під час обчислення спектрограми та частоти формування вектора ознак у вихідному потоці. Експерименти показали, що перекриття вікон на 80% забезпечує достатню стійкість відбитків до зсувів. Збільшення частоти генерації елементів відбитку зменшує вплив невідповідності положень порівнюваних послідовностей, але збільшує розмір відбитку, що знижує ефективність пошукової системи. Зазвичай достатньо частоти 20 субвідбитків за секунду.

На рисунку 3.12 зображено залежність коефіцієнта бітових розбіжностей від ступеня зсуву акустичних сигналів. Для відбитків у лінійній шкалі часу коефіцієнт становить $BER = 0,22$, а для логарифмічної шкали – $BER = 0,13$, що свідчить про підвищення стійкості алгоритму до зсувів на 61% при використанні логарифмічної шкали.

Варто зазначити, що вплив зсувів на рівень помилок значно знижується за наявності шуму та спотворень, які маскують ефект десинхронізації відбитків. Проте в задачах оцінки якості аудіосигналів за коефіцієнтом бітових розбіжностей порівняно з еталонним відбитком похибка через неузгодженість може бути значною, особливо для високоякісних сигналів (MP3 64 кбіт/с і вище).

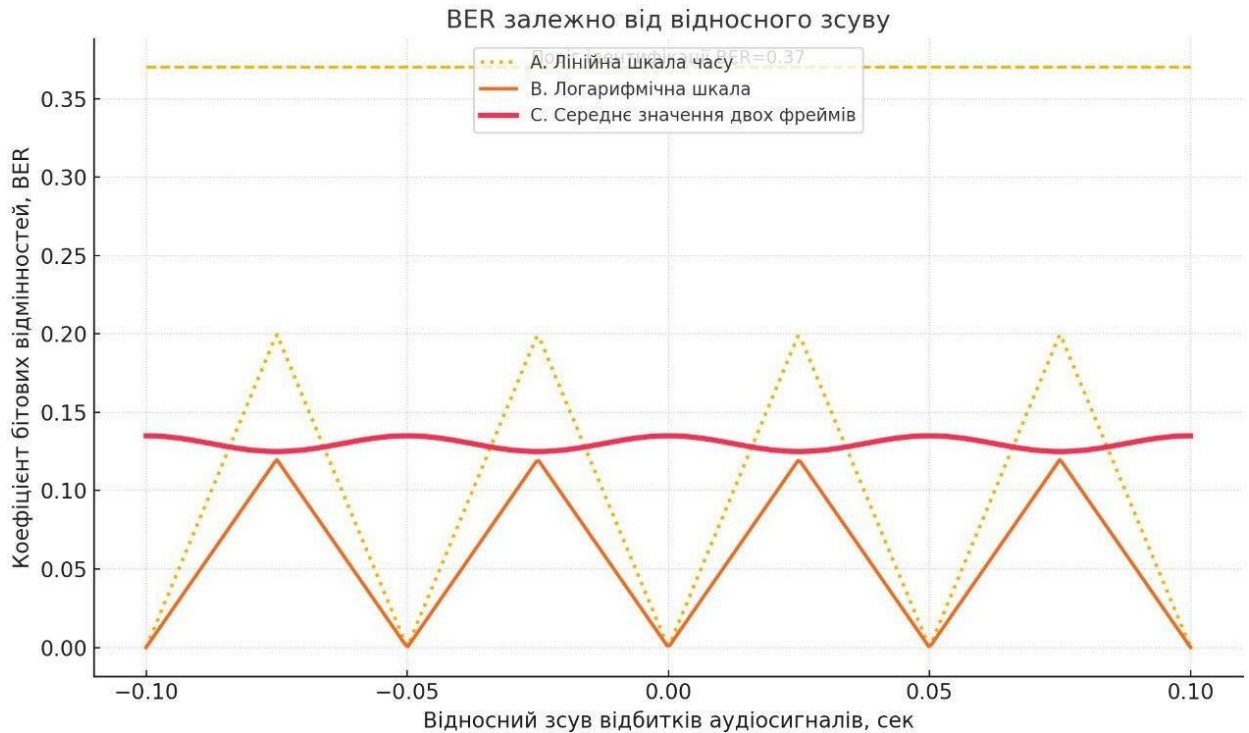


Рисунок 3.12 – Залежність коефіцієнта бітових відмінностей від відносного зсуву цифрових відбитків в часі

Для зниження впливу зсувів застосовується метод обчислення коефіцієнта бітових розбіжностей, усередненого за двома сусідніми елементами відбитку. При цьому відхилення коефіцієнта не перевищують 5,1%, що уможливорює використання цього методу для перевірки наявності сигналу та оцінки рівня спотворень аудіосигналів у каналі передачі. Ефективність методу аудіовідбитків забезпечує контроль якості передачі звукової інформації в радіотелефонних, кабельних і супутникових каналах зв'язку в реальному часі.

У певних ситуаціях виникає потреба в ідентифікації змішаних аудіосигналів, що складаються з кількох компонентів. Змішання може відбуватися, наприклад, під час аналогової передачі через взаємний вплив двох частотно близьких каналів або при студійному міксуванні (наприклад, дикторський голос на фоні музики).

Особливою складністю є розпізнавання монофонічних записів стерео- чи багатоканальних творів, відео або фільмів. При створенні монофонічного цифрового відбитку сигнали різних каналів сумуються з урахуванням їхніх ваг. Однак у монофонічному записі багатоканальних сигналів, наприклад, у форматах DTS (для кінотеатрів) або Dolby (для побутової техніки), переважає канал, на який спрямована відеокамера. Через вузьку спрямованість мікрофона камери зазвичай домінує центральний канал із діалогами, тоді як внесок бічних каналів значно менший.

На рисунку 3.13 показано залежність коефіцієнта бітових розбіжностей між двома сигналами та їхньою сумішю від співвідношення амплітуд компонентів. При порозі $BER = 0,37$ надійна ідентифікація сигналів у суміші забезпечується за співвідношення рівнів від -6 дБ до +6 дБ.

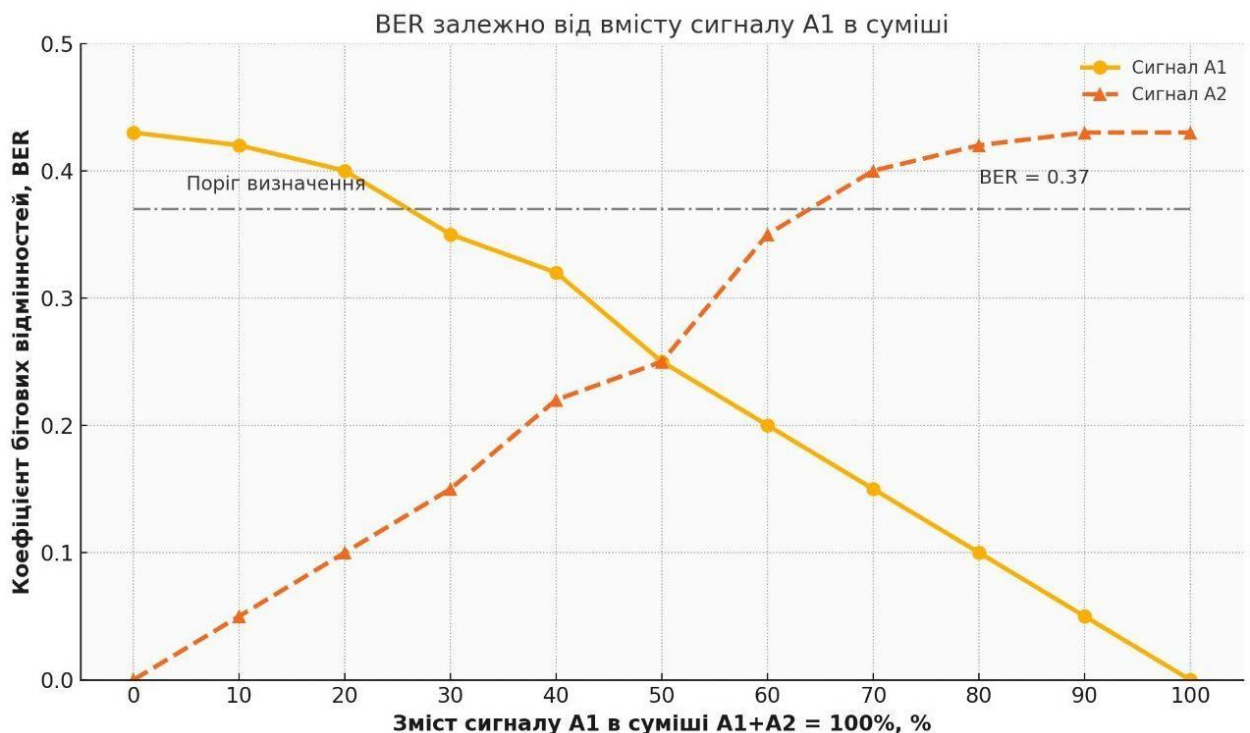


Рисунок 3.13 – Ідентифікація компонентів в двох змішаних аудіосигналах

Отже, розпізнавання змішаних сигналів можливе навіть за наявності перешкод, рівень яких удвічі перевищує рівень основного сигналу.

3.5 Висновки до розділу 3

Експерименти підтвердили, що застосування метамерів для створення зображень спричиняє порушення колірного балансу в копіях, що уможливорює використання цього методу для захисту аналогових зображень від несанкціонованого копіювання. Рівень колірних спотворень може сягати CIE Lab 22 для тілесних відтінків і 45 для нейтральних сірих кольорів.

Дослідження показали, що цифрові відбитки, створені на основі аналізу спектра модуляції в логарифмічній шкалі, стійкі до часового масштабування сигналу в межах 40%. Це значно зменшує вплив змін швидкості відтворення на точність розпізнавання сигналів із тимчасовими спотвореннями, зокрема під час конвертації фільмів, аналогових записів і мультимедійних файлів між різними форматами.

Доведено, що запропонована система пошуку та ідентифікації аудіоматеріалів ефективна для розпізнавання аудіосигналів, що зазнали різних спотворень і перетворень (зміна частоти дискретизації, фільтрація, цифрове чи динамічне стиснення, зміна швидкості відтворення тощо). Система може застосовуватися для моніторингу файлообмінних P2P-мереж, контролю поширення контенту в Інтернеті, перевірки радіо- та телепередач, оцінки якості звуку, виявлення несанкціонованих змін і підтвердження автентичності аудіоматеріалів.

ВИСНОВКИ

У роботі удосконалено метод захисту аудіоінформації від несанкціонованого копіювання та пошуку, ідентифікації, контролю за поширенням матеріалів на основі методу цифрових аудіовідбитків. Основні результати роботи сформульовані таким чином.

Проведений аналіз існуючих систем захисту аудіоматеріалів виявив такі недоліки:

- системи DRM не забезпечують надійного захисту матеріалів;
- системи DRM призначені для захисту матеріалів в цифровому форматі і не перешкоджають отриманню аналогових копій;
- створення єдиної копії в обхід DRM робить систему захисту неефективною, а поширення несанкціонованих копій неконтрольованим;
- відсутня реалізація ефективних методів контролю за поширенням інформації в мережі Інтернет;
- існуючі системи ідентифікації і пошуку не забезпечують точність, продуктивність і масштабованість, необхідну для пошуку аудіоматеріалів в мережі Інтернет.

Для захисту акустичних сигналів від несанкціонованого копіювання запропонований метод, заснований на модифікації аудіосигналу з метою зміни його характеристик, непомітних для людини, але які викликають помітні спотворення в процесі цифрового стиснення. Застосування даного методу призводить до зменшення ефективності алгоритмів стиснення на 9-29%.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Fink M., Covell M., Baluja S. Social- and Interactive-Television Applications Based on Real-Time Ambient-Audio Identification // Proc. EuroITV'06 Conf. – Israel: Center for Neural Computation, Hebrew University of Jerusalem, 2006.
2. Palfrey J., Bambauer D., Gasser U., Slater D., Smith M. Content and Control: Assessing the Impact of Policy Choices on Potential Online Business Models in the Music and Film Industries [Електронний ресурс]. – Режим доступу: http://cyber.law.harvard.edu/media/files/content_control.pdf.
3. Liu Y., Lu L., Jin J., Sun L., Fanelli A. XAttnMark: Learning Robust Audio Watermarking with Cross-Attention [Електронний ресурс] // arXiv preprint arXiv:2502.04230, 2025. – Режим доступу: <https://arxiv.org/abs/2502.04230>.
4. Cano P., Gómez E., Batlle E., Gomes L., Bonnet M. Audio Fingerprinting: Concepts and Applications // Proc. 2002 Int. Conf. on Fuzzy Systems Knowledge Discovery. – Singapore, 2002.
5. YouTube [Електронний ресурс]. – Режим доступу: <http://www.youtube.com>.
6. Philips. Audio Fingerprinting technology [Електронний ресурс]. – Режим доступу: <http://www.research.philips.com/initiatives/contentid/audiofp.html>.
7. Microsoft. Communication, Collaboration, and Signal Processing [Електронний ресурс]. – Режим доступу: <http://research.microsoft.com/research/ccsp/>.
8. Google [Електронний ресурс]. – Режим доступу: <http://www.google.com>.
9. Secure Digital Music Initiative [Електронний ресурс]. – Режим доступу: <http://www.sdmi.org>.

10. Battle E., Masip J., Guaus E., Cano P. Scalability issues in an HMM-based audio fingerprinting // Int. Conf. on Multimedia Computing and Systems. – 2004. – Vol. 1. – P. 735–738.
11. O'Reilly P., Jin Z., Su J., Pardo B. Deep Audio Watermarks are Shallow: Limitations of Post-Hoc Watermarking Techniques for Speech [Электронный ресурс] // arXiv preprint arXiv:2504.10782, 2025. – Режим доступа: <https://arxiv.org/abs/2504.10782>.
12. Chang S., Lee D., Park J., Lim H., Lee K., Ko K., Han Y. Neural Audio Fingerprint for High-specific Audio Retrieval based on Contrastive Learning [Электронный ресурс] // arXiv preprint arXiv:2010.11910, 2020. – Режим доступа: <https://arxiv.org/abs/2010.11910>.
13. Liu C., Zhang J., Zhang T., Yang X., Zhang W., Yu N. Detecting Voice Cloning Attacks via Timbre Watermarking [Электронный ресурс] // arXiv preprint arXiv:2312.03410, 2023. – Режим доступа: <https://arxiv.org/abs/2312.03410>.
14. Kurth F., Ribbrock A., Clausen M. Identification of Highly Distorted Audio Material for Querying Large Scale Databases // Proc. AES 112th Int. – Munich (Germany), 2002.
15. Allamanche E., Herre J., Helmuth O., Fröba B., Kasten T. Content-Based Identification of Audio Material Using MPEG-7 Low Level Description // Proc. Int. Symposium of Music Information Retrieval. – Indiana (USA), 2002.
16. Haitsma J. Highly Robust Audio Fingerprinting System // Proc. of the 3rd Int. Symposium on Music Information Retrieval. – 2002. – P. 144–148.
17. Seo S. J. S., Haitsma J., Kalker T. Linear speed-change resilient audio fingerprinting // Proc. IEEE Benelux Workshop on Model Based Processing and Coding of Audio. – Leuven (Belgium), 2002.
18. Rhebergen K. S. A Speech Intelligibility Index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners // J. Acoust. Soc. Am. – 2005. – Vol. 117. – P. 2181–2192.

19. Marrakchi-Mezghani I., Turki-Hadj Alouane M., Jaidane-Saidane M. Robustness of audio fingerprinting systems for connected audio applications // Proc. Second Int. Symposium on Communications, Control and Signal Processing. – 2006.
20. Dittmann J., Steinebach M., Lang A., Zmudzinsky S. Advanced audio watermarking benchmarking // SPIE. – 2004. – Vol. 5306.
21. Rhodes B. Just-in-time information retrieval agents // IBM Systems Journal. – 2003. – Vol. 39, No. 4. – P. 685–704.
22. ISO/IEC JTC1/SC29/WG11 (MPEG). Multimedia content description interface – Part 4: Audio. International Standard 15938-4. – ISO/IEC, 2001.
23. Haitsma J. Speed-change resistant audio fingerprinting using auto-correlation // Conf. on Acoustics, Speech and Signal Processing. – 2003. – Vol. 4. – P. 728–731.
24. Якименко І. З., Ковтун Н. В. Захист та ідентифікація аудіо матеріалів на основі методу цифрового відбитку // Сучасні комп'ютерні інформаційні технології: матеріали VI Всеукр. школи-семінару молодих вчених і студентів ACIT 2016. – Тернопіль: ТНЕУ, 2016. – С. 178–180.
25. Kumari I. R., Babu K. Designing Highly Secured Speaker Identification with Audio Fingerprinting using MODWT and RBFNN // Int. J. of Intelligent Systems and Applications in Engineering. – 2024. – Vol. 12, No. 16s. – С. 25–30. – Режим доступу: <https://www.ijisae.org/index.php/IJISAE/article/view/4779>.
26. Altwlkany K., Delalić S., Alihodžić A., Selmanović E., Hasić D. Application of Audio Fingerprinting Techniques for Real-Time Scalable Speech Retrieval and Speech Clusterization [Електронний ресурс] // arXiv preprint arXiv:2410.21876, 2024. – Режим доступу: <https://arxiv.org/abs/2410.21876>.
27. Patil A. J., Shelke R. An effective digital audio watermarking using a deep convolutional neural network with a search location optimization algorithm

- for improvement in Robustness and Imperceptibility // High-Confidence Computing. – 2023. – Vol. 4. – Article 100153. – DOI: 10.1016/j.hcc.2023.100153.
28. Wazirali R., Ahmad R., Al-Amayreh A., Al-Madi M., Khalifeh A. A novel adaptive watermark embedding approach for enhancing security of biometric voice/speech systems // Telecommunication Systems. – 2024. – DOI: 10.1007/s11235-024-01133-6.
29. Методичні вказівки до виконання бакалаврських дипломних робіт (проектів) для студентів спеціальностей 126 – «Інформаційні системи та технології», 151 – «Автоматизація та комп'ютерно-інтегровані технології» / Уклад. Р. Н. Кветний, О. М. Бевз, О. В. Бісікало. – Вінниця : ВНТУ, 2019. – 26 с.

ДОДАТКИ

Вінницький національний технічний університет
Кафедра автоматизації та інтелектуальних інформаційних технологій

ЗАТВЕРДЖУЮ

завідувач кафедри АІТ
д.т.н., професор Олег БІСІКАЛО

(підпис)

« 23 » березня 2025 р.

ТЕХНІЧНЕ ЗАВДАННЯ

на бакалаврську кваліфікаційну роботу
«Розробка програмного забезпечення для автоматизованої системи
ідентифікації та захисту аудіоконтенту»
08-31.БКР.001.02.000 ТЗ

Керівник бакалаврської кваліфікаційної роботи
доц. Володимир ГАРМАШ

(підпис)

« 23 » травня 2025 р.

Розробив студент гр. 1АКІТ-216

Антон ГНАТЕНКО

(підпис)

« 23 » травня 2025 р.

Додаток А (обов'язковий)

Технічне завдання на бакалаврську кваліфікаційну роботу

1 Підстава для проведення робіт

Підставою для виконання бакалаврської кваліфікаційної роботи на тему: «Розробка програмного забезпечення для автоматизованої системи ідентифікації та захисту аудіоконтенту» є наказ № 97 від 20.03.2025 р.

Термін виконання робіт:

початок 03.09.2024 р.

кінець 10.06.2025 р.

2 Мета та вихідні дані для проведення робіт

Метою бакалаврської кваліфікаційної роботи є розробка програмного забезпечення для автоматизованої системи ідентифікації, захисту та пошуку аудіоконтенту на основі методу цифрового відбитка.

Вихідними даними для проведення робіт є індивідуальне завдання на бакалаврську кваліфікаційну роботу від 20.03.2025 р.

3 Етапи виконання робіт

Виконавцем всіх перерахованих в даному розділі етапів є: студент групи ІАКІТ-216 Гнатенко Антон Валерійович факультету інтелектуальних інформаційних технологій та автоматизації Вінницького національного технічного університету, а замовником є кафедра автоматизації та інтелектуальних інформаційних технологій.

№ Етапу	Зміст етапу	Строки виконання
Е1	Загальні відомості про системи захисту та ідентифікації аудіоконтенту	07.03 – 19.03
Е2	Розробка системи ідентифікації та захисту аудіоконтенту	20.04 – 26.04
Е3	Розробка програмного забезпечення	27.04 – 09.04
Е4	Аналіз результатів роботи та підготовка роботи до захисту	02.06 – 20.06

4 Призначення і галузь застосування

Розроблене програмне забезпечення призначене для автоматизованої ідентифікації та захисту аудіоконтенту в комп'ютерно-інтегрованих системах, які використовуються в медіаіндустрії, телекомунікаціях і системах безпеки. Основною метою ПЗ є забезпечення надійного розпізнавання аудіоданих та їх захисту від несанкціонованого використання, що дозволяє ефективно контролювати розповсюдження контенту в цифрових мережах, наприклад, у потокових сервісах або системах моніторингу мовлення.

Галузь застосування охоплює створення систем захисту авторських прав, моніторингу радіо- та телевізійного ефіру, а також розробку рішень для ідентифікації аудіо в реальному часі, наприклад, для автоматичного виявлення піратського контенту або аналізу звукового середовища в системах безпеки. Програмне забезпечення може інтегруватися в хмарні платформи, системи управління цифровими активами (DAM) та медіааналітичні комплекси, забезпечуючи швидку обробку великих обсягів аудіоданих із збереженням високої точності ідентифікації та захисту.

5 Технічні дані

- 5.1 тип аудіоконтенту – монофонічний та стереофонічний;
- 5.2 формат аудіоданих – WAV, MP3, AAC;
- 5.3 мінімальна частота дискретизації – 16 кГц;
- 5.4 максимальна частота дискретизації – 48 кГц;
- 5.5 точність ідентифікації аудіоконтенту – від 90% до 99% залежно від рівня шуму та якості запису.

6 Джерела розробки

- 6.1 Liu Y., Lu L., Jin J., Sun L., Fanelli A. XAttnMark: Learning Robust Audio Watermarking with Cross-Attention [Електронний ресурс] // arXiv preprint arXiv:2502.04230, 2025. – Режим доступу: <https://arxiv.org/abs/2502.04230>.
- 6.2 Chang S., Lee D., Park J., Lim H., Lee K., Ko K., Han Y. Neural Audio Fingerprint for High-specific Audio Retrieval based on Contrastive Learning [Електронний ресурс] // arXiv preprint arXiv:2010.11910, 2020. – Режим доступу: <https://arxiv.org/abs/2010.11910>.
- 6.3 Якименко І. З., Ковтун Н. В. Захист та ідентифікація аудіо матеріалів на основі методу цифрового відбитку // Сучасні комп'ютерні інформаційні технології: матеріали VI Всеукр. школи-семінару молодих вчених і студентів ACIT 2016. – Тернопіль: ТНЕУ, 2016. – С. 178–180.

6.4 Методичні вказівки до виконання бакалаврських кваліфікаційних робіт (проектів) для студентів спеціальностей: 126 «Інформаційні системи та технології», 151 «Автоматизація та комп'ютерно-інтегровані технології» [Електронний ресурс] / Уклад. К. В. Овчинников, О. В. Бісікало. – Вінниця : ВНТУ, 2022. – 50 с. URL: <https://iq.vntu.edu.ua/fm/fdb/940/bdr/bdrmetod.pdf>.

Здобувач ст. гр. 1АКІТ-216 _____ Антон ГНАТЕНКО

Додаток Б (обов'язковий)

ІЛЮСТРАТИВНА ЧАСТИНА

**РОЗРОБКА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ АВТОМАТИЗОВАНОЇ
СИСТЕМИ ІДЕНТИФІКАЦІЇ ТА ЗАХИСТУ АУДІОКОНТЕНТУ**

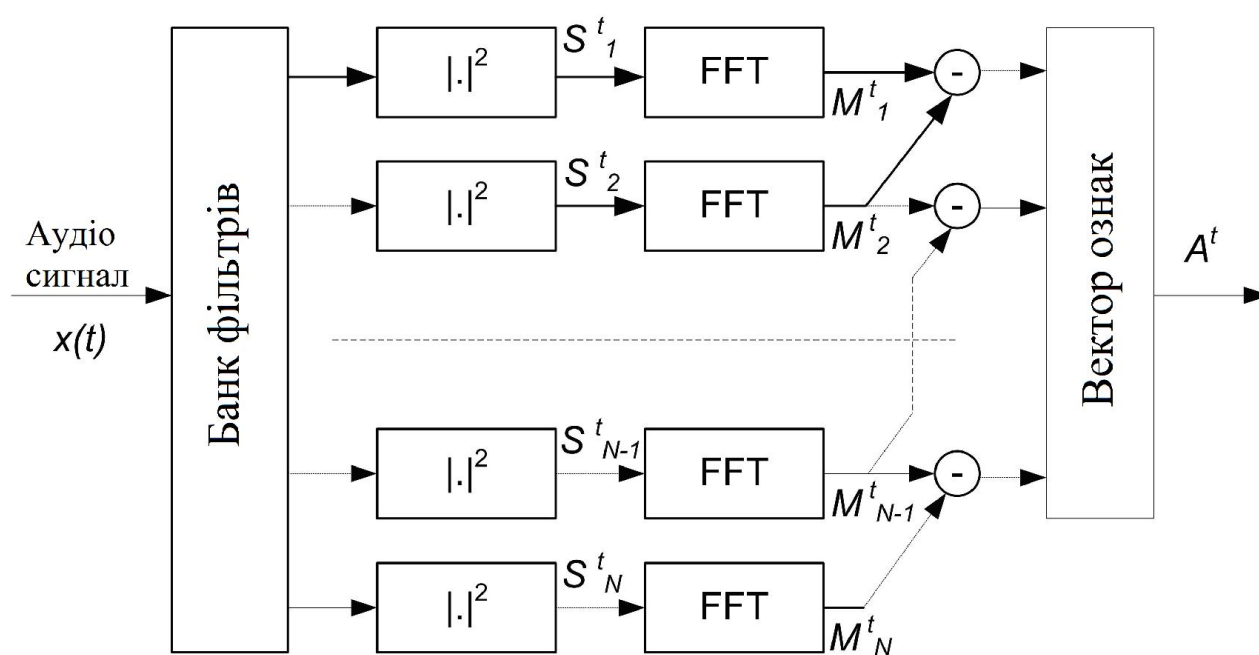


Рисунок Б.1 – Схема запропонованого алгоритму отримання вектора на підставі ознак спектра модуляції

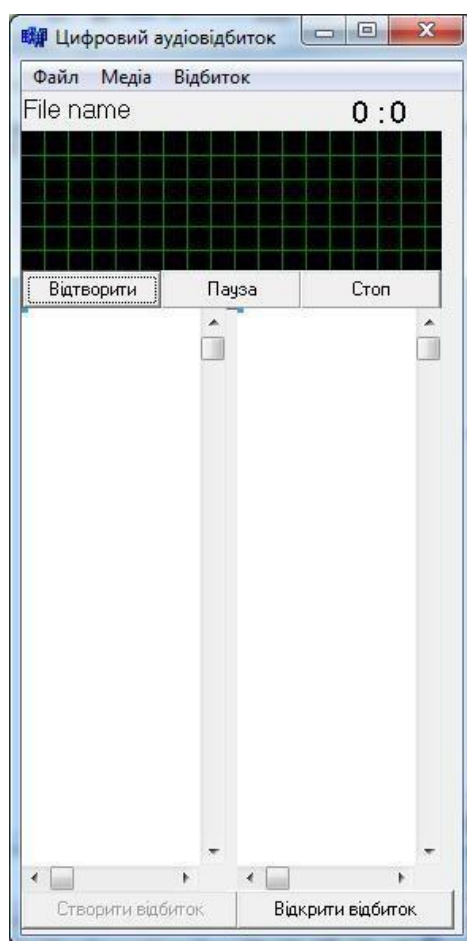


Рисунок Б.2 – Головне вікно робочої програми



Рисунок Б.3 – Блок представлення цифрових аудіовідбитків

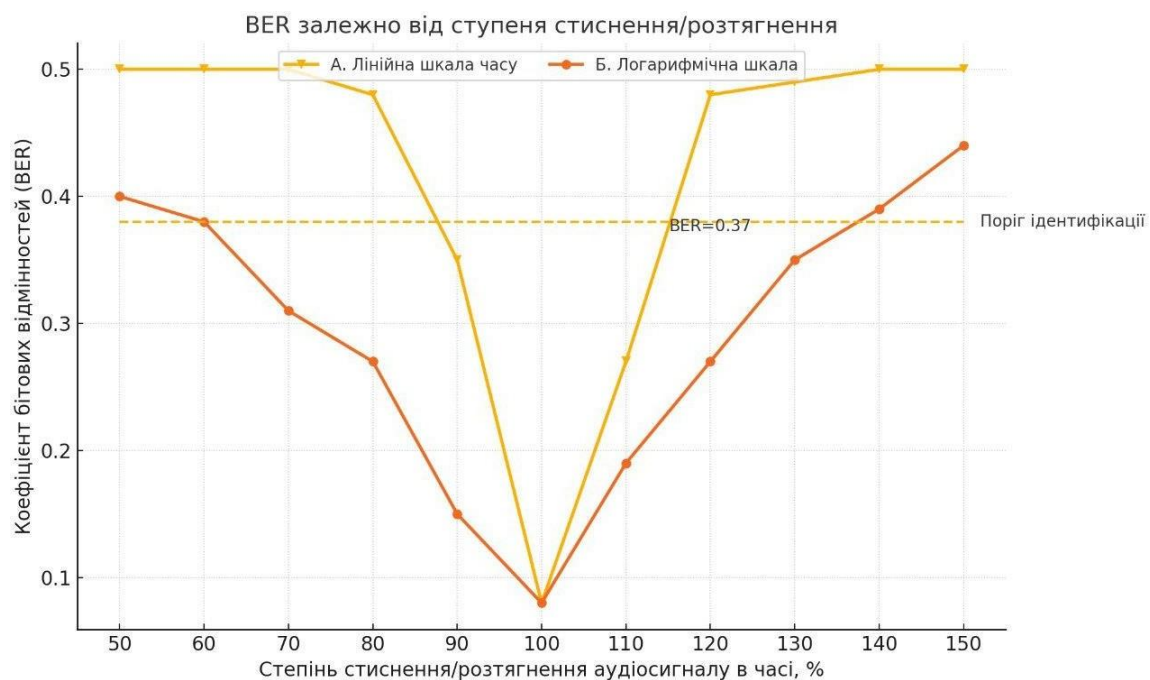


Рисунок Б.4 – Залежність коефіцієнта бітових відмінностей від зсуву при порівнянні музичного відбитка з вихідною послідовністю (А), в присутності шуму SNR = 0db (Б) і масштабування в часі на 15% (В)

Додаток В (обов'язковий) Лістинг модуля програми

```
#include <vcl.h>
#pragma hdrstop
#include "bass.h"

// Ініціалізація аудіопристрою
BASS_Init(-1, 44100, 0, 0, NULL);

// Створення потоку з файлу
HSTREAM stream = BASS_StreamCreateFile(false, path, 0, 0, BASS_SAMPLE_FLOAT);

// Запуск відтворення потоку
BASS_ChannelPlay(stream, false);

// Встановлення позиції потоку
BASS_ChannelSetPosition(stream, set, BASS_POS_BYTE);

// Фреймування аудіопотоку та отримання спектральних даних
int set = 0;
float data[256];
float buffer[128][128];
for (int str = 0; str < 128; str++) {
    set = str * 10001;
    BASS_ChannelSetPosition(stream, set, BASS_POS_BYTE);
    BASS_ChannelGetData(stream, data, BASS_DATA_FFT256);
    for (int pos = 0; pos < 128; pos++) {
        buffer[str][pos] = data[pos];
    }
}

// Функція швидкого перетворення Фур'є
void FFT(double *AVal, double *FTvl, int Nvl, int Nft) {
    // Реалізація FFT (заглушка, деталі в Додатку А)
}

// Поділ на частотні смуги
float step;
float band[128][32];
for (int i = 0; i < 128; i++) {
    for (int j = 0; j < 128; j += 4) {
        // Обробка смуг
    }
}

// Обчислення енергії сигналу
float sum[128];
for (int a = 0; a < 128; a++) {
    for (int b = 0; b < 128; b++) {
        sum[a] += pow(buffer[a][b], 2);
    }
}
```

```

// Обчислення чисельника для субвідбітків
for (int i = 0; i < 128; i++) {
    for (int j = 0; j < 128; j += 4) {
        step = pow(buffer[i][j], 2) + pow(buffer[i][j+1], 2) +
            pow(buffer[i][j+2], 2) + pow(buffer[i][j+3], 2);
        buffer[i][j] = step;
    }
}

// Нормалізація субвідбітків
for (int i = 0; i < 128; i++) {
    for (int j = 0; j < 32; j++) {
        band[i][j] = buffer[i][j*4] / sum[i];
    }
}

// Формування бітової карти
int BitMap[128][32];
float bit;
for (int m = 0; m < 127; m++) {
    for (int n = 1; n < 31; n++) {
        bit = band[m][n] - band[m+1][n] - (band[m][n-1] - band[m+1][n-1]);
        BitMap[m][n] = (bit > 0) ? 1 : 0;
    }
}

// Відображення бітової карти на DrawGrid
void __fastcall TForm1::DrawGrid1DrawCell(TObject *Sender, int ACol, int ARow, TRect &Rect,
TGridDrawState State) {
    if (BitMap[ARow][ACol] == 1) {
        DrawGrid1->Canvas->Brush->Color = clBlack;
    } else {
        DrawGrid1->Canvas->Brush->Color = clWhite;
    }
    DrawGrid1->Canvas->FillRect(Rect);
}

// Візуалізація спектрограми
void __fastcall TForm1::Timer2Timer(TObject *Sender) {
    DWORD level = BASS_ChannelGetLevel(stream);
    PerformanceGraph1->DataPoint(clBlue, level);
    PerformanceGraph1->Update();
}

// Збереження бітової карти
int iFileHandle;
if (SaveDialog1->Execute()) {
    try {
        iFileHandle = FileCreate(SaveDialog1->FileName);
        FileWrite(iFileHandle, BitMap, sizeof(BitMap));
    } catch(...) {
        Application->MessageBox("Помилка запису", "Помилка", MB_ICONSTOP + MB_OK);
    }
}

```

```

    FileClose(iFileHandle);
}

// Завантаження бітової карти
int BitMap2[128][32];
if (OpenDialog2->Execute()) {
    try {
        iFileHandle = FileOpen(OpenDialog2->FileName, fmOpenRead);
        FileSeek(iFileHandle, 0, 0);
        FileRead(iFileHandle, BitMap2, sizeof(BitMap2));
    } catch(...) {
        Application->MessageBox("Помилка відкриття", "Помилка", MB_ICONSTOP + MB_OK);
    }
    FileClose(iFileHandle);
}

// Порівняння відбитків
int a = 0;
for (int m = 0; m < 128; m++) {
    for (int n = 0; n < 32; n++) {
        if (n != 0 && n != 31) {
            if (BitMap[m][n] == BitMap2[m][n] &&
                BitMap[m-1][n-1] == BitMap2[m-1][n-1] &&
                BitMap[m+1][n+1] == BitMap2[m+1][n+1]) {
                a++;
            }
        } else {
            if (n == 0) {
                if (BitMap[m][n] == BitMap2[m][n] &&
                    BitMap[m+1][n+1] == BitMap2[m+1][n+1]) {
                    a++;
                }
            }
            if (n == 31) {
                if (BitMap[m][n] == BitMap2[m][n] &&
                    BitMap[m-1][n-1] == BitMap2[m-1][n-1]) {
                    a++;
                }
            }
        }
    }
}
float b = a * 100 / 4096;
ShowMessage("Схожість: " + FloatToStr(b) + "%.");

```

Додаток Г (обов'язковий)
ПРОТОКОЛ ПЕРЕВІРКИ КВАЛІФІКАЦІЙНОЇ РОБОТИ

Назва роботи: Розробка програмного забезпечення для автоматизованої системи ідентифікації та захисту аудіоконтенту

Тип роботи: бакалаврська кваліфікаційна робота
 (бакалаврська кваліфікаційна робота / магістерська кваліфікаційна робота)
 Підрозділ кафедра АІТ, ФІТА, ІАКІТ-216
 (кафедра, факультет, навчальна група)

Коефіцієнт подібності текстових запозичень, виявлених у роботі системою StrikePlagiarism 36,4 %

Висновок щодо перевірки кваліфікаційної роботи (відмітити потрібне)

- ☒ Запозичення, виявлені у роботі, є законними і не містять ознак плагіату, фабрикації, фальсифікації. Роботу прийняти до захисту
- ☐ У роботі не виявлено ознак плагіату, фабрикації, фальсифікації, але надмірна кількість текстових запозичень та/або наявність типових розрахунків не дозволяють прийняти рішення про оригінальність та самостійність її виконання. Роботу направити на доопрацювання.
- ☐ У роботі виявлено ознаки плагіату та/або текстових маніпуляцій як спроб укриття плагіату, фабрикації, фальсифікації, що суперечить вимогам законодавства та нормам академічної доброчесності. Робота до захисту не приймається.

Експертна комісія:

Олег БІСІКАЛО, зав. кафедри АІТ
 (прізвище, ініціали, посада)

Любов ЯНЧУК, секретар кафедри АІТ
 (прізвище, ініціали, посада)

(підпис)

(підпис)

Особа, відповідальна за перевірку

(підпис)

Костянтин ОВЧИННИКОВ
 (прізвище, ініціали)

З висновком експертної комісії ознайомлений(-на)

Керівник

(підпис)

Гармаш В.В.
 (прізвище, ініціали, посада)

Здобувач

(підпис)

Гнатенко А.В.
 (прізвище, ініціали)