

Міністерство освіти і науки України  
Тернопільський національний економічний університет  
Факультет комп'ютерних інформаційних технологій  
Кафедра комп'ютерної інженерії

Ковтун (Ільба) Наталія Василівна.

**Автоматизована система захисту та ідентифікації  
аудіоматеріалів на основі методу цифрового відбитку /  
Automated protection and identification system audio  
based on the digital print method**

**Спеціальність 8.091501 – Комп'ютерні системи та мережі**

**Дипломна робота за освітньо-кваліфікаційним рівнем «магістр»**

Науковий керівник  
к.т.н., доцент Якименко І.З.

---

Дипломну роботу допущено до захисту

«\_\_» \_\_\_\_\_ 20\_\_ р.

Зав. кафедри КІ

Березький О.М. \_\_\_\_\_

**Тернопіль – 2017**

Міністерство освіти і науки України  
Тернопільський національний економічний університет  
Факультет комп'ютерних інформаційних технологій  
Кафедра комп'ютерної інженерії

Освітньо-кваліфікаційний рівень магістр  
Спеціальність 8.05010201 комп'ютерні системи та мережі

«Затверджую»  
завідувач кафедри  
д.т.н., проф. Березький О.М.

“ \_\_\_\_\_ ” \_\_\_\_\_ 20\_\_ р.

**ЗАВДАННЯ  
НА ДИПЛОМНУ РОБОТУ СТУДЕНТА**

Ковтун Наталії Василівни

1. Тема дипломної роботи " Автоматизована система захисту та ідентифікації аудіоматеріалів на основі методу цифрового відбитку " затверджена наказом №484 від "22" жовтня 2014 р.
2. Термін здачі закінченої дипломної роботи “ \_\_\_\_\_ ” \_\_\_\_\_ 20\_\_ р.
3. Об'єкт дослідження: поширення несанкціонованих копій аудіоматеріалів в мережі Інтернет.
4. Предмет дослідження: методи захисту аудіоматеріалів від несанкціонованого копіювання.
5. Перелік задач, які мають бути вирішені:
  - дослідити сучасний стан проблеми захисту мультимедійних матеріалів;
  - удосконалити метод отримання і порівняння цифрових відбитків для максимального широкого спектру різних типів аудіосигналів;
  - провести тестування та оптимізацію запропонованих методів для досягнення надійності захисту аудіофайлів від несанкціонованого розповсюдження.

6. Перелік ілюстративного матеріалу (назва слайдів 13):

- тема;
- актуальність;
- мета і завдання роботи;
- об'єкт і предмет дослідження;
- наукова новизна, практичне значення;
- публікація та апробація;
- недоліки існуючих систем;
- схема отримання аудіовідбитків;
- енергія спектральних діапазонів;
- зміна енергії в частотних діапазонах;
- алгоритм роботи програми;
- цифровий аудіовідбиток;
- висновки.

7. Консультанти по роботі

Розділ	Консультант	Підпис
Нормо-контроль	Мельник Г.М.	

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів дипломної роботи	Строк виконання етапів роботи	Примітка
1	Стан питання та задачі дослідження		
2	Система методів ідентифікації аудіо матеріалів з допомогою цифрових відбитків		
3	Реалізація удосконаленого методу захисту аудіоматеріалів з допомогою цифрових відбитків		
4	Нормоконтроль, попередній захист	24.01.2016 – 25.01.2016	
5	Захист	01.02.2016	

Завдання прийняв до виконання \_\_\_\_\_

( підпис )

\_\_\_\_\_ (прізвище та ініціали)

Керівник дипломної роботи \_\_\_\_\_

( підпис )

\_\_\_\_\_ (прізвище та ініціали)

## РЕЗЮМЕ

Дипломна робота на тему “Автоматизована система захисту та ідентифікації аудіоматеріалів на основі методу цифрового відбитку ” на здобуття освітньо-кваліфікаційного рівня “Магістр” зі спеціальності “Комп’ютерні системи та мережі” написана обсягом 94 сторінки і містить 28 ілюстрацій, 2 таблиці, 5 додатки та 52 джерел за переліком посилань.

Метою роботи є створення автоматизованої системи захисту, ідентифікація і пошук аудіоматеріалів на основі методу цифрових відбитків.

Методи досліджень. В роботі використані методи захисту інформації, зокрема, ідентифікація аудіоматеріалів за допомогою цифрових відбитків; обробка цифрових сигналів (фільтрації, перетворення Фур’є, Меліна, Хартлі, косинусне перетворення); експертних оцінок; програмування та оптимізації.

Результати дослідження: розроблена автоматизована інформаційно-пошукова система для ідентифікації і контролю над поширенням аудіоматеріалів на основі методу цифрових відбитків, котра використовується для моніторингу матеріалів.

Результати роботи можуть бути використані для ідентифікації та пошуку широкого класу акустичних сигналів, включаючи мову, музику, структурований і неструктурований шум; ідентифікації матеріалів, що містять аудіо канал (тобто практично всі комерційні відео передачі, фільми, і музичні кліпи); моніторингу об’єктів, що володіють унікальним акустичним портретом.

Орієнтовні напрямки розвитку досліджень: розробити ефективний метод нечіткого пошуку цифрових відбитків в базі даних, що містять великий об’єм інформації.

**КЛЮЧОВІ СЛОВА:** АЛГОРИТМ, АУДІОМАТЕРІАЛИ, ЦИФРОВИЙ ВІДБИТОК, АНАЛІЗ, СИСТЕМА ЗАХИСТУ, ІДЕНТИФІКАЦІЯ АУДІОМАТЕРІАЛІВ.

## RESUME

Diploma work: “Automated systems of protection and identification of audio content on the basis of the method digital fingerprint ” for obtaining the educational qualifying level “Master” in specialty “Computer systems and networks” written by a volume of 94 pages and contains 28 illustrations, 2 tables, 5 appendices, and 52 sources in the list of references.

The aim of this work is to create an automated system for the protection, identification and search of audio content on the basis of the method of digital prints.

Research methods. We used the methods of information protection, in particular, identification of audio using digital prints; digital signal processing (filtering, Fourier transform, Melina, Hartley, cosine transform); expert assessments; programming and optimization.

The results of the study: the developed automated information retrieval system to identify and control distribution of audio content on the basis of the method of digital prints, which is used for monitoring materials.

The results can be used for identification and retrieval of a wide class of acoustic signals, including speech, music, structured and unstructured noise; identification of materials containing an audio channel (i.e. almost all commercial video transfer, movies, and music videos); monitoring of objects with unique acoustic portrait.

The approximate directions of research: to develop an effective method of fuzzy search for digital fingerprints in the database that contain a large amount of information.

**KEY WORDS: ALGORITHM, AUDIO, DIGITAL FINGERPRINT ANALYSIS SECURITY SYSTEM, IDENTIFICATION OF AUDIO MATERIAL.**

## ЗМІСТ

Вступ.....	7
1 Стан питання та задачі дослідження.....	11
1.1 Вимоги до систем захисту та ідентифікації аудіоматеріалів .....	11
1.2 Огляд компонентів комплексної системи захисту аудіоматеріалів від копіювання та розповсюдження.....	13
1.3 Основні принципи одержання та використання цифрових відбитків в системах ідентифікації аудіоматеріалів.....	20
2 Система методів ідентифікації аудіоматеріалів з допомогою цифрових відбитків .....	33
2.1 Експериментальне порівняння і вибір дескрипторів аудіоматеріалів.....	33
2.2 Дослідження стійкості дескрипторів на основі моделей сприйняття акустичних сигналів людиною .....	40
2.3 Метод цифрових аудіовідбитків .....	49
3 Реалізація удосконаленого методу захисту аудіоматеріалів з допомогою цифрових відбитків.....	63
3.1 Обґрунтування вибору середовища для програмування.....	63
3.2 Реалізація додатку для створення цифрових аудіовідбитків.....	65
3.3 Тестування розробленої програми.....	73
3.4 Властивості підсистеми пошуку та ідентифікації аудіоматеріалів на основі цифрових аудіовідбитків.....	80
Висновки.....	89
Список використаних джерел.....	90
Додаток А Алгоритм роботи програми.....	95
Додаток Б Лістинг програми.....	96
Додаток В Приклад цифрового відбитку.....	107
Додаток Г Світлокопії виданих публікацій.....	108
Додаток Д Довідка про впровадження.....	111

## ВСТУП

**Актуальність теми.** Впровадження цифрових методів стиснення і простота копіювання матеріалів в цифровому форматі призвели до широкого і неконтрольованого поширення несанкціонованих копій різноманітних ліцензованих аудіоматеріалів. Відомо, що існують методи захисту і контролю за поширення аудіовізуальної інформації, які мають вагомі недоліки і не забезпечують необхідного рівня надійності і продуктивності. Поява хоч би однієї несанкціонованої копії і її неконтрольоване розповсюдження в мережі Інтернет робить всі спроби захисту такого матеріалу даремними.

Швидкий ріст кількості нелегальних копій в першу чергу пов'язано з появою і поширенням мереж для обміну файлами (peer-to-peer networks, P2P) і сервісів для обміну аматорськими аудіовізуальними матеріалами, найбільш відомими з яких є YouTube [9]. YouTube забезпечує доступ приблизно до 40 мільйонів файлів і контролює 45% ринку аматорських фільмів в Інтернеті, серед яких зустрічаються і несанкціоновані копії телевізійних програм, музичних кліпів і фільмів, які були відзняті за допомогою мобільних приладів. Згідно інформації, наданої медіа компанією Viacom [2] (власник каналу MTV) в своєму судовому позові, в 2006 році на сайті YouTube було розміщено 160 тисяч фрагментів телевізійних програм і музичних кліпів, які переглянуті 1,5 мільярди разів, а збитки оцінені в 1 мільйон доларів США.

Згідно інформації Міжнародної Федерації Фонографічної Індустрії (International Federation of The Phonographic Industry, IFPI), в 2005 році 94% всіх нелегальних копій музичних творів розпоширювалися за допомогою мережі Інтернет [3]. Всього було поширено 20 мільярдів музичних творів (з них 12 мільярдів з допомогою peer-to-peer мереж), а збитки склали 4,5 мільярди доларів США. Згідно даних Асоціації Кінокомпаній Америки (Motion Picture Association of America, МРАА) безпосередні втрати американських кінокомпаній від несанкціонованого поширення фільмів склали 6,1 мільярда доларів, з яких 38% припадають на Інтернет [4]. При цьому 23% нелегальних копій фільмів було

зроблено мобільними пристроями під час перегляду в кінотеатрах, а проміжок між прем'єрою фільму та появленням його копії в мережі Інтернет складає 1-2 дня. Згідно з тими ж даними, прямий збиток від використання піратських копій аудіовізуальних матеріалів в Україні і Росії складають приблизно 750 мільйонів доларів, проте повна сума збитків може бути значно більша.

Для вирішення проблеми захисту мультимедійних матеріалів необхідна розробка нових методів ідентифікації і контролю за поширенням аудіовізуальної інформації. Одним з ефективних способів ідентифікації мультимедійних матеріалів є технологія, широко відома як «метод цифрових відбитків», названий аналогічно до методу ідентифікації людини по його відбитках пальців. Останнім часом технологія відбитків пальців і, зокрема, аудіовідбитків отримала широке розповсюдження для монітору P2P мереж, контролю програм радіо- і телепередач, реклами і пошуку несанкціонованих копій ліцензійних матеріалів. Технологія аудіовідбитку лежить в основі міжнародних стандартів, включаючи MPEG-7. Велике зацікавлення до даної технології проявляють численні компанії, що надають сервіс в області доставки, пошуку і захисту мультимедійних матеріалів та інтелектуальної власності, зокрема з використанням мережі Інтернет [10] [12].

Разом з цим відомо, що існуючі технології цифрових відбитків і пошуку в базі даних не спроможні забезпечити ідентифікацію навіть частини мультимедійних матеріалів, поширених в мережі Інтернет, що і визначає актуальність даної роботи.

**Мета і задачі дослідження.** Метою роботи є створення автоматизованої системи захисту, ідентифікація і пошук аудіоматеріалів на основі методу цифрових відбитків.

Для досягнення мети в роботі поставлені наступні основні задачі:

- дослідити сучасний стан проблеми захисту мультимедійних матеріалів;
- удосконалити метод отримання і порівняння цифрових відбитків для максимального широкого спектру різних типів аудіосигналів;



- провести тестування та оптимізацію запропонованих методів для досягнення надійності захисту аудіофайлів від несанкціонованого розповсюдження.

**Об'єкт дослідження:** поширення несанкціонованих копій аудіоматеріалів в мережі Інтернет.

**Предмет дослідження:** методи захисту аудіоматеріалів від несанкціонованого копіювання.

**Методи дослідження.** В роботі використані методи захисту інформації, зокрема, ідентифікація аудіоматеріалів за допомогою цифрових відбитків; обробка цифрових сигналів (фільтрації, перетворення Фур'є, Меліна, Хартлі, косинусне перетворення); експертних оцінок; програмування та оптимізації.

**Наукова новизна отриманих результатів:**

1) розроблений алгоритм моніторингу і захисту аудіоматеріалів від несанкціонованого розповсюдження, який передбачає захист сигналів з використанням цифрових відбитків;

2) удосконалений метод і математична модель цифрових аудіо відбитків, яка відрізняється від існуючих представленням спектру модуляції в нелінійній шкалі часу, метод стійких до шуму, масштабування в часі, частотних і нелінійних спотворень, цифрового стиснення і фільтрації.

**Практичне значення отриманих результатів:**

1) розроблений метод цифрових відбитків може бути використаний для:

- ідентифікації та пошуку широкого класу акустичних сигналів, включаючи мову, музику, структурований і неструктурований шум;
- ідентифікації матеріалів, що містять аудіо канал (тобто практично всі комерційні відео передачі, фільми, і музичні кліпи);
- моніторингу об'єктів, що володіють унікальним акустичним портретом.

2) розроблена автоматизована інформаційно-пошукова система для ідентифікації і контролю над поширенням аудіоматеріалів на основі методу цифрових відбитків, котра використовується для моніторингу матеріалів.

**Публікація та апробація ДР.** За результатами наукових досліджень, проведених у магістерській роботі, підготовлено тези доповіді «Захист та ідентифікація аудіо матеріалів на основі методу цифрового відбитку» обсягом 3 сторінки на Всеукраїнській школі-семінарі «Сучасні комп'ютерні інформаційні технології» (АСІТ-2016).

**Структура та об'єм магістерської роботи.** Магістерська робота складається з вступу, трьох розділів, висновку, списку використаних джерел з 52 найменувань та п'яти додатків. Робота містить 28 рисунків і 2 таблиці. Об'єм роботи складає 94 сторінки основного тексту.

## 1 СТАН ПИТАННЯ І ЗАДАЧІ ДОСЛІДЖЕННЯ

В першому розділі здійснено аналіз існуючого стану проблеми розробки систем, які призначені для захисту інформації. Розглянуті недоліки існуючих систем управління правами. Запропоновано для ідентифікації матеріалів використовувати методи цифрових відбитків і проаналізовано особливості використання цього методу в системах захисту і пошуку.

### 1.1 Вимоги до системи захисту та ідентифікації аудіоматеріалів

На сьогоднішній день, аудіо- та відеоматеріали в основному поширюються і зберігаються в цифровому форматі. Цифрове представлення даних дозволяє значно покращити якість зображення та звуку при передачі даних по існуючих каналах зв'язку. Разом з тим, цифрове представлення даних спрощує і процеси створення копій, які майже не відрізняються від оригіналу. Доступні комерційні методи захисту аудіо і відеоінформації не є серйозною перешкодою для створення несанкціонованих копій. Простота процесу створення несанкціонованих копій і наявність швидкісних каналів зв'язку призвели до того, що в Інтернеті можливо знайти майже будь-який фільм чи музикальний твір. Новий імпульс для піратства був наданий широким поширенням файлообмінних P2P мереж, які представляють широкі можливості для збереження і доставки цифрових матеріалів. Загальна кількість зареєстрованих учасників файлообмінних мереж P2P в цілому світі складає близько 100 мільйонів, а кількість музичних файлів в цих мережах складає порядку 1 мільярду [13].

Одним з перших про впровадження систем моніторингу вмісту своєї P2P мережі заявила компанія Napster [14]. Згідно рішення суду в 2001 році робота P2P мережі Napster була призупинена до тих пір, поки компанія не зможе ефективно блокувати доступ до матеріалів, які захищені авторськими правами.

Функціональність мережі була відновлена тільки після того, як Napster ліцензувала та впровадила систему акустичних відбитків TRM фірми Relatable [8]. Проте питання про реальну ефективність цієї системи залишається відкритим.

Актуальність розробки і впровадження методів автоматичного пошуку та ідентифікації аудіовізуальних матеріалів значно виросла після появи Інтернет сервісів для обміну відеофільмами, найбільш популярними з яких представлені фірмою YouTube [9], яка є дочірньою компанією фірми Google [12]. На сьогоднішній день для ідентифікації і контролю аудіовізуальних матеріалів компанією Google використовується система фільтрів, які засновані на аналізі метаданих, що надходять на сайт матеріалів, які надаються користувачами. Поряд з відносно високою імовірністю помилок схожих систем, пов'язаних з відмінностями в написанні ключових слів, не виключається можливість і навмисних спотворень метаданих. Великий об'єм інформації і відносно низький рівень контролю за вмістом мережі призвели до того, що ряд великих мультимедійних компаній почали вимагати від YouTube видалити всі несанкціоновані копії, що доступні на цьому сайті. Проте ця вимога так і не була виконана, в тому числі і через те, що відсутні ефективні технології ідентифікації копій матеріалів.

Дослідницький центр фірми Google анонсував ряд додатків для систем ідентифікації аудіосигналів [9]. Зокрема, встановлення подібних систем на комп'ютер, який володіє звуковою картою, мікрофоном і виходом в Інтернет, що дозволяє проводити моніторинг радіо- і відео каналів, що переглядаються і прослуховуються в приміщенні. В той час, як комерційна і практична цінність даної системи викликає певні сумні, відсутність адаптації такого методу до конкретної задачі моніторингу матеріалів на сайті дочірньої компанії YouTube доводить його недостатню ефективність.

Для успішного вирішення проблеми ідентифікації та пошуку цифрових копій аудіо та відеоматеріалів в Інтернеті необхідно, щоб система надавала можливість здійснювати ефективний пошук в базі даних, що містить цифрові відбитки аудіовізуальних матеріалів, загальна тривалість яких може складати мільярд секунд і більше. На даний час, не існує жодної робочої пошукової

системи, яка б була здатна забезпечити ефективний пошук аудіовізуальних матеріалів в базах даних такого розміру. Наприклад, технологія фірми Shazam, одна із лідерів на ринку ідентифікації музичних матеріалів за допомогою музичних творів [2]. Використання цієї технології для пошуку матеріалів тільки на одному сайті YouTube вимагало б приблизно 2000 комп'ютерів. При цьому для опрацювання постійно нових матеріалів, що надходять потрібне додаткове встановлення 5 комп'ютерів щоденно.

## 1.2 Огляд компонентів комплексної системи захисту аудіовізуальних матеріалів від копіювання і поширення

### 1.2.1 Цифрові системи управління правами

Цифрові системи управління правами (Digital Rights Management, DRM) в першу чергу призначені для захисту матеріалів від несанкціонованого копіювання і поширення. Системи DRM засновані на кодуванні матеріалів за допомогою різноманітних (за звичайно симетричних) методів шифрування. При цьому не враховуються специфічні параметри сигналів, що кодуються і одні і ті ж алгоритми можуть використовуватися для кодування будь-яких типів матеріалів, включаючи аудіо, відео і текст. Для повного декодування матеріалів при відтворенні необхідні спеціальні ключі або ліцензії, котрі прив'язані до серійного номеру пристрою. При цьому передача захищених матеріалів і ліцензій допускається по відкритих каналах зв'язку, проте, відтворення можливе тільки на пристроях, які мають відповідні серійні номери чи ключі. До подібних систем відносяться системи шифрування даних на оптичних носіях (Sony BMG CD Extended Copy Protection, MediaMax CD-3, CDDA, DVD CSS, HD DVD и Blue-Ray Disk AACS), системи захисту для комп'ютерів і портативних цифрових відтворювачів (Windows Media DRM, Apple FairВідтворити) та інші. Розширений список технологій і систем захисту інтелектуальної власності наведений в [17].

Законодавство підтримує використання систем DRM. Акт про Авторське Право Цифрового Тисячоліття (Digital Millennium Copyright Act, DMCA), що був прийнятий в 1998 році, забороняє розробку і розповсюдження пристроїв, які не відповідають умовам захисту, обходять системи DRM та надають можливість створювати копії захищених матеріалів. Федеральна Комісія з Комунікації (Federal Communications Commission, FCC) почала вимагати, щоб всі пристрої, здатні приймати цифрові телевізійні програми, що були випущені після липня 2006 року, розпізнавали захищені матеріали і контролювали можливість їх копіювання. Ці пристрої повинні були копіювати захищені матеріали в спотвореному вигляді і бути стійкими щодо спроб модифікації та взлому. Виробники, також, повинні отримувати дозвіл від FCC на виробництво і поширення таких пристроїв [18].

Однак, існує точка зору, що DRM забезпечує ілюзію захисту від копіювання [3]. Відомо, що всі перелічені системи були розкриті і, якщо на взлом DVD CSS потрібно було близько 3 років, то система HD DVD AACS була розкрита майже одночасно з появою її на ринку. Також відомо, що витрати на взлом будь-якої системи DRM значно менші, ніж вартість розробки самої системи. Наприклад, відключити систему CD Extended Copy Protection фірми можна простим замальовуванням сектора на оптичному диску за допомогою темного маркеру. Більш детальний огляд існуючих методів цифрового захисту і способів їх взломів виходить за межі даної роботи. На думку автора, жодна з існуючих систем DRM не виконує своїх функцій і, відповідно, не запобігає роз поширенню захищених аудіо- і відеоматеріалів. Варто також відмітити, що DRM не захищає цифрові матеріали на стадії відтворення, а також не створює перешкод для того, щоб отримати аналогові копії. Крім цього, в даний час на частину матеріалів, захищених за допомогою системи DRM, припадає не більше 10% всіх музикальних творів, що продаються, в той час, коли решту 90% постачають на ком пакт-дисках, що не мають ніяких систем захисту [3].

## 1.2.2 Прийоми і методи захисту аналогових сигналів

Існує ряд методів захисту, що викликають порушення нормального функціонування апаратури, що здійснює аудіо- і відеозаписи, і заснованих на недоліках її реалізації. Наприклад, сильні електромагнітні поля можуть викликати наведення в вхідних ланцюгах пристроїв і викликати їх перенавантаження. Апаратура для створення подібних перешкод в основному призначена для захисту від підслуховування і прихованого спостереження в невеликих приміщеннях і не розрахована на довготривалу експлуатацію.

До цього ж класу можна віднести і широко розповсюджений метод фірми Macrovision [20], що заснований на низькому захисті від перешкод і особливостях роботи системи автоматичного регулювання яскравості (АРЯ) в перших моделях відеомагнітофонів фірми JVC. Ці конструкційні недоліки в подальшому стали частиною стандарту VHS і остаточно закріплені в Digital Millennium Copyright Act. Подібні методи захисту легко долаються достатньо простою модифікацією обладнання і, незважаючи на певний комерційний успіх, не являють собою серйозної перешкоди для отримання несанкціонованих копій. Наприклад, система Macrovision припиняє роботу, якщо VHS відеомагнітофон не має системи АРЯ, якщо система АРЯ відключена чи змінена стала часу RC-ланцюжка, що визначається швидкістю реакції системи на зміну рівня яскравості сигналу.

Основна частина методу захисту відео зображень заснована на ефекті биття, що виникає у випадку, коли частота зміни параметрів зображення не співпадає з відповідними характеристиками пристрою, що здійснює запис [21][22]. Відомо, що відео зображення не є в повному обсязі аналоговим. Зміни зображення в часі проходять не плавно, а скачками, що пов'язані зі зміною окремих кадрів зображення. Варто відмітити, що ефект биття помітний для людського ока, якщо їх частота відносно невелика. При значній різниці в частоті кадрів пристроїв, що здійснюють відтворення та запис (наприклад при відеозаписі кінофільму, використання відтворюючого і записуючого обладнання різних стандартів, збільшення часу витримки) ефект биття може бути усунутий майже повністю. На сьогоднішній день не існує працездатної реалізації системи захисту відеозображень на основі ефекту биття.

### 1.2.3 Метод ідентифікації матеріалів на основі водяних знаків

Для ідентифікації аудіовізуальних матеріалів може використовуватися метод водяних знаків – спеціальних закодованих сигналів, які, будучи додані в оригінальному сигналі, практично не помітні для людини, але які в подальшому можуть бути витягнуті та декодовані. В цьому випадку різноманітні водяні знаки можуть бути додані в різні версії одного і того ж аудіо матеріалу і в подальшому бути використані для визначення джерела нелегальних копій. Хоч і позначені водяними знаками версії будуть звучати майже однаково, аналіз водяних знаків дає можливість визначити, який конкретний варіант використовувався для створення копії [23][24].

На практиці виявилось, що методи водяних знаків мають ряд недоліків і їх використання обмежене. Було встановлено, що важко, а інколи і неможливо отримати маркери, які справді не були чутні для людини при прослуховуванні високоякісного оригіналу, але і в цей же час достатньо стійкі до різноманітних спотворень, включаючи стиснення. Крім цього, цей підхід не може бути використаний для ідентифікації матеріалів, котрі були створені та випущені раніше і не містять жодних маркерів. Аналогічно, як і системи DRM, метод водяних знаків втрачають будь-яку ефективність після взламування методу чи при появі і розповсюдженню копій захищених матеріалів з інших джерел. Найпростішим методом видалення водяних знаків, відомом як Collusion Attack, є складання та усереднення декількох копій аудіо сигналів, які отриманих із різних джерел і містять різні маркери, в результаті чого ці маркери розмиваються [25].

Цей підхід, зокрема, був запропонований асоціацією компаній для області захисту ліцензійних матеріалів (Secure Digital Music Initiative, SDMI) [13]. В своєму відкритому заклику, відомому як «SDMI Challenge», асоціація звернулася до всіх бажаючих спробувати взламати розроблену систему захисту аудіо файлів на основі методу водяних знаків. В результаті система захисту була взламана декількома групами, і SDMI через кілька років зупинила своє активне існування.

Варто додати, що метод водяних знаків існує вже більше 50-ти років, і не дивлячись на існування великої кількості робіт і розроблених алгоритмів, значного поширення за цей час не отримав [4].



#### 1.2.4 Метод ідентифікації на основі цифрових відбитків

Цифровий відбиток представляє собою компактну цифрову послідовність, яка описує характерні і найбільш важливі з точки зору психології сприйняття компоненти аудіосигналу [10]. Використання методу цифрових відбитків дозволяє проводити порівняння вмісту матеріалів без їх модифікації і впровадження водяних знаків чи додаткової передачі метаданих, що описують матеріали, які передаються [29].

Основною важливою властивістю цифрових відбитків є можливість їх використання для ефективною класифікації і порівняння сигналів, виходячи з їх перцепційної подібності. Цей підхід дозволяє замінити порівняння двох великих за об'ємом і можливо значно спотворених сигналів на порівняння відповідних більш компактних і стійких до спотворення відбитків. В процесі обробки вихідний сигнал ділиться на послідовні фрагменти, для кожного з яких генерується цифровий відбиток. Отримані відбитки зазвичай зберігаються в базі даних разом з інформацією про фрагмент і можуть бути використані, як контекстний індекс для пошуку метаданих.

Використання відбитків замість вихідних сигналів має значні переваги, зокрема, висуває менші вимоги до розміру необхідної пам'яті і сховища даних і дозволяє здійснювати більш ефективний пошук в базі даних. Значною особливістю відбитків є можливість використання їх для порівняння сигналів до яких були застосовані спотворення і які містять шуми. Зокрема, одним з застосувань є визначення ступеня спотворення і цілісності матеріалів, а також для виявлення модифікацій і маніпуляцій, що були здійснені над вихідним сигналом [12].

Основним джерелом проблем при отриманні якісних цифрових відбитків є велика різниця характеристик сигналів для перцепційно схожих і навіть ідентичних з точки зору людини матеріалів. Найпростішим способом отримання відбитку медіафайла могло би бути пряме порівняння цифрових послідовностей. Ефективна реалізація цього методу полягає в виявленні хеш-функцій, наприклад MD5 (Message Digest), SHA-1 (Secure Hash Algorithm) чи CRC (Cyclic Redundancy Checking) для отримання компактного представлення файлу. Отриманий таким

чином відбиток описує формат, в якому містяться дані і не може виступати критерієм для ідентифікації вмісту цього файлу. Цей спосіб не може бути використаний для ідентифікації навіть несильно спотворених чи модифікованих сигналів.

### 1.2.5 Область використання цифрових відбитків і водяних знаків

Методи цифрових відбитків і водяних знаків мають принципові відмінності. Очевидною відмінністю, що призводить до різниці в характеристиках цих методів є необхідність, в випадку водяних знаків, впровадження інформаційного сигналу в матеріали, в той час, як аудіо відбитки не потребують модифікації аудіо матеріалів. Це призводить до наступних особливостей використання цих методів:

- Для використання водяних знаків необхідний доступ до матеріалів до початку їх розповсюдження.
- Водяні знаки дозволяють помітити та розрізнити різні копії одного і того ж матеріалу. Метод цифрових відбитків не надає такої можливості.
- Впровадження водяних знаків вносить спотворення в сигнал, що може бути помітним для людини.
- Зміни в алгоритмах водяних знаків, наприклад в процесі їх удосконалення, потребує заміни всіх раніше розповсюджених матеріалів на модифікації. Таких обмежень немає у випадку вдосконалення нових методів відбитків для впровадження яких достатньо оновити значення полів в базі даних відбитків.
- Методи водяних знаків не вимагають зберігання яких-небудь специфічних характеристик матеріалів в базах даних і здатні розпізнавати будь-які матеріали, відмічені конкретних методом. У випадку відбитків, для ідентифікації матеріалів необхідно, щоб відповідний відбиток зберігався в базі даних. При цьому кількість відбитків, що зберігаються в базі даних може складати десятки мільйонів.

Характерні особливості використання водяних знаків і цифрових відбитків призводять до того, що ці два методи гарно доповнюють один одного і можуть бути спільно використані в системах ідентифікації аудіовізуальних матеріалів.

В даний час існує велика кількість різних застосувань для методу аудіо відбитків, деякі з яких перераховані нижче:

– Ідентифікація матеріалів і пошук асоційованих метаданих. Технології цифрових відбитків дозволяють легко ідентифікувати невідомі аудіовізуальні матеріали і знайти відповідні метадані, які можуть включати права та обмеження на використання, що асоціюються з кожним окремим твором (MusicBrainz [38], Gracenote[12]).

– Використання відбитків в системах online продажів. Користувач може ідентифікувати фрагмент твору, що його цікавить за допомогою мобільного телефона чи іншого електронного пристрою та купити його цифрову копію (Shazam [10], Gracenote [12]).

– Захист інтелектуальної власності. Пошук та ідентифікація ліцензійних матеріалів здійснюється за допомогою сканування сайтів Інтернету і P2P мереж з подальшим використанням технологій цифрових відбитків (Audible Magic [9]).

– Моніторинг радіо- і телевізійних каналів. Застосування цифрових відбитків дозволяє визначити точний час, тривалість, швидкість програвання передач і рекламних оголошень (Shazam [10], Audible Magic [9]).

– Контроль цілісності та якості. Аудіовізуальні відбитки можуть використовуватися в якості цифрового підпису мультимедійних матеріалів. Порівняння цифрових відбитків дозволяє виявити різницю в характерних ознаках матеріалів, здійснювати моніторинг якості сигналу в каналі передачі, аналізувати рівень шуму та спотворень, степінь проведених змін, в тому числі навмисних фальсифікацій [10][12].

– Моніторинг обладнання. Аналіз вібрацій і акустичних шумів дозволяє встановити степінь зносу і здійснювати діагностику обладнання [32].

### 1.3 Основні принципи отримання та використання цифрових відбитків в системах ідентифікації аудіоматеріалів.

Ідеальна схема отримання аудіо відбитків повинна бути здатна точно та надійно ідентифікувати матеріали незалежно від рівня стиснення, шуму і спотворень сигналу в каналі передачі. В залежності від додатків, алгоритм повинен ідентифікувати матеріали на основі невеликих випадково обраних фрагментів тривалістю декілька секунд і володіти стійкістю до варіацій швидкості відтворення, частотної фільтрації і цифровому стисненню з використанням алгоритмів, призначених як для стиснення мови (GSM), так і музики (MPEG - 2 Layer 3, MP3).

#### 1.3.1 Вимоги до цифрових відбитків

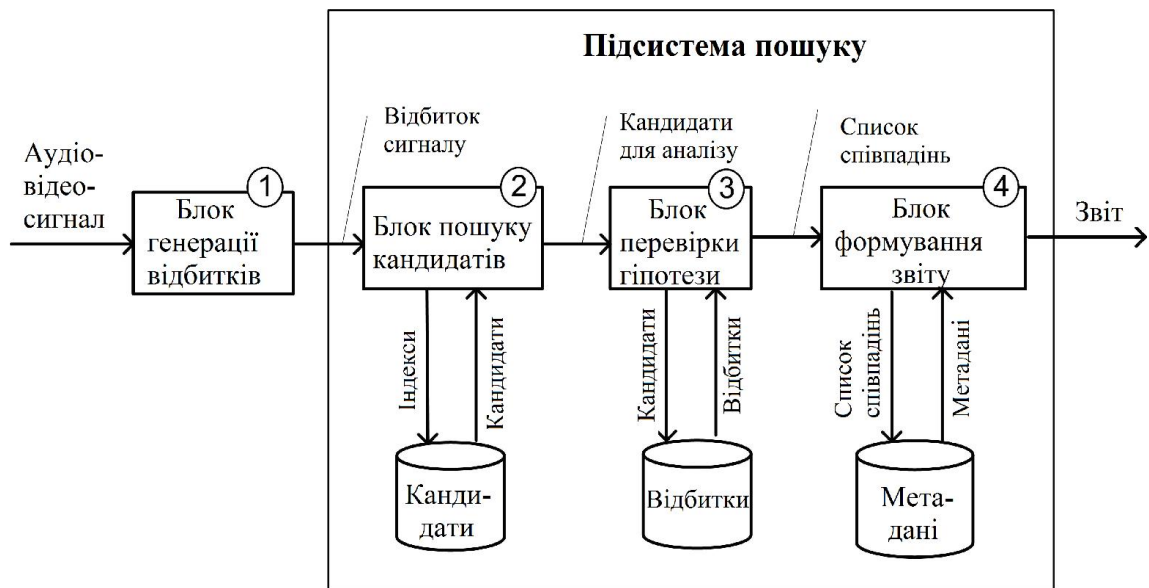
До методів для отримання цифрових відбитків висуваються наступні вимоги:

- Максимум значущою (яка сприймається людиною) інформації. Відбиток повинен бути достатньо унікальним для того, щоб відрізнитися при пошуку в базі даних, яка містить велику кількість відбитків різних матеріалів, найчастіше, які мають схожі характеристики (наприклад, музичні твори одного жанру).

- Нечутливість до різних спотворень. Відбитки, отримані при аналізі одного і того ж сигналу з різних джерел, до яких були застосовані цифрові (стиснення з втратами, фільтрація) та аналогові (запис на магнітну стрічку, мікрофон, ревербація) спотворення, а також при наявності шуму повинні бути максимально схожі.

- Компактність, надійність і простота обчислень. Отримані відбитки повинні мати відносно невеликий розмір і забезпечувати швидкий пошук в базі даних, яка містить значний об'єм інформації (мільйони записів загальною тривалістю більше мільярда секунд).

Система обробки аудіо відбитків складається з двох основних компонентів: блока генерації відбитків і підсистеми пошуку. Блок-схема запропонованої системи ідентифікації цифрових відбитків представлена на рисунку 1.1.



Рисунк 1.1 – Блок-схема системи ідентифікації матеріалів з використання цифрових відбитків

Система включає наступні блоки:

1. Блок генерації відбитку вхідного аудіовізуального сигналу;
2. Блок індексації цифрових відбитків і пошуку кандидатів;
3. Блок перевірки гіпотези і перевірки співпадінь відбитків;
4. Блок пошуку метаданих і формування відбитків.

Система ідентифікації матеріалів може бути розподіленою в цьому випадку обчислення відбитків може бути реалізовано на портативних пристроях в процесі запису сигналу в реальному масштабі часу, наприклад, на мобільний телефон. Створений відбиток, який має відносно невеликий розмір, передається від клієнта на сервер, де здійснюється пошук відповідних метаданих, які можуть бути передані клієнту назад.

При пошуку аудіовідбиток порівнюється з даними в базі даних, для чого розраховується відстань між заданим відбитком і найближчим відбитком в базі даних. Так як кількість порівнянь відбитків може бути дуже великою, а розрахунок відстані вимагає значних обчислювальних ресурсів, пошук

розбивається на декілька етапів. На першому етапі за рахунок індексації або спрощеного методу порівняння відбитків здійснюється швидкий пошук кандидатів. На наступному етапі здійснюється більш детальний аналіз обраного набору алгоритмів. Очевидною вимогою до систем пошуку та ідентифікації аудіовідбитків є:

- швидкість. Система повинна забезпечувати швидкий пошук кандидатів в базах даних, які містять мільйони записів;
- точність. Підсистема порівняння кандидатів повинна забезпечувати низький рівень помилок, які включають як не ідентифіковані матеріали (false negative rate, FNR), так і матеріали, які ідентифіковані неправильно (false positive rate, FPR);
- мінімальні вимоги до пам'яті. Більшість систем здійснюють первинний пошук кандидатів з використанням таблиць індексів, які займають весь доступний об'єм оперативної пам'яті. Об'єм пам'яті, що використовується, значно впливає на ефективність пошуку і максимальний розмір бази даних.

### 1.3.2 Методи обробки сигналів в блоці генерації аудіовідбитків

Існуючі методи отримання аудіовідбитків діляться на три категорії [10]:

- Локальні характеристики сигналу. Послідовності векторів ознак (траєкторії в евклідовому просторі) обчислюються через рівні проміжки часу. Прикладом може виступати величина спектральної рівномірності, яка використовується в стандарті MPEG-7. для обчислення відстані між елементами відбитків зазвичай використовують відстань Хеммінга. Пошук і порівняння послідовності елементів відбитків може здійснюватися за допомогою метода динамічного програмування (Dynamic Time Warping, DTW).
- Довгострокові характеристики аудіосигналу. Наприклад, середнє значення і дисперсія параметрів, отримані для всього музичного твору таких, як ритм чи частота домінантних тонів. Ці характеристики не потребують точного позиціонування при порівнянні, проте володіють низькою стійкістю до шуму і спотворення [7]. До того ж, вони не дозволяють порівнювати окремі фрагменти, а також матеріали, що передаються у вигляді неперервного потоку.

– Приховані Марківські моделі. Не дивлячись на високу ефективність і поширеність прихованих Марківських моделей в алгоритмах розпізнавання мовлення [8] і здійснених спробах їх використання для ідентифікації випадкових аудіосигналів [35][14], в даний час не існує ефективних систем цифрових відбитків, заснованих на цьому методі.

На основі аналізу приведених категорій методів генерації відбитків встановлено, що найбільш повно проблему ідентифікації аудіовізуальних матеріалів вирішують методи, засновані на аналізі локальних параметрів сигналу, що визначають структуру процесу обчислення набору ознак аудіосигналу. Схема цього процесу представлена на рисунку 1.2.

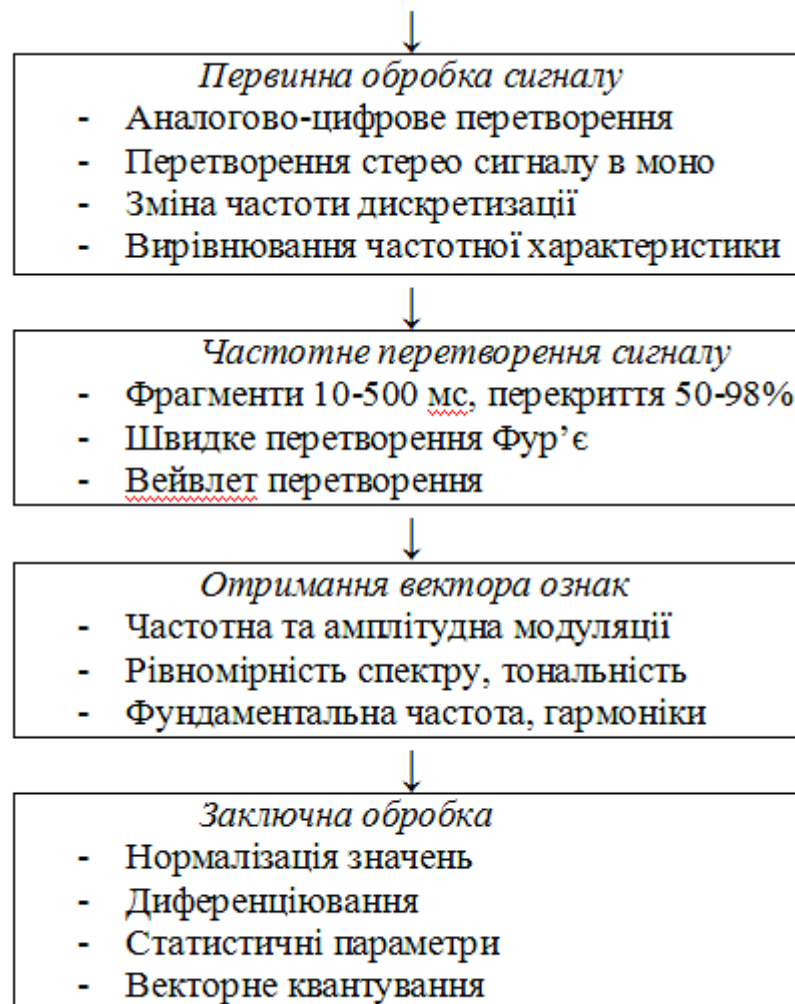


Рисунок 1.2 – Блок-схема обробки сигналу в блоці отримання цифрових аудіовідбитків

Первинна обробка. На першому етапі акустичні сигнали перетворюються в цифрове представлення і конвертуються в основний формат для подальшої обробки (зазвичай 16-бітний моно сигнал з частотою дискретизації 5-44.1 КГц). Інколи аудіо піддається специфічним перетворенням з метою моделювання або компенсації спотворень каналу передачі. Інші типи перетворень включають GSM кодування в системах телефонії, амплітудну нормалізацію і фільтрацію [37].

Частотне перетворення сигналу. Перетворення дозволяють виділити характерні параметри акустичних сигналів і зменшити надлишковість отриманих даних. Існують перетворення, що володіють оптимальними властивостями при стисненні і деко реляції такі, як перетворення Карунена-Лоева (Karhunen Loeve Transform, KLT) і метод головних компонентів (Principal Component Analysis, PCA) [68]. Ці перетворення дозволяють значно зменшити розмірність векторів ознак, але володіють істотними недоліками. Основний недолік полягає в тому, що параметри перетворень обчислюються динамічно і залежать від структури вхідного сигналу і не можуть бути відтворені, якщо сигнал був пошкоджений спотворенням і модифікаціями. Наявність більшої кількості шуму призводить до того, що корисний сигнал може бути практично втрачений. По цій причині більшість перетворень, що використовуються, мають фіксовані ортогональні базові функції і перебувають в перетворенні даних в частотний домен. Методи, засновані на аналізі змін потужності сигналу [15][16] можна розглянути як спрощений варіант частотно-часового розподілення з однією частотною смугою. Перетворення, яке найбільш широко використовується, є швидке перетворення Фур'є (Fast Fourier Transform, FFT). Також використовується дискретне конусне перетворення (Discrete Cosine Transform, DCT), перетворення Хаара (Haar Transform), модельоване комплексне перетворення (Modulated Complex Transform, MCLT).

Припускається, що сигнал  $X^t$  є стаціонарним в відносно невеликому інтервалі часу і для аналізу розбиваються 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 в частотне представлення, сигнал піддається додатковій обробці для виділення характерних акустичних параметрів сигналу [17]. Основне призначення цього етапу полягає в зменшенні розмірності вектора ознак і в цей же час збереження стійкості до спотворень. Для цього часто використовуються алгоритми, які моделюють аудиторну систему людини для виявлення перцепційно важливих параметрів, в більшості випадків за допомогою використання банку смугових фільтрів, що моделюють критичні смуги аудиторної системи людини. Широко використовуються наступні 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 :

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

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

$$SFM_{i,n} = \frac{\left[ \prod_{m=a_n}^{b_n} |f_i(m)|^2 \right]^{\frac{1}{b_n - a_n + 1}}}{\frac{1}{b_n - a_n + 1} \sum_{m=a_n}^{b_n} |f_i(m)|^2} \quad (1.5)$$

– 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 (Mel Frequency Cepstrum Coefficients, MFCC) [18]. Для отримання MFCC, коефіцієнти амплітудно-частотного спектру Фур'є, згладжуються і розбиваються на частотні діапазони, розташовані відповідно, так званої, Mel-шкали. Mel-шкала була розроблена для моделювання можливостей людського вуха розпізнавати різні тональності і є лінійною для частот менше 1000 Гц і логарифмічною для більш високих частот. На наступному

етапі здійснюється обчислення алгоритму амплітуд сигналу в кожному діапазоні з наступним частотним перетворенням отриманих значень. В якості частотного перетворення зазвичай використовують дискретне косинусне перетворення (DCT), застосування якого дозволяє сконцентрувати спектральні дані в області коефіцієнтів з низькими частотами.

Для моделювання музикальних характеристик також використовуються такі параметри, як фундаментальна частота, список гармонік, гучність і ритм, а також методи класифікації типів аудіо сигналів [19].

Заключна обробка і кодування. Перелічені вище характеристики акустичних сигналів є абсолютними величинами. Для більш повного і стійкого представлення сигналів, що змінюються в часі (амплітудна модуляція) і по частоті (частотна модуляція) використовують похідні відповідних параметрів. Наприклад, в роботі [44] автори для отримання векторів ознак використання MFCC коефіцієнти і їх похідні. Застосування похідних параметрів (швидкість і прискорення) часто призводить до посилення шуму, але разом з тим мінімізує спотворення в каналі передачі. Для цієї ж цілі може використовуватися нормалізація спектру модуляції [45] і кепстральних коефіцієнтів (Cepstrum Mean Normalization, CMN) [20], які дозволяють зменшувати вплив лінійних спотворень, які повільно змінюються в часі каналів передачі. На останньому етапі використовується квантування значень векторів ознак з низьким розширенням, що дозволяє ще більше збільшити стійкість до шуму та мінімізувати розмір отриманих відбитків.

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

### 1.3.3 Методи порівняння та індексації відбитків в підсистемі пошуку

Методи розрахунку відстані між елементами цифрових відбитків залежать від типу використаної моделі представленого вектора ознак. Найбільш часто використовуваною мірою схожості окремих елементів відбитків є відстань в

багатовимірному Евклідовому просторі і відстані Хеммінга, засновані на кількості бітових відмінностей (Bit Error Rate, BER) [10]. Для обчислення відстані між послідовністю елементів використовується відстань Левенштейна. Відстань Левенштейна [47] між двома послідовністю символів визначається, як кількість операцій, необхідних для конвертування однієї з них в іншу. Допустимими операціями є вставка, видалення і заміна символу в послідовності. В залежності від конкретних додатків, різними операціями можуть бути присвоєні різні вагові коефіцієнти. У випадку, якщо єдиною допустимою операцією є тільки заміна одного символу на інший, то в цьому випадку ця відстань ідентична відстані Хеммінга. У випадку, якщо допустимими операціями є вставка і видалення з ваговим коефіцієнтом 1, то така відстань задає найдовшу загальну послідовність (Longest Common Subsequence, LCS).

Для розрахунку відстані Левенштейна  $D_{ik}$  між послідовністю  $q$  довжиною  $N$  символів і послідовністю  $p$ , довжиною  $M$  символів для випадку, коли всі операції редагуються (видалення, вставка, заміна) мають вагу рівну 1, використовуються наступні рівняння:

$$\begin{aligned}
 D_{i,0} &= i \quad \forall 0 \leq i \leq N \\
 D_{0,k} &= k \quad \forall 0 \leq k \leq M \\
 D_{i,k} &= \begin{cases} D_{i-1,k-1}, & q_i = p_k \\ \min[D_{i-1,k-1}, D_{i,k-1}, D_{i-1,k}], & q_i \neq p_k \end{cases}
 \end{aligned} \tag{1.8}$$

Класичним методом обчислення відстані Левенштейна є метод динамічного програмування [21][22].

Фундаментальною проблемою при використанні цифрових відбитків є розробка ефективних методів індексації і пошуку в базі даних, яка містить мільйони відбитків. Представлення вектора ознак в багатовимірному просторі дозволяє використовувати ефективні просторові методи індексації відбитків [50], які дозволяють мінімізувати кількість операцій прямого порівняння послідовності ознак. Цей метод індексації для нечіткого пошуку дозволяє виділити невелику

групу кандидатів близько розташованих до елементів послідовності, яку шукають, в багатовимірному просторі векторів ознак. Більшість методів індексації засновані на припущенні, що з високим ступенем ймовірності серед великої кількості індексів, які використовуються, знайдеться хоча б один, який не містить помилок і має однакові значення, як для вектора ознак, для якого здійснюється пошук, так і для послідовностей, що зберігаються в базі даних [23].

На наступному етапі здійснюється пряме порівняння і відсіюються кандидати, для яких сумарна відстань між елементами перевищує допустиму встановлену границю. Для порівняння відбитків також може використовуватися метод Вітербо, методи, що використовуються в біології для аналізу ДНК [24], а також нейронні мережі [26]. Індексція елементів заснована на ключових послідовностях символів, при цьому індекси елементів можуть містити інформацію про розташування відповідного елемента в послідовності, що дозволяє поєднати ділянки відбитків, що порівнюються, і здійснювати порівняння за лінійний час [16][25].

Це призводить до необхідності для порівняння окремих елементів відбитків використовувати відстань Хеммінга.

#### 1.3.4 Стандарт MPEG-7

Викладені вище принципи використовуються в більшості існуючих методів цифрових відбитків. Для забезпечення сумісності різних методів було здійснено ряд спроб їх стандартизації. Одним з таких стандартів був запропонований групою MPEG (Moving Pictures Experts Group, ISO/IEC JTC1/SC29/WG11), яка спеціалізується на розробці і стандартизації мультимедійних технологій. Поряд з розробкою таких відомих стандартів стиснення і кодування аудіо - і відеоматеріалів як MPEG-2 і MPEG-4, був запропонований стандарт MPEG-7, призначений для опису характеристик мультимедійних матеріалів [57]. Стандарт MPEG-7 надає безліч методів для опису мультимедійних даних і використовує XML в якості мови опису вмісту мультимедійних матеріалів. Головними елементами стандарту MPEG-7 є [57] [58]:

- Дескриптори (D) – визначають синтаксис і семантику подання кожної з характеристик.

- Схеми опису DS (Description Scheme), які специфікують структуру і семантику взаємодії між компонентами. Ці компоненти в свою чергу можуть бути дескрипторами і схемами опису.

- Мова опису визначень DDL (Description Definition Language), що дозволяє створювати нові схеми опису і забезпечує розширення і модифікацію існуючих схем.

Розділ стандарту, призначений для опису аудіосигналів містить поряд з метаданими також специфічні параметри, які безпосередньо характеризують акустичний сигнал [59]. Стандарт MPEG-7 містить опис структур, які можуть бути виділені в два відмінних класу параметрів – основні характеристики акустичних сигналів і методи, специфічні для конкретних додатків. Останні містять інструменти для виконання наступних функцій:

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

Стандарт MPEG-7 надає різні структурні елементи для побудови аудіовідбитків. Один з них, AudioSpectrumFlatness Descriptor, містить набір низькорівневих описувачів сигналу (Low Level is invalid, LLD) і є універсальним для даного типу додатків. Конкретні характеристики для низькорівневих описувачів задаються за допомогою схеми, яка описує, зокрема, часові і частотні параметри відбитка. Згідно стандарту, розмір вікна для спектральних обчислень становить 30 мс а діапазон частот від 240 Гц до 4 кГц розбитий на 16 діапазонів шириною 1/4 октави, для кожного з яких обчислюється середнє значення і дисперсія спектральних характеристик. В якості основного дескриптор використовується спектральна рівномірність SFM (вираз 1.5).

Обчислювальна ефективність алгоритму досить висока і складає більш 100х (тобто в 100 разів швидше, ніж швидкість програвання). Згідно з даними розробників [27] система стійка до типових спотворень сигналу і здатна ідентифікувати довільні фрагменти незалежно від їх зміщення в тестовому сигналі. Для сигналів, що піддаються невеликим викривленням, точність розпізнавання зазвичай становить приблизно 98-99% [27].

Незважаючи на велику кількість публікацій, присвячених технології MPEG-7, на даний момент не існує ефективної реалізації даного методу, що забезпечує надійну ідентифікацію і прийнятну швидкість пошуку, необхідні для його практичного використання. Високий рівень помилкових ідентифікацій в реальних додатках, низька швидкість пошуку і обмеження на розмір бази даних роблять цей метод недостатньо ефективним для ідентифікації і пошуку аудіоматеріалів в Інтернеті. В той час, як стандарт MPEG-7 створювався як універсальна мова опису параметрів аудіовізуальних сигналів, застосування цього стандарту обмежене з тієї причини, що недостатньо гнучкості для конкретних задач ідентифікації матеріалів.

Висновки до розділу 1.

1. Визначено, що існуючі системи захисту аудіовізуальної інформації орієнтовані на запобігання копіювання матеріалів в цифровому форматі і не дозволяють захищати матеріали після їх перетворення в аналоговий формат в процесі перегляду користувачем. Користувач має широкі можливості для отримання аналогових копій зображення і звуку, подальшого стиснення і розповсюдження їх в мережі Інтернет.

2. Запропоновано для контролю над розповсюдженням матеріалів використовувати пошук та ідентифікацію несанкціонованих копій за допомогою цифрових відбитків.

3. Встановлено, що цифрові відбитки дозволяють знаходити схожі аудіовізуальні матеріали за відносно невеликих взірців, при чому надійна ідентифікація можлива незалежно від формату представлення даних, рівня стиснення, а також в присутності шуму і різних типів спотворень.

4. Проведений аналіз літературних даних показав, що всі існуючі методи аудіо відбитків у випадку використання в системах ідентифікації аудіо матеріалів мають достатньо високий сумарний рівень похибок.

Таким чином, є всі підстави вважати, що на сьогоднішній день не існує систем ідентифікації, які були би в значній мірі стійкими до всіх типів спотворень та шуму, а тим більше до їх комбінацій. Для забезпечення необхідної ефективності і точності систем пошуку та ідентифікації розроблені нові методи цифрових відбитків, які представленні в наступному розділі.



## 2 СИСТЕМА МЕТОДІВ ІДЕНТИФІКАЦІЇ АУДІОМАТЕРІАЛІВ ЗА ДОПОМОГОЮ ЦИФРОВИХ ВІДБИТКІВ

Основною проблемою при реалізації систем пошуку та ідентифікації аудіовізуальних матеріалів є великий обсяг інформації, яку необхідно обробити, а також недостатня продуктивність і точність існуючих технологій [28][64]. В даному розділі експериментально досліджено стійкість різних типів дескрипторів акустичних сигналів до спотворень з метою вибору оптимальних параметрів, які забезпечують точне і швидке порівняння цифрових відбитків. Представлені нові методи пошуку та ідентифікації матеріалів з використанням цифрових аудіо - і відеовідбитків на основі аналізу локальних ознак сигналу. Запропоновано спосіб забезпечення сумісності потоків аудіо - і відеовідбитків для їх комбінованого використання в системі ідентифікації аудіовізуальних матеріалів.

### 2.1 Експериментальне порівняння і вибір дескрипторів аудіоматеріалів

В даному розділі представлені отримані експериментальні результати і підсумки тестування розглянутих раніше дескрипторів (1.1-1.7) з метою вибору найбільш ефективного способу подання вектора ознак. Були реалізовані і експериментально оцінені властивості таких методів:

1. Muscle Fish. Одна з перших систем ідентифікації аудіоматеріалів на підставі їх вмісту [65]. Ця система виділяє основні перцепційні параметри звуку такі, як гучність, тональність, тембр, смуга частот, ритм, темп і аналізує, як вони змінюються в часі. Для кожного з цих параметрів обчислюється середнє значення, дисперсія та автокореляція. При порівнянні музичних фрагментів у якості критерію подібності використовується відстань між отриманими векторами в евклідовому просторі [66]. Надалі, на цій основі була створена технологія, яка використовується компанією Audible Magic [9]. Додатково до перерахованих

вище параметрів сигналу в цій технології використовуються значення Мел-частотних кепструм коефіцієнтів (MFCC).

2. AudioDNA. Процедура отримання АудіоДНК в основному полягає на аналізі спектральних характеристик сигналу та використанні MFCC [24]. Схема отримання акустичного вектора ознак AudioDNA представлена на рисунку 2.1.

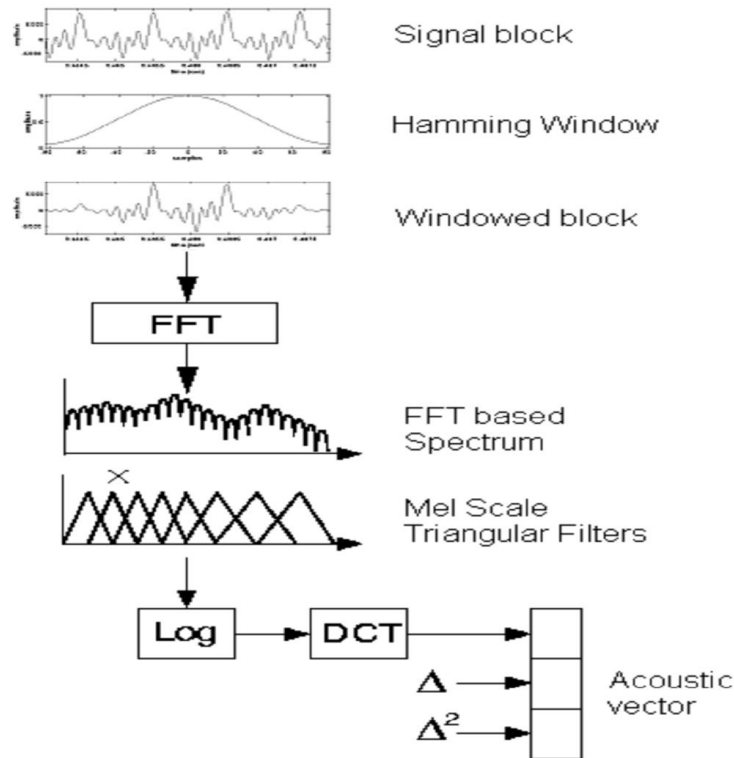


Рисунок 2.1 – Схема обчислення вектора ознак AudioDNA

Кожен індивідуальний вектор ознак сигналу ("Аудіо Ген") [29] представлений у вигляді прихованої Марківської моделі (Hidden Markov Model, HMM) і на остаточному етапі порівняння найбільш ймовірна послідовність АудіоГенів визначається за допомогою алгоритму Вітербо [8].

3. RARE (Robust Audio Recognition Engine). Метод розроблений в лабораторії фірми Microsoft [12][68]; його схема представлена на рисунку 2.2. Для частотно-часового подання даних у методі використовується модульоване комплексне перетворення (MCLT).

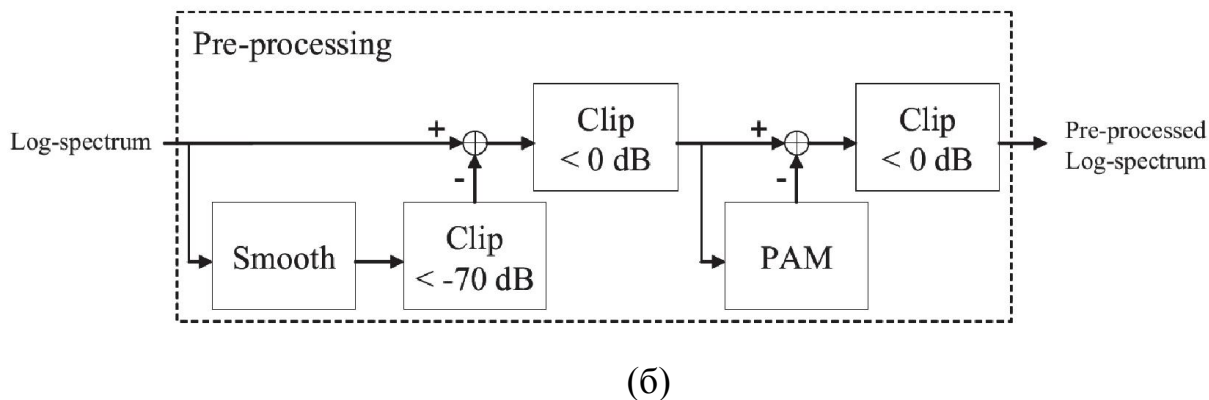
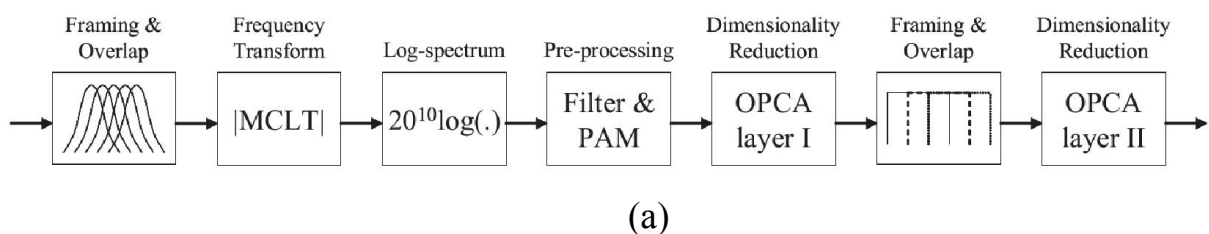


Рисунок 2.2 – Метод RARE фірми Microsoft генерації відбитків (а) і попередньої обробки аудіо сигналів (б)

Попередня обробка спектральних даних дозволяє виключити частотні спотворення і полягає в обчисленні різниці між MCLT коефіцієнтами і згладженою версією спектру, з подальшим застосуванням психоакустичної моделі (Psycho Acoustic Model, PAM) для видалення компонентів сигналу, які людина не здатна почути. Для зменшення розмірності аудіовідбитків, використовується модифікація методу аналізу основних компонентів (Oriented Principle Component Analysis, OPCA), що застосовується до сигналу послідовно в два етапи. На першому етапі OPCA застосовується до амплітудно-частотного представлення сигналу в вікнах, що накладаються тривалістю 372 мс і для тренування OPCA використовується матриця кореляції між спектром сигналу і його різними спотвореними варіантами. На другому етапі результати 32 перетворень першого рівня, що охоплюють фрагмент аудіосигналу тривалістю 6.1 секунди додатково стискаються, в результаті чого виходить 64-мірний вектор ознак сигналу.

4. Philips. Одним з методів, який не вимагає великого обсягу обчислень, є метод запропонований J. A. Naitsma (Philips) [30]. Відбиток для аудіофрагмента довжиною 3 секунди складається з 256 послідовних 32-бітних суб-відбитків,

виконаних з інтервалом в 11.7 мілісекунд. Схема алгоритму представлена на рисунку 2.3.

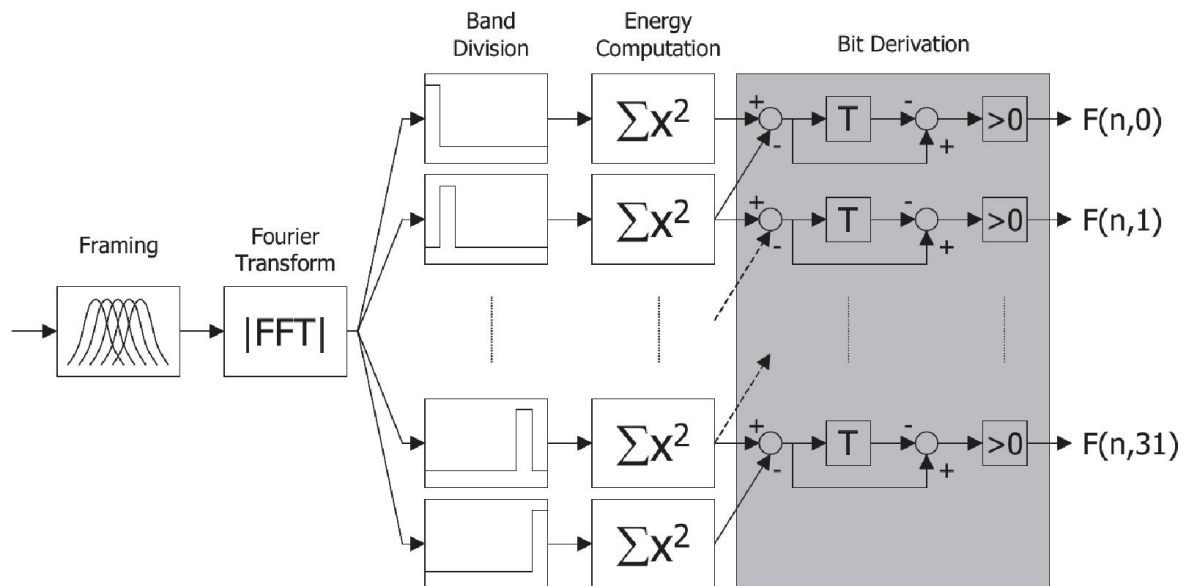


Рисунок 2.3 – Схема отримання аудіовідбитків згідно методу J.A. Naitsma (Philips)

Для обчислення суб-відбитків сигнал розбивається на фрагменти, що накладаються довжиною 0,37 секунд і фактором перекриття 31/32 і перетворюються в частотно-часове розширення (спектрограму), використовуючи перетворення Фур'є (FFT). Діапазон частот від 300 Гц до 2000 Гц розбивається на 33 смуги, які не накладаються одна на одну, з логарифмічним розподілом. Знак зміни енергії в кожному частотному діапазоні одночасно в часі і по частоті (формула 1.6) є унікальною і стійкою до спотворень характеристикою аудіосигналу.

Для аналізу схожості двох відбитків використовується критерій, заснований на кількості бітових відмінностей (Bit Error Rate, BER). Відбитки вважаються однаковими, якщо кількість бітових відмінностей лежить нижче межі, яка визначає співвідношення правильно і неправильно ідентифікованих відбитків. У ході експериментів поріг коефіцієнта бітових спотворень BER був обраний 0.35.

Слід зазначити, що даний метод досить широко використовується в різних системах ідентифікації аудіоматеріалів і в силу своєї простоти та відтворюваності

результатів де-факто є стандартом при порівнянні ефективності різних методів аудіовідбитків.

5. Philips (modified). Для компенсації ефектів, пов'язаних з масштабуванням частотно-амплітудного спектру сигналу при стисненні/розтягуванні сигналу в часі, була запропонована модифікація описаного вище методу, заснована на отриманні відбитків у масштабно-інваріантному просторі, використовуючи перетворення Фур'є-Мелліна (Fourier-Mellin Transform, FMT)[31]. Змінений алгоритм включає додатковий частотний логарифмічний масштаб спектра потужності з подальшим перетворенням Фур'є. З наведених даних випливає, що модифікований алгоритм приблизно в 5 разів менш чутливий до зміни швидкості програвання аудіоматеріалів, ніж вихідний метод. Аналогічні результати були отримані при заміні перетворення Фур'є на функцію автокореляції [71].

6. MusicDNA. У методі MusicDNA, розробленому фірмою Cantamatrix і згодом придбаному фірмою Gracenote [12], використовуються середні значення та варіації енергії сигналу в 15 частотних барк діапазонах 12 послідовних фрагментах сигналу, які накладаються, тривалістю 3 секунди кожен. Отримані два вектора ознак нормалізуються і формують відбиток сигналу, що містить 30 компонентів.

7. Google. В роботі [32] автори використовували методи, які широко використовуються для розпізнавання графічних образів, зокрема, для розпізнавання обличчя людини. Особливістю даного підходу є уявлення спектрограми аудіосигналу у вигляді графічного зображення. Конкретна реалізація даного методу являє собою модифікований метод запропонований J. A. Naitzma [30], в якому замість фільтра з ядром, поданим виразом 1.6, використовується набір, що складається з кількох Хаар фільтрів. Із загальної кількості усіх можливих комбінацій Хаар фільтрів у процесі тренування системи була обрана підмножина, що забезпечує одночасно компактне представлення даних і надійне розпізнавання та ідентифікацію навіть значно спотворених сигналів. Подальшим розвитком цього підходу є метод, розроблений фірмою Google, в якому замість Хаар фільтрів використовується вейвлет перетворення сегментів спектрограми [73].

Отримані експериментальні дані виявили такі недоліки вказаних методів:

- MFCC коефіцієнти методів 1 і 2 чутливі до шуму (відношення сигнал/шум не менше 6 db) і спотворень сигналу в результаті цифрового стиснення (MP3, 32 kbps);
- стійкість методу основних компонентів 3 визначається умовами тренування; при цьому практично неможливо забезпечити стійкість одночасно до кількох різних типів спотворень сигналу і шуму;
- параметри, засновані на різниці енергій частотних діапазонів (методи 4 і 5) чутливі до частотних спотворень, реверберації (час затримки > 30 мс) і кодування алгоритмами, призначеними для стиснення мови (GSM);
- приховані Марківські моделі не вирішують проблему індексації даних і можуть бути використані тільки для порівняння кандидатів на етапі перевірки гіпотези;
- усі розглянуті методи чутливі до масштабування сигналу в часі (<5%) і зміні швидкості (<2%).

Слід зазначити, що стійкість до зміни масштабу часу має важливе значення для моніторингу радіо - і телепередач, де ступінь стиснення/розтягування матеріалів може досягати 10-15%. На рисунку 2.4 показано пороги ідентифікації для неспотворених сигналів і сигналів, стиснених цифровими методами MP3 і GSM. Згідно з даними, представленими на рисунку 2.4, навіть модифікований метод Philips, автори якого заявляють, що вирішили проблему масштабування, не забезпечує надійну ідентифікацію таких сигналів. Згідно з отриманими експериментальними даними діапазон масштабування, при якому цей метод здатний ідентифікувати матеріали, стислі цифровими методами, насправді не перевищує  $\pm 5\%$ .

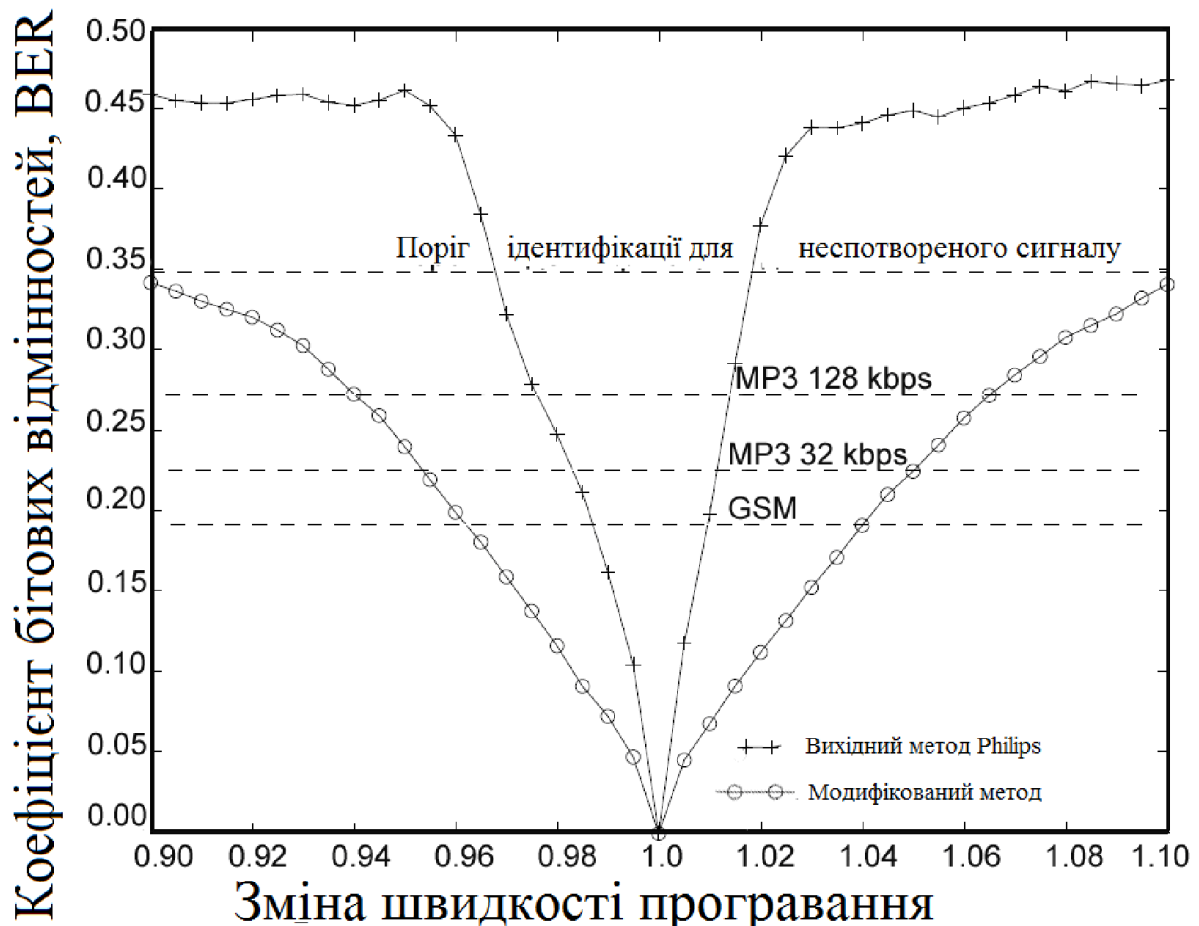


Рисунок 2.4 – Залежність коефіцієнта бітових відмінностей від швидкості програвання аудіосигналу для методу цифрових відбитків фірми Philips і його модифікації на основі перетворення Фур'є-Мелліна

Можна сказати, істотним недоліком при розробці всіх розглянутих методів було ігнорування особливостей сприйняття звуку людиною і, відповідно, спотворень, що вносяться в процесі цифрового стиснення сигналу. Слід зазначити, що практично всі аудіовізуальні матеріали подані в стислом вигляді з допомогою алгоритмів, які базуються на психоакустичному моделюванні аудиторної системи людини і видаленні маскованих компонент сигналу, які людина не чує. Запропоновано при розробці методів ідентифікації цифрових аудіовізуальних матеріалів враховувати особливості їх сприйняття людиною і забезпечувати відповідність параметрів дескрипторів сигналу психоакустичним моделям, які використовуються в алгоритмах цифрового стиснення.

Таким чином, було експериментально показано, що існуючі системи не здатні забезпечити надійну ідентифікацію та пошук аудіоматеріалів з метою

контролю за їх розповсюдженням. У наступному пункті буде запропоновано стійкі дескриптори сигналу та досліджено їх параметри на основі аналізу властивостей аудиторної системи людини.

## 2.2 Аналіз стійкості дескрипторів на основі моделей сприйняття акустичних сигналів людиною

Людина здатна розпізнавати мову в складних умовах, зокрема в присутності шуму, реверберації та при наявності комплексних спотворень. Можна припустити, що в процесі еволюції людини вдосконалення розпізнавання мови відбувалося як за рахунок поліпшення параметрів системи слуху, так і за рахунок формування найбільш ефективного методу подання і кодування лінгвістичної (фонетичної) інформації. Розгляд особливостей слухової системи людини представляє інтерес в першу чергу для визначення параметрів мовлення, які містять максимальний обсяг лінгвістичної інформації. Використання таких параметрів дозволяє виділити характерні властивості акустичних сигналів найбільш стійких до різних перетворень, зокрема, фільтрації і цифрового стиснення. Відомо, що більшість алгоритмів стиснення звуку в системах телефонії такі, як GSM і CELP засновані на моделюванні вокального тракту та слухової системи людини і вносять суттєві викривлення в невербальні акустичні сигнали (шум, музика). З урахуванням того, що аудіоматеріали зберігаються і передаються в стислому вигляді, стійкість відбитків до цифрового кодування є одним з найважливіших критеріїв при виборі параметрів дескрипторів акустичних сигналів.

### 2.2.1 Вибір параметрів частотного перетворення з урахуванням частотного розширення аудиторної системи

Первинний частотний аналіз акустичних сигналів в аудиторній системі людини здійснюється базальною мембраною, преобразовующою звук в



частотно-часове подання [33]. Базальна мембрана діє аналогічно банку смугових фільтрів, смуга пропускання яких залежить від частоти. Людське вухо здатне розрізняти незначні (менше 0.2%) зміни частоти індивідуальних тональних складових сигналів. Для подання тональності сигналів використовується одиниця вимірювання крейда; 1000 крейда відповідають тональності сигналу з частотою 1000 Гц на рівні 40 db. Іншою характеристикою частотного дозволу аудиторної системи є критичні смуги, ширина яких відповідає мінімальній різниці частот сигналів, при якій ці сигнали сприймаються і обробляються аудиторної системою незалежно. Зокрема, при згладжуванні спектральних характеристик сигналу розбірливість мови зберігається, якщо розмір вікна згладжування не перевищує ширину критичної смуги [75].

Існує також зв'язок між частотними дозволами аудиторної системи і сприйняттям музичних творів, зокрема консонансних співзвуч. У загальному випадку консонанс можна охарактеризувати як злине гармонічне співзвуччя, яке звучить приємно для людини. Зв'язок між гармонійним сприйняттям консонансних співзвуч і критичними діапазонами представлена в роботі [34], в якій показано, що два чистих тональних сигналу сприймаються як гармонійне співзвуччя тоді, коли різниця їх частот трохи перевищує ширину відповідної критичної смуги. Ця залежність представлена на рис. 2.5. Графік показує, що консонанс має мінімальне значення, коли два тони рознесені по частоті на величину рівну чверті ширини відповідної критичної смуги і досягає максимального значення на межі критичної смуги.

В роботі [77] показано, що існує тісний зв'язок між тональним консонансом і частотними інтервалами, використовуваними в європейській тональній музичній системі. Передбачається, що в європейській тональній музиці для стійких співзвуч в кожному критичному діапазоні повинен бути тільки один домінуючий тональний компонент.

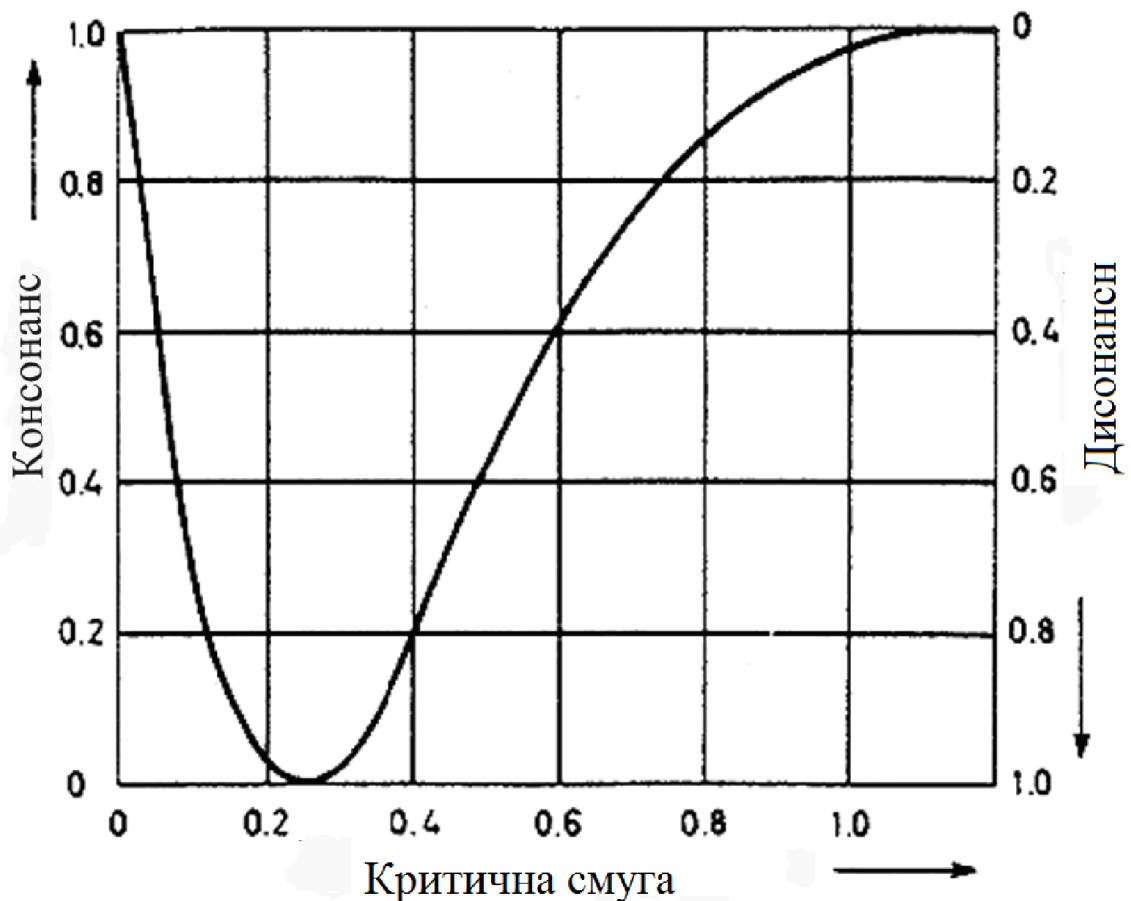


Рисунок 2.5 - Залежність співзвучності двох тональностей від відстані між ними щодо ширини критичної смуги в частотному домені

Поняття критичних смуг лежить в основі барк-шкали, що використовується при моделюванні аудиторної системи за допомогою набору смугових фільтрів, а також в різних системах стиснення аудіосигналів (MPEG-1,2). З огляду на зв'язок критичних діапазонів з особливостями побудови і сприйняття музики і мови, а також, враховуючи, особливості архітектури існуючих алгоритмів стиснення аудіосигналів, заснованих на їхньому уявленні і кодуванні в критичних діапазонах, можна припустити, що для стійкої ідентифікації аудіосигналів досить використовувати банк смугових фільтрів з критичним розподілом частот. Ширина критичних діапазонів залежить від частоти і з достатньою для методу відбитків точністю може бути апроксимована за допомогою набору фільтрів з пропускнуою здатністю 1/3 октави.

Таким чином для формування банку фільтрів в методі аудіовідбитків нами запропоновано використовувати набір смугових фільтрів, що перекриваються з логарифмічним розподілом частот і смугою пропускання  $1/3$  октави.

### 2.2.2 Вибір дескрипторів акустичних сигналів з урахуванням особливостей сприйняття людської мови

Важливою властивістю людської мови є те, що велика частина лінгвістичної інформації міститься у відносно повільних змінах структури мовного сигналу. Доказом цього можуть бути експерименти з канальним вокодером, розробленим в 1928 році в лабораторії Bell Labs [78] для моделювання і стиснення мови. Вокодер кодує мову, витягуючи параметри моделі мовного сигналу, і потім синтезує її на основі цих параметрів (рисунок 2.6). Вокодер складається з 10 аналогових смугових фільтрів, що перекривають діапазон частот 250-3000 Гц. На етапі аналізу, в кожному частотному діапазоні вироблялося виділення основної частоти і обвідної сигналу.

На етапі синтезу, огинаюча сигналу в кожному частотному діапазоні накладається на несучу частоту і сигнали всіх діапазонів складаються. В ході експериментів було встановлено, що розбірлива мова може бути отримана при виділенні огинаючої з використанням фільтра низьких частот з пропускну здатністю 0-25 Гц. При цьому основна частина сигналу була представлена частотами 10 Гц і нижче [35]. У більш пізніх роботах було показано, що частоти вище 16 Гц не впливають на розбірливість мови [36].

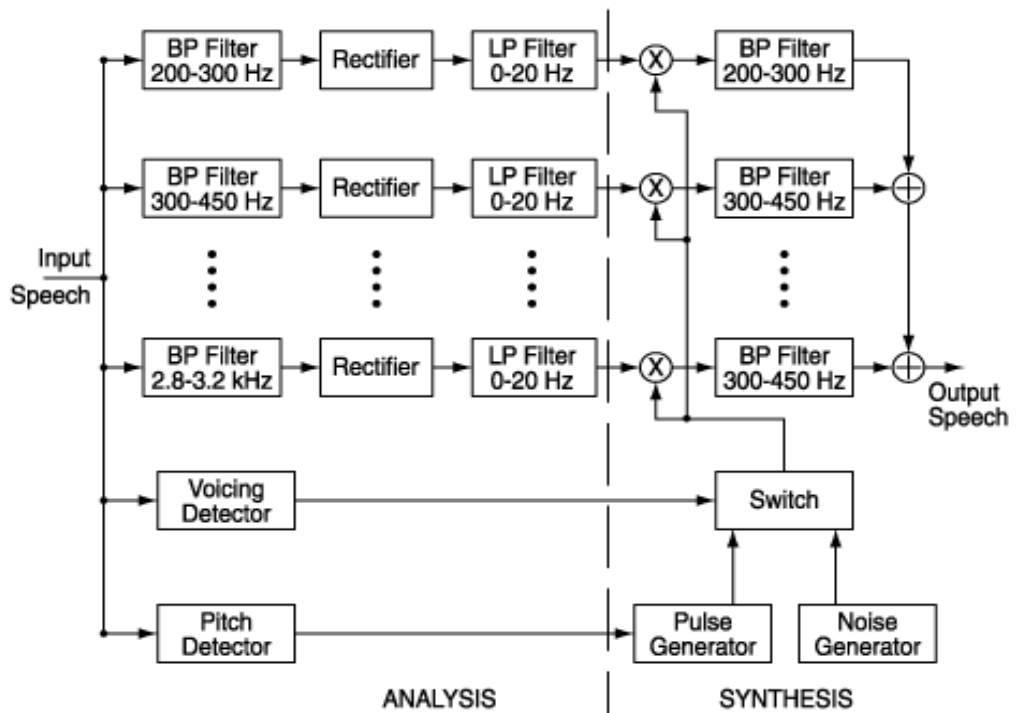


Рисунок 2.6 – Блок-схема вокодера

Зміни спектральних характеристик мови в часі можуть бути представлені у вигляді спектра модуляції. Спектр модуляції обчислюється на основі спектрального аналізу огибаючої сигналу з подальшою нормалізацією значень [78] [35]:

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

де,  $s(t)$  являє собою енергію кривої сигналу і  $\langle s(t) \rangle$  є середнім значенням  $s(t)$ . Огибаюча сигналу може бути отримана фільтрацією детектованого сигналу за допомогою фільтра низьких частот. Для складних сигналів, таких як мова, різні частотні діапазони можуть мати низький рівень кореляції і відмінні параметри спектра модуляції. Таким чином, для складних сигналів аналіз спектра модуляції доцільно проводити окремо для кожного діапазону. Модуляція мови викликається двома принципово відмінними причинами. Перша причина викликана змінами в загальному рівні енергії сигналу, які відповідають зміні фонем і однаково

впливають на всі діапазони. Друга - змінами в спектральних розподілах формант, специфічних для кожного частотного діапазону.

Рівень спотворень сигналу в каналі передачі має велике значення для розпізнавання мови. Якісною характеристикою каналу передачі є функція передачі модуляції (Modulation Transfer Function, MTF). Ця передавальна функція являє собою відношення спектра модуляції отриманого сигналу до спектру модуляції вихідного сигналу і враховує різні ефекти і форми акустичних спотворень в каналі передачі.

Для каналу, що містить шум, модуляційна передавальна функція є константою, що визначає ступінь ослаблення коефіцієнтів модуляції, яка не залежить від частоти модуляції [35]:

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

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:

$$H_m(\omega) = \left| \frac{\int_0^{\infty} h^2(t) e^{-j\omega t} dt}{\int_0^{\infty} h^2(t) dt} \right| \quad (2.3)$$

Відлуння в приміщеннях призводить до того, що MTF має яскраво виражений підйом на низьких модуляційних частотах. Для покращення сприйняття мови в приміщенні з великим часом реверберації було запропоновано використовувати фільтрацію сигналів в модуляційно-частотному домені із застосуванням смугових фільтрів, що мають підсилення в 10-15 db на частоті 4-6 Гц [81].

Функція передачі модуляції може використовуватися для обчислення якості мови, зокрема, для обчислення індексу передачі мови (Speech Transmission Index,

STI) [37]. STI являє собою чисельну оцінку MTF каналу для частот 125-8000 Гц і частот модуляції 0.625-12.5 Гц. Було показано, що передбачені значення STI добре відповідають експериментальним даним по сприйняттю мови в умовах реверберації на різних мовах для комунікаційних каналів, що вносять шум і нелінійні спотворення [83]. Широке використання STI, заснованого на аналізі спектра модуляції в ряді стандартів якості сприйняття мови підтверджує, що повільні зміни в спектральних характеристиках мови несуть істотний обсяг лінгвістичної інформації [38].

Пізніші дослідження показали, що для надійного розпізнавання мови в умовах реверберації діапазон частот модуляції може становити всього 8 Гц [85]. При цьому було відзначено особливе значення частот модуляції в районі 4 Гц. Для вимірювання якості сприйняття мови був запропонований індекс частотно-часової модуляційний (Spectro-Temporal Modulation Index, STMI) [39] [40]. STMI враховує зміни в уявленні сигналу як в часі, так і по частоті, а також враховує індекси як амплітудної (AM), так і частотної (FM) модуляції.

Існує також припущення, що в аудиторній системі існують канали, налаштовані на сприйняття модуляційних частот аналогічно тому, як це відбувається при сприйнятті спектральних частот в критичних діапазонах [88]. Було показано, що особливості сприйняття людиною тимчасової динаміки акустичних сигналів можуть бути пояснені фільтрацією в модуляційно-частотних каналах аудиторної системи [41].

В результаті аналізу особливостей сприйняття аудіоінформації людиною, за доцільне, використовувати дескриптори на основі аналізу амплітудної і частотної модуляції акустичних сигналів в частотних діапазонах з логарифмічним розподілом частот. Аналіз стійкості запропонованого методу до цифрового стиску в порівнянні з іншими типами дескрипторів (формули 1.1-1.7) наведено в наступному розділі.

### 2.2.3 Аналіз впливу цифрового стиснення на стійкість дескрипторів аудіосигналів

Більшість аудіовізуальних матеріалів поширюються в цифровому вигляді. Тип алгоритму і ступінь стиснення суттєво впливає на якість цифрових сигналів. У мультимедійних додатках використовуються алгоритми стиснення, що забезпечують високу якість звуку, а саме стиснення здійснюється з досить високим бітрейтом. Ідентифікація таких сигналів не викликає труднощів для більшості існуючих методів цифрових аудіовідбитків. У той же час, в ряді програм, зокрема в IP і стільникової телефонії, використовуються спеціалізовані алгоритми стиснення, оптимізовані для відтворення специфічних типів аудіосигналів і забезпечують максимально високий рівень стиснення. Широко поширений в телефонії стандарт GSM оптимізований для ефективного стиснення мови, заснований на моделюванні мовного тракту людини і погано пристосований для кодування складних багатотональних сигналів, в тому числі музичних творів.

Стійкість алгоритмів цифрових відбитків до GSM-стиску має важливе практичне значення для додатків ідентифікації аудіоматеріалів, що передаються по цифрових каналах, в тому числі по каналах IP і стільникової телефонії.

В роботі [42] представлені результати досліджень щодо впливу спотворень, що вносяться GSM каналом передачі на дескриптори різних типів аудіосигналів. Проведені дослідження показали, що повільні зміни сигналу, зокрема з частотою 4 Гц значно меншою мірою схильні до спотворень в процесі GSM стиснення, ніж дескриптори, засновані на спектральних параметрах 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) і їх похідні для різних частотних діапазонів.

Рівень шуму SNR (Signal to Noise Ratio) для кожного типу дескрипторів визначався за такою формулою:

$$SNR = 10 \log_{10} \frac{\sum_{m=1}^M (F_m)^2}{\sum_{m=1}^M (F_m - F_m^c)^2} \quad (2.4)$$

де  $F_m$  – значення дескриптора в вихідному сигналі і значення цього дескриптора після GSM кодування/декодування. Результати експерименту наведені в таблиці 2.

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	3.55	2.66	2.85	2.12	1.25	0,55
SFM	9.43	6.30	6.85	4.20	4.46	2.70
SC	5.00	6.00	-3.28	-8.00	-2.52	-3.46
MFCC	-2.40	-3.00	-2.00	1.00	0,36	0.08
ZCR	5.90	4.70	8.20	14.20	7.13	8.66
<b>4 Гц</b>	<b>26.37</b>	<b>24.44</b>	<b>23.96</b>	<b>23.70</b>	<b>19.46</b>	<b>19.80</b>

Згідно з даними, наведеними в таблиці 2.1, співвідношення сигнал/шум в каналі GSM для дескриптора, заснованого на аналізі спектра модуляції з частотою 4 Гц, в середньому на 20 db перевищує відповідні значення для спектральних параметрів.

Наведені дані підтверджують, що повільні коливання амплітуди акустичних сигналів набагато стійкіші до цифрового GSM стиску, ніж акустичні спектральні параметри (формули 1.1-1.7).

Таким чином, для ідентифікації аудіосигналів запропоновано використовувати дескриптори на основі аналізу спектру модуляції акустичних сигналів.



#### 2.2.4 Результати аналізу та висновки щодо вибору дескрипторів

Сприйняття мови і музичних творів людиною засноване на спільному часовому, частотному і модуляційному аналізі акустичних сигналів. При цьому частотний аналіз відбувається в критичних смугах, аналіз спектра модуляції відбувається в модуляційних каналах, найбільш значущі з яких розташовані в діапазоні 0..8 Гц. Повільні зміни огинаючої акустичного сигналу мають велике значення для розпізнавання мови в складних умовах, в присутності шуму, спотворень і реверберації, а також більш стійкі до цифрового стиску з низьким бітрейтом. Разом з тим, для розпізнавання мови істотне значення має не тільки амплітудна модуляція (АМ) сигналу, а й коливання фундаментальної частоти, викликані частотною модуляцією (FM) в голосовому тракті. З огляду на властивості аудиторної системи людини і результатів аналізу існуючих додатків, запропонований новий метод ідентифікації матеріалів, заснований на аналізі спектру АМ/FM модуляції акустичних сигналів в діапазоні 0.5-8 Гц. Реалізація методу приведена в наступному параграфі.

#### 2.3 Метод цифрових аудіо відбитків

Грунтуючись на властивостях повільних коливань сигналу, розглянутих в попередньому параграфі, зокрема, стійкості до шуму, перекручувань і цифровому стисненню, було запропоновано новий метод ідентифікації аудіофрагментів на основі аналізу спектру модуляції [44]. В основі запропонованого методу лежить поняття спектрограми амплітуди модуляції звукового сигналу, яка представляє частотний розподіл амплітуди повільних змін сигналу в часі. Спектрограма амплітуди модуляції обчислюється на основі звичайної спектрограми сигналу і представляє залежність спектральної енергії від часу і частоти. У цьому параграфі розглянуті властивості частотно-часових уявлень, особливості їх використання та схема отримання цифрових відбитків на основі спектра модуляції.

### 2.3.1 Метод аудіо відбитків на основі частотно-часового аналізу аудіо сигналів

При аналізі сигналів часто передбачається, що сигнал, який аналізують є стаціонарним. У той же час, це припущення рідко буває вірним в разі акустичних сигналів. Частота і амплітуда музичних творів і мовних сигналів може значно змінюватися в часі. Аналіз змінних в часі сигналів здійснюється за допомогою локального перетворення Фур'є (Short Time Fourier Transform, STFT) [95] і частотно-тимчасового розподілу (Time-Frequency Distribution, TFD) [96].

Локальне перетворення Фур'є засноване на розбитті сигналу на невеликі фрагменти, що перекриваються, розмір яких досить малий для того, щоб сигнал в них міг бути представлений як стаціонарний. Щоб обмежити розмір фрагмента і мінімізації спотворень, викликаних розривами на кінцях ділянки інтегрування на фрагмент накладається віконна функція  $w(t)$ :

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

Розмір фрагмента, який визначається розміром і формою віконної функції  $w$ , впливають на частотне і тимчасове розширення.

Іншим способом отримання частотно-часових уявлень є частотно-часовий розподіл, який теоретично дозволяє отримати більш високе розширення, ніж локальне перетворення Фур'є. Найбільш простим 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 те, що вони вносять суттєві артефакти, викликані взаємодією множинних частотних компонентів. Інші типи частотно-часових розподілів такі, як розподіл Чоя-Вільямса (Choi-Williams),

дозволяють зменшити інтеркомпонентні артефакти, але все одно їх розширення обмежено принципом невизначеності [96].

Як було показано в параграфі 2.1.1, для задач ідентифікації і порівняння аудіосигналів високе частотне розширення не потрібне. У той же час інтеркомпонентні артефакти, властиві TFD є вкрай небажаними і можуть істотно знизити реальну роздільну здатність алгоритму, особливо в разі наявності сильних перешкод і спотворень. Тому в даній роботі запропоновано застосування локального перетворення Фур'є, в якому для забезпечення необхідного частотного розширення спектра модуляції 8 Гц розмір вікна становить 62.5 мс. В цьому випадку частотне розширення локального перетворення Фур'є складає 32 Гц, що менше ширини критичних смуг і, як було показано раніше в параграфі 2.2.1, є достатнім для систем ідентифікації аудіосигналів. Крім цього, перевагою локального перетворення є його простота реалізації і передбачуваність результатів.

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

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

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  $M(t, \omega, \nu)$  обчислюється за допомогою другого перетворення Фур'є для кожного частотного діапазону:

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

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

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

Експериментально встановлено, що отримана спектрограма стійка до різних спотворень і навіть після грубого квантування може бути використана для порівняння та ідентифікації різних типів аудіосигналів [30].

Як було зазначено вище, метод аналізу спектру модуляції акустичних сигналів забезпечує хороші результати в системах автоматичного розпізнавання мови [91]. На малюнку 2.7 приведена блок-схема методу отримання спектрограми модуляції, використаного в роботі [43], в якій вхідний сигнал розбивається на частотні діапазони за допомогою набору смугових фільтрів з мінімальним частотним перекриттям. Після детектування сигнал в кожному діапазоні фільтрується, нормалізується, і спектр потужності модуляції обчислюється за допомогою перетворення Фур'є.

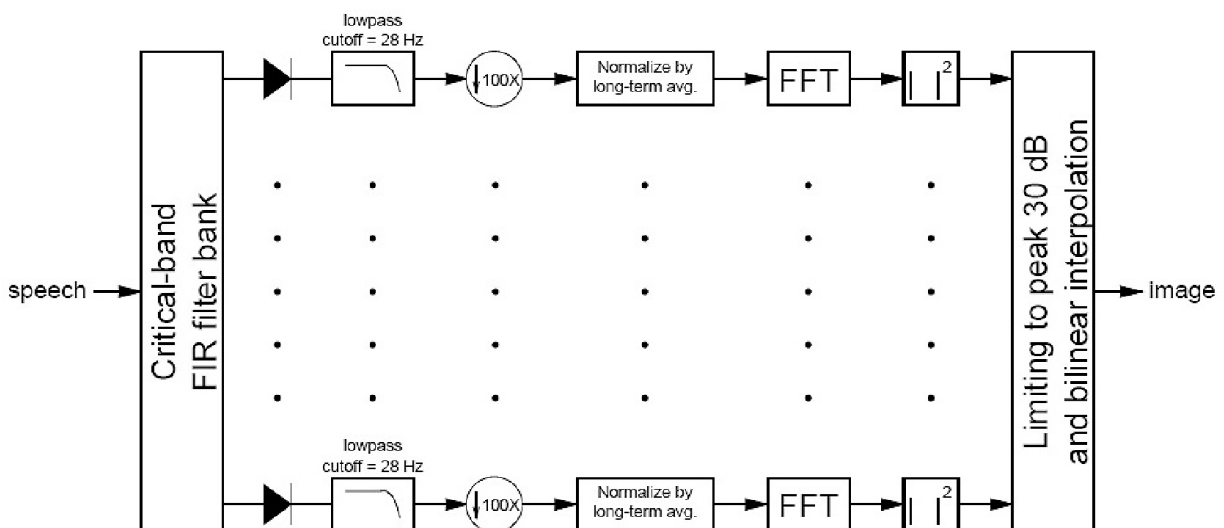


Рисунок 2.7 – Блок-схема алгоритму обчислення спектрограми амплітуди модуляції відеосигналу в роботі [43]

Аналогічний підхід використаний в методі цифрових відбитків на основі спектра модуляції, запропонованому Sukittanon and Atlas [97], Основною відмінністю цього методу від описаного вище методу розпізнавання мови [43] є спосіб отримання параметрів модуляції, в якому замість другого перетворення Фур'є використовується вейвлет перетворення, що забезпечує краще розширення на низьких частотах. У реалізованому методі на основі наведеного алгоритму [97] дескриптор сигналу заснований на оцінці геометричного середнього спектрограми модуляції аудіосигналу тривалістю 12 секунд, розбитого на фрагменти, що перекриваються довжиною 4 секунди в часі і 19 барків по частоті. Отримані експериментальні результати показали, що за сукупністю властивостей даний метод не перевершує метод фірми Philips [30], а по ряду параметрів навіть поступається йому. Наприклад, алгоритм дуже чутливий до частотних спотворень. Зокрема, частотна корекція за допомогою еквалайзера призводить до значного зменшення повноти розпізнавання до 90%, що є повністю неприйнятним в разі практичного використання даного алгоритму в системах пошуку та ідентифікації аудіовізуальних матеріалів. Даний алгоритм також має високу чутливість до спотворень, що виникають при цифровому стисненні. Для аудіосигналів, підданих MP3-стисненню з бітрейтом 96 kbps, яке забезпечує досить високу якість при кодуванні монофонічних сигналів, повнота розпізнавання становила лише 97%, в той час як алгоритм фірми Philips здатний стійко ідентифікувати аудіоматеріали, стислі MP3 до рівня 32 kbps. Стійкість до цифрового стиснення має важливе практичне значення і безпосередньо впливає на продуктивність систем пошуку, заснованих на обробці сигналів, які передаються по цифрових каналах.

Слід також зазначити, що перераховані вище параметри отримані для часу інтегрування 10 секунд. У разі зменшення часу інтегрування (зменшення тривалості тестового фрагмента) до 5 секунд, середня повнота розпізнавання падає до 80%. Слід також враховувати, що в реальних додатках акустичні сигнали піддаються не одному конкретному типу спотворень, а їх комбінації, що включає всі перераховані вище перетворення і шум.

Зазначених недоліків позбавлений запропонований нами новий алгоритм отримання цифрових відбитків, схема якого представлена на рисунку 2.8. У ньому спектрограма вихідного сигналу, отримана за допомогою локального перетворення Фур'є, розбивається на  $N=13$  частотних діапазонів з логарифмічним частотним розподілом. Спектр модуляції кожного діапазону, отриманий за допомогою другого перетворення Фур'є, нормалізується і стискається з використанням диференціального кодування з наступним квантуванням. При надходженні чергового аудіофрагмента тривалістю 50 мс на вхід аналізатора, процес обчислення повторюється, в результаті чого генерується вихідна послідовність відбитків зі швидкістю 240 байт в секунду.

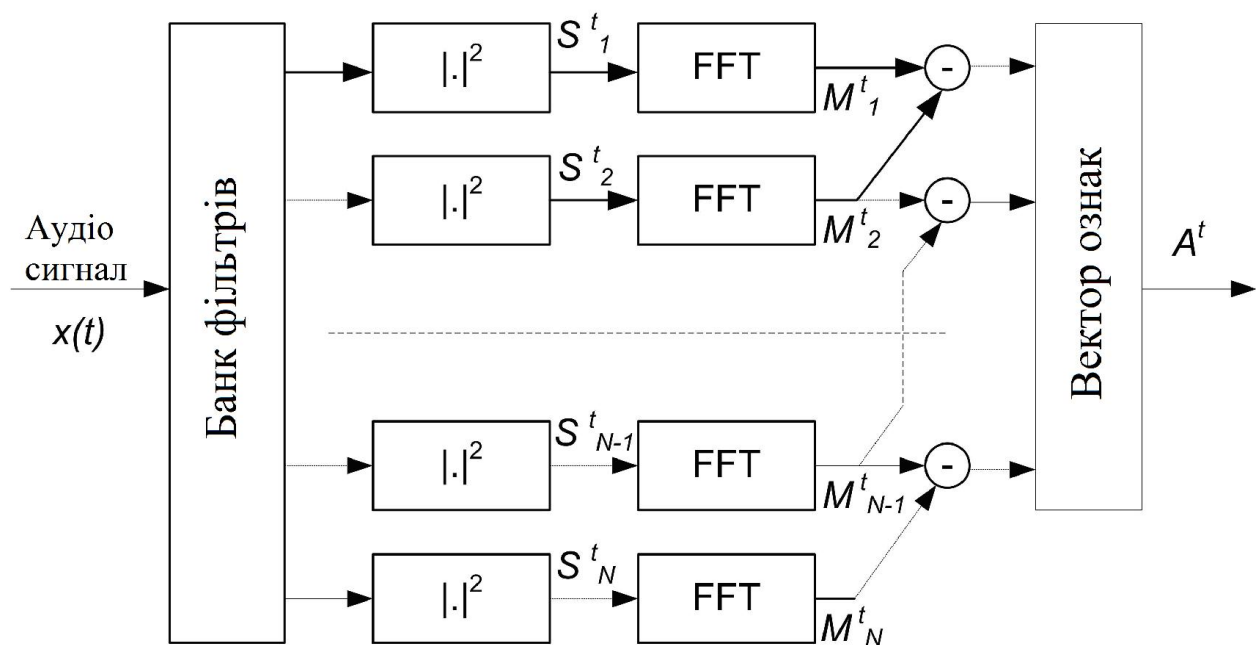


Рисунок 2.8 – Блок-схема запропонованого алгоритму отримання вектора на підставі ознак спектра модуляції

Частотний поділ здійснюється за допомогою банк фільтрів, що складаються з кусочно-безперервних смугових фільтрів, що перекриваються з лінійною крутизною характеристики, що перекривають смугу частот 200-3200 Гц (рисунок 2.9).

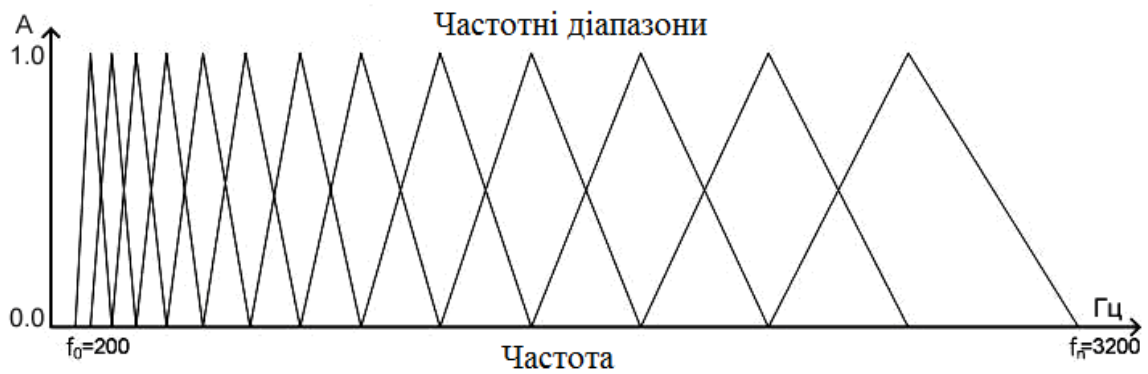


Рисунок 2.9 – Частотні характеристики банку смугових фільтрів

Лінійність косих фільтрів забезпечує мінімальний рівень спотворень при демодуляції частотно-модульованих FM сигналів і ефективний розподіл AM і FM компонентів в спектрі модуляції. Виходячи з частотного розширення аудиторної системи людини і особливостей побудови європейської музичної системи, можна припустити, що для переважної більшості реальних сигналів в кожному частотному діапазоні, ширина якого не перевищує критичну смугу, може бути присутнім єдиний домінуючий тональний сигнал. Для перевірки гіпотези досліджена модель, за якої акустичний сигнал  $x(t)$  представлений у вигляді несучої частоти  $\omega_c$ , модульованої по частоті сигналом  $f_m(t)$ , а по амплітуді сигналом  $a_m(t)$  :

$$x(t) = [1 + \alpha a_m(t)] \cdot A_c \cdot \sin(\omega_c t + \varphi_0 + \beta f_m(t)) \quad (2.10)$$

де  $\alpha$  і  $\beta$  є коефіцієнтами AM і FM модуляції, а  $(A_c, \omega_c, \varphi_0)$  - амплітудою, частотою і початковою фазою несучої акустичної хвилі відповідно. 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:

$$x(t) = [1 + \alpha \sin(\nu_a t)] \cdot \sin(\omega_c t + \beta \sin(\mathcal{G}_f t)) \quad (2.11)$$

де  $\mathcal{G}_a$  і  $\mathcal{G}_f$  є частотами амплітудної і частотної модуляції.

Спільна демодуляція АМ і FM складових сигналу (2.11) здійснювалася за допомогою дискримінатора, що представляє собою смуговий фільтр, налаштований на частоту вище частоти сигналу з подальшим детектуванням і виділенням огинаючої. На рисунку 2.10А представлений алгоритм детектування тестового FM/AM сигналу на схилі смугового фільтра, що входить до складу банку фільтрів.

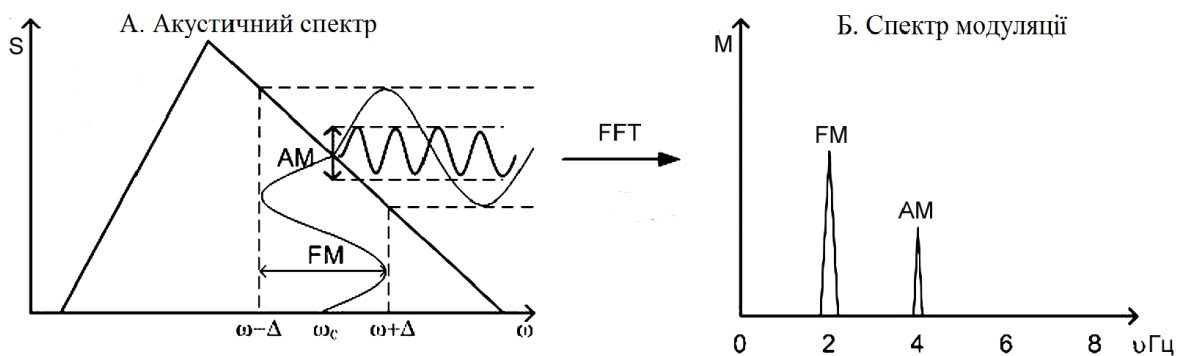


Рисунок 2.10 – Схема детектування АМ/FM сигналів за допомогою смугового фільтра (А) і спектр отриманого детектованого сигналу (М), в якому чітко помітні АМ і FM компоненти, з частотою 4Гц і 2Гц відповідно

Після детектування спектр модуляції для кожного частотного діапазону обчислюється за допомогою перетворення Фур'є. На рисунку 2.6Б наведено спектр модуляції  $M(t, \omega, \nu)$  тестового сигналу з частотою  $\omega_c = 440$  Гц, АМ/FM-модульованого синусоїдальними сигналами з частотою  $\nu_a = 4$  Гц і  $\nu_f = 2$  Гц, відповідно.

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



$$A^t(i, k) = \begin{cases} 1, \text{если } M^t(i, k) - M^t(i + 1, k) > 0 \\ 0 \text{ в противном случае} \end{cases} \quad (2.12)$$

де  $i$ ,  $1 \leq i < N$  є індексом частотного діапазону в банку фільтрів;

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

Експериментальні дані показали, що незважаючи на відносну простоту реалізації, запропонований алгоритм не поступається існуючим методам в точності і повноті розпізнавання аудіосигналів, особливо при наявності спотворень і шумів. Разом з тим, одним з недоліків запропонованого методу (як втім і більшості методів, розглянутих раніше) є його відносна чутливість до масштабування аудіосигналів в часі. Запропонований метод дозволяє ідентифікувати сигнали, стислі і розтягнуті в часі на 5%, в той час, як деякі матеріали, які транслюються по радіо і телевізійним каналам (зокрема рекламні оголошення) можуть бути навмисно стиснуті на 10% і більше для повного використання ефірного часу.

### 2.3.2 Використання перетворення Мелліна для аналізу спектра модуляції акустичних сигналів.

Як уже зазначалося, слух людини здатний надійно ідентифікувати аудіофрагменти, що значно відрізняються як за тривалістю (зміна масштабу в часі), так і по висоті звучання (зміна масштабу в частотному домені). Ряд методів, що широко застосовуються такі, як MFCC і перетворення Фур'є-Мелліна інваріантні до зміни масштабу і засновані на обробці даних в логарифмічно частотній шкалі, аналогічні до того, як це відбувається в слуховій системі людини [45]. Однак ці методи використовуються для короткострокового аналізу в інтервалі, заданому віконною функцією локального перетворення, і не враховують тривалих змін, які мають важливе значення для ідентифікації акустичних сигналів [99]. Довгострокові методи такі, як аналіз спектра модуляції [97] мають ряд переваг, але, в той же час, чутливі до зміни швидкості відтворення. У даній роботі представлений метод ідентифікації аудіосигналів, заснований на

аналізі спектру модуляції за допомогою швидкого перетворення Мелліна, інваріантного до зміни масштабу.

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:

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

Аналогічною властивістю по відношенню до зміни масштабу сигналу при його стисненні або розтягуванні володіє і перетворення Мелліна. Перетворення Мелліна  $D_f$  для функції  $f(t)$  визначається наступним виразом:

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

де  $s = -jc + \beta$ ,  $\beta \in R$  є комплексною змінною.

Стосовно до цифрової обробки сигналів, перетворення Мелліна більш відоме як масштабне перетворення (Scale Transform) [100], [46], яке є окремим випадком перетворення Мелліна при  $\beta = 1/2$ :

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

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  $t = \ln \tau$ :

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

За аналогією з локальним перетворенням Фур'є, локальне перетворення Мелліна може бути представлено в наступному вигляді:

$$D_t(c) = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} f(\tau) w_m(\tau-t) e^{(-jc+\beta-1)\ln\tau} d\tau \quad (2.17)$$

У даній роботі швидке перетворення Мелліна реалізовано з використанням швидкого перетворення Фур'є сигналу, експоненціально масштабованого в часі в інтервалі, в якому віконна функція  $h(\tau)$  має ненульові значення. Необхідну зміну масштабу в часі здійснено шляхом інтерполяції значень локального сигналу  $f(\tau)h(\tau-t) = f(\tau)w_s(\tau-t)e^{\beta\tau}$  в точках, розташованих з експоненціальним кроком.

Важливою особливістю масштабного перетворення є його інваріантність до зміни масштабу сигналу  $f(t)$ . Як відомо, масштабування – це стиснення або розтягнення сигналу в часі зі збереженням енергії. Для сигналу  $f(t)$  його масштабована версія  $g(t)$  може бути отримана в такий спосіб:

$$g(t) = \sqrt{\alpha} f(\alpha t) \text{ де } \alpha \in R^+ \text{ представляє масштаб перетворення.}$$

Відповідність між масштабним перетворенням вихідного сигналу  $f(t)$  і його стислій/розтягнутій в часі версією  $g(t)$  може бути представлено в наступному вигляді:

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

Таким чином, масштабування призводить до зсуву фаз, в той час як амплітуди коефіцієнтів масштабного перетворення вихідного і масштабованого сигналів не змінюються:

$$|D_g(c)| = |D_f(c)| \quad (2.19)$$

Ефективність розробленої реалізації швидкого перетворення Мелліна може бути продемонстрована на наступному прикладі. На рисунку 2.11 представлений фрагмент акустичного сигналу, відповідний звуку "О" в слові "один" і його масштабована версія. На графіку 2.11А представлений вихідний сигнал, а в наступній діаграмі 2.11Б його версія, стисла в часі в 2 рази. При стисненні сигналу його спектр Фур'є розтягується і навпаки – при розтягуванні сигналу, його спектр стискається.

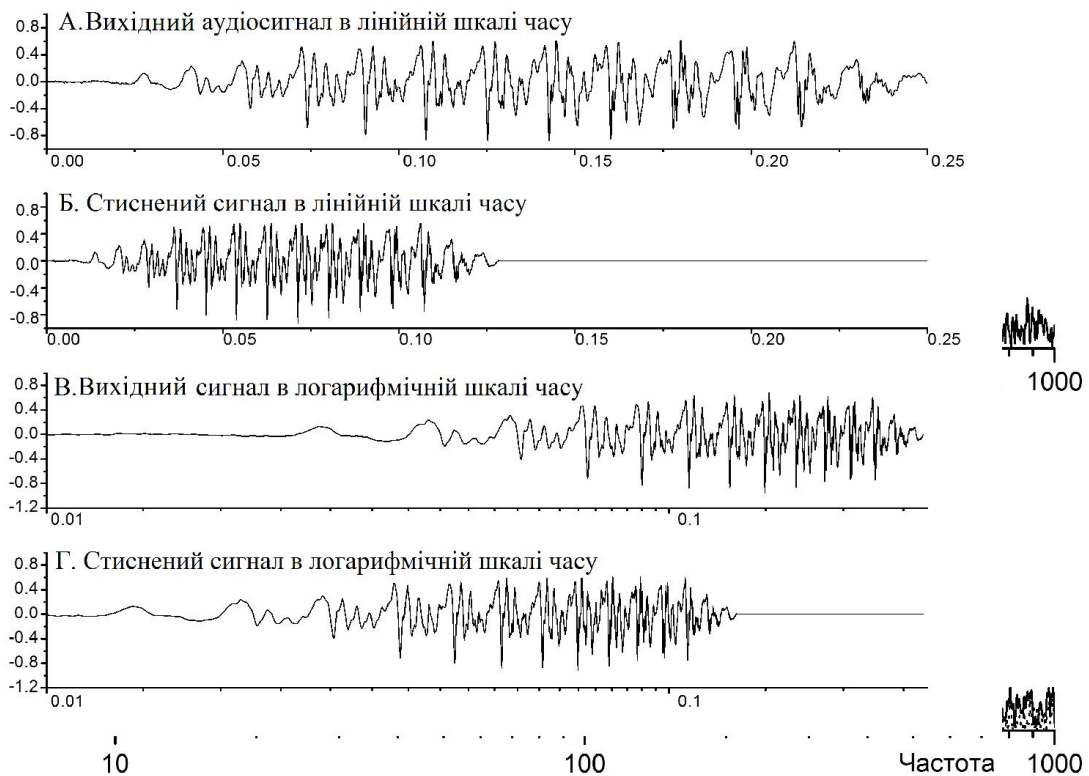


Рисунок 2.11 – Вхідний аудіосигнал (А), сигнал, стиснений в 2 рази (Б) і їх подання до логарифмічної шкали часу

Незважаючи на те, що амплітудні спектри вихідного і стисненого сигналів відрізняються тільки масштабом, їх пряме зіставлення можливо тільки в тому випадку, якщо відомий фактор масштабування. Для реальних ідентифікованих сигналів масштаб невідомий, пряме порівняння амплітудних спектрів неможливе.

На рисунках 2.11В і 2.11Г наведені вихідний і, відповідно, стиснений сигнали в логарифмічній шкалі. Видно, що в логарифмічній шкалі часу обидва сигнали мають один і той же масштаб і відрізняються тільки величиною зсуву.

На рисунку 2.12 представлені амплітудно-частотні спектри вихідного і стисненого в часі сигналів, отримані в логарифмічній шкалі часу, а також їх різниця. На низьких і середніх частотах спектри практично ідентичні (помилка не перевищує 0.05). Відмінності в високочастотних складових можуть бути мінімізовані шляхом обмеження максимальної частоти сигналу, або збільшенням частоти вибірки.

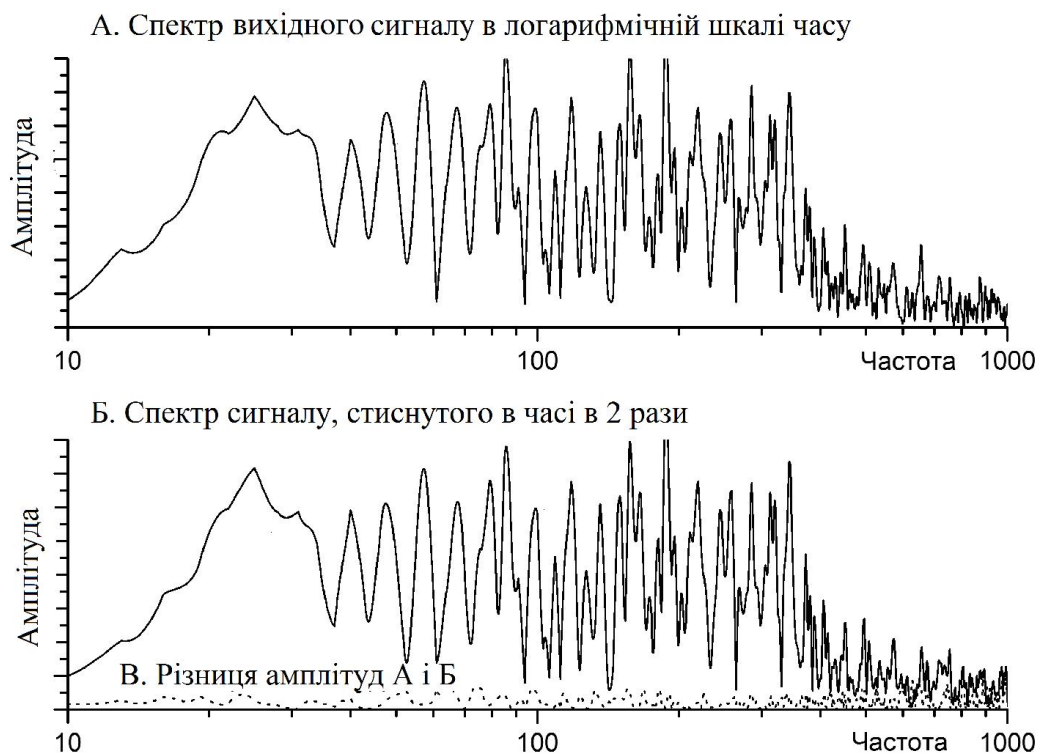


Рисунок 2.12 – Амплітудно-частотні спектри вхідного (А) та стисненого в 2 рази (Б) сигналів в логарифмічній шкалі, а також їх різниця (В)

Слід зазначити, що в той час, як перетворення Мелліна інваріантне з масштабуванням, воно не є інваріантним до зрушення. Тому в практичних додатках залежність цифрового відбитка від зсуву і вимоги до синхронізації сигналів можуть бути мінімізовані за рахунок більшої міри перекриття вікон локального перетворення Фур'є в процесі обчислення спектрограми.

Висновки до розділу 2.

1. Експериментально встановлено, що існуючі методи не забезпечують надійну ідентифікацію акустичних сигналів, стислих у часі і, таким чином непридатні для моніторингу радіо- і телевізійних передач.

2. Розглянуто та проаналізовано властивості різних типів дескрипторів акустичних сигналів. На підставі експериментальних даних показано, що найбільшу стійкість до спотворень і цифрового стиснення має метод, заснований на довгостроковому аналізі змін енергії в частотних діапазонах.

3. Удосконалений метод ідентифікації акустичних сигналів, заснований на аналізі спектру модуляції в частотних діапазонах. З метою підвищення стійкості алгоритму до масштабування сигналу в часі для обчислення параметрів спектра модуляції запропоновано використовувати перетворення Мелліна і його модифікований варіант, заснований на дискретному косинусном перетворенні. Запропонований алгоритм має високу стійкість до зміни масштабу аудіосигналів і дозволяє здійснювати надійну ідентифікацію аудіовізуальних матеріалів, що мають аудіоканал.

## 3 РЕАЛІЗАЦІЯ УДОСКОНАЛЕНОГО МЕТОДУ ЗАХИСТУ АУДІОМАТЕРІАЛІВ З ДОПОМОГОЮ ЦИФРОВИХ ВІДБИТКІВ

### 3.1 Обґрунтування вибору середовища для програмування

Створення додатку здійснюється на базі мови програмування C++, яка є універсальною алгоритмічною мовою та використовується для розробки системних та складних прикладних програм. Вона дозволяє працювати не тільки зі змінними, але і з їх адресами, розміщувати дані як у пам'яті, так і в регістрах, використовувати непряму адресацію (завдання адреси комірки, в якій зазначена адреса даного), автоматично змінювати адресу.

Бурхливий розвиток інформаційних технологій, без яких неможливо уявити сучасне людство, призвело до створення систем програмування, що забезпечують швидку розробку додатків для найпоширенішої операційної системи Windows. До таких систем програмування відносяться Microsoft Visual Basic, Borland Delphi і система, з'явилася пізніше Borland C ++ Builder. Системи швидкої розробки ґрунтуються на технології візуального проектування і подієвого програмування. Програмування в подібних системах зводиться в основному до наочного будівництва додатку з готових компонентів, які містить саме середовище і до написання функцій обробки тих чи інших подій, на які здатна реагувати Windows. У програміста вивільняється багато часу через те, що навколишнє середовище автоматично генерує більшу частину програмного коду. Цей час можна використати на творчість в логіці програми - отже, зробити додаток більш корисним і цікавим.

Мова C++ є мовою високого рівня і основою багатьох систем програмування: Borland C++, Visual C++, Borland C++ Builder. Найбільш популярною з них вважається Borland C++ Builder. За допомогою цієї системи візуального об'єктно-орієнтованого програмування як користувач-початківець, так і програміст-професіонал мають можливість створювати інтерфейс користувача до прикладних програм різноманітних класів, що виглядає однаково професійно.

C++ Builder — середовище швидкої розробки (RAD), що випускається компанією Codegear, дочірньою фірмою компанії Embarcadero (раніше Borland). Призначена для написання програм на мові програмування C++. C++ Builder об'єднує Бібліотеку візуальних компонентів і середовище програмування (IDE), написане на Delphi з компілятором C++. Цикл розробки аналогічний Delphi, але з істотними покращеннями, доданими в C++ Builder.[9] Більшість компонентів, розроблених в Delphi, можна використовувати і в C++ Builder без модифікації, але, на жаль, зворотне твердження не вірне.

C++ Builder містить інструменти, які дозволяють здійснювати справжню візуальну розробку Windows-програм методом drag-and-drop, спрощуючи програмування завдяки WYSIWYG редакторові інтерфейсу, вбудованому в його середовище розробки.

В ході реалізації додатку було використано такі компоненти таких вкладок, як Standard, Additional, System, Dialog, Samples:

1. Label (Етикетка) - створює на формі текстову мітку або напис.
2. Edit (Редагування) - створює на формі поле для редагування текстового рядка.
3. Memo (Поле) - відображає на формі поле для редагування текстових рядків. Зазвичай служить для створення редакторів і полів для виведення блоків даних.
4. Button (Кнопка) - є найпоширенішим компонентом. Служить для створення в додатку різних прямокутних кнопок з текстовим написом.
5. DrawGrid – використовується для створення в додатку таблиці, яка може містити графічні зображення.
6. Timer – дозволяє задавати в додатку інтервали часу.
7. Компоненти OpenFileDialog - діалог «Відкрити файл» і SaveDialog - діалог «Зберегти файл як ...» використовуються найчастіше, в більшості додатків.
8. PerformanceGraph – побудова графіка в залежності від семплу.

Синтаксиси, що були використані для реалізації алгоритму створення цифрових аудіо відбитків представлено у питанні 3.2.



## 3.2 Реалізація додатку для створення цифрових аудіо відбитків

Для реалізації додатку для створення цифрових аудіо відбитків окрім середовища для програмування C++ Builder було додатково використано бібліотеку функцій bass.dll.

Цифровий аудіо відбиток формується за наступним алгоритмом (додаток А). Перш за все необхідно заданий аудіо файл розбити на частини, здійснити так зване фреймування. Адже кожен кадр буди містити свою вибірку даних для опису композиції.

Як було описано в питанні 2.1, нам достатньо фрагмента довжиною 3 секунди для ідентифікації файлу та створення відбитку на основі цих даних.

Спершу потрібно було ініціалізувати вихідний пристрій.

```
BASS_Init(-1, 44100, 0, 0, NULL);
```

Ця функція повинна бути викликана перед використанням будь-якого семплу, потоку або музичного файлу. В параметрах даної функції зазначаються пристрій, який буде використовуватися для прослуховування звуку, частота звуку, додаткові параметри відтворення такі, як кількість каналів, розширення і т.д.

Наступним кроком є створення потоку даних. Для цього було використано функцію:

```
BASS_StreamCreateFile(false, path, 0, 0, BASS_SAMPLE_FLOAT);
```

За допомогою даної функції, створюється потік даних. В параметрах вказуємо зчитуємо файл з оперативної пам'яті чи пристрою, шлях до необхідної нам композиції, точку з якої починаємо відтворювати файл і точку завершення програвання, а також параметри подачі даних.

Для запуску композиції використовується функція:

```
BASS_ChannelPlay(stream, false);
```

Тут вказується створений потік даних для відтворення та можливість програвання композиції з початку або з моменту зупинки.

Для того, щоб здійснити фреймування було використано функцію бібліотеки bass.dll

```
BASS_ChannelSetPosition(stream, set, BASS_POS_BYTE);
```

Першим параметром тут виступає вже вище створений потік даних, з яким потрібно працювати. Наступний параметр вказує інтервал поділу. І третій параметр зазначає одиницю вимірювання інтервалу, в даному випадку, фреймування відбувається по байтах. Отже, в загальному алгоритм поділу файлу на семпли має наступний вигляд:

```
int set=0;
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];
}
}
}
```

Стає зрозумілим, що параметр set вказує на точку початку вибраного інтервалу.

Як видно з попередньої частини коду, ще одна функція, що використовується BASS\_ChannelGetData. Дана функція призначення для отримання вибірки даних, що характеризують композицію. Цей набір даних в подальшому проходить перетворення Фур'є. Даний крок необхідний для побудови спектрограми та цифрового відбитку. Отже, функція BASS\_ChannelGetData має наступний синтаксис:

```
DWORD BASS_ChannelGetData(
    DWORD handle,
    void *buffer,
    DWORD length);
```

Перший параметр `handle` вказує на те, який потік даних використовується, тобто з якою композицією ми будемо працювати. В масив `buffer` функція, яка представлена вище запише отриману вибірку з якою ми в подальшому будемо працювати. Параметр `length` вказує довжину вибірки в байтах, що отримується чи на яку встановлюється прапорець-вказівник кількості семплів необхідних користувачеві, або на вибірку даних для кожного каналу зокрема, якщо це стерео канал, або ж для того, щоб переглянути кількість даних композиції.

Після того, як ми отримаємо таку вибірку, необхідно здійснити над нею швидке перетворення Фур'є. Для цього створиться функція FFT:

```
void FFT(double *AVal, double *FTv1, int Nv1, int Nft), де
```

`AVal` – масив даних, які аналізують,

`Nv1` – довжина масиву (повинна бути кратна 2),

`FTv1` – масив отриманих даних,

`Nft` – довжина масиву, що повинна бути рівна значенню `Nv1`

Під час отримання вибірки даних командою `BASS_ChannelGetData`, вказується прапорець `BASS_DATA_FFT256`, в результаті чого отримується 256 семплів, які пройшли швидке перетворення Фур'є, а масив `buffer` міститиме 128 значень, так як перетворення Фур'є складається з двох частин, а саме: реальної та уявної, що є однаковими за значенням, але різними за знаками. Тому отримані значення коливатимуться в межі від 0 до 1.

На рисунку 3.1 представлено результат використання даних функцій, тобто набір даних, які отримуються у вибірці.

Для того, щоб побудувати спектрограму необхідний таймер за допомогою якого кожна секунду передавалися значення, які б відображалися на графіку. Тому було використано клас `Timer`, а також компонент `PerformanceGraph` закладки `Sample`. Реалізація побудови спектрограми представлена в коді поданому нижче.

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

```
PerformanceGraph1->Update();
```

```
}
```

```
5.8348767595362E-8  
2.20395097727533E-8  
5.33134425495518E-8  
1.05873887434882E-7  
1.10314765322528E-7  
8.74115002602593E-8  
9.9462887703794E-8  
1.17582189318455E-7  
8.33516438092374E-8  
2.55665533188676E-8  
5.8361067090118E-9  
2.85997288074213E-8  
6.02953846851051E-8  
7.9779042039263E-8  
8.16350151922052E-8  
6.83758472064255E-8  
5.07250490500155E-8  
3.90323755539157E-8  
2.82368262105592E-8  
1.3681928479059E-8  
1.40109976953795E-8  
2.45608884341664E-8  
3.02569986843082E-8
```

Рисунок 3.1 – Дані вибірки після застосування швидкого перетворення Фур’є

Наступним кроком для побудови цифрового відбитку аудіо файлу є поділ отриманих даних на смуги. Для цього використали цикл for:

```
float step;  
for (int i=0; i<128; i++)  
{  
for (int j=0; j<128; j=j+4)  
{  
}}
```

В результаті отримується 33 смуги, що не перекриваються.

На наступному етапі необхідно визначити енергію сигналу. Для цього була використана формула 1.2. Спершу сумуються рядки, так як вони виступають константою для кожного субвідбитку. Таким чином позбулися зайвих виконань

команд, що дозволило пришвидшити час роботи програми в побудові цифрового відбитку.

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

Наступний крок у визначенні енергії сигналу є розрахунок кожного значення субвідбитку згідно формули 1.2. Код реалізації розрахунку чисельника в формулі 1.2 представлено нижче:

```
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;
```

Для виконуємо просте ділення, кожне значення субвідбитку ділимо на квадрат суми всіх значень 32-бітного субвідбитку, тобто на значення отримані в масиві sum:

```
band[i][j]=buffer[i][j*4]/sum[i];
```

Повторюється дана процедура для кожного субвідбитку, тобто 128 разів. Для цього використовується цикл представлений нижче:

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

В масиві `band` зберігається енергія для кожного біту субвідбитку. Дані значення необхідні для побільшої побудови бітової карти цифрового відбитку аудіо файлу.

Отже, спершу для того, щоб отримати цифровий відбиток потрібно визначити, яке значення має кожен біт окремого субвідбитку. Для цього використовується формула 1.6, яка реалізується в додатку. Якщо результат значення 0 або від'ємне значення, то біт фарбується у білий колір, якщо більше 0 – то у чорний.

```
bit=band[m][n]-band[m+1][n]-(band[m][n-1]-band[m+1][n-1]);
```

Щоб реалізувати перевірку кожного значення, використовується логічний оператор `if`. Алгоритм використання даного оператора представлено у наступному синтаксисі:

```
if (bit>0)
{BitMap[m][n]=1;
}
else
{BitMap[m][n]=0;
```

Для відображення цифрового відбитку, було використано компонент `DrawGrid` із закладки `Additional`. Для цього кожна клітинка даного компонента фарбується у відповідний колір.

```
DrawGrid1->Canvas->Brush->Color=clBlack;
DrawGrid1->Canvas->FillRect(DrawGrid1->CellRect(n,m));
DrawGrid1->Canvas->Brush->Color=clWhite;
DrawGrid1->Canvas->FillRect(DrawGrid1->CellRect(n,m));}
```

Для порівняння відбитків, виникла необхідність створення функції, яка б дозволяла їх зберігати. Для цього було використано команду `FileWrite`. Додатково також використався компонент `SaveDialog` закладки `Dialogs` для того, щоб користувач зміг обирати куди саме записуватиметься той чи інший цифровий відбиток аудіо файлу. Змінна `iFileHandle` - дескриптор файлу, в якій вказувати повний шлях або ім'я файлу.

Також передбачається випадок, якщо виникла помилка через яку не вдалося зберегти файл, програма сповістить користувача про це повідомленням вигляду MessageBox. Код реалізації даної функції представлений нижче.

```
int iFileHandle;
if (!SaveDialog1->Execute()) return;
try {
    iFileHandle = FileCreate(SaveDialog1->FileName);
    FileWrite(iFileHandle, BitMap, sizeof(BitMap));
}
catch(...) {
    Application->MessageBox("Під час запису сталася помилка",
        "Помилка запису",
        MB_ICONСТОП+MB_ОК);
}
FileClose(iFileHandle);
```

Для того, щоб можна було порівняти цифрові відбитки композиції, потрібна була можливість завантажувати раніше уже створені відбитки. Для цього створена функція, яка схожа з функцією збереження. Проте у даному випадку використовується компонента OpenFileDialog закладки Dialogs та команда FileRead для зчитування даних про існуючий цифровий аудіо відбиток. Аналогічно існує змінна в межах цієї функції iFileHandle де зберігається повний шлях або ім'я файлу.

Крім цього, тут також передбачений випадок виникнення помилки у випадку, коли не вдалося відкрити файл, що містить інформацію про цифровий аудіовідбиток. Програма про це також сповістить користувача відповідним повідомленням. Реалізація коду функції завантаження цифрового відбитку:

```
int iFileHandle;
if (!OpenDialog2->Execute()) return;
try {
    iFileHandle = FileOpen(OpenDialog2->FileName, fmOpenRead);
    FileSeek(iFileHandle,0,0);
```

```

FileRead(iFileHandle, BitMap2, sizeof(BitMap2));
}
catch(...) {
Application->MessageBox("Під час відкриття сталася помилка",
    "Помилка відкриття",,
    MB_ICONСТОП+MB_OK);
}
FileClose(iFileHandle);

```

Отже, отримані даною програмою цифрові відбитки потрібно порівняти для визначення їхньої унікальності. Для цього створюється функція порівняння, яка співставляє по три біти поспіль. Якщо порівнювати конкретно один біт, тобто одну точку на бітовій мапі цифрового відбитку, то буде надто велике співпадіння, так як один біт не дає реальної картини ідентифікації цифрового відбитку.

Тому, якщо три біти поспіль співпадають, то тимчасова змінна збільшується на одиничку. Всього кількість таких точок є 4096.

В результаті, отримане значення ділиться на кількість точок у відбитку, тобто на 4096 і множиться на 100%. Таким чином отримується відсоток схожості двох композицій. Користувачеві ця інформація надається програмою через повідомлення типу MessageBox.

```

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("Композиції схожі на: "+FloatToStr(b)+"%.");

```

Окрім основних функцій, які потрібні для роботи з цифровими аудіо відбитками, також є функції для завантаження аудіо файлу, кнопками керування відтворення, зупинки композиції, тощо. Код реалізації всієї програми представлено у додатку Б.

### 3.3 Тестування розробленої програми

Результатом даної дипломної роботи є програма, котра створює цифрові відбитки для захисту аудіофайлів будь-якого розширення. На основі отриманих вибірок даних розроблений додаток формує двовимірну матрицю, котра описує унікальний відбиток для кожної окремої композиції.

При запуску додатку перед користувачем відкривається вікно програми, котре зображене на рисунку 3.2. Інтерфейс головного вікна є типовим, як і для більшості програм, що робить програмою зрозумілої та легкою до освоєння. Перше, на що звернемо увагу – кнопки керування вікном. Вони дозволяють згорнути програму в панель задач. Розкрити на весь екран чи зменшити до попередніх розмірів та завершити роботу програми і закрити вікно. Умовно інтерфейс програми можна розділити на блоки, що слідуєть один за одним, а саме: головне меню, поле для відображення назви аудіофайлу та секундоміра, який відображає момент програвання композиції. Далі розміщений блок для

відображення осцилограми, що показує графік програвання звуку, тобто залежність між фізичними величинами, що швидко змінюються. Нижче розташовані кнопки для керування процесами над аудіо файлами такими, як запуск, пауза та стоп.

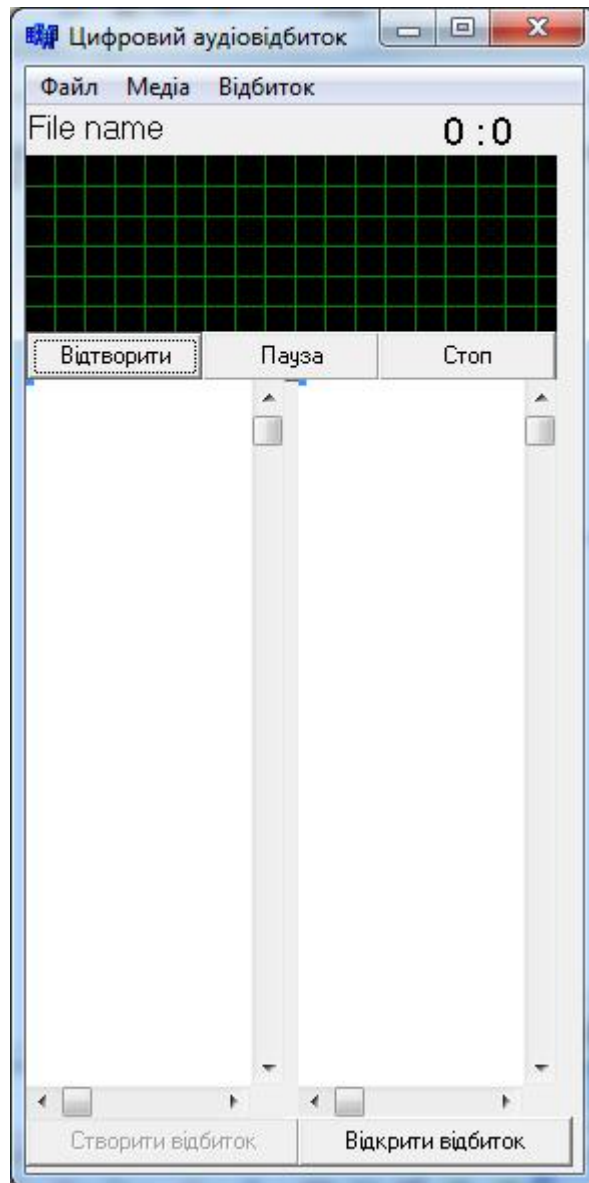


Рисунок 3.2 – Головне вікно робочої програми

Наступних два блоки, найбільші за розміром призначені для відображення цифрових відбитків. В першому зображується сформований цифровий відбиток для композиції, що відтворюється в даний момент часу, а в другому завантажується уже раніше створений цифровий відбиток для порівняння. Нижче пари цих блоків розміщені кнопки для керування процесами над аудіо

відбитками, що найчастіше використовуються, а це створення відбитку та завантаження раніше сформованого. На цьому інтерфейс головного вікна завершується.

Головне меню програми, що зображено на рисунку 3.3 містить команди групи «Файл», «Медіа» та «Відбиток». Пункт «Файл» надає користувачеві можливість завантажити необхідний йому аудіофайл для подальшої роботи з ним, а саме для формування цифрового відбитку і порівняння з вже існуючими. Також цей пункт меню надає можливість завершення роботи з додатком.

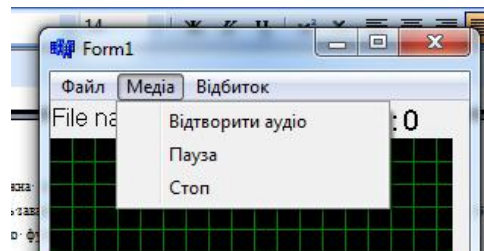


Рисунок 3.3 – Головне меню програми

Для зручності завантаження необхідного аудіофайлу чи цифрового відбитку використано діалогове вікно. Воно дозволяє користувачеві легко та швидко знайти шлях до бажаної композиції чи відбитку та обрати відповідний файл.

Пункт меню «Медіа» надає можливість керування відтворення композиції. Як видно на рисунку 3.3, цей пункт містить команди «Відтворити аудіо», «Пауза» та «Стоп». Для зручності такі ж команди виведені на головне вікно програми.

Третій пункт головного меню «Відбиток», як стає зрозуміло з назви, призначений для роботи з цифровими відбитками. Перший пункт меню це команда «Зберегти відбиток». Дана команда використовується для збереження вже сформованого відбитку. Збереження інформації про цифровий відбиток відбувається шляхом запису даних у файл формату txt. Це необхідно для того, щоб в подальшому користувач зміг швидко та легко порівняти вже існуючі відбитки для визначення не ліцензованих копій та збереженню авторських прав.

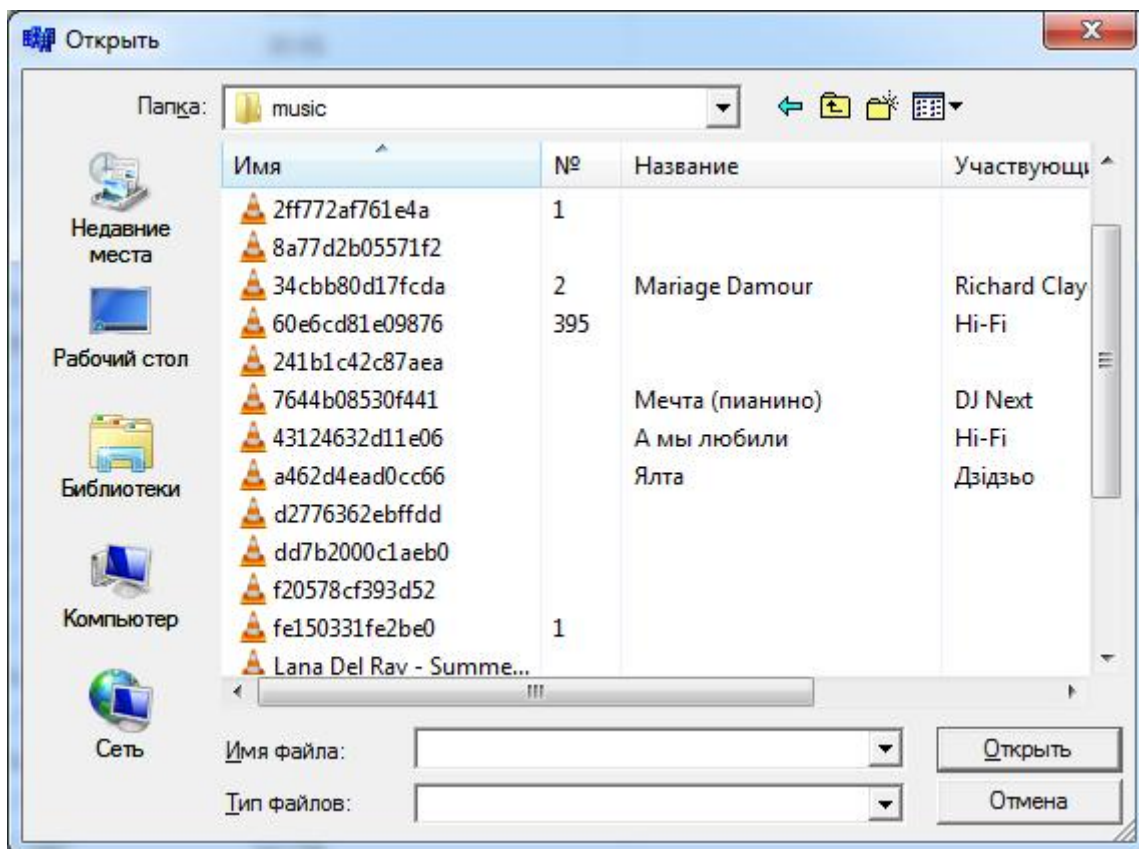


Рисунок 3.4 – Діалогове вікно завантаження файлу

Наступний пункт меню «Відбиток» це команда «Відкрити відбиток». Дана команда призначена для завантаження уже раніше створених цифрових аудіо відбитків з метою їх порівняння для визначення унікальності заданої композиції.

Останні, третій пункт «Порівняти відбиток» використовується безпосередньо для запуску процесу порівняння цифрових аудіо відбитків. Дана команда порівнює між собою уже створений відбиток з тими, що були раніше збережені. Обидва цифрові відбитки є завантажені у відповідні блоки для роботи з ними.

Нижче головного меню розташована стрічка «Інформаційний блок» для виведення назви композиції, з якою працюють в даний момент і таймером, котрий відображає точку відтворення аудіо файлу. Даний блок представлено на рисунку 3.5.

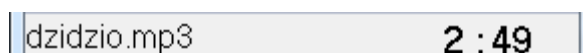


Рисунок 3.5 – Інформаційний блок

Наступний елемент додатку призначений для відображення осцилограми, котра описує аудіо файл, що відтворюється в конкретний момент часу. Даний блок використовується для графічного відображення сигналу, що подається на вхід програми.

Традиційно при цифровому звукозаписі аудіодоріжка представляється у вигляді осцилограми, що відображає форму звукової хвилі, тобто залежність амплітуди звуку від часу. Осцилограма дозволяє побачити основні події, які відбуваються під час відтворення звуку, а саме такі як зміни гучності, паузи між частинами твору і часто навіть окремі ноти в сольному записі інструменту.

На основі цих даних формується унікальний опис аудіо файлу. Це необхідно для створення цифрового аудіо відбитку.

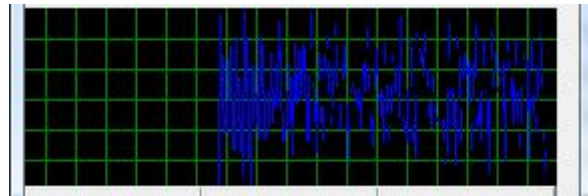


Рисунок 3.6 – Блок відображення осцилограми

Отже, наступний блок, що розташований за графічною частиною для представлення осцилограми є блок для керування відтворення композиції, що зображений на рисунку 3.7. Він містить 3 кнопки. Кнопка «Відтворити» призначена для старту відтворення аудіо файлу. Кнопка «Пауза» використовується для призупинення програвання композиції та кнопка «Стоп» для повної зупинки відтворюваного аудіо файлу.

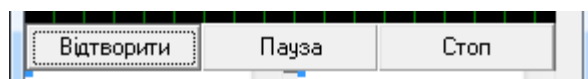


Рисунок 3.7 – Блок керування програвання аудіо файлу

На рисунку 3.8 зображено блок представлення цифрових відбитків. З назви стає зрозуміло, що в даній частині вікна відображаються цифрові відбитки, що

генеруються в програмі або були уже раніше сформовані. Цифровий відбиток представляє собою набір бітів, 32 біти в рядку, 256 рядків в одному відбитку.



Рисунок 3.8 – Блок представлення цифрових аудіовідбитків

Як уже було відзначено вище, звуковий відбиток (сигнатура) являє собою короткий опис аудіофайлу, що враховує його звукові особливості. Причому повинні враховуватися такі особливості, які є не тільки стійкими до різних модифікацій звукового файлу (стиску, аналого-цифрового і цифро-аналоговому перетворенню, фільтрування і т.п.), але і мають відношення до його сприйняття.

Цифровий відбиток відображається хеш-таблицею, в якій роль ключа виконує значення частоти для піку інтенсивності і для його опорної точки, навпроти яких ставиться час в секундах від початку треку.

В роботі для формування відбитків звукового файлу використовуються добре відомі характеристики звуку, стійкі до спотворень, такі як коефіцієнти Фур'є, коефіцієнти косинусного перетворення Фур'є (MFCC), нерівномірність спектра (spectral flatness).

Дві кнопки що представлені на рисунку 3.9 призначені для пришвидшення та полегшення роботи з цифровими аудіо відбитками. Перша кнопка «Створити відбиток» використовується для того, щоб згенерувати цифровий відбиток для

аудіо файлу, що відтворюється в даний момент часу. З моменту запуску відтворення аудіокомпозиції, програма передає набір даних, так звані вибірки для кожного окремого моменту на основі яких в подальшому і формується бітова карта, що описує цифровий відбиток. Натиснувши цю кнопку, користувач подасть сигнал для початку генерування цифрового аудіовідбитку, котрий з'явиться у лівій частині блоку, що зображений на рисунку 3.8.

Кнопка «Відкрити відбиток» призначена для того, щоб завантажити вже раніше створений відбиток у праву частину блоку для відображення цифрових відбитків. Це необхідно для того, щоб було можливо порівняти отримані відбитки та перевірити, таким чином, їх на унікальність, щоб забезпечити авторські права та надійний захист від нелегального використання та розповсюдження аудіо композицій.

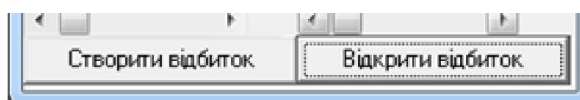


Рисунок 3.9 – Кнопки для керування аудіо відбитками

Для того, щоб користувач міг представити свою аудіо композицію в інтернеті, вона спершу має пройти тестову перевірку на визначення авторських прав. Якщо згенерований відбиток аудіо файлу виявиться надто схожим до вже існуючого, то користувач не матиме права розповсюджувати дану композицію. Інформацію про схожість композиції програма видасть у повідомленні, вивівши результат схожості аудіо файлу. Таке повідомлення представлено на рисунку 3.10.

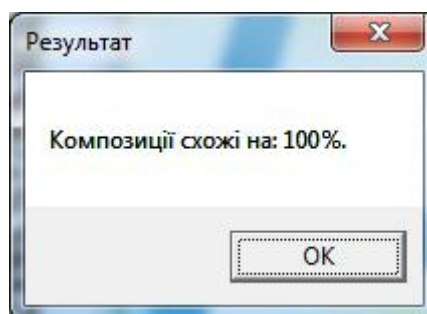


Рисунок 3.10 – Інформаційне повідомлення

Основним призначенням цифрових аудіо відбитків, як писалося раніше, є захист аудіо композицій від виникнення піратських копій, і таким чином від нелегального розповсюдження. Тому результат розробки даної програми є генерування цифрових аудіо відбитків на основі отриманих вибірок даних, що описують кожну композицію зокрема.

В результаті виконання дипломної роботи можна зробити наступний висновок: запропонований метод захисту аудіофайлів від несанкціонованого копіювання та розповсюдження цілком може бути використаний для поставленого завдання. Але неможливо вмонтувати ідентифікаційну мітку в усі копії файлів. При використанні такого способу відстеження аудіофайлу по його цифровому відбитку можна відстежити як завгодно модифіковані копії файлу-оригіналу. Особливість використання цифрових відбитків для ідентифікації файлів полягає в тому, що рішення про подібність або відмінність аудіофайлів видається з деякою часткою ймовірності, тобто результат розпізнавання файлів оцінюється двома помилками: помилкою першого роду (файл не знайдено при його наявності) і помилкою другого роду (помилково знайдений інший файл). Помилка другого роду у методі спектрограм приблизно в 2,5 рази вище, ніж у методі покадрового обчислення відбитків. Для поліпшення імовірнісних показників виявлення передбачається введення деякої надмірності, що може збільшити час пошуку.

### 3.4 Властивості підсистеми пошуку та ідентифікації аудіоматеріалів на основі цифрових аудіовідбитків.

#### 3.4.1 Визначення порогу коефіцієнта бітових відмінностей при порівнянні аудіо відбитків

Для порівняння двох аудіофрагментів відповідної пари відбитків обчислювався коефіцієнт бітових відмінностей BER. Фрагменти збігаються, якщо коефіцієнт бітових відмінностей BER між відповідними відбитками знаходиться



нижче певного порогового рівня. На рисунку 3.11 приведена залежність коефіцієнта бітових відмінностей від відносного зсуву при порівнянні відбитка музичного твору з вихідним неспотвореним відбитком (рисунок 3.11А), з відбитком цього ж сигналу в присутності шуму ( $SNR = 0\text{db}$ , рисунок 3.11Б) і з відбитком сигналу, стисненого в часі зі збереженням тональності (рисунок 3.11В).

Видно, що коефіцієнт бітових відмінностей між незбіжними відбитками знаходиться в діапазоні 0,45-0,55, в той час як при збігу взаємного положення порівнюваних відбитків його величина менше 0,40.

Згідно з експериментальними даними, для надійної ідентифікації різних послідовностей відбитків пороговий рівень коефіцієнта бітових відмінностей повинен складати  $BER = 0,37$ . В цьому випадку гарантується повна відсутність помилок навіть у випадку коротких послідовностей.

У практичній реалізації методу, використовується динамічно обчислений поріг бітових відмінностей, величина якого залежить від довжини послідовності, що збігається. Таким чином, алгоритм дозволяє ідентифікувати короткі слабо спотворені послідовності тривалістю 5 секунд з порогом  $BER = 0,37$ .

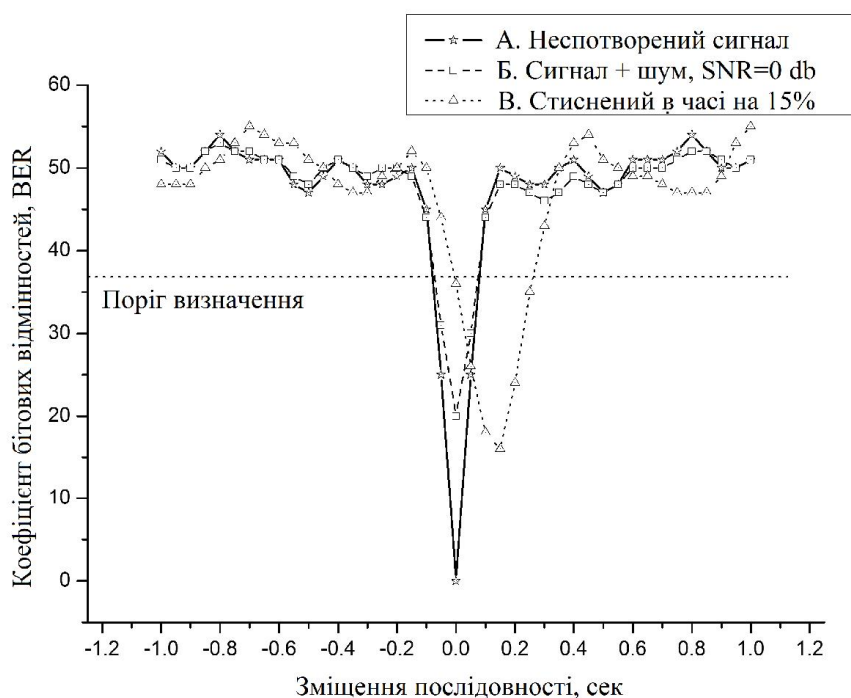


Рисунок 3.11 – Залежність коефіцієнта бітових відмінностей від зсуву при порівнянні музичного відбитка з вихідною послідовністю (А), в присутності шуму  $SNR = 0\text{db}$  (Б) і масштабування в часі на 15% (В)

У той самий час для більш тривалих співпадаючих послідовностей, поріг визначення піднімається до  $BER = 0,40$ , що дозволяє ідентифікувати сигнали в присутності значного шуму без зменшення надійності методу.

#### 3.4.2 Стійкість відбитків до масштабування

Чутливість існуючих методів цифрових відбитків до масштабування в часі (зі змінами швидкості програвання) є істотним недоліком, що обмежує застосування цих методів для ідентифікації аудіоматеріалів, що транслюються по радіоканалам. Компанії, що надають і транслюють матеріали по кабельних мережах і радіоканалах, часто навмисно модифікують швидкість програвання матеріалів для того, щоб більш ефективно використовувати ефірний час і більш точно слідувати програмі передач. Зміни швидкості програвання також відбувається при перетворенні аудіоматеріалів з одного цифрового формату в інший. Зокрема, при перетворенні телевізійного сигналу зі стандарту PAL в NTSC відбувається за рахунок уповільнення частоти кадрів з 25 до 23.976 кадрів в секунду з подальшим перетворенням групи з 4-х кадрів в групу, що складається з 5-ти кадрів (3: 2 pulldown). В результаті перетворення швидкість відтворення матеріалів зменшується на 4%, що практично непомітно для людини, але викликає певні труднощі для існуючих алгоритмів ідентифікації.

На рисунку 3.12 приведена залежність коефіцієнта бітових відмінностей запропонованого методу від ступеня стиснення/розтягування сигналу в часі. У порівнянні з вихідним методом, представленим раніше на рисунку 2.13А, використання перетворення Мелліна дозволило збільшити стійкість алгоритму до масштабування сигналу в часі в 2,5 рази, що для порогу ідентифікації  $BER = 0,37$  відповідає максимально можливому прискоренню/уповільненню матеріалів, що дорівнює 40%.

Слід зазначити, що в реальних додатках, ступінь масштабування рідко перевищує 15%, проте для надійної ідентифікації матеріалів, що піддаються комплексним спотворенням потрібний істотний запас продуктивності.

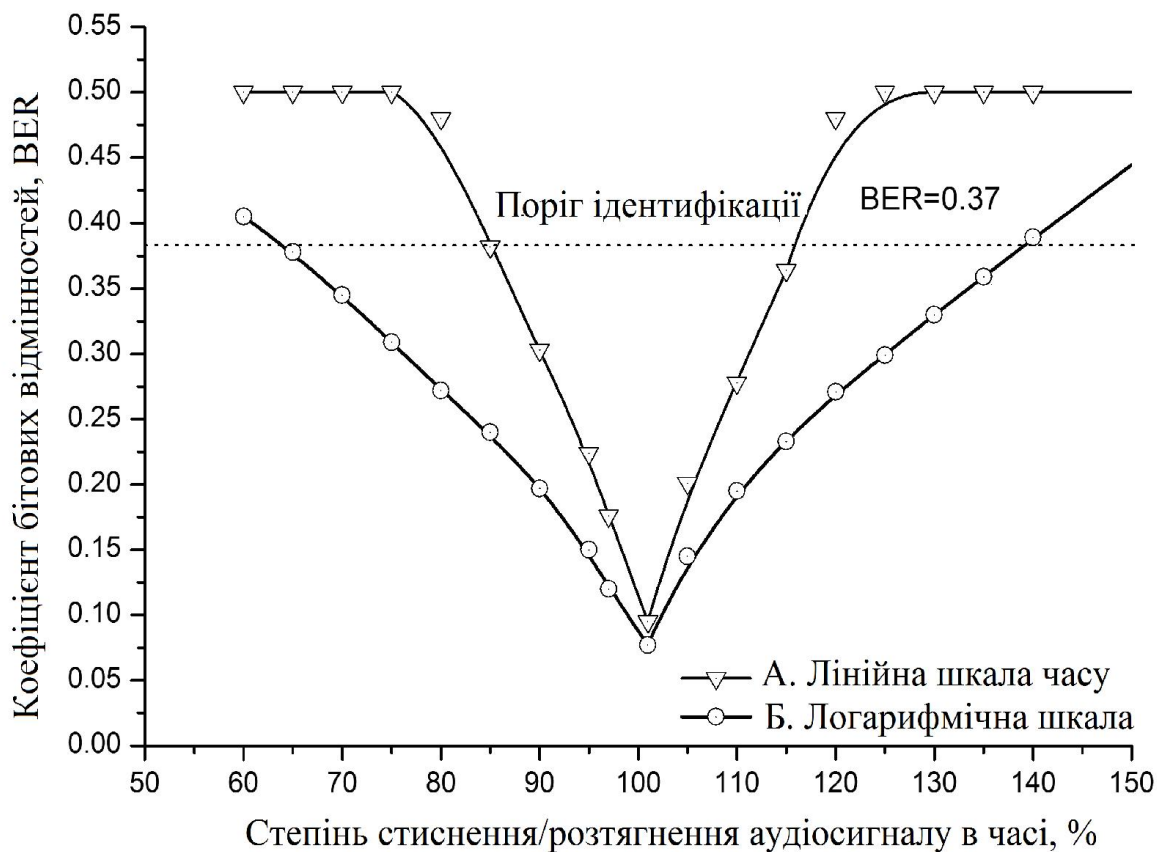


Рисунок 3.12 – Залежність коефіцієнта бітових відмінностей від ступеня масштабування сигналу в часі для відбитків, отриманих на підставі аналізу спектра модуляції в лінійній і логарифмічній шкалі

Слід також зазначити, що навмисне внесення тимчасових спотворень є одним з варіантів атаки на системи ідентифікації аудіосигналів [47]. Підвищення стійкості до такого роду атак має практичне значення для систем пошуку та ідентифікації несанкціонованих копій аудіо- і відеоматеріалів в мережі Інтернет і не тільки.

Функціональні характеристики приймача (Receiver Operation Characteristics, ROC) для двох зазначених методів у присутності помірного шуму (співвідношення сигнал/шум 12 db) наведені на рисунку 3.13. Функціональна характеристика приймача являє собою залежність рівня коректних результатів ідентифікації сигналу (True Positive Rate, TPR) від рівня помилок (False Positive Rate, FPR) при різних порогових значеннях коефіцієнта бітових спотворень BER.

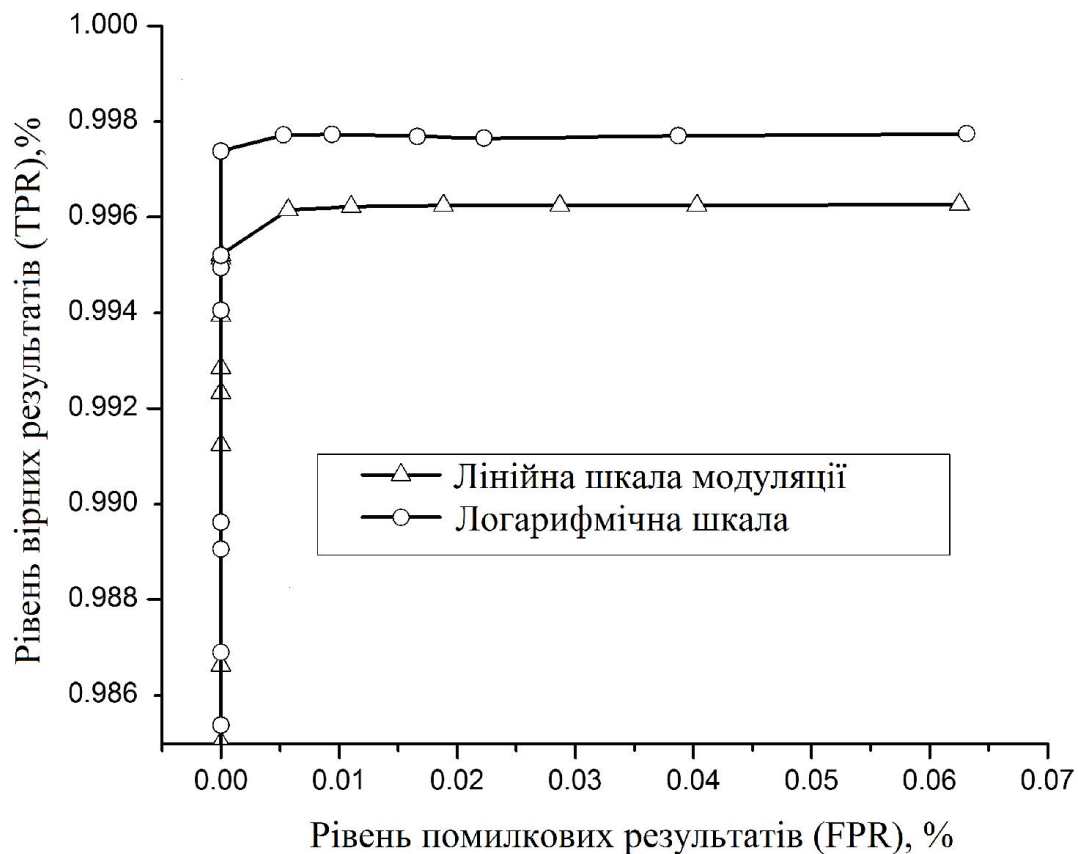


Рисунок 3.13 – Функціональні характеристики приймача аудіосигналу (SNR = 12db) для відбитків, отриманих на підставі аналізу спектра модуляції в лінійній і логарифмічній шкалі

Таким чином, аудіовідбитки, отримані за допомогою аналізу спектру модуляції в логарифмічній шкалі мають у 2 рази менший рівень помилок ніж відбитки, отримані в лінійній шкалі (0.002% і 0.004% відповідно).

### 3.4.3 Стійкість відбитків до зсуву

Стійкість алгоритму до зсуву в значній мірі визначається ступенем перекриття вікон локального перетворення Фур'є в процесі обчислення спектрограми і частоти генерації вектора ознак, що зберігаються в вихідному потоці. Експериментально було встановлено, що 80% перекриття вікон спектрограми цілком достатньо для забезпечення стійкості відбитків до зсуву. Збільшення частоти проходження елементів цифрового відбитку також дозволяє значно зменшити вплив неузгодженості положень послідовностей, що порівнюються, проте збільшує розмір відбитка і відповідно погіршує

продуктивність системи пошуку. У більшості випадків достатньо, щоб частота генерації елементів становила 20 субвідбітків в секунду.

На рисунку 3.14 представлений графік залежності коефіцієнта бітових відмінностей від ступеня неузгодженості (зсуву) акустичних сигналів. Видно, що девіація коефіцієнта бітових відмінностей становить  $BER = 0.21$  для відбітків, отриманих в лінійній шкалі часу і  $BER = 0.12$  для відбітків, отриманих в логарифмічній шкалі, використання якої збільшило стійкість алгоритму до зсуву на 60%.

Слід зазначити, що вплив зсуву на рівень помилки значно зменшується при наявності шуму і спотворень, які маскують ефект десинхронізації відбітків. Разом з тим, для ряду додатків, зокрема для задач визначення якості аудіосигналів по величині коефіцієнта бітових відмінностей в порівнянні з еталонним відбітком, рівень похибки, що вноситься ефектом неузгодженості може бути досить високий, особливо для якісних сигналів (MP3 64 kbps і вище).

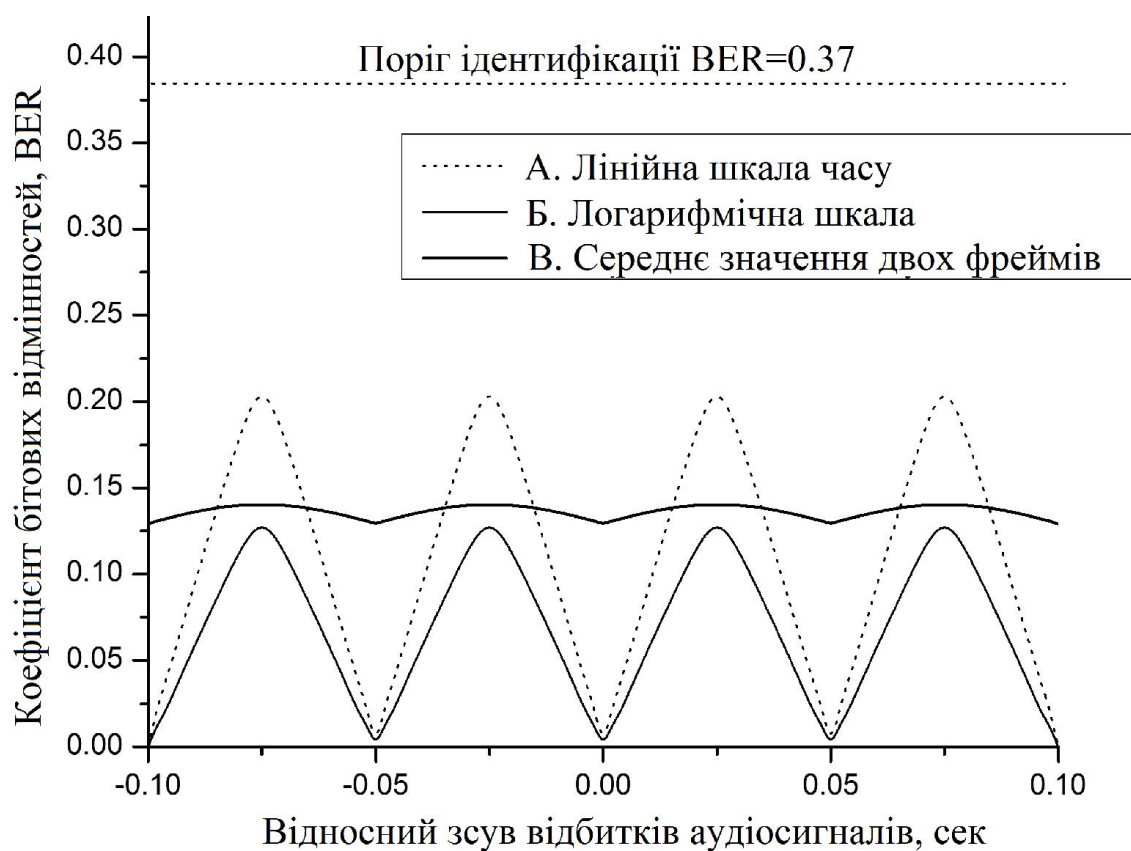


Рисунок 3.14 – Залежність коефіцієнта бітових відмінностей від відносного зсуву цифрових відбітків в часі

Для зменшення впливу зсуву може використовуватися метод обчислення коефіцієнта бітових відмінностей, усередненого по двох сусідніх елементах відбитку. В цьому випадку варіації коефіцієнта бітових відмінностей становлять не більше 5% (рисунок 4.9В), що дозволяє використовувати даний метод для контролю наявності сигналу і рівня спотворень аудіосигналів в каналі передачі. Продуктивність методу аудіовідбитків дозволяє контролювати якість передачі звукової інформації в радіотелефонних, кабельних і супутникових каналах зв'язку в реальному масштабі часу.

#### 3.4.4 Аналіз використання аудіовідбитків для ідентифікації змішаних сигналів

В деяких випадках виникає необхідність ідентифікації змішаних сигналів, що складається з декількох аудіокомпонентів. Змішання сигналів може статися, наприклад, при аналоговій передачі сигналів за рахунок взаємного впливу двох близько розташованих за частотою каналів, а також в разі мікшування аудіосигналів в студії (дикторський текст на тлі музики).

Окремою проблемою є ідентифікація монофонічних записів стерео і багатоканальних музичних творів, відео- і кінофільмів. У той час як при побудові монофонічного цифрового відбитку, сигнали різних каналів підсумовуються з урахуванням їх ваг, при монофонічному записі багатоканальних акустичних сигналів, наприклад, представлених в стандартах DTS (використовується в кінотеатрах) і Dolby (звуковідтворення використовується в побутовій техніці), домінуючим каналом є той, на який спрямована відеокамера. В силу вузької діаграми спрямованості мікрофона камери зазвичай домінує центральний канал, призначений для відтворення діалогів, а внесок бічних каналів значно менше.

На рисунку 3.15 представлена залежність коефіцієнта бітових відмінностей між двома сигналами і їх сумішшю від співвідношення амплітуд компонентів в суміші. Видно, що для обраного порогу  $BER = 0,37$  надійна ідентифікація сигналів у суміші гарантується при співвідношенні їх рівнів в діапазоні від -6 db до +6 db.

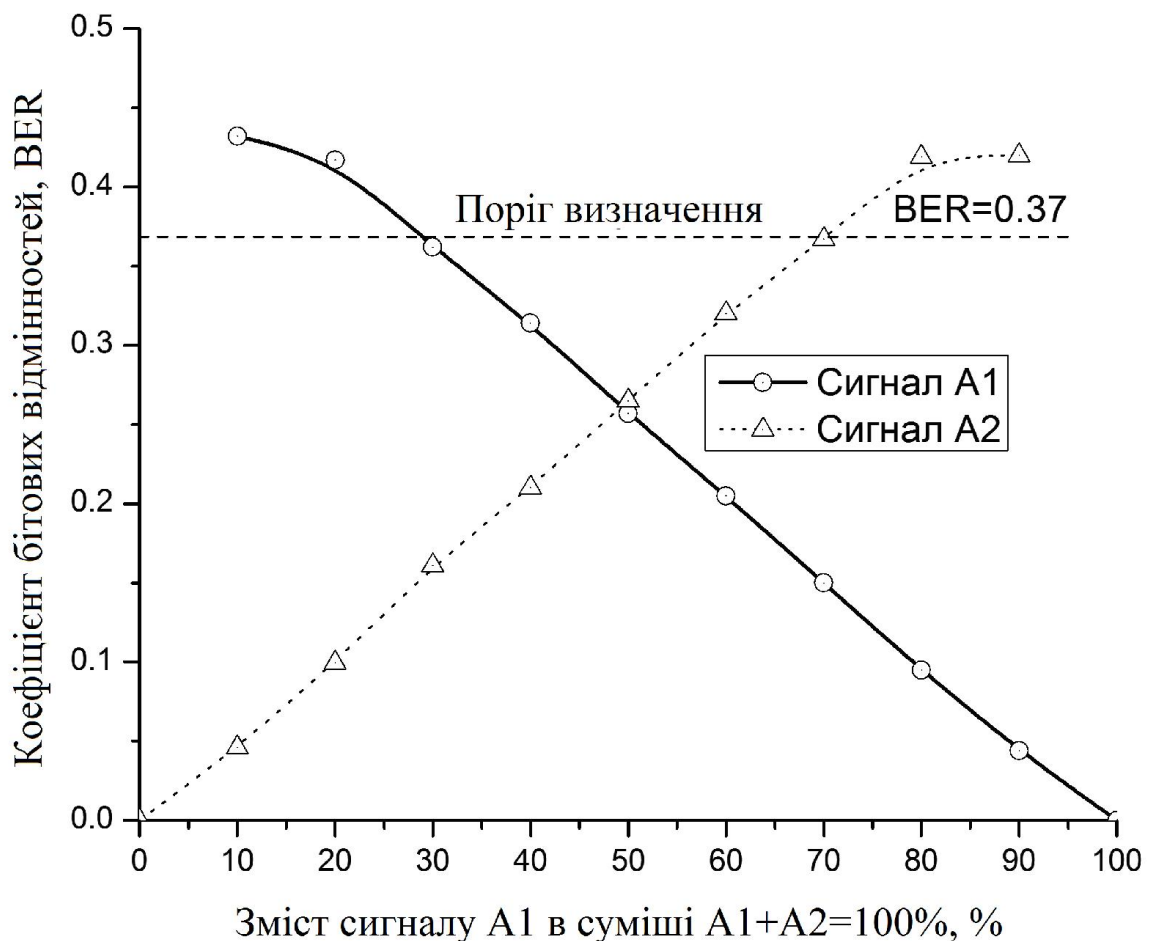


Рисунок 3.15 – Ідентифікація компонентів в двох змішаних аудіосигналах

Таким чином, ідентифікація змішаних сигналів забезпечується навіть у разі присутності перешкод, рівень яких може перевищувати рівень основного сигналу в два рази.

Висновки до розділу 3.

1. Експериментально підтверджено, що використання метамерів для формування зображень призводить до спотворень колірного балансу в копіях, що дозволяє використовувати запропонований метод для захисту аналогових зображень від несанкціонованого копіювання. При цьому ступінь колірних спотворень може досягати значення CIE Lab 22 для тілесних і 45 для нейтральних сірих кольорів.

2. Експериментально встановлено, що цифрові відбитки, засновані на аналізі спектра модуляції в логарифмічній шкалі стійкі до масштабування сигналу

в часі в діапазоні 40%. Це дозволило практично виключити вплив швидкості відтворення на точність розпізнавання сигналів, що піддаються тимчасовим спотворенням, зокрема, при перетворенні фільмів, аналогових записів і мультимедійних файлів з одного формату в інший.

3. Показано, що запронована система пошуку та ідентифікації аудіоматеріалів може бути застосована для ідентифікації аудіосигналів, що підлягали різним спотворенням і перетворенням (зміни частоти вибірки, фільтрація, цифрове і динамічне стиснення, зміна швидкості програвання і ін.). Зокрема, розроблена система може бути використана для контролю трафіку файлообмінних P2P мереж і поширення матеріалів в Інтернеті, моніторингу радіо- і телепередач, оцінки якості звучання, виявлення несанкціонованих змін і аутентифікації аудіоматеріалів.



## ВИСНОВКИ

У дипломній 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.

1. 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.

2. 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 10-30%.

3. Удосконалено метод ідентифікації і пошуку аудіоматеріалів з використанням цифрових аудіовідбитків, який відрізняється від існуючих представленням спектру модуляції в нелінійній шкалі часу, стійких до шуму, масштабування в часі, частотних спотворень, цифрового стиснення і фільтрації

4. Проведено комп'ютерне моделювання та експерименти по визначенню ефективності і надійності алгоритму ідентифікації акустичних сигналів за допомогою розробленого методу цифрових відбитків. Показано, що метод забезпечує надійну ідентифікацію аудіоматеріалів.

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Michael Fink Social- and Interactive-Television Applications Based on Real-Time Ambient-Audio Identification: Fink M., Covel M., Baluja S., Proc. EuroITV'06 Conf., Israel, Center for Neural Computation, Hebrew University of Jerusalem, 2006.
2. Jaap Haitsma, Antonius Kalker, "A Highly Robust Audio Fingerprinting System", International Symposium on Music Information Retrieval (ISMIR) 2002, pp. 107-115.
3. Content and Control: Assessing the Impact of Policy Choices on Potential Online Business Models in the Music and Film Industries. [Електронний ресурс] / John Palfrey, Derek Bambauer, Urs Gasser, Derek Slater, Meg Smith Режим доступу: [http://cyber.law.harvard.edu/media/files/content\\_control.pdf](http://cyber.law.harvard.edu/media/files/content_control.pdf).
4. Cox I.J., The first 50 years of electronic watermarking. Cox I.J., Miller M.L., Journal of Applied Signal Processing, V.2. - 2002, P 126-132.
5. Cano P., Audio Fingerprinting: Concepts and Applications. / Cano P., Gómez E., Batlle E., Gomes L., Bonnet M. // Proceedings of 2002 International Conference on Fuzzy Systems Knowledge Discovery. – Singapore. – 2002.
6. Gomez E. Mixed Watermarking-Fingerprinting Approach for Integrity Verification of Audio Recordings. / Cano P., Gómez E., Batlle E., Gomes L., Bonnet M. // Proceedings of the International Telecommunications Symposium, Natal (Brazil). - 2002.
7. Ананьев А.Б. Спектральное сопоставление музыкальных произведений / Ананьев А.Б., Просвиров Д.В. // Акустичний вісник. – 2004. - Т.7, №3.- С.7-13
8. Rabiner L. R. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition // Proc. of the IEEE. – 1989.- Vol.77, No.2.- P. 257-286 .
9. YouTube / [Електронний ресурс] // Режим доступу: <http://www.youtube.com>.
10. Philips. Audio Fingerprinting technology / [Електронний ресурс] // Режим доступу: <http://www.research.philips.com/initiatives/contentid/audiofp.html>.

11. Microsoft. Communication, Collaboration, and Signal Processing / [Электронный ресурс] // Режим доступа: <http://research.microsoft.com/research/ccsp/>.
12. Google / [Электронный ресурс] // Режим доступа: <http://www.google.com>.
13. Secure Digital Music Initiative / [Электронный ресурс] // Режим доступа: <http://www.sdmi.org>.
14. Battle E. Scalability issues in an HMM-based audio fingerprinting / Battle E., Masip J., Gaus E., Cano P. // International Conference on Multimedia Computing and Systems.- 2004.- Vol. 1. – P. 735-738.
15. Lourens J. Detection and Logging Advertisements Using Its Sound // Proc. COMSIG.- Johannesburg (SAR).- 1990.
16. Kurth F. Identification of Highly Distorted Audio Material for Querying Large Scale Databases / Kurth F., Ribbrock A., Clausen M. // Proc. AES 112th Int. – Munich (Germany). – 2002.
17. Battle E. Feature Decorrelation Methods in Speech Recognition. A Comparative Study / Battle E., Nadeu C., Fonollosa J. // Proc. of International Conference on Speech and Language Processing. – 1998. - Sydney (Australia).- P. 951-954.
18. Lu L. A robust audio classification and segmentation method / Lu L., Jiang H., Zhang H., Proc. ACM Multimedia (MM'01).– Ottawa (Canada). - 2001. P. 203–211.
19. Battle E. Automatic song identification in noisy broadcast audio / Battle E., Masip J., Gaus E. // Proc. of the SIP. - 2002.
20. Morishima M. Phonetically adaptive cepstrum mean normalization for acoustic mismatch compensation / Morishima M., Isobe T., Takahashi J. // Proc. Automatic Speech Recognition and Understanding. – Santa Barbara (USA). - 1997.- P. 436-441.
21. Haitsma J. Robust audio hashing for content identification / Haitsma J., Kalker T., Oostveen J. // Proc. of the Content-Based Multimedia Indexing - Firenze (Italy) .- 2001.

22. Cano P. Robust sound modeling for song detection in broadcast audio / Cano P., Batlle E., Mayer H., Neuschmied H. //Proc. AES 112th Int. Conv.- Munich (Germany). – 2002.
23. Zhang T., Hierarchical classification of audio data for archiving and retrieving // Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing.– 1999. – Phoenix (USA). - Vol. 6, pp. 3001–3004.
24. Жирков А.О. Графический метод представления и нейросетевое распознавание частотно-временных векторов речевой информации / Жирков А.О., Корчагин Д.Н., Лукин А.С., Крылов А.С., Баяковский Ю.М. // Программирование.- 2003.- №4.- С.41-52.
25. Allamanche E. Content-Based Identification of Audio Material Using MPEG-7 Low Level Description / Allamanche E., Herre J., Helmuth O., Fröba B., Kasten T. // Proc. International Symposium of Music Information Retrieval. - Indiana (USA). - 2002.
26. Cano P. A review of algorithms for audio fingerprinting. / Cano P., Batlle E., T. Kalker, Haitsma J. //Proc. IEEE Workshop on Multimedia Signal Processing. - 2002. - P. 169-173.
27. Neuschmied H. Content-based identification of audio titles on the internet / Neuschmied H., Mayer H., Batlle E. // Proc. International Conference on Web Delivering of Music.- 2001.
28. Haitsma J. Highly Robust Audio Fingerprinting System// Proc. of the 3rd Int. Symposium on Music Information Retrieval. – 2002. - P. 144-148.
29. Sd Jin Soo Seo. Linear speed-change resilient audio fingerprinting // Sd Jin Soo Seo, Haitsma J., Kalker T., Proc. IEEE Benelux Workshop on Model based Processing and Coding of Audio. – Leuven (Belgium). – 2002.
30. Zwicker E., Fastl H. Psychoacoustics: Facts and Models / Springer Verlag, 2nd ed. - 1999.
31. Plomp, R., & Levelt, W. J. M. Tonal consonance and critical bandwidth // Journal of the Acoustical Society of America. – 1965. - Vol. 37, P.548-560.
32. Dudley H. Remaking speech // Journal of the Acoustical Society of America. – 1939. – Vol. 11. No 2. – P.169-177.

33. Drullman R. Effect of temporal envelope smearing on speech reception / Drullman R., Festen J.M., Plomp R. // JASA. – 1994. – Vol. 95. No.2. - P.1053-1064.
34. H. J. M. Steeneken. A physical method for measuring speech transmission quality. H. J. M. Steeneken and T. Houtgast. Journal of the Acoustical Society of America, 67(1):318-326, January 1980.
35. 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. 117, - 2005, P. 2181-2192.
36. T. Chi. "Spectrotemporal modulation transfer functions and speech intelligibility," / T. Chi, Y. Gao, M. Guyton, P. Ru, and S. Shamma // Journal of Acoustical Society of America, vol. 106, no. 5, pp. 2719–2732, 1999.
37. S. Sheft "Temporal integration in amplitude modulation detection,"/ S. Sheft and W. Yost // Journal of Acoustical Society of America, vol. 88, - 1990. pp. 796–805.
38. Marrakchi-Mezghani I. Robustness of audio fingerprinting systems for connected audio applications / Marrakchi-Mezghani I., Turki-Hadj Alouane M., Jaidane-Saidane M. // Proc. Second International Symposium on Communications, Control and Signal Processing.- 2006.
39. Greenberg S. The Modulation Spectrogram: in Pursuit of an Invariant Representation of Speech / Greenberg S., Kingsbury B. // ICASSP.- 1997.- P. 1647-1650.
40. С.В.Билобров. Метод ідентифікації аудіоматеріалів на основі аналізу спектра модуляції сигналу // Вісник Донецького університету, Сер.А: Природничі науки.-2005.-Вип.2.- С.387-391.
41. Altes R.A. The Fourier-Mellin transform and mammalian hearing // J. Acous Am. – 1978.- Vol.63. No. 1. - P.174-183.
42. Cohen L. Instantaneous Scale and the Short-Time Scale Transform // IEEE Signal Processing Int. Symp. Time-Frequency and Time-Scale Anal.– 1992.- P. 383-386.
43. Dittmann J. Advanced audio watermarking benchmarking / Dittmann J., Steinebach M., Lang A., Zmudzinsky S. // SPIE.- 2004, Vol. 5306

44. Rhodes, B. Just-in-time information retrieval agents. // IBM Systems Journal, 39(4): - 2003. P. 685-704.
45. Logan B. Mel Frequency Cepstral Coefficients for Music Modelling // Proc. International Symposium on Music Information Retrieval. - Plymouth (USA). - 2000.
46. Morishima M. Phonetically adaptive cepstrum mean normalization for acoustic mismatch compensation / Morishima M., Isobe T., Takahashi J. // Proc. Automatic Speech Recognition and Understanding. – Santa Barbara (USA). - 1997.- P. 436-441.
47. ISO/IEC JTC1/SC29/WG11 (MPEG). Multimedia content description interface - part 4: Audio. International Standard 15938-4, ISO/IEC, 2001.
48. Билобров С.В., Каргин А.А. Обзор методов идентификации аудиоматериалов с использованием аудиоотпечатков // Вісник Донецького університету, Сер.А: Природничі науки.-2005.-Вип.1.- С.450-454.
49. Wold E. Content-Based Classification, Search, and Retrieval of Audio / Wold E., Blum T., Keislar D., Wheaton J. // IEEE Multimedia. – 1996.- Vol. 3, No.3.- P. 27-36.
50. Haitsma J. Speed-change resistant audio fingerprinting using auto-correlation // Conf. on Acoustics, Speech and Signal Processing. – 2003. - Vol.4, P. 728-731.
51. Методичні рекомендації до виконання дипломної роботи з освітньо-кваліфікаційного рівня “Магістр”. Спеціальність "Комп'ютерні системи та мережі" / О.М. Березький, Л.О. Дубчак, Г.М. Мельник /Під ред. О.М. Березького – Тернопіль: ТНЕУ, 2016.– 47 с.
52. Якименко І.З., Ковтун Н.В. Захист та ідентифікація аудіо матеріалів на основі методу цифрового відбитку. Сучасні комп'ютерні інформаційні технології: матеріали VI Всеукраїнської школи-семінару молодих вчених і студентів АСІТ 2016. – Тернопіль: ТНЕУ, 2016. – 178-180.