Міністерство освіти і науки України

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «КИЄВО-МОГИЛЯНСЬКА АКАДЕМІЯ»

Факультет інформатики

Кафедра інформатики

Магістерська робота

освітній ступінь – магістр

На тему: «**AI-СИСТЕМА ДЛЯ АВТОМАТИЗОВАНОГО СТВОРЕННЯ ВІДЕОКЛІПІВ НА ОСНОВІ АНАЛІЗУ МУЗИЧНИХ ХАРАКТЕРИСТИК АУДІО»**

Виконав: студент 2-го року навчання,

освітньо-наукової програми «Інженерія програмного забезпечення», 121

Плисюк Антон Валерійович

Керівник Нагірна А.М., _____

кан.ф.-м.н., доцент Нагірна А.М.

Рецензент _____

(прізвище та ініціали)

Магістерська робота захищена/не захищена

з оцінкою _____

Секретар ЕК _____

«____» _____ 2026 р.

Київ – 2026

ГРАФІК ПІДГОТОВКИ МАГІСТЕРСЬКОЇ РОБОТИ ДО ЗАХИСТУ

№з/п	ПЕРЕЛІК РОБІТ	Термін виконання	Примітки
1.	Отримання завдання на кваліфікаційну роботу	20.09.2025	
2.	Вивчення джерел літератури за темою.	30.12.2025	
4.	Розробка архітектури та програмування системи	28.02.2026	
5.	Написання текстової частини до кваліфікаційної роботи	01.04.2026	
7.	Аналіз та коригування роботи з керівником	07.04.2026	
8.	Створення слайдів для доповіді та написання доповіді.	26.04.2026	
9.	Попередній захист кваліфікаційної роботи	27.04.2026	
10.	Коригування роботи за результатами попереднього захисту	10.05.2026	
11.	Попередній захист кваліфікаційної роботи на засіданні кафедри	10.05.2026	
12.	Остаточне оформлення роботи та слайді	20.05.2026	
13.	Захист магістерської роботи	Згідно з розкладом ЕІ	

Графік узгоджено «__» _____ 20__ р.

Науковий керівник Нагірна Алла Миколаївна

Виконавець кваліфікаційної роботи Плисюк Антон Валерійович

ЗМІСТ

ГРАФІК ПІДГОТОВКИ МАГІСТЕРСЬКОЇ РОБОТИ ДО ЗАХИСТУ 2
АНОТАЦІЯ 5
ПЕРЕЛІК СКОРОЧЕНЬ ТА ТЕРМІНІВ 6
ВСТУП 7
1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ 9
1.1 Загальний огляд предметної області 9
1.2 Аналіз інструментів генерації відео 10
1.3 Аналіз існуючих рішень 12
1.4 Аналіз методів визначення музичних характеристик аудіо 14
Висновки до розділу 16
2 РОЗРОБКА МЕТОДУ АНАЛІЗУ МУЗИЧНИХ ХАРАКТЕРИСТИК 17
2.1 Обґрунтування вибору методу аналізу аудіосигналу 17
2.2 Виділення ознак аудіосигналу 19
2.2.1 Часові ознаки 20
2.2.2 Спектральні ознаки 20
2.2.3 Ритмічні ознаки 22
2.2.4 Формування вектора ознак 23
2.3 Алгоритм аналізу аудіосигналу 24
2.4 Структурна схема системи 26
2.5 Формування параметрів відео на основі аудіо 28
2.5.1 Принцип формування параметрів відео 28
2.5.2 Перетворення аудіо у відеопараметри 29
2.5.3 Особливості підходу 31
Висновки до розділу 32
3 РОЗРОБКА AI-СИСТЕМИ СТВОРЕННЯ ВІДЕОКЛІПІВ НА ОСНОВІ АНАЛІЗУ МУЗИЧНИХ ХАРАКТЕРИСТИК АУДІО 33
3.1 Архітектура системи 33
3.1.1 Вхідні дані та аналіз аудіо 34

3.1.2 Генерація теми та опису треку 34
3.1.3 Генерація сцен і промптів 34
3.1.4 Генерація візуального контенту 35
3.1.5 Генерація відео 35
3.2 Реалізація системи 35
3.2.1 Аналіз аудіо та формування опису 36
3.2.2 Генерація сцен і зображень 37
3.3 Приклад роботи системи 40
3.3.1 Аналіз аудіофайлу 40
3.3.2 Створення сцен та prompt-ів 41
3.3.3 Генерація відеофрагментів 42
3.3.4 Формування фінального відео 43
Висновки до розділу 44
4 РЕЗУЛЬТАТИ ТА ПЕРСПЕКТИВИ РОЗВИТКУ 45
ВИСНОВКИ 47
СПИСОК ЛІТЕРАТУРИ 48

АНОТАЦІЯ

Ця робота присвячена розробці системи автоматизованого створення відеокліпів на основі аналізу музичних характеристик аудіосигналу. У роботі досліджено сучасні підходи до генерації відеоконтенту з використанням методів обробки аудіо та технологій штучного інтелекту. Також було проведено аналіз існуючих рішень та визначено їхні переваги і недоліки.

Розглянуто методи аналізу аудіосигналу, зокрема визначення темпу, ритму, енергії та спектральних характеристик, що використовуються для подальшої синхронізації відео з музикою. Запропоновано метод автоматизованої генерації відео на основі аналізу музичних характеристик аудіосигналу, який використовує BPM, spectral features та segmentation для адаптивного формування сцен і синхронізації переходів.

Також описано архітектуру системи та деталі її програмної реалізації із використанням сучасних бібліотек та AI-моделей, та моделі для генерації текстового опису та сцен, а також інструментів для генерації зображень і генерацію відео.

Наостанок продемонстровано приклад роботи системи, який демострує можливість автоматизованого створення відеоконтенту, на основі аудіосигналу. Та наведені можливі напрями подальшого розвитку системи.

ПЕРЕЛІК СКОРОЧЕНЬ ТА ТЕРМІНІВ

AI (Artificial Intelligence) — це галузь інформатики, яка створює комп'ютерні системи, здатні імітувати людські когнітивні функції.

API (Application Programming Interface) — інтерфейс програмування застосунків, для взаємодії між різними програмними компонентами.

BPM (Beats Per Minute) — кількість ударів на хвилину, показник темпу музичної композиції.

FFT (Fast Fourier Transform) — швидке перетворення Фур’є, алгоритм для аналізу частотного складу сигналу.

JSON (JavaScript Object Notation) — текстовий формат обміну даними, який використовується для збереження структурованої інформації.

GPT (Generative Pre-trained Transformer) — модель штучного інтелекту, яка використовується для генерації тексту.

Prompt — текстовий опис або інструкція, яка передається AI-моделі для генерації контенту (зображень або тексту).

Text-to-Image модель — модель штучного інтелекту, яка генерує зображення на основі текстового опису.

Pipeline — послідовність етапів обробки даних, де результат одного етапу є вхідними даними для наступного.

Rendering (рендеринг) — процес формування фінального відео або зображення з окремих елементів.

Librosa — бібліотека Python для аналізу аудіосигналів.

MoviePy — бібліотека Python для створення та обробки відео.

ВСТУП

Актуальність теми. З стрімким розвитком цифрових технологій та штучного інтелекту з’являється потреба у створенні систем, які допоможуть автоматизувати процес створення мультимедійного контенту. Штучний інтелект уже використовується у багатьох сферах і продовжує набувати популярності. Тому використання AI може значно прискорити процес виробництва відеоконтенту та підвищити його адаптивність до індивідуальних характеристик музики.

Наразі сучасні підходи до генерації відео базуються на використанні нейронних мереж та генеративних моделей, які дають змогу аналізувати вхідні дані та генерувати відповідний візуальний контент. Зокрема AI аналізує характеристики аудіо, такі як ритм, темп, енергія тощо, і дозволяє синхронізувати відеоряд із музичною доріжкою створюючи узгоджений і динамічний контент.

Існуючі рішення мають ряд проблем та обмежень, які не дозволяють отримати якісний результат, зокрема вони мають такі недоліки як:

- погана синхронізації відеоряду з аудіосигналом;
- обмежена адаптація до індивідуальних характеристик музики (ритму, темпу, енергії);
- обмежена гнучкість до генерації відео;

Тому актуальною є задача розробки ефективної системи генерації відео на основі аудіо, яка забезпечує гнучке налаштування та контроль з боку користувача, високу якість згенерованого відео, точну синхронізацію з аудіосигналом і адаптацію до індивідуальних характеристик музики.

Мета і завдання дослідження. Проаналізувати сучасні підходи та методи генерації відео на основі аудіо; визначити особливості аналізу аудіосигналу (ритму, темпу, енергії та спектральних характеристик) для подальшої генерації відеоконтенту.

Обрати оптимальні методи та алгоритми з урахуванням вимог до якості та продуктивності системи; розробити систему автоматизованої генерації відео на основі аудіо; реалізувати програмний прототип системи та провести його тестування для оцінки якості згенерованого відеоконтенту.

Об'єкт дослідження. Процес автоматизованої генерації відеоконтенту на основі аудіосигналів.

Методи дослідження. Аналіз наукової літератури та існуючих рішень у сфері генерації відео та обробки аудіо; аналіз та обробка аудіосигналів із використанням сучасних алгоритмів; порівняння методів генерації відеоконтенту; програмна реалізація системи з використанням обраних підходів; експериментальне тестування та оцінка якості отриманих результатів.

Джерела дослідження. Наукові публікації та статті у сфері генерації відео та обробки аудіосигналів; відкриті набори аудіо- та відеоданих; документація та програмні бібліотеки для аналізу аудіо і генерації відеоконтенту; інтернет-ресурси, що містять приклади реалізації подібних систем та алгоритмів.

Наукова новизна одержаних результатів. Удосконалено підхід до автоматизованої генерації відео на основі аудіо за рахунок урахування характеристик музичного сигналу, що забезпечує покращення якості синхронізації та адаптивності відеоконтенту.

Практичне значення одержаних результатів. Розроблена система може бути використана для автоматизованого створення відеоконтенту на основі музики у сферах медіа, маркетингу та розваг. Отримані результати можуть бути застосовані для скорочення часу виробництва відео та підвищення його якості.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Загальний огляд предметної області

Відео це один із найпопулярніших способів передачі інформації у сучасному світі. Відеоконтент використовується у сферах розваг, реклами, освіти та соціальних медіа, тощо. Окремо можна виділити відеокліпи, які поєднують аудіо та візуальний ряд для передачі інформації. Приклади використання відеоконтенту в різних сферах наведено на рисунку 1.1.

Рисунок 1.1 — Приклади відеоконтенту у різниз сферах.

Музичний кліп є достатньо складним мультимедійним продуктом, який повинен вкдлючати багато різних елементів таких як темп, ритм, емоційний зміст тощо. Тому створення якісного кінцевого продукту потребує значних ресурсів: часу, фінансових витрат та роботи професійної команди.

Але з розвитком штучного інтелекту з’являється багато нових підходів, які допомагають спростити та оптимізувати рутинні процеси або і зовсім автоматизувати генерацію відео.

1.2 Аналіз інструментів генерації відео

Сучасні інструменти для створення відеоконтенту активно розвиваються разом із цифровими технологіями. Сьогодні користувач має доступ до великої кількості рішень, від професійних програм для монтажу до простих онлайн-сервісів і сучасних систем на основі штучного інтелекту. Вони відрізняються рівнем складності, функціоналом та ступенем автоматизації, тому їх доцільно поділити на три основні категорії: традиційні відеоредактори, шаблонні онлайн-сервіси та системи, що використовують штучний інтелект.

До першої категорії належать класичні відеоредактори, такі як Adobe Premiere Pro, Final Cut Pro та DaVinci Resove. Вони дають користувачу повний контроль над процесом створення відео: можна вручну монтувати сцени, додавати ефекти, працювати зі звуком і кольором. Проте такий підхід потребує часу, досвіду і детального розуміння процесу відеомонтажу. Інтерфейс класичного відеоредактора показано на рисунку 1.2.

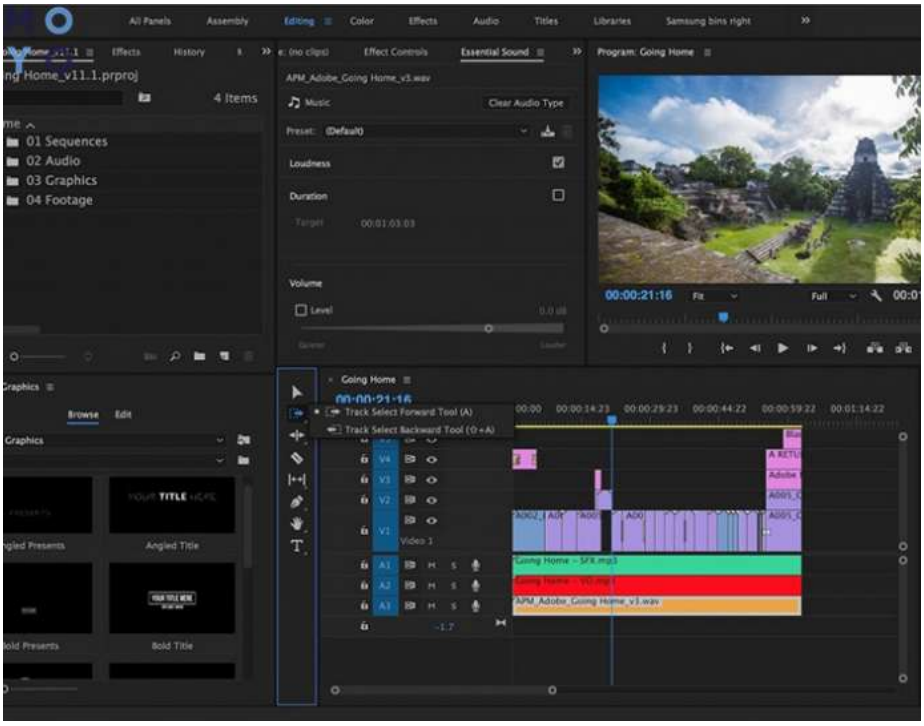


Рисунок 1.2 — інтерфейс класичного відеоредактора

Шаблонні онлайн-сервіси, наприклад Canva, Animoto або Renderforest, орієнтовані на швидкість і простоту. Користувач обирає готовий шаблон, додає свої матеріали і система автоматично створює відео. Це зручно, але такі сервіси обмежені у налаштуваннях і не завжди дозволяють отримати унікальний результат.

Сучасні інструменти з штучним інтелектом, такі як Runway[1], Pika та Synthesia використовують генеративні моделі для створення відео на основі вхідних даних, такі як текст, зображення та аудіо. Використання генеративних моделей дозволяє автоматизувати процес створення мультимедійного контенту та скоротити час його формування, проте часто не забезпечують достатньої точності синхронізації з аудіо або зовсім не мають такого функціоналу. Приклад платформи Runway для AI-генерації відео наведено на рисунку 1.3.

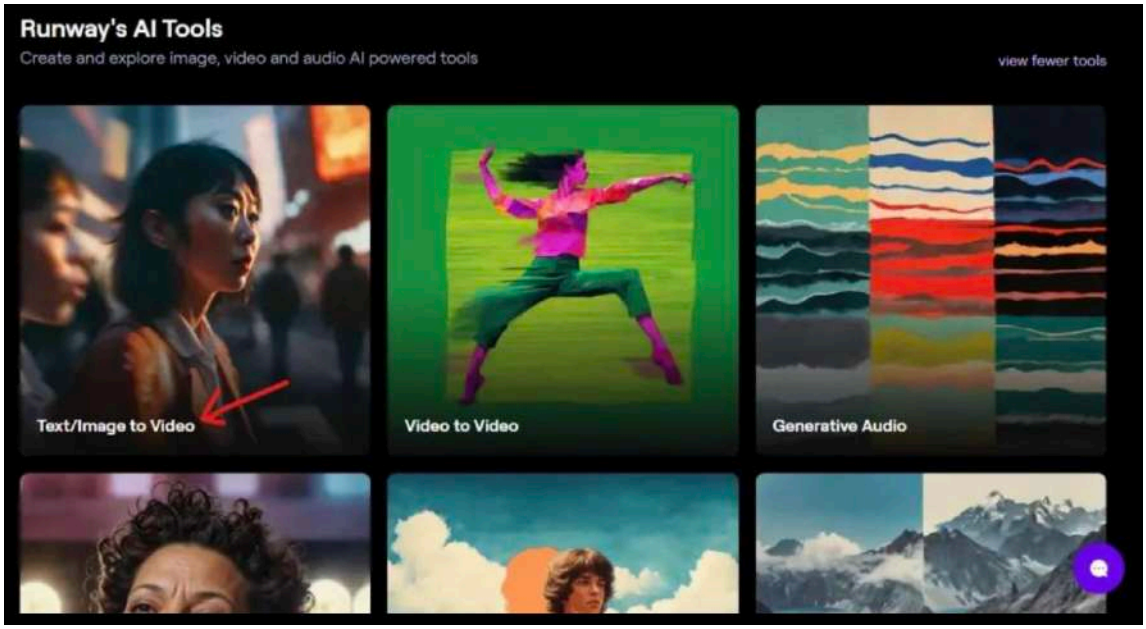


Рисунок 1.3 — Інтерфейс платформи Runway

1.3 Аналіз існуючих рішень

Аналізуючи існуючі рішення у сфері автоматизованого створення відеокліпів на основі аудіо, можна відзначити активний розвиток систем, що використовують аналіз музичних характеристик. Такі системи застосовують алгоритми обробки аудіосигналів та методи машинного навчання для визначення ключових параметрів, зокрема ритму, темпу, енергії та спектральних особливостей. На основі отриманих даних формується візуальний контент, узгоджений із музичним супроводом, що дозволяє автоматично синхронізувати відеоряд із аудіо та створювати відеокліпи з мінімальною участю користувача.

До найбільш популярних рішень у цій сфері належать платформи, що дозволяють створювати відео на основі аудіо з урахуванням його характеристик. Зокрема, платформа Rotor Videos[2] що спеціалізується на автоматичному створенні музичних відеокліпів, аналізуючи структуру треку та підбираючи відповідні відеофрагменти. Платформа Kaiber використовує генеративні моделі для створення відео що дозволяє отримувати динамічні та стилізовані результати. Інтерфейс системи Rotor Videos показано на рисунку 1.4.

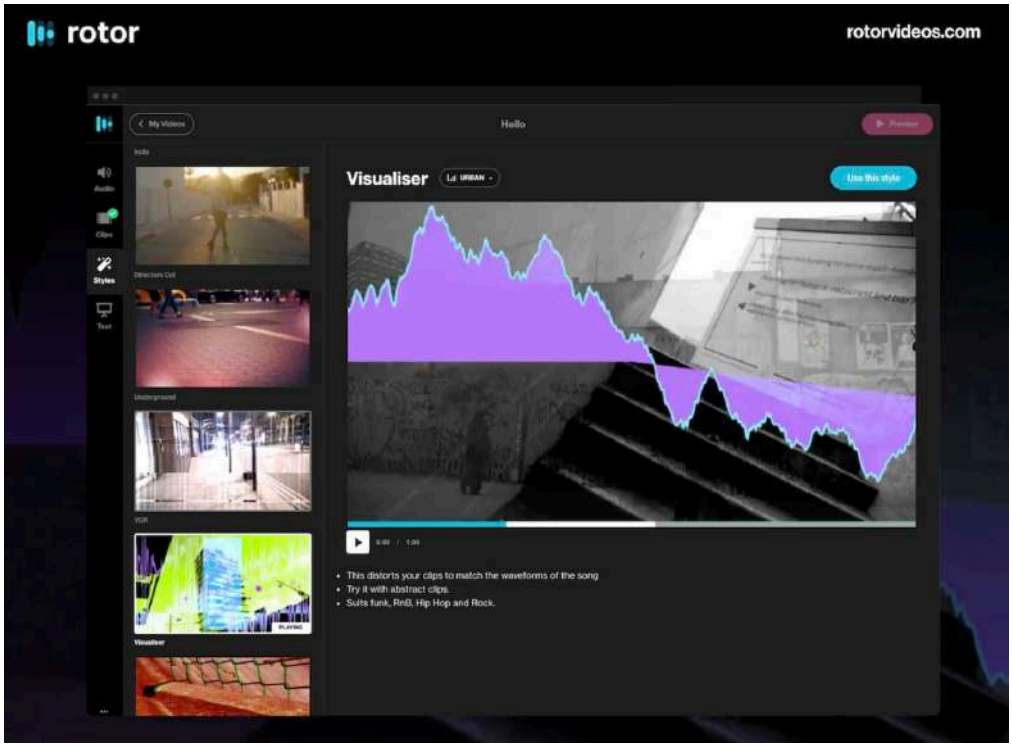


Рисунок 1.4 — інтерфейс програми Rotor Videos

У таблиці 1.1 наведено порівняння існуючих AI-систем генерації відеоконтенту на основі музичного супроводу.

Таблиця 1.1 — Порівняння існуючих AI-систем генерації відеоконтенту

	Аналіз аудіо	AI генерація	Автосинхронізація	Гнучкість
Runway	частково	так	середня	висока
Rotor	так	частково	добра	низька
Kaiber	частково	так	середня	середня

Основними перевагами таких систем є:

- автоматичний аналіз аудіосигналу (ритм, темп, енергія)
- генерація музики для відеоряду
- можливість генерації унікальних візуальних ефектів
- використання сучасних AI моделей

Водночас такі системи мають ряд обмежень:

- погана синхронізації відеоряду з аудіосигналом;
- обмежена адаптація до індивідуальних характеристик музики (ритму, темпу, енергії);
- обмежена гнучкість до генерації відео;
- використання загальних шаблонів або стилів генерації
- не зручний інтерфейс

1.4 Аналіз методів визначення музичних характеристик аудіо

Для того щоб відео було узгоджене з музикою, необхідно попередньо проаналізувати аудіосигнал і визначити його основні характеристики. Саме від цього етапу залежить, наскільки точно відеоряд буде відповідати ритму та настрою музичного супроводу.

До основних характеристик аудіо належать темп (BPM), ритм, енергія сигналу та спектральні параметри. Вони дозволяють описати як змінюється музика в часі та які частоти в ній переважають.

Одним із базових методів аналізу є швидке перетворення Фур'є (FFT)[3]. Воно дозволяє розкласти звук на окремі частоти та побачити, які з них домінують у сигналі. Даний метод дозволяє визначити частотний склад сигналу, з яких складових формується звук, і використовується для аналізу його структури. Приклад частотного аналізу аудіосигналу наведено на рисунку 1.5.

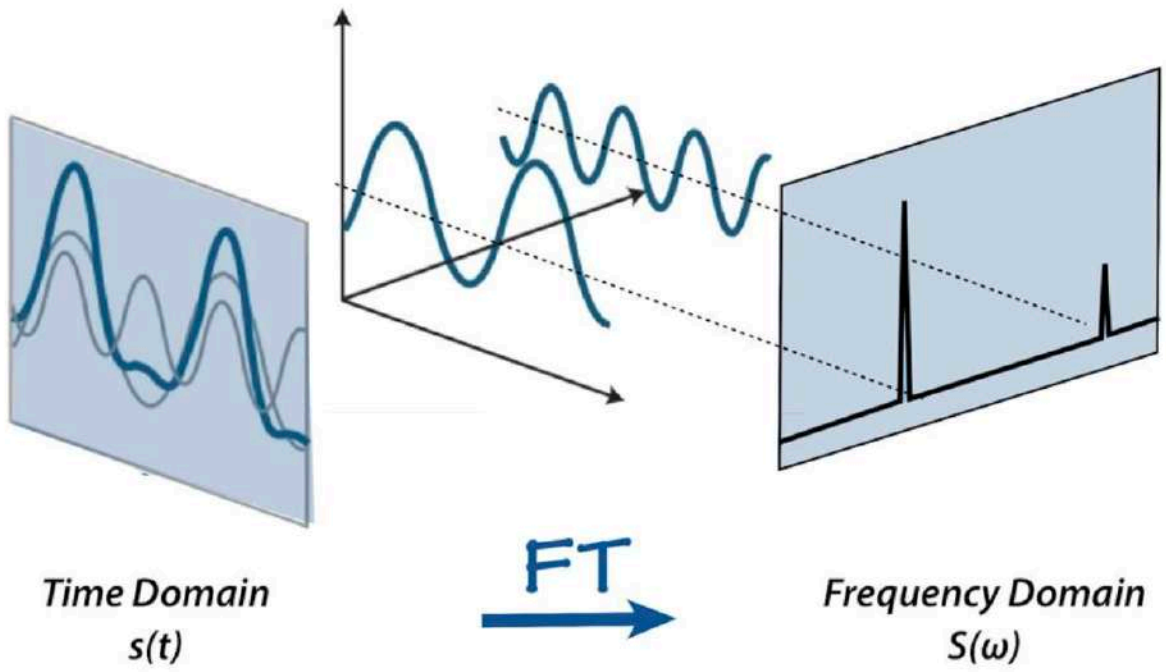


Рисунок 1.5 — Частотний аналіз аудіо

Окрім частотного аналізу, важливу роль відіграє визначення ритму, так званий beat detection. Ці алгоритми знаходять моменти часу[4], коли відбуваються музичні удари. Це дозволяє синхронізувати зміну кадрів або ефектів у відео з ритмом музики, що робить відео більш динамічним і природним для сприйняття. Процес визначення ритму (beat detection) показано на рисунку 1.6.

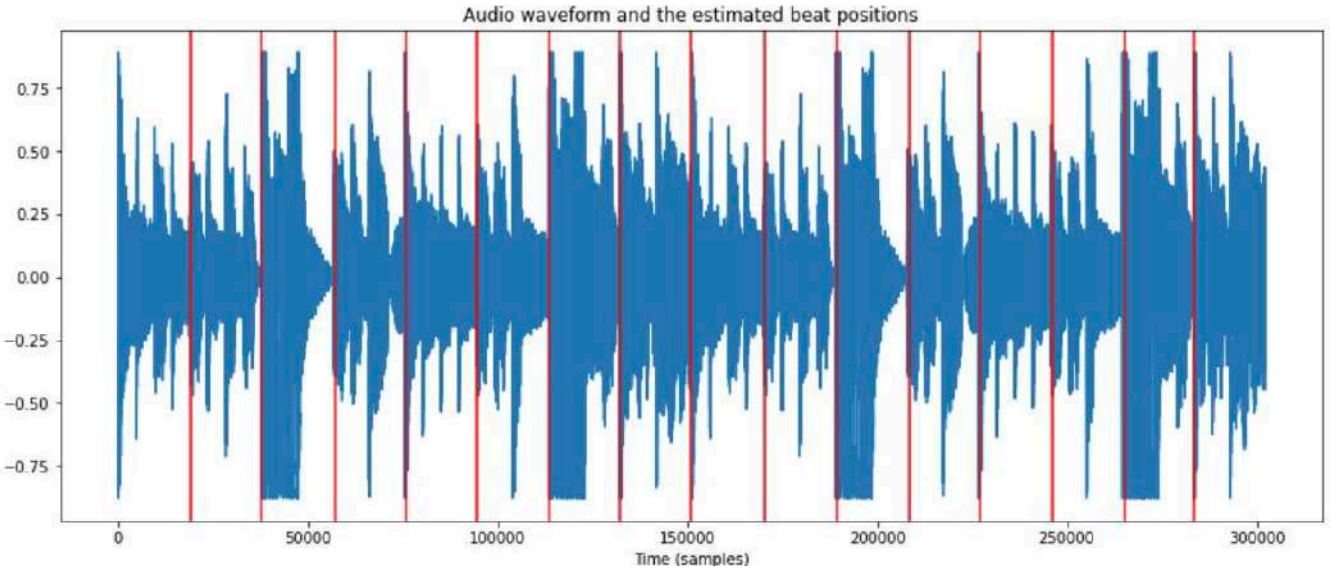


Рисунок 1.6 — Визначення ритму (beat detection)

Сьогодні для аналізу аудіо активно використовуються спеціальні бібліотеки, такі як Librosa[5], Essentia та PyDub. Вони дозволяють автоматично визначати темп, енергію, спектральні характеристики та інші параметри без необхідності реалізовувати складні алгоритми вручну. Застосування таких методів і інструментів дозволяє створювати системи генерації відео, які краще відповідають музичним характеристиками та більш точно відтворюють її структуру у візуальному вигляді.

Висновки до розділу

В цьому розділі розглянули предметну область та інструменти для генерації відео, також було проведено порівняння різних підходів до генерації відео та переваги і недоліки основних конкурентів на ринку.

Аналіз методів обробки аудіо показав можливість визначення ключових параметрів, які можуть бути використані для синхронізації відео з музикою. Таким чином, розробка системи генерації відео є актуальною задачою, яка враховує характеристики аудіосигналу та забезпечує більш точну синхронізацію відеоряду.

2 РОЗРОБКА МЕТОДУ АНАЛІЗУ МУЗИЧНИХ ХАРАКТЕРИСТИК

2.1 Обґрунтування вибору методу аналізу аудіосигналу

У задачах автоматизованого створення відеоконтенту на основі музики використовуються різні підходи до аналізу аудіосигналу. Найбільш поширеними є евристичні методи та методи, що базуються на використанні нейронних мереж.

Евристичні методи ґрунтуються на застосуванні заздалегідь визначених правил, які встановлюють відповідність між характеристиками аудіосигналу та параметрами відеоряду. Перевагою таких методів є відносно невисока складність реалізації та невисокі обчислювальні витрати. Недоліком є обмежена точність та відсутність адаптації до різних типів музичних сигналів[6]. Приклад використання евристичних методів наведено на рисунку 2.1.

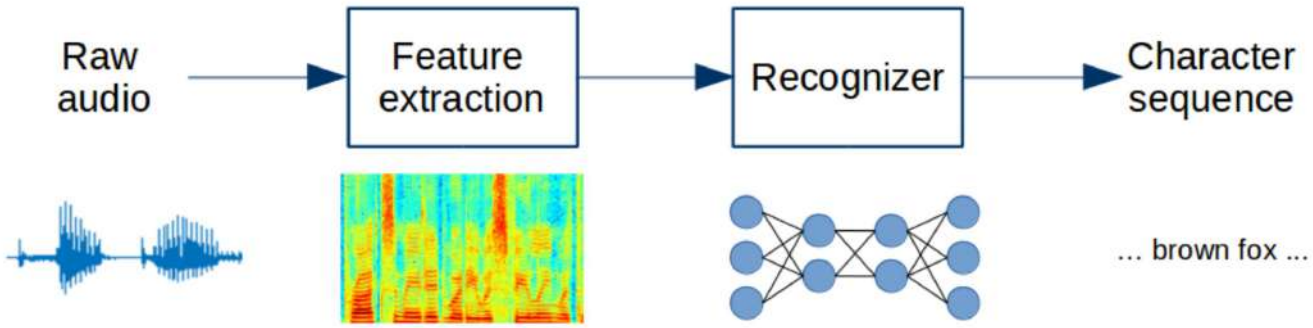


Рисунок 2.1 — Евристичні методи

Методи, які базуються на використанні нейронних мереж[7, 8], дозволяють автоматично визначати залежності між аудіо та відео. Однак їх застосування потребує значних обчислювальних ресурсів, наявності даних для навчання та складності реалізації системи. На рисунку 2.2, методи на основі нейронних мереж дозволяють враховувати складні залежності між аудіо та відео.

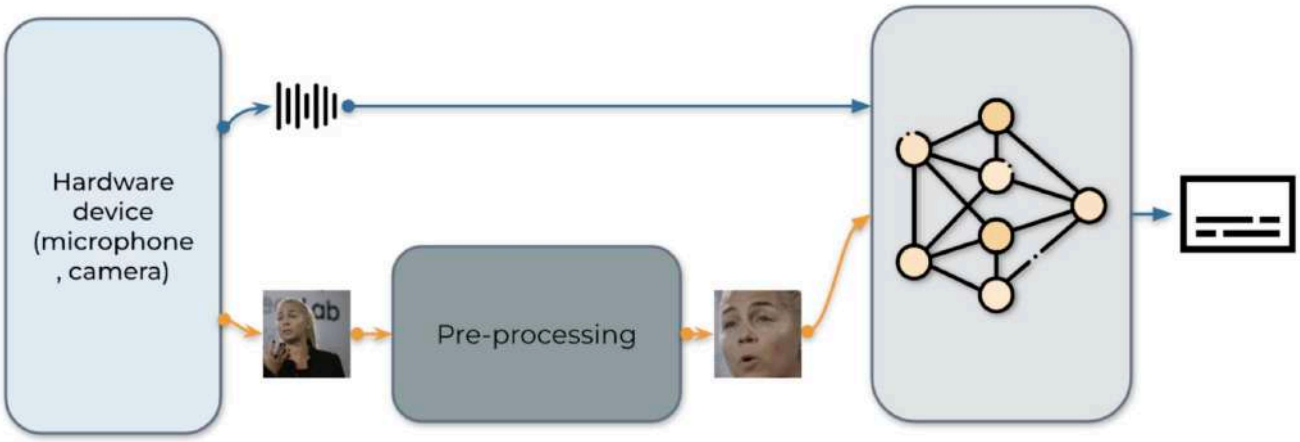


Рисунок 2.2 — Методи на основі нейронних мереж

Обидва підходи мають свої недоліки. Евристичні методи є простими у реалізації, але вони працюють за заздалегідь заданими правилами, тому погано адаптуються до складної або нестандартної музики і можуть давати менш якісний результат. У той же час методи на основі нейронних мереж здатні краще враховувати складні залежності між аудіо та відео, але потребують великих обчислювальних ресурсів, значної кількості навчальних даних і є складнішими у реалізації та налаштуванні.

У даній роботі використовується підхід, що базується на методах цифрової обробки сигналів. Обраний підхід передбачає виділення основних характеристик аудіосигналу та їх подальше використання для формування параметрів відеоконтенту.

До основних характеристик віднесено:

- енергію сигналу.
- спектральні характеристики.
- ритмічні параметри.
- емоційне сприйняття.

Вибір саме цих характеристик зумовлений тим, що вони відображають ключові властивості музики, такі як гучність, тембр і ритм. Ці параметри безпосередньо впливають на сприйняття аудіосигналу та можуть бути ефективно використані для керування процесом генерації відеоряду.

Обраний метод дозволяє забезпечити хорошу точність аналізу при невисоких обчислювальних витратах. Крім того, він не потребує попереднього навчання моделі та може бути реалізований у вигляді модульної системи.

Таким чином, використання методів цифрової обробки сигналів є доцільним для розв’язання поставленої задачі.

2.2 Виділення ознак аудіосигналу

Аналіз аудіосигналу в розроблюваній системі базується на виділенні інформативних ознак, які дозволяють описати його основні властивості та використовуються для подальшого формування параметрів відеоконтенту. Відповідно до вимог до побудови основної частини роботи, метод повинен мати аналітичний характер та містити обґрунтовані розрахункові залежності. [9]

Аудіосигнал розглядається як дискретна послідовність відліків:

$$x[n], n = 0,1,2,\ldots, N.$$

(2.1)

де N — кількість відліків сигналу.

Для забезпечення локального аналізу сигнал розбивається на короткі часові інтервали (фрейми). Такий підхід дозволяє визначати зміну характеристик сигналу

в часі та використовується у більшості сучасних систем обробки аудіо. У даній роботі виділяються три основні групи ознак: часові, спектральні та ритмічні.

2.2.1 Часові ознаки

Часові характеристики дозволяють оцінити, наскільки інтенсивним є сигнал у певний момент часу. Найпростішою і водночас важливою характеристикою є енергія сигналу, яка обчислюється як сума квадратів його значень [10]:

$$E = n = 0 \sum Nx2[n] .$$

(2.2)

де x[n] — значення сигналу, а N — кількість відліків.

Дана характеристика дозволяє оцінити рівень гучності сигналу на окремих ділянках. У розроблюваній системі енергія використовується для визначення динаміки аудіосигналу та подальшого керування параметрами відеоряду.

2.2.2 Спектральні ознаки

Для аналізу частотного складу сигналу використовується перетворення Фур'є[11], яке дозволяє представити сигнал не у вигляді змін у часі, а у вигляді набору частот:

$$X[k] = n = 0 \sum N - 1x[n]e - jN2\pi kn .$$

(2.3)

де X[k] — спектр сигналу, x[n] — дискретний сигнал, N — кількість відліків.

Отриманий спектр сигналу використовується для визначення його тембрових характеристик. На основі спектру можуть обчислюватися узагальнені параметри, що характеризують розподіл енергії по частотах. Перетворення Фур’є до аудіосигналу можна побачити на рисунку 2.3.

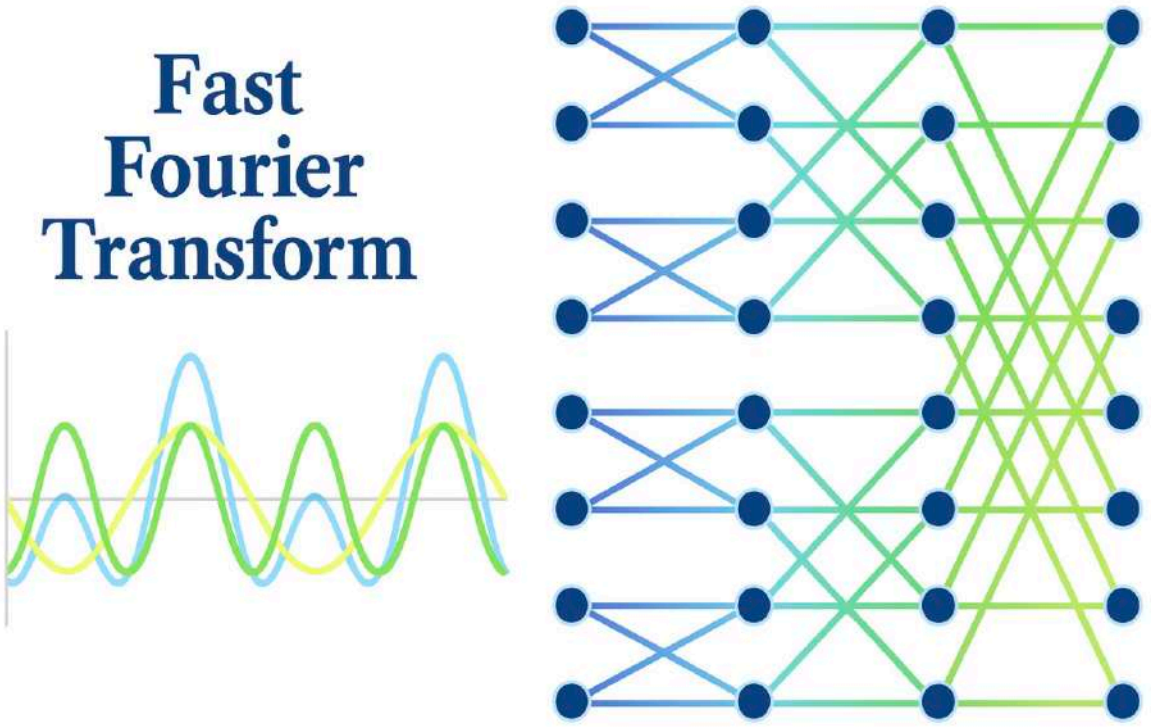


Рисунок 2.3 — Перетворення Фур’є

Спектральні ознаки дозволяють враховувати особливості звучання аудіосигналу, зокрема його тембр та насиченість. У нашій системі ці характеристики використовуються при формуванні візуальних параметрів відео, наприклад, кольорової гами або стилю відображення.

2.2.3 Ритмічні ознаки

Ритмічні характеристики визначають часову структуру сигналу. Для їх оцінки використовується аналіз змін спектру у часі. Зокрема, визначаються моменти різких змін сигналу, які відповідають ударам (beats)[12].

Для цього використовується функція різниці спектрів:

$$D[n] = k \sum | Xn[k] - Xn - 1[k] | . \tag{2.4}$$

де X n [k] — спектр сигналу для поточного фрейму.

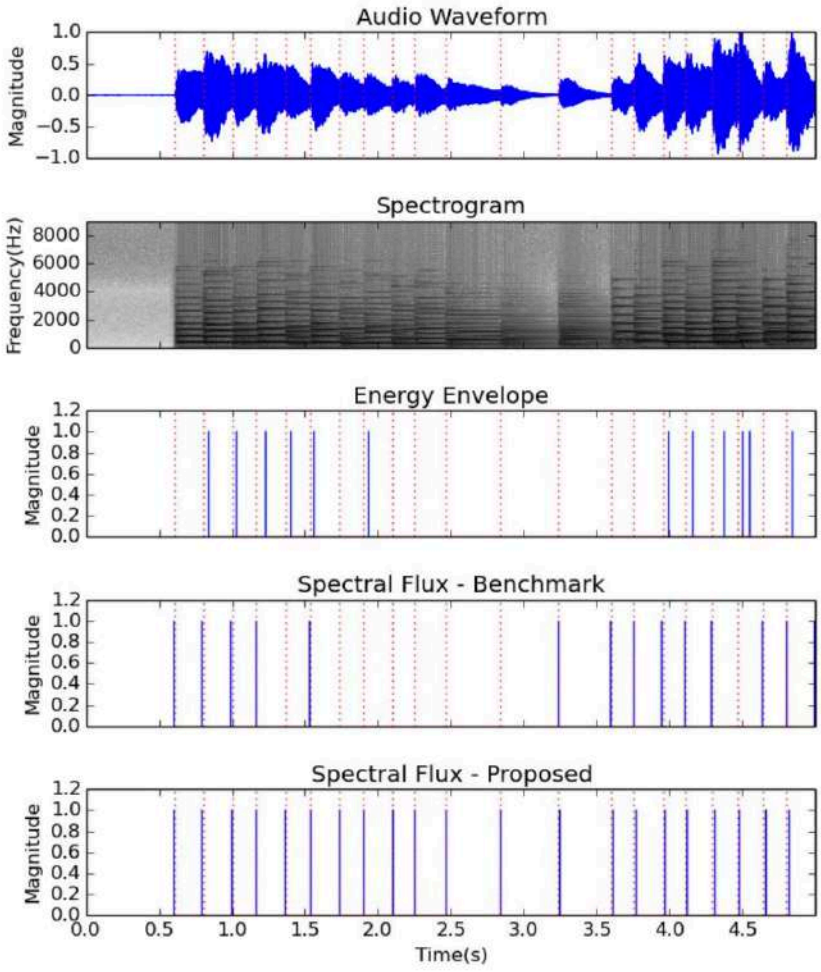


Рисунок 2.4 — Виявлення ритму та ударів у аудіосигналі

На рисунку 2.4 показано приклад аналізу ритму аудіосигналу. У верхній частині зображено audio waveform, де видно зміну його амплітуди. Нижче наведено спектрограму, яка відображає, як енергія сигналу розподіляється по частотах у часі.

Також на рисунку показано energy envelope, яка дає уявлення про загальну динаміку сигналу. Найбільш важливими для визначення ритму є графіки spectral flux, які відображають, наскільки сильно змінюється сигнал між сусідніми моментами часу.

Піки на цих графіках відповідають різким змінам у сигналі та визначаються як ритмічні удари. Вертикальні лінії показують знайдені моменти ударів, які узгоджуються з реальною структурою сигналу. Тому на основі цього визначається темп сигналу, який використовується для синхронізації відеоряду з музичним супроводом.

2.2.4 Формування вектора ознак

Результатом аналізу є формування вектора ознак, який описує аудіосигнал у компактному вигляді[13]:

$$f = [f1, f2, ..., fn] \tag{2.5}$$

У розроблюваній системі вектор ознак включає значення енергії, спектральних та ритмічних характеристик. Отриманий вектор використовується на наступному етапі для побудови алгоритму аналізу та формування параметрів відео.

2.3 Алгоритм аналізу аудіосигналу

Враховуючи інформацію у попередньому розділах можна сформувати ознаки алгоритмічного аналізу аудіосигналу. Алгоритм забезпечує послідовне перетворення вхідного аудіофайлу у вектор характеристик, який використовується для подальшого формування параметрів відеоконтенту. Відповідно до вимог до основної частини роботи, алгоритм повинен мати чітку структуру та відображати логіку обробки даних

Обробка аудіосигналу виконується поетапно. На першому етапі здійснюється завантаження аудіофайлу та приведення його до єдиного формату. Це дозволяє забезпечити коректність подальших обчислень незалежно від параметрів вхідного сигналу.

Далі виконується попередня обробка сигналу, яка включає нормалізацію амплітуди та, за необхідності, фільтрацію шумів. Після цього сигнал розбивається на короткі часові інтервали, що дозволяє виконувати локальний аналіз його характеристик.

Для кожного інтервалу обчислюються часові, спектральні та ритмічні параметри. Отримані значення об’єднуються у вектор ознак, який є результатом роботи алгоритму. Такий підхід дозволяє зменшити обсяг даних та виділити найбільш інформативні характеристики сигналу.

2.1 — Алгоритм аналізу аудіосигналу можна подати у вигляді псевдокоду:

Вхід: аудіофайл A

Вихід: вектор ознак F

1. Завантажити аудіофайл A
2. Перетворити сигнал у дискретний вигляд x[n]

3. Виконати нормалізацію сигналу

4. Розбити сигнал на фрейми

5. Для кожного фрейму виконати:

    1. Обчислити енергію сигналу

    2. Обчислити спектральні характеристики

    3. Визначити ритмічні параметри

6. Об’єднати отримані значення у вектор ознак F

7. Передати F до наступного етапу обробки

8. Завершити роботу алгоритму

Псевдокод відображає загальну послідовність виконання операцій та може бути використаний як основа для програмної реалізації. У розроблюваній системі передбачається модульна реалізація алгоритму, що дозволяє окремо реалізувати обробку сигналу, виділення ознак та їх подальше використання. Алгоритм аналізу аудіосигналу у вигляді блок-схеми представлено на рисунку 2.5.

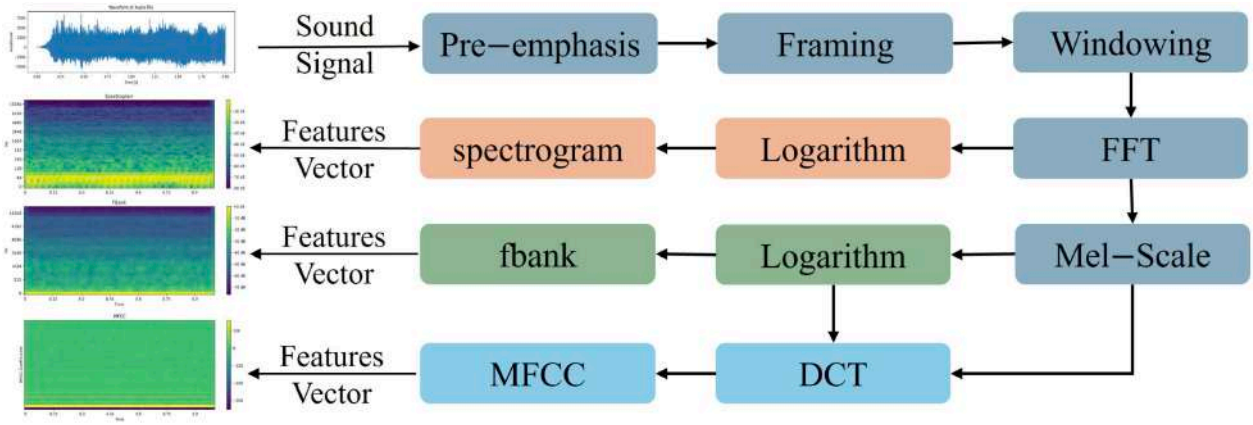


Рисунок 2.5 — Алгоритм аналізу аудіосигналу

Таким чином, запропонований алгоритм забезпечує універсальний підхід до аналізу аудіосигналу та може бути застосований для різних типів музичних композицій.

2.4 Структурна схема системи

У процесі розробки програмної системи необхідно визначити її структурну схему, яка відображає склад основних компонентів та взаємозв’язки між ними.

На рисунку 2.6 наведено блок-схему процесу аналізу аудіосигналу та генерації відеоконтенту.

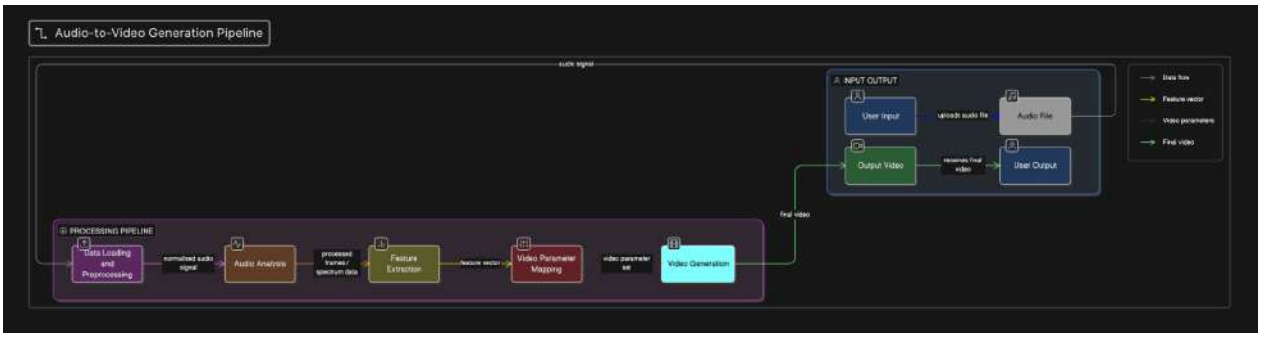


Рисунок 2.6 — Блок-схема процесу аналізу аудіосигналу та генерації відеоконтенту

Система автоматизованого створення відеоконтенту реалізована у вигляді послідовного обчислювального процесу, що складається з ряду функціональних модулів. Кожен модуль виконує окремий етап обробки даних та передає результати на наступний етап.

На початковому етапі користувач завантажує аудіофайл, який подається на вхід системи. Вхідний сигнал надходить до модуля завантаження та попередньої

обробки даних, у якому здійснюється перевірка формату та нормалізація сигналу. Результатом роботи цього модуля є нормалізований аудіосигнал.

Далі сигнал передається до модуля аналізу аудіо, у якому виконується його розбиття на короткі часові інтервали та визначення спектральних характеристик. На цьому етапі формується представлення сигналу у вигляді послідовності фреймів та відповідних спектральних даних.

Наступним етапом є модуль виділення ознак, у якому обчислюються основні характеристики аудіосигналу. До них відносяться енергія сигналу, спектральні параметри та ритмічні характеристики. Результатом цього етапу є вектор ознак, який використовується для подальшої обробки.

Отриманий вектор ознак передається до модуля формування параметрів відео, у якому визначається відповідність між характеристиками аудіосигналу та параметрами відеоряду. У результаті формується набір параметрів, що визначає візуальні характеристики майбутнього відео.

На завершальному етапі модуль генерації відеоконтенту формує кінцевий відеофайл на основі отриманих параметрів. Сформований відеофайл передається користувачу як результат роботи системи.

Структурна схема має модульну організацію, що забезпечує гнучкість системи та можливість її подальшого розвитку. Зокрема, окремі модулі можуть бути вдосконалені або замінені без зміни загальної структури системи.

Таким чином, розроблена структурна схема відображає логіку роботи системи та забезпечує основу для її програмної реалізації.

2.5 Формування параметрів відео на основі аудіо

На завершальному етапі аналізу аудіосигналу здійснюється формування параметрів відеоконтенту на основі виділених аудіо ознак. Основною задачею цього етапу є встановлення відповідності між характеристиками аудіосигналу та візуальними параметрами відеоряду.

2.5.1 Принцип формування параметрів відео

У системі використовується параметро-орієнтований підхід до формування відеоконтенту. Це означає, що кожна характеристика аудіосигналу впливає на певні параметри відео.

Енергія сигналу використовується для визначення яскравості або інтенсивності відеоряду, що узгоджується зі сприйняттям гучності людиною. Ритмічні характеристики, зокрема темп і моменти ударів, застосовуються для визначення частоти зміни сцен або кадрів у відео. Спектральні параметри дозволяють враховувати темброві особливості сигналу та можуть використовуватися для формування кольорової гами або загального стилю візуалізації.

Основні відповідності наведено нижче:

- енергія сигналу визначає яскравість та інтенсивність відео;
- ритмічні характеристики визначають частоту зміни сцен;
- спектральні параметри впливають на кольорову палітру та стиль відео.

Такий підхід дозволяє забезпечити узгодженість між аудіо та відео та створити візуальний ряд, синхронізований із музикою. На рисунку 2.6 показано типи звукових хвиль в залежності від енергії та ритму.

Рисунок 2.6 — типи звукових хвиль в залежності від енергії та ритму

2.5.2 Перетворення аудіо у відеопараметри

Для створення відповідності між аудіо та відео використовуються функції нормалізації та масштабування. Наприклад, значення енергії сигналу можуть бути приведені до діапазону яскравості відео [14]:

$$B = k \cdot E \tag{2.6}$$

де B — яскравість, E — енергія сигналу, k — коефіцієнт масштабування.

Аналогічно, темп сигналу може використовуватися для визначення швидкості зміни кадрів або сцен у відео.

Процес формування параметрів відео можна описати таким алгоритмом:

Алгоритм 2.2 — Формування параметрів відео

Вхід: вектор ознак F

Вихід: набір параметрів відео V

1. Отримати вектор ознак F
2. Нормалізувати значення ознак
3. Визначити параметри:
    1. яскравість на основі енергії
    2. частоту зміни сцен на основі ритму
    3. кольорову палітру на основі спектральних характеристик
4. Сформувати набір параметрів V
5. Передати V до модуля генерації відео

Цей алгоритм забезпечує узгоджене перетворення аудіохарактеристик у параметри відео та дозволяє реалізувати гнучке керування процесом генерації.

2.5.3 Особливості підходу

Запропонований підхід має низку переваг. Насамперед, він є відносно простим у реалізації та не потребує використання складних обчислювальних моделей. Крім того, відсутня необхідність у попередньому навчанні, що спрощує впровадження системи та зменшує вимоги до обчислювальних ресурсів. Також підхід забезпечує можливість гнучкого налаштування відповідностей між характеристиками аудіосигналу та параметрами відео.

Разом з тим, даний підхід може бути розширений шляхом використання більш складних методів, зокрема нейронних мереж або генеративних моделей. Це дозволить враховувати складніші залежності між аудіо та відео та підвищити якість сформованого відеоконтенту.

Висновки до розділу

У даному розділі розроблено метод аналізу аудіосигналу та формування параметрів відеоконтенту для системи автоматизованого створення відео на основі музики. Запропонований підхід базується на використанні методів цифрової обробки сигналів, що дозволяє досягти достатньої точності аналізу при відносно невисокій складності реалізації.

У процесі дослідження визначено основні інформативні характеристики аудіосигналу, зокрема енергетичні, спектральні та ритмічні параметри. На їх основі сформовано компактне представлення сигналу у вигляді вектора ознак, який використовується для подальшої обробки.

Розроблено алгоритм аналізу аудіосигналу, що забезпечує послідовне перетворення вхідних даних у набір характеристик, придатних для використання у системі генерації відео. Алгоритм має чітку структуру та може бути реалізований у вигляді окремих функціональних модулів.

Побудовано структурну схему системи, яка відображає взаємодію її основних компонентів та загальну логіку обробки даних. Запропонована архітектура є модульною, що забезпечує можливість подальшого розширення та вдосконалення системи.

Також визначено принцип формування параметрів відеоконтенту на основі характеристик аудіосигналу. Встановлено відповідність між аудіо параметрами та візуальними параметрами, що дозволяє забезпечити узгодженість аудіо та відеоряду.

Таким чином, у розділі сформовано цілісний підхід до аналізу аудіосигналу та генерації відеоконтенту, який є основою для подальшої програмної реалізації системи у наступному розділі.

3 РОЗРОБКА AI-СИСТЕМИ СТВОРЕННЯ ВІДЕОКЛІПІВ НА ОСНОВІ АНАЛІЗУ МУЗИЧНИХ ХАРАКТЕРИСТИК АУДІО

3.1 Архітектура системи

Загальну архітектуру застосунку схематично показано на рисунку 3.1.

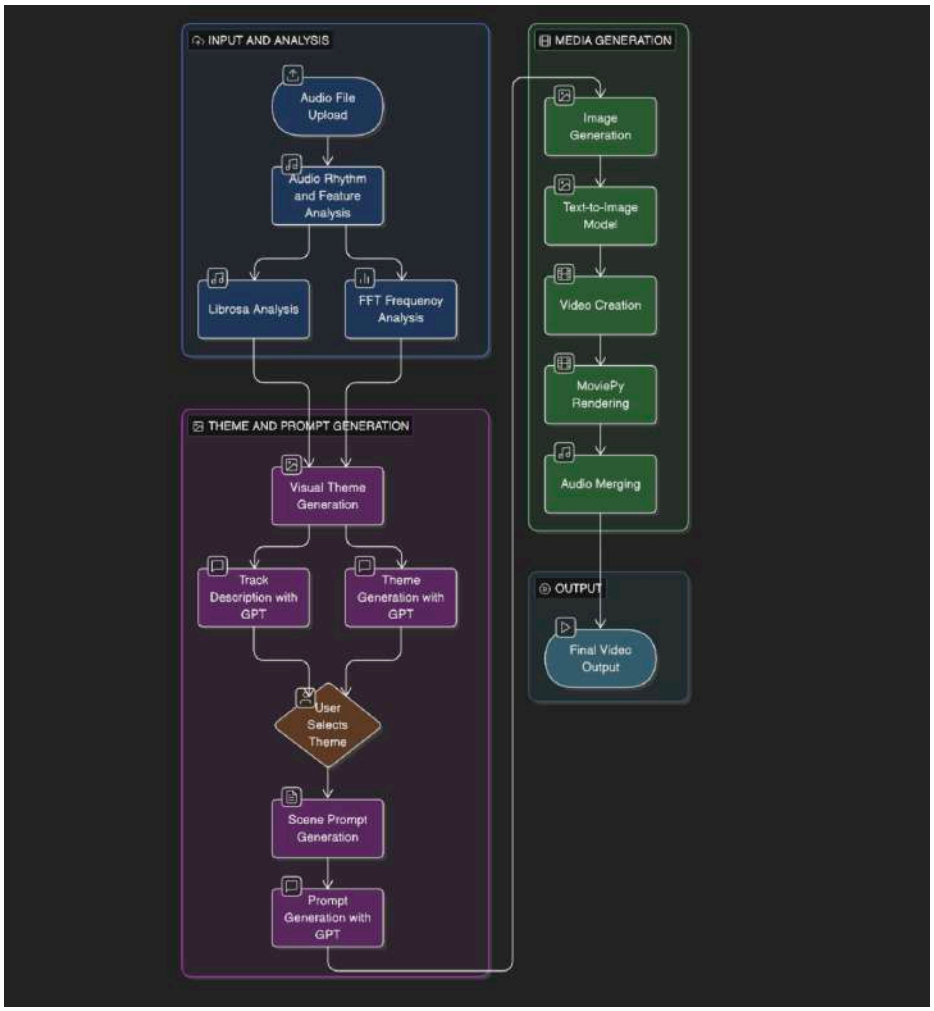


Рисунок 3.1 — Загальна архітектура системи

На рисунку представлено загальну архітектуру системи автоматизованого створення відеокліпів на основі аудіофайлу. Діаграма відображає основні етапи обробки даних та взаємодію між модулями системи. Система побудована за модульною архітектурою, що дозволяє розділити процес генерації відео на логічно

незалежні частини. Загальна логіка роботи виглядає як послідовність етапів: аналіз аудіо, формування змісту, генерація візуального контенту та створення фінального відео.

3.1.1 Вхідні дані та аналіз аудіо

На першому етапі користувач завантажує аудіофайл у систему. Після цього виконується аналіз аудіосигналу, який включає визначення ритму, темпу та інших характеристик. Для цього використовується бібліотека Librosa, щоб визначити темп на ритм. Та методи спектрального аналізу, зокрема FFT, для аналізу частотних характеристик. Результатом цього етапу є набір параметрів, які описують структуру музичного треку та використовуються для синхронізації відео.

3.1.2 Генерація теми та опису треку

Після аналізу аудіо система переходить до інтерпретації отриманих даних. За допомогою AI-моделі формується текстовий опис треку, який включає такі праметри як жанр, настрій та загальну атмосферу композиції. Паралельно генеруються варіанти візуальних тем, які визначають стиль майбутнього відеокліпу. Користувач обирає одну з запропонованих тем, що дозволяє контролювати результат генерації.

3.1.3 Генерація сцен і промптів

Після вибору теми система формує набір сцен, які відповідають структурі аудіо. Кількість сцен визначається на основі ритмічних моментів композиції. Для кожної сцени створюється текстовий опис (prompt), який використовується для

генерації зображень. Цей етап забезпечує зв’язок між музикою та візуальним рядом.

3.1.4 Генерація візуального контенту

На основі сформованих промптів система генерує зображення, які виступають як окремі кадри майбутнього відео. Генерація здійснюється за допомогою text-to-image моделей. Кожне зображення відповідає певному фрагменту аудіо, що дозволяє досягти узгодженості між візуальним рядом і музикою.

3.1.5 Генерація відео

На фінальному етапі система об’єднує всі згенеровані зображення у відеоряд. Тривалість кожного кадру визначається відповідно до ритму аудіо. Для створення відео використовується бібліотека MoviePy, яка дозволяє: сформувати відеоряд із зображень, виконати рендеринг, додати аудіодоріжку. Після цього формується фінальний відеофайл, який передається користувачу.

3.2 Реалізація системи

Реалізація системи виконана у вигляді програмного прототипу мовою Python. Вибір цієї мови обумовлений наявністю бібліотек для обробки аудіосигналів, роботи з AI-моделями та генерації відео.

У процесі розробки використовувались такі основні технології:

- Librosa - для аналізу аудіосигналу (темп, ритм, тривалість);
- NumPy - для роботи з масивами даних[15];

- OpenAI GPT API - для генерації текстового опису треку, тем і сцен;
- text-to-image модель - для генерації зображень на основі промптів;
- MoviePy - для формування фінального відео.
- Sora 2 - генерація відео

Система реалізована як послідовний pipeline, де кожен етап використовує результати попереднього.

3.2.1 Аналіз аудіо та формування опису

На першому етапі система завантажує аудіофайл (у форматі MP3 або WAV) та виконує його аналіз за допомогою бібліотеки Librosa.

У процесі аналізу:

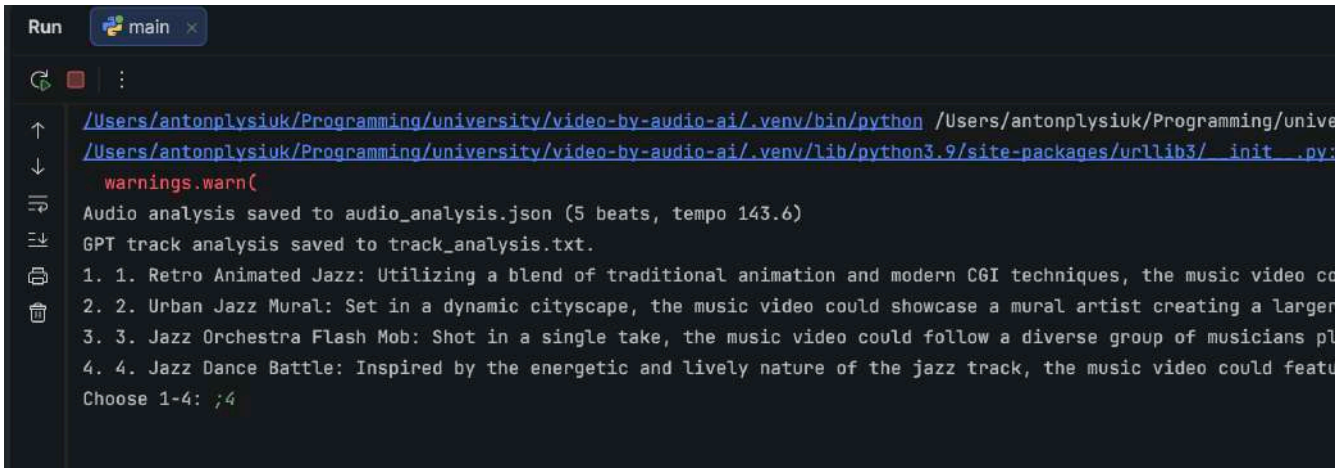
- визначається темп композиції (BPM) за допомогою функції librosa.beat.beat_track;
- знаходяться ритмічні удари (beats);
- визначається тривалість аудіо.

Результати аналізу зберігаються у форматі JSON, що дозволяє передавати їх між модулями системи.

Після цього формується текстовий опис треку за допомогою OpenAI GPT API[16]. Модель отримує інформацію про аудіо та генерує опис, який включає настрій, стиль і загальну атмосферу композиції. На основі цього опису також генеруються кілька варіантів візуальних тем, з яких користувач обирає одну.

3.2.2 Генерація сцен і зображень

На рисунку 3.2 показано приклад згенерованого опису та варіантів тем.



3.2 — приклад відповіді GPT, опис та теми.

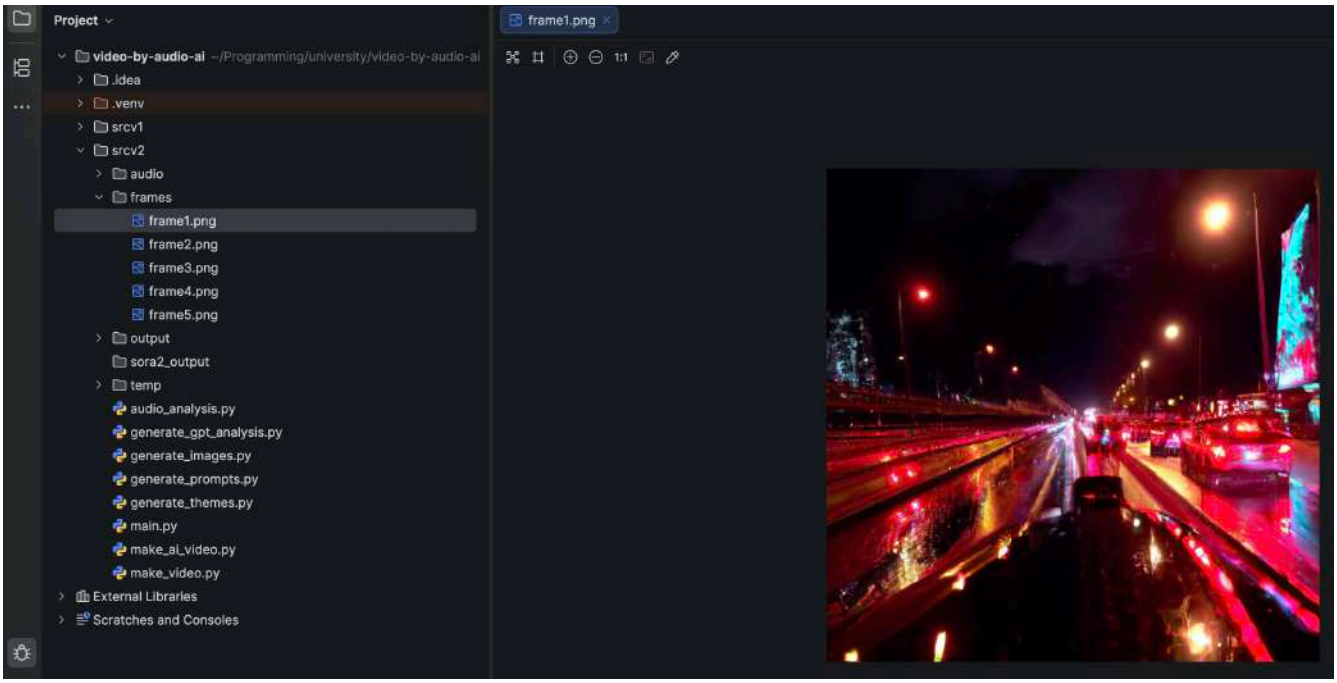
Після того як користувач обирає візуальну тему, система переходить до формування сцен майбутнього відео. На цьому етапі основна задача розбити аудіо на логічні частини та створити для кожної з них візуальний опис.

Кількість сцен визначається на основі ритмічної структури аудіо. Система використовує знайдені біти (ритм), які були отримані під час аналізу, і на їх основі формує ключові моменти, де повинна змінюватися сцена. У поточній реалізації кількість сцен обмежується для того, щоб зменшити час генерації та навантаження на API.

Для кожної сцени формується текстовий опис prompt. Генерація цих описів виконується за допомогою OpenAI GPT API. Модель отримує інформацію про:

- обрану тему;
- загальний настрій треку;
- позицію сцени в часі (наприклад, початок, середина або кінець);

На рисунку 3.3 наведено приклади фреймів, створених на основі вибраної теми.



3.3 — Згенеровані фрейми на основі вибраної теми

На основі цих даних створюються описи сцен, які виглядають як короткі текстові інструкції для генерації зображень. Наприклад, це може бути опис сцени з певною атмосферою, кольорами або стилем.

Після формування промптів система переходить до генерації зображень. Для цього використовується text-to-image[17] модель. Кожен prompt передається в модель, яка генерує відповідне зображення.

Згенеровані зображення зберігаються у вигляді окремих файлів (frame1.png, frame2.png тощо). Кожне зображення відповідає певній сцені та буде використане як кадр у фінальному відео. Такий підхід дозволяє зв’язати музику та візуальний контент.

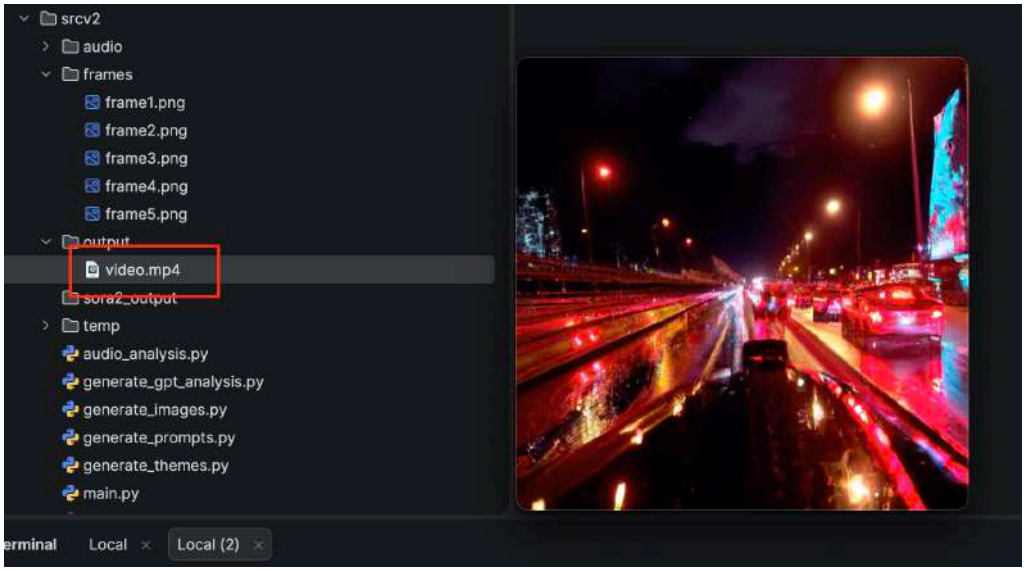
3.2.3 Генерація відео

На фінальному етапі система переходить до створення відеофрагментів. На відміну від попередніх етапів, де створювались окремі зображення, тут система вже формує повноцінні відеосцени за допомогою AI-моделі Sora 2.

Для кожного prompt-а система надсилає окремий запит до моделі генерації відео. Prompt містить опис сцени, атмосфери, руху камери та візуального стилю. На основі цього опису Sora 2 автоматично створює короткий відеофрагмент, який відповідає настрою композиції та обраній темі.

Після завершення генерації всі отримані відеофрагменти автоматично зберігаються у локальній директорії системи. Далі виконується етап об’єднання відео та аудіо.

Для формування фінального відеокліпу використовується бібліотека MoviePy [18]. Система послідовно об’єднує всі згенеровані фрагменти. Після цього до відео додається початкова аудіодоріжка. Згенеровані фрейми використовуються для створення фінального відео, що показано на рисунку 3.4.



3.4 — Згенероване відео на основі фремів

3.3 Приклад роботи системи

3.3.1 Аналіз аудіофайлу

Для демонстрації роботи системи було використано музичний трек у форматі MP3. Після запуску програми аудіофайл автоматично передається до модуля аналізу.

На цьому етапі система визначає:

- темп композиції;
- ритмічні удари;
- тривалість треку;
- загальні характеристики аудіо.

Для аналізу використовується бібліотека Librosa. Отримані параметри зберігаються у файл audio_analysis.json та використовуються на наступних етапах роботи системи.

Після завершення аналізу система автоматично формує текстовий опис композиції за допомогою OpenAI GPT API. На основі цього опису генерується кілька можливих візуальних тем для майбутнього відеокліпу.

Користувач обирає одну із запропонованих тем. У даному прикладі було обрано тему Electric Vibe, яка описує неоновий стиль та динамічну атмосферу відео.

На рисунку 3.5 показано результат аналізу аудіо та приклад згенерованих тем.

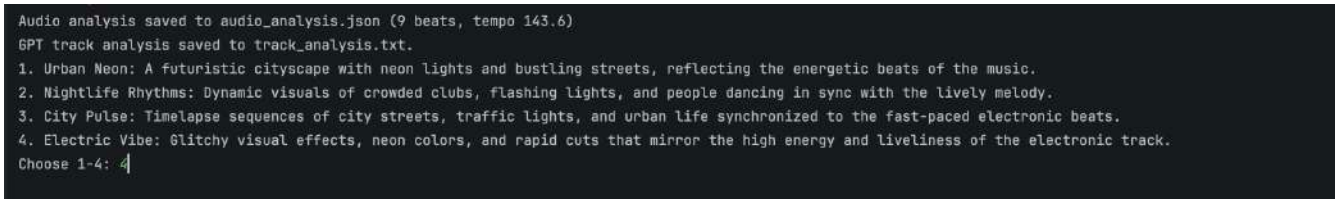


Рисунок 3.5— Аналіз аудіофайлу та генерація варіантів тем

3.3.2 Створення сцен та prompt-ів

Після вибору теми система переходить до створення сцен для майбутнього відеокліпу.

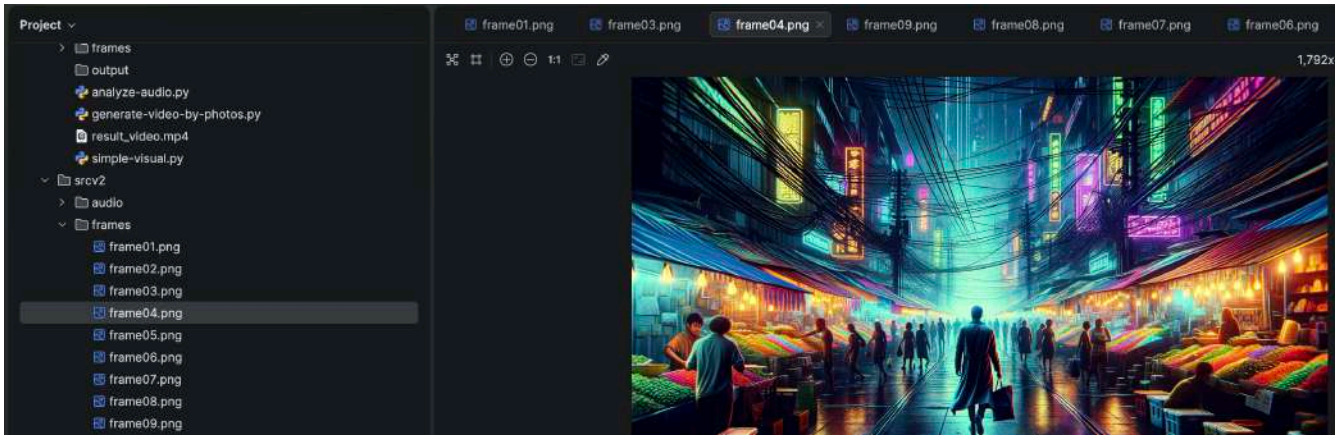
Кількість сцен визначається на основі ритму композиції та кількості виявлених ударів. Для кожної сцени система автоматично формує окремий prompt, який містить опис кадру, атмосфери, руху камери та візуального стилю.

У даному прикладі система сформувала 9 окремих сцен та зберегла їх у файл prompts.txt.

На рисунку 3.6 показано процес генерації сцен та створення prompt-ів.

```
Theme 'Electric Vibe: Glitchy visual effects, neon colors, and rapid cuts that mirror the high energy and liveliness of the electronic track.'
9 prompts saved to prompts.txt.
Generating: frames/frame01.png (A young woman wearing a vibrant neon jacket dances in a down...)
Generating: frames/frame02.png (As the camera pans out, the cityscape emerges around her, sk...)
Generating: frames/frame03.png (Cut to a rapid montage of close-up shots: hands tapping on g...)
Generating: frames/frame04.png (Transition to a bustling market street, vendors selling colo...)
Generating: frames/frame05.png (Zoom in on her face, pure joy and excitement evident in her ...)
Generating: frames/frame06.png (A sudden glitch effect distorts the scene briefly before sna...)
Generating: frames/frame07.png (The camera follows her as she ascends a spiral staircase, ea...)
Generating: frames/frame08.png (Close-up on her silhouette against the neon skyline, her fig...)
Generating: frames/frame09.png (Final shot of her standing tall at the edge of the rooftop, ...)
All frames generated.
```

a)



b)

3.6 — Генерація сцен та prompt-ів: a) генерація фреймів; b) згенеровані фремів за допомогою Chat GPT

3.3.3 Генерація відеофрагментів

Після створення prompt-ів система переходить до генерації відеофрагментів за допомогою AI-моделі Sora 2.

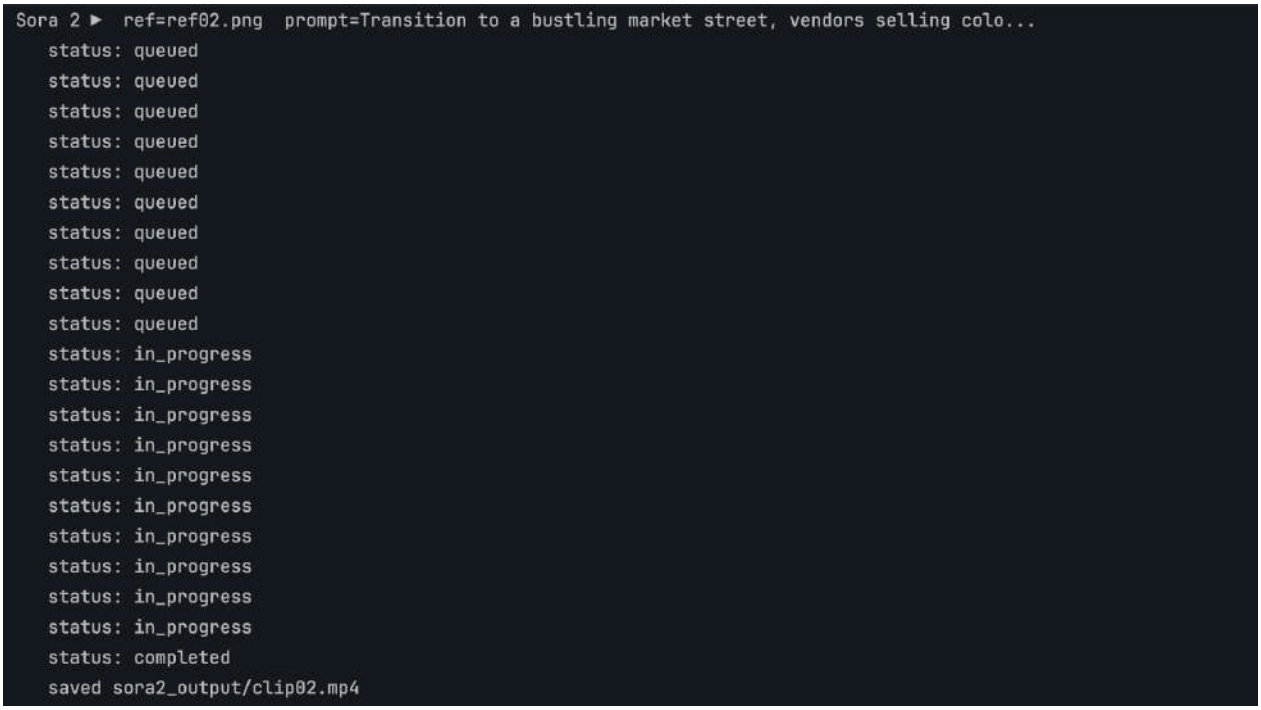
Для кожної сцени формується окремий запит до моделі, після чого система автоматично створює короткі відеофрагменти, які відповідають обраній темі та настрою музики.

Під час генерації відображається поточний статус виконання:

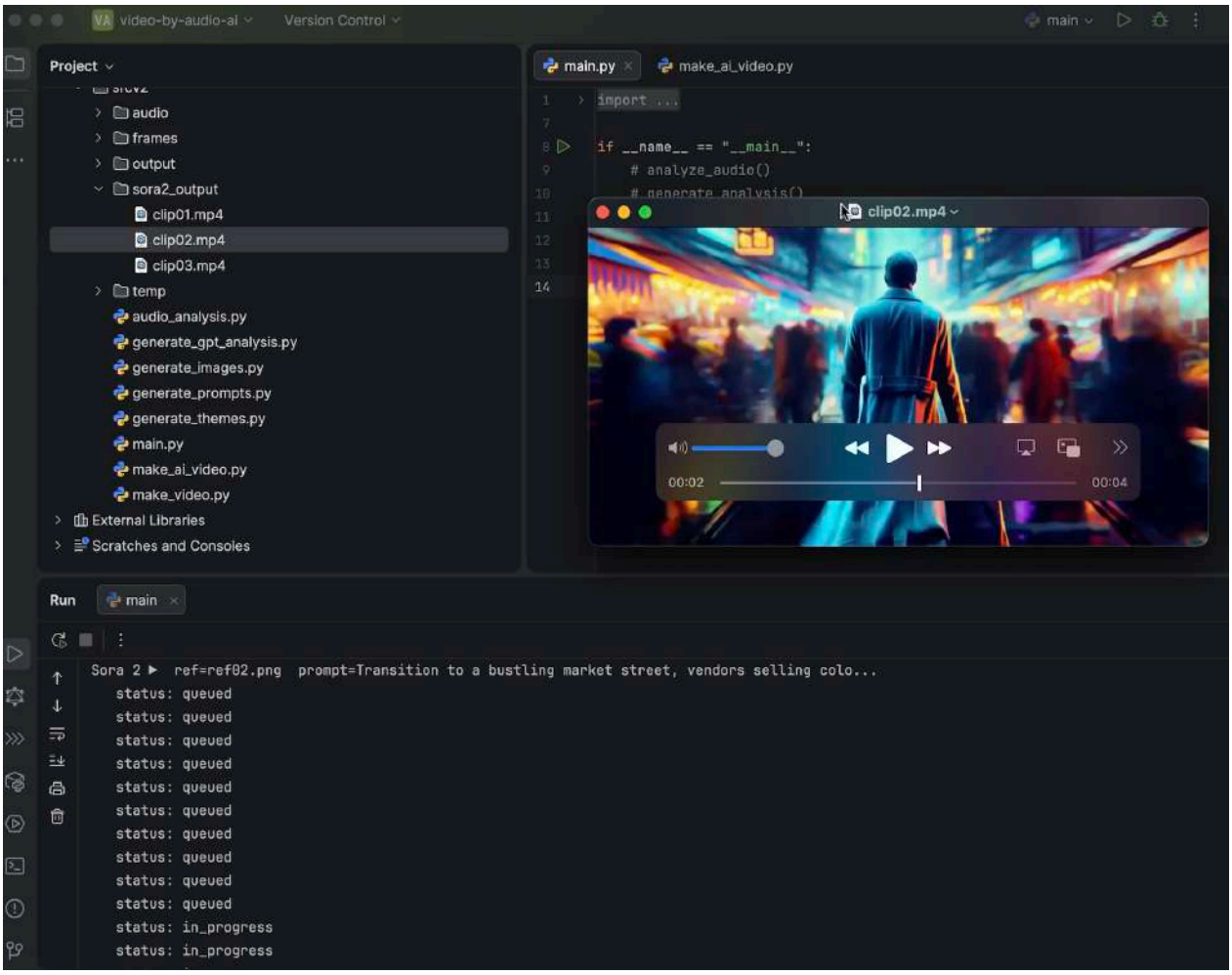
- queued — запит знаходиться у черзі;
- in_progress — виконується генерація;
- completed — генерацію завершено.

Після завершення генерації готові відеофрагменти автоматично зберігаються у локальній директорії системи.

На рисунку 3.7 показано процес генерації відеофрагмента та зміну статусів виконання.



а)



b)

Рисунок 3.7 — Генерація відеофрагмента за допомогою Sora 2: a) Процес генерації відео; b) Згенеровані фрагменти відео

3.3.4 Формування фінального відео

На фінальному етапі система об’єднує всі згенеровані відеофрагменти у єдиний відеоряд. Для цього використовується бібліотека MoviePy.

Система синхронізує зміну сцен із ритмом композиції, після чого додає початкову аудіодоріжку. Завдяки цьому фінальний відеокліп виглядає цілісним та узгодженим із музикою.

Після завершення рендерингу система формує фінальний відеофайл у форматі MP4.

У результаті тестування було підтверджено, що система успішно виконує повний цикл автоматизованого створення музичного відеоконтенту: від аналізу аудіо до генерації готового відеокліпу.

Висновки до розділу

У даному розділі було розглянуто реалізацію системи автоматизованого створення відеокліпів на основі аудіо. Описано загальну архітектуру системи, а також детально розглянуто основні етапи її роботи від аналізу аудіосигналу до формування фінального відеофайлу.

У процесі реалізації було використано сучасні програмні інструменти та бібліотеки, зокрема Librosa для аналізу аудіо, OpenAI GPT API для генерації текстового опису та сцен, text-to-image моделі для створення зображень і MoviePy для формування відео. Це дозволило об’єднати різні технології в єдину систему.

На прикладі роботи було продемонстровано, що система здатна автоматично створювати відеоконтент, узгоджений з аудіо. Зміна сцен відбувається відповідно до ритму, а візуальний стиль відповідає настрою музичного треку.

Таким чином, розроблена система підтверджує можливість ефективного поєднання методів обробки аудіо та AI-технологій для автоматизованої генерації відеокліпів. Отримані результати можуть бути використані як основа для подальшого розвитку більш складних мультимедійних систем.

4 РЕЗУЛЬТАТИ ТА ПЕРСПЕКТИВИ РОЗВИТКУ

У результаті виконання дипломної роботи було розроблено систему автоматизованого створення музичних відеокліпів на основі аналізу аудіо та використання AI-технологій. Система реалізує повний цикл обробки від завантаження аудіофайлу до формування готового відео.

Під час тестування було підтверджено, що система коректно виконує:

- аналіз музичних характеристик аудіо;
- генерацію текстового опису композиції;
- створення візуальних тем;
- формування сцен;
- генерацію відеофрагментів;
- об’єднання відео та аудіо у фінальний результат.

Отримані результати показують, що використання AI-моделей дозволяє значно спростити процес створення музичного відеоконтенту та автоматизувати більшість етапів роботи.

Разом із цим під час розробки та тестування було виявлено ряд обмежень системи.

Однією з основних проблем є нестабільність генерації відеофрагментів. AI-модель не завжди створює сцени однакової якості або стилю. У деяких випадках між окремими фрагментами можуть виникати різкі зміни атмосфери, освітлення або композиції кадру.

Також присутні обмеження, пов’язані з аналізом аудіо. Визначення ритмічних ударів працює коректно не для всіх типів музики. Наприклад, для

повільних або атмосферних композицій система може неточно визначати beats, що впливає на синхронізацію відео зі звуком.

Окремою проблемою є формування сцен. На даний момент система генерує сцени переважно на основі ритму та загального опису композиції. Через це відео може виглядати недостатньо зв’язним або мати повторювані елементи.

Перспективи подальшого розвитку системи пов’язані з покращенням якості генерації та більш глибоким аналізом аудіо.

У майбутньому систему можна вдосконалити шляхом:

- використання більш точних алгоритмів аналізу аудіо;
- додавання аналізу структури композиції (куплет, приспів, перехід);
- покращення синхронізації сцен із музикою;
- використання кращих AI-моделей генерації відео;
- створення більш зв’язних та послідовних сцен;
- додавання можливості редагування результату користувачем;
- генерації повноцінного відео без поділу на окремі фрагменти.

Таким чином, розроблена система демонструє можливість автоматизованого створення музичних відеокліпів та може використовуватись як основа для подальших досліджень у сфері AI-генерації мультимедійного контенту.

ВИСНОВКИ

У ході виконання магістерської роботи практично доведено можливість автоматизованої генерації музичних відеокліпів за допомогою сучасних методів аналізу аудіосигналів та генеративних моделей штучного інтелекту. Розроблено AI-систему, яка виконує аналіз музичної композиції, визначає її основні характеристики та формує відеоряд, синхронізований із музичним супроводом.

У рамках виконання дослідження було проаналізовано предметну область генерації мультимедійного контенту, сучасні інструменти генерації відео та методи аналізу музичних характеристик аудіо. Сформовано функціональні вимоги до системи та визначено підходи до реалізації основних компонентів.

Досліджено та порівняно методи аналізу аудіосигналу. Для реалізації системи обрано підхід на основі часових, спектральних і ритмічних ознак. Розроблено алгоритм аналізу музичних характеристик, який дозволяє автоматично формувати параметри для генерації сцен і візуальних елементів.

Розроблено архітектуру AI-системи та реалізовано основні функціональні модулі: аналіз аудіо, генерацію описів сцен, генерацію зображень і відеофрагментів та формування фінального відеокліпу. Проведено тестування системи на композиціях різних жанрів, результати якого підтвердили можливість створення відеоряду, що відповідає темпу та емоційній складовій музики.

Наостанок описано практичні сфери застосування розробленої системи, її поточні обмеження та перспективи розвитку. Подальшими напрямами розвитку можуть стати покращення синхронізації відео з музичними подіями, використання більш сучасних моделей генерації відео та реалізація генерації сцен у режимі реального часу.

СПИСОК ЛІТЕРАТУРИ

1) Runway ML. *Runway AI Tools* [Електронний ресурс]. — Режим доступу до ресурсу: https://runwayml.com/

2) Rotor Videos. *Music Video Generator* [Електронний ресурс]. — Режим доступу до ресурсу: https://rotorvideos.com/

3) Oppenheim A. V., Schafer R. W. *Discrete-Time Signal Processing*. — Pearson, 2010.

4) Smith S. W. The Scientist and Engineer's Guide to Digital Signal Processing. — California Technical Publishing.

5) McFee B., Raffel C., Liang D. et al. Librosa: Audio and Music Signal Analysis in Python. — 2015.

6) Müller M. Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications. — Springer, 2015.

7) Goodfellow I., Bengio Y., Courville A. *Deep Learning*. — MIT Press, 2016.

8) Chollet F. *Deep Learning with Python*. — Manning Publications, 2018.

9) TutorialsPoint. *Representation of a Discrete Time Signal* [Електронний ресурс]. — Режим доступу до ресурсу: https://www.tutorialspoint.com/signals_and_systems/representation_of_a_discrete_time_signal.htm

10) TutorialsPoint. *Energy and Power Signals* [Електронний ресурс]. — Режим доступу до ресурсу: https://www.tutorialspoint.com/signals_and_systems/energy_and_power_signals.htm

11) GeeksforGeeks. *Fourier Transform* [Електронний ресурс]. — Режим доступу до ресурсу: https://www.geeksforgeeks.org/maths/fourier-transform/

12) Müller M. *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications.* — Springer, 2015.

13) Wikipedia. *Feature Vector* [Електронний ресурс]. — Режим доступу до ресурсу: https://en.wikipedia.org/wiki/Feature_(machine_learning)

14) NumPy Developers. *NumPy Documentation* [Електронний ресурс]. — Режим доступу до ресурсу: https://numpy.org/doc/

15) Wikipedia. *Feature Scaling* [Електронний ресурс]. — Режим доступу до ресурсу: https://en.wikipedia.org/wiki/Feature_scaling

16) OpenAI. *GPT Models Documentation* [Електронний ресурс]. — Режим доступу до ресурсу: https://platform.openai.com/docs

17) OpenAI. *DALL·E API Documentation* [Електронний ресурс]. — Режим доступу до ресурсу: https://platform.openai.com/docs/guides/images

18) MoviePy Developers. *MoviePy Documentation* [Електронний ресурс]. — Режим доступу до ресурсу: https://zulko.github.io/moviepy/