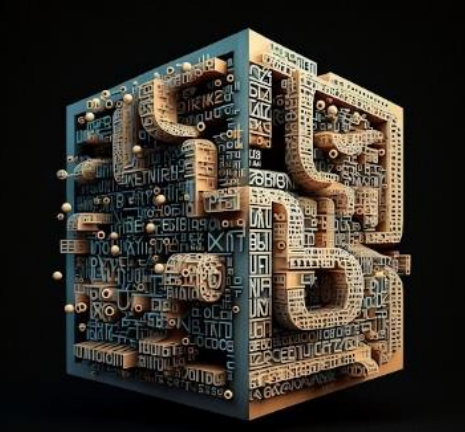


**Міністерство освіти та науки України
Харківський національний економічний
університет імені Семена Кузнеця**



**НАВЧАЛЬНО – НАУКОВИЙ ІНСТИТУТ
ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ**
кафедра: Мультимедійних систем і технологій

дисципліна
"ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ"
викладач: доцент, к.т.н.
Токарєв Володимир Володимирович



Тема. Технології обробки мультимедіа. Редагування аудіоконтенту.

АУДІОКОНТЕНТ



АУДІОКОНТЕНТ



Музика



Аудіокниги



Подкасти



Мовні курси



Радіо



Аудіошоу



Музика



Радіо



Звукові ефекти

Аудіоконтент — це сукупність звукових даних, представлених у цифровій формі та призначених для:

- відтворення;
- обробки;
- або поширення через різні медіаплатформи.

До аудіоконтенту належать:

- музичні твори;
- подкасти;
- радіопрограми;
- аудіокниги;
- саунд-дизайн;
- рекламні аудіоролики;
- озвучування відеоінтерактивний звук (ігри,

AR/VR).

AR-доповнена реальність — це технологія накладання цифрових об'єктів:

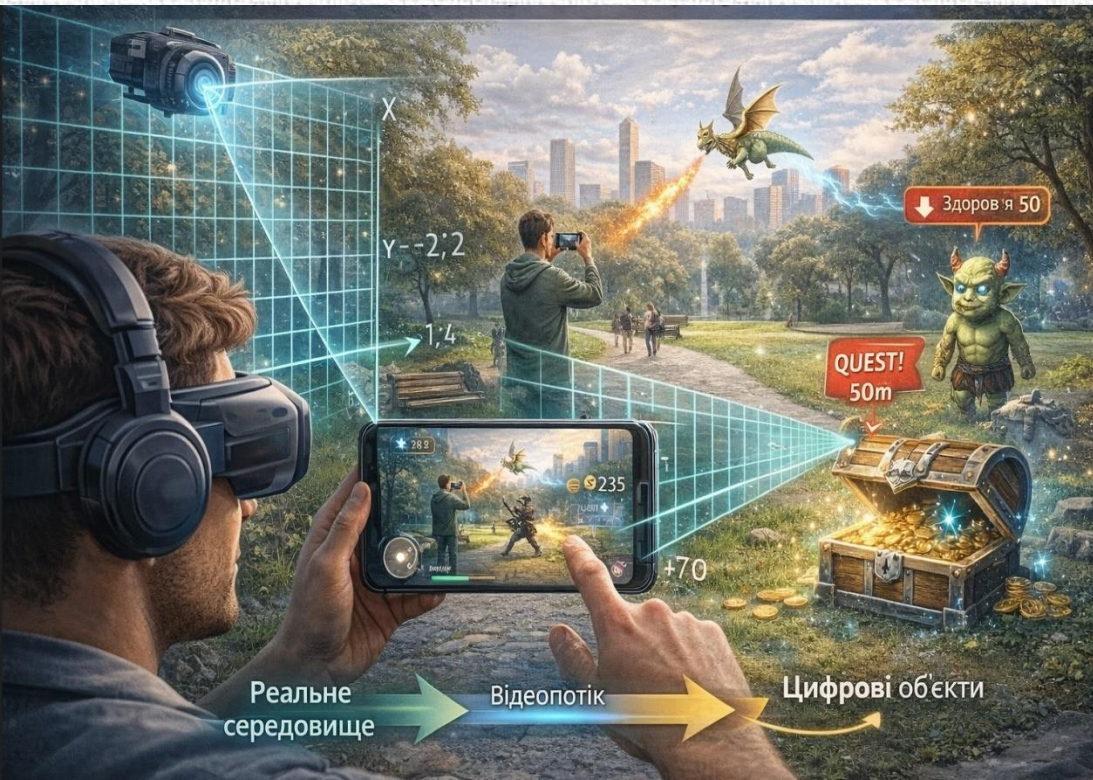
- 3D-моделей;
- тексту;

-анімації на реальний світ у режимі реального часу.

Принцип роботи.

1. Камера зчитує реальне середовище.
2. Система визначає просторові координати.
3. Цифрові об'єкти інтегруються у відеопотік.
4. Користувач бачить поєднання реального та віртуального.

Парк/міський пейзаж — це фізичний світ, який існує незалежно від **AR**. Він є "опорною сценою" для прив'язки віртуальних об'єктів. **AR** не створює світ з нуля як **VR**, а накладає цифрові елементи на реальний світ.



Камера/оптичний сенсор і просторова сітка (grid). Це символізує:

- захоплення відеокадру;
- оцінку глибини структури;
- побудову карти простору та системи координат.

На сітці з'являються осі X, Y та числа — умовні координати/положення.

Принцип роботи.

Технічний зміст зображення - **AR-система** виконує трекінг камери (позиція + орієнтація) та сприйняття сцени площини/глибина.

Це близько до задач - SLAM - Simultaneous Localization and Mapping - Одновременная локализация и построение карты.



Принцип роботи.

Користувач тримає смартфон або використовує **AR-гарнітуру**. Це означає:

- камера пристрою постійно знімає сцену;
- датчики - (акселерометр або гіроскоп) допомагають визначати рух;

- графічний процесор виконує обчислення й рендер.

Сенс: пристрій є одночасно **сенсором** (вхід) і **екраном** (вихід).



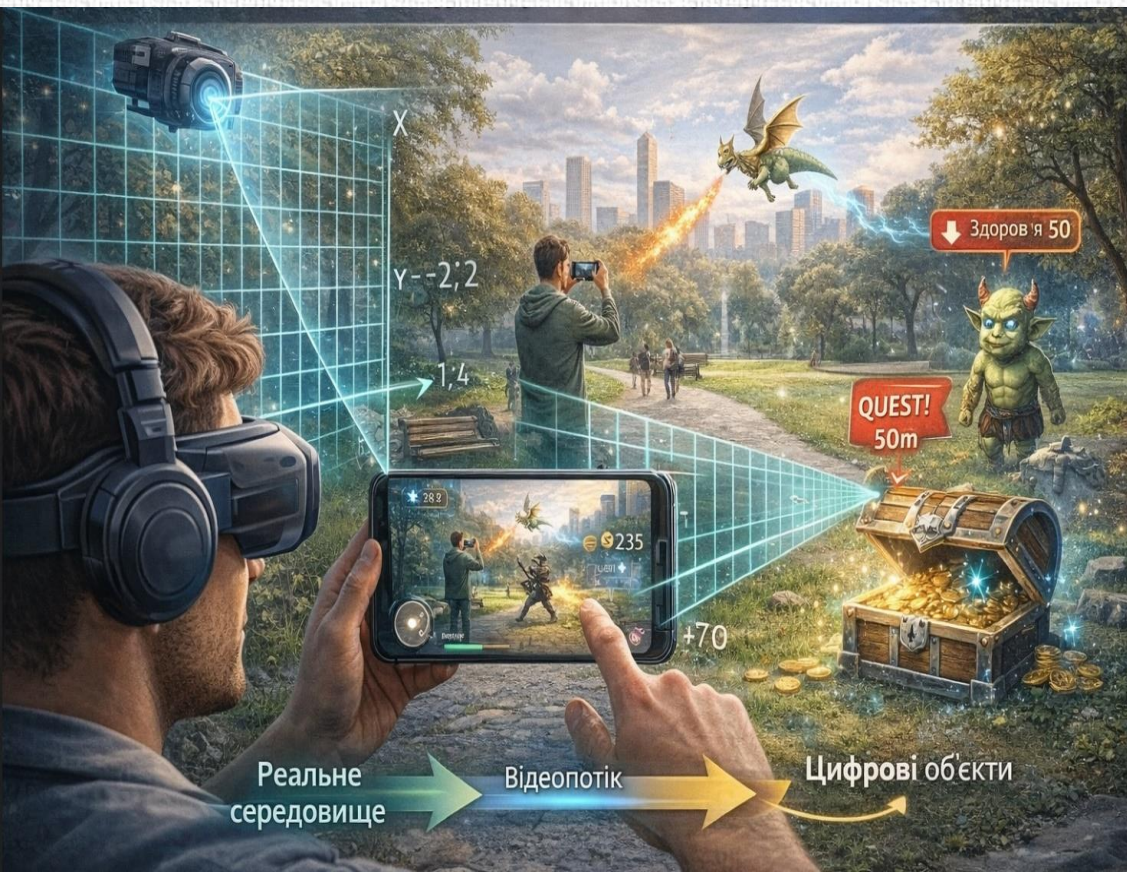
Принцип роботи.

“Реальне середовище” → “Відеопотік” → “Цифрові об’єкти”.

Відеопотік (екран смартфона)

Екран смартфона показує “живе” відео з камери — це і є відеопотік. На ньому видно:

- реальний пейзаж;
- віртуальні об’єкти, “вбудовані” у сцену;
- UI**-елементи - (бали, іконки, написи).



Технічний зміст

відбувається композиція:

- кадр з камери;
- 3D**-рендер цифрових об’єктів;
- накладання UI;
- відображення користувачу в реальному часі з частотою **30–60 FPS**.

Користувач бачить поєднання реального та віртуального.

Принцип роботи.

Цифрові об'єкти (права частина: "скриня", персонажі, маркери)

Праворуч видно ігрові, віртуальні елементи: скриня, "квест 50 м", індикатор "здоров'я", істоти тощо. Це приклади **цифрових об'єктів**, які:



- мають **3D-позицію** у просторі сцени;
 - масштабуються або обертаються відповідно до перспективи;
 - можуть взаємодіяти з користувачем (торкання, жест, наведення).
- Щоб об'єкти виглядали "в сцені", система повинна правильно:
- обчислити позу камери;
 - виконати проєкцію **3D→2D**;
 - узгодити освітлення/тіні (на якісних **AR**-сценах).

Принцип роботи.

Які технології використовуються:

1. Зчитування середовища - комп'ютерний зір.
2. Координати – **SLAM**, трекінг.
3. інтеграція у відео - **3D**-рендеринг, композиція кадру.
4. сприйняття користувачем - **UX/UI**, освітлення.



(VR)-віртуальна реальність - це повністю штучне цифрове середовище. Користувач через **VR-гарнітуру** занурюється у повністю штучне цифрове середовище, яке не є "накладкою" на реальний світ, а повністю його заміщує.



На фоні зображено фантастичний світ:

- замок;
- інша планета;
- дракон;
- лицар;
- портал;
- космічний корабель.

Це підкреслює ключову відмінність VR:

VR це повна заміна реальної сцени на **3D**-сцену, згенеровану комп'ютером.

На відміну від **AR**, тут немає відеопотоку з камери як основи — реальність "відрізана" від сприйняття.



Користувач у VR-гарнітурі

Людина в кадрі одягнула **VR-шолом/гарнітуру** (**HMD** — Head-Mounted Display), тримає два контролери.

Це означає дві фундаментальні підсистеми VR:

-виведення (display) - два окремі зображення для лівого/правого ока → стереозір → відчуття глибини.

-трекінг (tracking) - система відстежує голову і руки/контролери → користувач може оглядатися і взаємодіяти.



Контролери символізують інтерфейс взаємодії:

- рухи рук переносяться у віртуальний світ;
- натискання кнопок/тригерів → дії (взяти предмет, стріляти, меню, тощо);
- разом із трекінгом голови це створює відчуття присутності.

Піксельна/цифрова "рамка" навколо сцени — художній прийом, що показує:

- світ складається з цифрових даних/рендерингу;
- це "перехід" від реальності до симуляції;
- акцент на тому, що середовище синтетичне.





Віртуальна сцена та об'єкти:

- замок;
- гори;
- інша планета;
- дракон;
- лицар;
- портал.

Це віртуальні об'єкти **3D**-моделі з анімацією та фізикою, які реагують на дії користувача. Портал — типовий елемент **VR**-сценаріїв. Дракон/вогонь — демонстрація динамічних ефектів (VFX). Космічний корабель - підкреслює, що можна моделювати будь-які світи й події.



Які технології використовуються:

1. Трекінг пози - голова/руки.
2. Рендеринг у реальному часі - дві проєкції (ліве/праве око), висока частота кадрів.
3. Низька затримка.
4. Стерео - широке поле зору формує ефект "я всередині".



Редагування аудіоконтенту — це процес корекції, обробки та структуризації звукового матеріалу з метою досягнення високої якості звучання та відповідності поставленому завданню:

- подкаст;
- музика;
- лекція;
- реклама тощо.



Спеціаліст за робочою станцією з навушниками.

Це підкреслює ключову практику аудіомонтажу:

- критичне прослуховування (critical listening);
- контроль артефактів (шум, кліки, спотворення);
- перевірка балансу частот і динаміки.

Навушники означають, що людина слухає деталі, які можуть губитися на колонках у кімнаті.



Інтерфейс **DAW** (монітори, доріжки, мікшер)



На основних екранах видно:

- хвильові форми (waveforms);
- кілька аудіодоріжок різного кольору (голос, музика, **SFX** - Sound Effects - звукові ефекти);
- мікшер/фейдери (зведення рівнів).

Це відповідає типовій цифровій схемі редагування:

- імпорт/запис матеріалу;
- монтаж по таймлайну;
- обробка (ефекти);
- зведення;
- фінальний експорт (render).



Узагальнена логіка, яку показує рисунок

Модель редагування аудіоконтенту:
корекція + обробка + структуризація

Базові кроки:

1. Структура - різання/монтаж;
2. Якість - шумоподавлення/корекції;
3. Звучання - ефекти;
4. Фінальний продукт - зведення і експорт.



Базовий процес редагування аудіоконтенту

Редагування аудіоконтенту



Базовий процес редагування аудіоконтенту



1. Запис (Recording / Capture / Voiceover).

Захоплюємо "сирий" аудіосигнал:

- голос;
- інструменти;
- атмосфера;
- інтерв'ю.

Налаштовуємо мікрофон, інтерфейс, рівні.

Якість запису визначає: 70–80% кінцевого результату. Поганий запис "не врятує" жоден плагін. Рівень піків під час запису: приблизно $-12...-6$ dBFS (без кліпінгу). Відстань до мікрофона – стабільна.

Зменшення шумів - вимкнути перешкоди. Формат - WAV, 24-bit, 48 kHz (для відео) або 44.1 kHz (для музики). Контроль якості - немає кліпінгу (0 dBFS), немає перевантаження входу.

Базовий процес редагування аудіоконтенту



Прибираємо дефекти, які заважають сприйняттю:

- шум фону;
- гул мережі;
- кліки;
- зайве дихання, тощо.

Підготувати матеріал до монтажу й подальшої обробки так, щоб ефекти/компресія не "підняли" шум.

Типові операції:

- прибрати клацання/тріск;
- прибрати гармоніки;
- спектральне шумозаглушення;
- приглушити паузи.

Після шумозаглушення голос не стає "металевим/роботизованим". Пауза звучить природно (часто краще залишити легкий фон, ніж "мертву тишу").

Базовий процес редагування аудіоконтенту



Розглянемо докладніше типові операції

1. Типова операція - прибрати клацання/тріск.

Що таке – клацання – це дуже короткий імпульс (1–10 мс), часто одиничний. Причини - дотик мікрофона, поганий контакт кабелю, цифровий глітч.

Що таке – тріск - серія дрібних імпульсів, "шурхіт" або "пісок" (як вініл/поганий конектор/ перевантаження).

Як працює алгоритм.

1. Виявляє аномалії з дуже крутим фронтом (надто швидка зміна амплітуди).

2. Визначає короткий інтервал навколо дефекту.

3. Замінює його на інтерпольований сигнал:

- часовою інтерполяцією (по сусідніх семплах);

- або в спектрі (як "заміна" по частотних сусідах).

2. Типова операція - прибрати гармоніки.

Що таке – прибрати гармоніки – стабільний низькочастотний гул електричної мережі.

Базовий процес редагування аудіоконтенту

Розглянемо докладніше типові операції

Причини - наводки, земляні петлі, погані блоки живлення, близькість до трансформаторів.

Найчастіше це **набір вузьких "піків"** у спектрі. Тому застосовують спеціальний модуль, який автоматично знаходить фундамент і гармоніки та пригнічує їх.

Clean-up



Видалення шуму,
кліків, дихання

3. Типова операція - спектральне шумозаглушення.

Постійний фон - вентиляція, кондиціонер, комп'ютерний шум, вуличний "шум міста". Як працює (класична схема). Береться профіль шуму з паузи. Будується спектральна модель шуму. Із кожного короткого вікна сигналу, шум віднімається/пригнічується, виконується - спектральне віднімання, потім застосовуються методи машинного навчання **ML-методи** (в сучасних інструментах).



Association for
Computing Machinery

Advancing Computing as a Science & Profession

Базовий процес редагування аудіоконтенту

Розглянемо докладніше типові операції

Причини - наводки, земляні петлі, погані блоки живлення, близькість до трансформаторів.

Найчастіше це **набір вузьких "піків"** у спектрі. Тому застосовують спеціальний модуль, який автоматично знаходить фундамент і гармоніки та пригнічує їх.

Clean-up



Видалення шуму,
кліків, дихання

3. Типова операція - спектральне шумозаглушення.

Постійний фон - вентиляція, кондиціонер, комп'ютерний шум, вуличний "шум міста". Як працює (класична схема). Береться профіль шуму з паузи. Будується спектральна модель шуму. Із кожного короткого вікна сигналу, шум віднімається/пригнічується, виконується - спектральне віднімання, потім застосовуються методи машинного навчання **ML-методи** (в сучасних інструментах).

4. Типова операція - приглушити паузи. Зменшує рівень тихих ділянок, але не обнуляє їх. Навіщо це роблять. Потрібно зменшити чутність шуму в паузах. Прибрати зайві звуки між фразами (шурхіт, віддалений фон).

Базовий процес редагування аудіоконтенту



Що робимо. Формуємо структуру - вирізаємо помилки, повтори, зайві паузи. Склеюємо дублі, переставляємо фрагменти, робимо "чисті" переходи.

Навіщо ми це робимо - щоб зробити матеріал логічним, динамічним і зрозумілим під задачу реклама ≠ подкаст.

Типові прийоми:

- точне обрізання по нульовому перетину;
- короткі кросфейди на стиках (щоб не було "клацання");
- підкладання кімнатного фону в місцях склейок;
- легке вкорочення пауз без втрати природності.

Після цього:

- немає "стрибків" фону на склейках;
- темп мовлення природний, не "перемотаний".

Базовий процес редагування аудіоконтенту



Що робимо. Налаштовуємо тембр і динаміку – еквайзер, компресор, сатурація, реверб (за потреби).

Навіщо ми це робимо - підняти розбірливість, вирівняти гучність, зробити звук "професійним" та цільовим для платформи.

Після цього:

- Голос не "плоский" і не "перекомпресований".
- Немає різкого свисту/сіблянтів після підсилення високих.
- Немає кліпінгу на виході плагінів.

Базовий процес редагування аудіоконтенту



Що робимо. Балансуємо всі доріжки - голос, музика, ефекти, інтерв'ю. Виставляємо гучності, панораму (якщо стерео), автоматизацію. Музика "присідає", коли говорить диктор.

Навіщо ми це робимо. Забезпечити чітку ієрархію - головний зміст (голос) завжди читабельний, решта - підтримує.

Типові правила. Голос це "центр", музика на $-10...-20$ dB відносно голосу (орієнтовно, залежить від жанру).

Контроль сумісності. моно/стерео, фаза, відсутність "провалів".

Після цього:

- На телефонах/ноутбуці голос не зникає під музику.
- Немає перевантаження на низьких частотах.

Базовий процес редагування аудіоконтенту



Що робимо. Виводимо готовий матеріал у потрібний формат, частоту дискретизації, бітність, кодек.

Навіщо ми це робимо. Платформи мають різні вимоги для аудіоформату та параметрів звуку, які застосовуються під час експорту мультимедійного матеріалу.

- стримінг;
- подкасти;
- як звукова доріжка у відео.

Типові варіанти:

- Для стримінгу - WAV 24-bit, 48 kHz.
- Для подкасту - MP3 128–192 kbps або AAC.
- Для відео - обов'язково частота дискретизації аудіо - 48 kHz.

Після цього:

- Перевірка початку/кінця немає обрізаних хвостів.
- Метадані (назва, автор, обкладинка) — якщо потрібно.

Базовий процес редагування аудіоконтенту

Приклад (YouTube)



Контейнер - MP4.

Video - H.264.

Audio:

-**AAC**- 48 kHz.

-**Stereo** - 320 kbps.

Розглянемо докладніше кожен пункт і що він означає на практиці

1.Контейнер - MP4 - це контейнерний формат (упаковка), який зберігає всередині:

- відеопотік (video stream);
- аудіопотік (audio stream);
- за потреби субтитри, метадані тощо.

Важливо - контейнер \neq кодек. **MP4** визначає "**коробку**", а не спосіб стиснення.

Практичний сенс. MP4 - найкраща сумісність із плеєрами, смартфонами, YouTube, соцмережами.

Базовий процес редагування аудіоконтенту

Приклад (YouTube)



Контейнер - MP4.

Video - H.264.

Audio:

-**AAC**- 48 kHz.

-**Stereo** - 320 kbps.

Розглянемо докладніше кожен пункт і що він означає на практиці

2.Video - H.264 - це відеокодек, тобто алгоритм стиснення відеоряду:

Що дає H.264:

- висока сумісність (практично всюди відтворюється);
- помірний розмір файлу при хорошій якості.

Що це означає для експорту:

-ви кодуєте картинку (кадри) у H.264 і кладете її в контейнер MP4.

-якість/розмір найбільше визначаються відеобітрейтом, роздільною здатністю, FPS, профілем/рівнем H.264.

Базовий процес редагування аудіоконтенту

Приклад (YouTube)



Контейнер - MP4.

Video - H.264.

Audio:

-**AAC**- 48 kHz.

-**Stereo** - 320 kbps.

Розглянемо докладніше кожен пункт і що він означає на практиці

3. Audio - AAC, 48 kHz, Stereo, 320 kbps.

Це параметри аудіодоріжки у відео. AAC (Advanced Audio Coding) - популярний кодек стиснення аудіо (lossy), зазвичай кращий за MP3 при тій самій швидкості. Чому AAC у відео — стандарт: добре підтримується в MP4, оптимальний для стримінгу та платформ.

48 кГц - частота дискретизації аудіо. Це стандарт у відеоіндустрії (телебачення/кіно/YouTube часто орієнтуються на 48 kHz).

Stereo означає **2** канали (**L/R**). Підходить для музики, більшості роликів і подкастів у відео. Якщо контент — суто "розмовний голос" і треба економити, інколи роблять **mono**, але для універсальності **stereo** нормально.

Базовий процес редагування аудіоконтенту

Приклад (YouTube)



Контейнер - MP4.

Video - H.264.

Audio:

-**AAC**- 48 kHz.

-**Stereo** - 320 kbps.

Розглянемо докладніше кожен пункт і що він означає на практиці

3. Audio - AAC, 48 kHz, Stereo, 320 kbps.

320 kbps - це бітрейт аудіокодування (скільки даних/секунду виділяється на звук). Для AAC це дуже високий бітрейт, тобто - звук буде майже без помітних втрат для більшості сценаріїв, але файл буде більший.

Практичний коментар:

-для голосу часто достатньо AAC 96–160 kbps (залежно від вимог);

-для музики/якісного саунду 192–320 kbps.

Як читати весь рядок разом

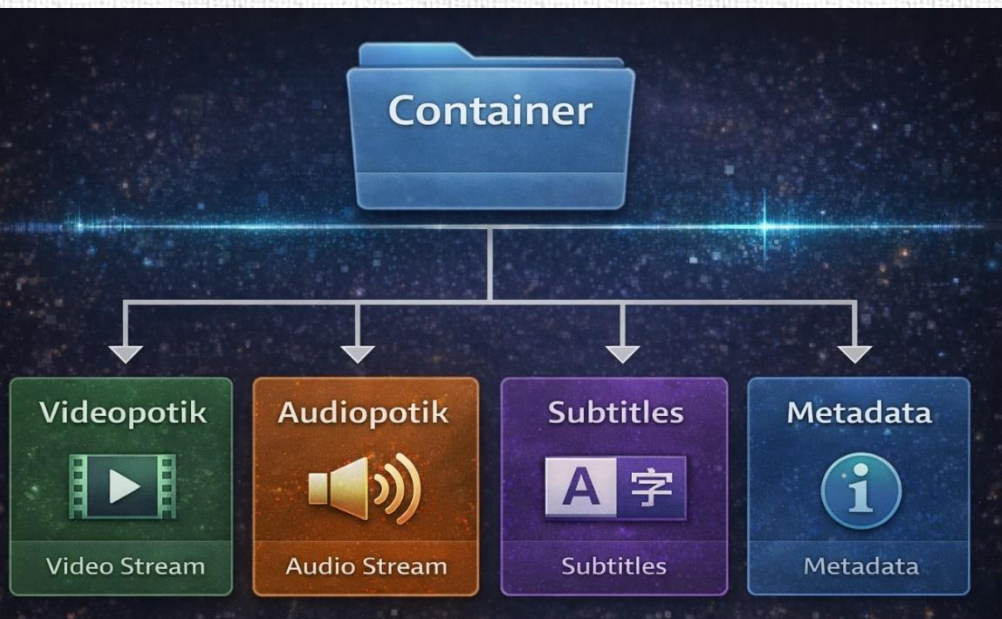
Файл MP4, всередині відео стиснене H.264, а звук - AAC із частотою 48 kHz, стерео, 320 kbps.

Контейнер (мультимедійний контейнер, media container) — це формат файлу-«упаковки», який зберігає один або кілька потоків (відео, аудіо, субтитри) разом із службовою інформацією (метадані, індекси, таймкоди). **Контейнер** не визначає, як саме стиснені відео/аудіо (це роблять кодеки), а визначає як усе це організовано в одному файлі.

Для чого потрібен контейнер

Об'єднати все в один файл. Наприклад:

- відео - (H.264/H.265/AV1);
- аудіо - (AAC/Opus/MP3);
- субтитри - (SRT/ASS/TTML/tx3g);
- глави, теги, обкладинка, мова доріжок.



Як контейнер "працює" всередині

Уявіть, що є окремі потоки - **video stream** - послідовність стиснених кадрів, **audio stream** - стиснені аудіофрейми, **subtitle stream** - текст/таймінги.

Контейнер робить три ключові речі:

-**мультиплексування**. Це "переплітання" фрагментів потоків у файлі: шматок відео, потім кілька аудіофреймів, потім знову відео, щоб плеєру було зручно читати послідовно.

-**таймстемпи та синхронізація**. Кожному фрагменту даються часові мітки: **PTS** - (коли показати/відтворити), **DTS** - (коли декодувати, важливо для **В-кадрів**). Саме вони забезпечують, що звук і картинка "йдуть разом".

-**індексація та метадані**. Контейнер містить таблиці/індекси - де в файлі лежить фрагмент для часу **T**, які доріжки є, їхня мова, тривалість, бітрейти, кодеки.



Мультимедійний Контейнер

Як контейнер працює всередині



Контейнер vs кодек - критично важливе розрізнення

1. Контейнер - MP4, MKV, MOV, AVI, TS, WebM, MXF.

2. Відеокодек - H.264, H.265, VP9, AV1.

3. Аудіокодек - AAC, Opus, MP3.

Той самий контейнер **MP4** може містити різні кодеки (в межах підтримки стандартом/плеєром). І навпаки - **H.264** може бути в **MP4, MKV, TS, MOV** тощо.



Які бувають контейнерні формати

1.MP4 (.mp4) - найпоширеніший для вебу й мобільних пристроїв.

Зазвичай: H.264/H.265 + AAC. Добре підходить для платформ.

2.MOV (.mov) - "Рідний" контейнер Apple/QuickTime. Дуже поширений у монтажі. Гнучкий для професійних пайплайнів.

3.MKV (.mkv) - дуже гнучкий контейнер. Зручно для фільмів - багато аудіодоріжок, субтитрів, глав. Іноді гірша сумісність із "простими" телевізорами/плеєрами, ніж MP4.

4.WebM (.webm) - веб-орієнтований контейнер Зручний для браузерів.



Які бувають контейнерні формати

5.MPEG-TS (.ts) - транспортний потік для мовлення/стримінгу. Дуже стійкий до помилок, зручний для передачі. Часто використовується в **HLS** (особливо історично).

6.fMP4 (fragmented MP4) - це **MP4**, але "порізаний" на фрагменти для стримінгу. Дуже важливий для MPEG-DASH і сучасного HLS (CMAF).

7.MXF (.mxf) - професійний broadcast-контейнер. Новини/ТБ/виробничі лінії. Надійний і стандартизований для індустрії.

8.AVI (.avi) - старий контейнер (**Microsoft**). Обмеження, проблеми з сучасними функціями (субтитри/таймінги/кодеки). Сьогодні здебільшого

Короткі практичні рекомендації

- 1.Для YouTube/соцмереж: MP4 (H.264 + AAC 48 kHz).
- 2.Для архіву з доріжками/субтитрами: MKV.
- 3.Для монтажу: MOV або MXF (залежно від студії).
- 4.Для вебу/браузера: WebM (VP9/AV1 + Opus).
- 5.Для стримінгу: fMP4.



MP4 — це один із найпоширеніших мультимедійних контейнерів. Технічно він базується на стандарті ISO Base Media File Format і організований як набір вкладених "блоків/атомів", які описують які потоки є, як їх відтворювати в часі і де лежать дані.

MP4-файл містить:

- відеопотік - (наприклад H.264/AVC, H.265/HEVC);
- аудіопотік - (часто AAC, інколи ALAC);
- субтитри - (наприклад - mov_text);
- метадані - (назва, автор, обкладинка);
- таймкоди, індекси, таблиці семплів (щоб плеєр знав, що і коли відтворювати).

MP4-файл - складається з блоків.

1.ftyp - "хто я" - вказує, що файл сумісний з MP4, які "бренди" підтримує. Плеєр одразу розуміє, як його трактувати.

2.moov — "інструкція відтворення". Це найкритичніший блок метаданих. Містить - опис доріжок (tracks), часову шкалу, таблиці семплів (де який кадр/аудіофрейм лежить), параметри кодеків. Усередині moov зазвичай - movie header — загальна таймшкала/тривалість, один або кілька trak — кожна доріжка окремо (відео, аудіо, субтитри), у sample table - "карта" семплів.



MP4-файл - складається з блоків.

3.mdat — "самі медіадані. "Тут лежать байти закодованого відео/аудіо (фрейми/семпли). **moov** каже плеєру, де саме в **mdat** знаходяться потрібні фрагменти.

Плюси і мінуси - MP4

Плюси:

- максимальна сумісність;
- добре для YouTube/соцмереж/телефонів;
- підтримка таймкодів/метаданих.

Мінуси:

- менш гнучкий, ніж **MKV** для "кінофайлів" з багатьма доріжками/субтитрами;
- деякі типи субтитрів/форматів краще живуть у MKV;
- для стримінгу потрібен fMP4/сегментація, а не "один великий файл".