

Міністерство освіти і науки України
Вінницький національний технічний університет

О. Ф. Грійо Тукало, Л. В. Крупельницький, О. М. Ткаченко

**ПОШУК ЗАДАНИХ ФРАГМЕНТІВ
В АРХІВІ АУДІОЗАПИСІВ
ІЗ ЗАСТОСУВАННЯМ KD-ДЕРЕВ**

Монографія

Вінниця
ВНТУ
2020

Замовити цю книгу <https://press.vntu.edu.ua/index.php/vntu/catalog/book/581>

Видавництво Вінницького національного технічного університету

<https://press.vntu.edu.ua/index.php/vntu/catalog>

УДК 004.934 + 004.627

Г48

Рекомендовано до друку Вченою радою Вінницького національного технічного університету Міністерства освіти і науки України (протокол № 5 від 23.12.2019 р.)

Рецензенти:

Г. Г. Бортник, доктор технічних наук, професор

Л. Б. Ліщинська, доктор технічних наук, професор

В. А. Лужецький, доктор технічних наук, професор

Грійо Тукало, О. Ф.

Г48 Пошук заданих фрагментів в архіві аудіо записів із застосуванням kd-дерев : монографія / О. Ф. Грійо Тукало, О. М. Ткаченко., Л. В. Крупельницький. – Вінниця : ВНТУ, 2020. – 148 с.

ISBN 978-966-641-789-6

В монографії розглянуто шляхи підвищення повноти і релевантності результатів та швидкості пошуку за рахунок розроблення інформаційної технології пошуку заданих фрагментів в архіві аудіозаписів із застосуванням kd-дерев. Запропоновано модель корпусу аудіозаписів, яка містить масив файлів, базу параметрів та метаданих і динамічне kd-дерево. На основі результатів теоретичних досліджень розроблено інформаційну технологію пошуку заданих аудіофрагментів в архіві аудіозаписів. Запропоновані методи та розроблені на їхній основі алгоритми та програмні засоби дозволяють проводити пошук заданого аудіофрагмента в реальному масштабі часу.

УДК 004.934 + 004.627

ISBN 978-966-641-789-6

© О. Грійо Тукало, О. Ткаченко., Л. Крупельницький, 2020

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ.....	6
ВСТУП	7
Розділ 1 АНАЛІЗ СУЧАСНИХ МЕТОДІВ І СИСТЕМ ПОШУКУ АУДІОЗАПИСІВ У ВЕЛИКИХ КОРПУСАХ.....	9
1.1 Актуальність задачі пошуку на основі аудіоконтенту	9
1.1.1 Сфери застосування пошуку на основі аудіоконтенту	9
1.1.2 Оцінювання подібності аудіозаписів.....	10
1.2 Характеристика основних етапів розв'язання задачі автоматичного пошуку аудіозаписів.....	13
1.2.1 Аналіз параметрів математичної моделі аудіосигналу	13
1.2.2 Аналіз методів класифікації та методів швидкого пошуку	18
1.3 Огляд основних підходів та методів оцінювання подібності мультимедійного контенту.....	26
1.4 Огляд сучасних систем пошуку аудіозаписів за аудіоконтентом.....	31
Розділ 2 РОЗРОБКА МАТЕМАТИЧНОЇ МОДЕЛІ ОПИСУ АУДІОСИГНАЛІВ ТА ФОРМУВАННЯ КОРПУСУ АУДІОЗАПИСІВ НА ОСНОВІ КЛАСТЕРНОГО АНАЛІЗУ.....	36
2.1 Розробка математичної моделі опису аудіосигналів.....	36
2.1.1 Обґрунтування вибору MFCC як параметрів опису аудіозаписів	36
2.1.2 Дослідження основних характеристик MFCC для розв'язання задачі пошуку аудіозаписів	37
2.1.3 Загальна схема обчислення MFCC.....	40
2.2 Обґрунтування застосування кластерного аналізу для формування корпусу аудіозаписів.....	42
2.3 Розробка методу кластеризації на основі послідовного запуску k-середніх з удосконаленим вибором кандидата на нову позицію вставки.....	46
2.4 Дослідження kd-дерев як засобу структуризації бази параметрів для пошуку аудіофрагмента в корпусі аудіозаписів.....	54
2.4.1 Задача пошуку найближчого сусіда. Обґрунтування виконання швидкого пошуку на основі kd-дерева.....	54
2.4.2 Аналіз прямої та зворотної фази пошуку найближчого сусіда на основі kd-дерева	55

2.4.3 Аналіз чинників впливу на продуктивність пошуку на основі kd-дерева.....	58
2.4.4 Вибір варіанта поділу простору для побудови kd-дерева	59
2.4.5 Дослідження залежності продуктивності пошуку від розмірності і коефіцієнта розповсюдження зворотної фази пошуку	61
Розділ 3 РОЗРОБКА МЕТОДІВ ШВИДКОГО ПОШУКУ ІЗ ЗАСТОСУВАННЯМ KD-ДЕРЕВ.....	
3.1 Визначення приведеної відстані (ПВ).....	66
3.2 Обґрунтування можливості визначення аудіозапису за його аудіофрагментом	67
3.2.1 Визначення мінімальної тривалості аудіофрагмента.....	70
3.2.2 Метод пошуку найближчого аудіозапису в корпусі за ПВ.....	73
3.2.3 Результати оцінювання близькості аудіозаписів за методом пошуку власного аудіозапису на основі ПВ.....	73
3.3 Метод пошуку аудіофрагмента на основі kd-дерева з оцінкою міри близькості за зваженою кількістю попадань в список найближчих центроїдів	76
3.3.1 Загальний підхід до реалізації пошуку в корпусі аудіозаписів, впорядкованому на основі kd-дерева	76
3.3.2 Оцінка міри близькості за приведеною відстанню	80
3.3.3 Оцінка міри близькості за кількістю попадань в список k найближчих центроїдів (Hits)	81
3.3.4 Оцінка міри близькості за зваженою кількістю попадань в список k найближчих центроїдів (Weighted Hits).....	81
3.3.5 Результати порівняння запропонованих оцінок міри близькості невідомого аудіофрагмента з аудіозаписами в корпусі.....	83
3.4 Дослідження чинників впливу на продуктивність пошуку із застосуванням kd-дерев	85
3.4.1 Залежність часу пошуку по kd-дереву від розміру бази параметрів корпусу	86
3.4.2 Залежність обчислювальної складності пошуку від розмірності параметрів	88
3.4.3 Залежність обчислювальної складності від коефіцієнта розповсюдження зворотної фази пошуку	91
3.4.4 Метод комбінованого пошуку аудіозапису за заданим аудіофрагментом	93

Розділ 4 РЕАЛІЗАЦІЯ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ДЛЯ РОЗВ'ЯЗАННЯ ЗАДАЧ ПОШУКУ ЗАДАНИХ АУДІОФРАГМЕНТІВ ТА МОНІТОРИНГУ РАДІОМОВЛЕННЯ.....	98
4.1 Розробка структури інформаційної технології пошуку заданих фрагментів в архіві аудіозаписів	98
4.2 Розробка програмного забезпечення інформаційної технології пошуку заданих фрагментів в архіві аудіозаписів	102
4.3 Застосування запропонованих методів пошуку для моніторингу радіомовлення	119
4.3.1 Особливості виконання пошуку кількох ідентичних аудіозаписів в архіві	119
4.3.2 Перевірка стійкості системи пошуку до спотворень аудіозаписів	124
4.4 Оцінювання ефективності інформаційної технології пошуку заданих фрагментів в архіві аудіозаписів.....	126
ВИСНОВКИ.....	131
ЛІТЕРАТУРА	133

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

ЕМ	Евклідова метрика
ПА	Інформаційний пошук аудіозаписів
МО	Математичне очікування
МТ	Музичний твір
ПВ	Приведена відстань
ПЗ	Програмне забезпечення
GMMs	Gaussian Mixture Models (Гауссові моделі сумішей)
LSF	Line Spectrum Frequencies (Лінійні частоти спектра)
MFCC	Mel-frequency cepstrum coefficients (мел-частотні кепстральні коефіцієнти)
RC	Метод кластеризації з визначенням позиції центроїда на основі вибору вектора-кандидата, який забезпечує мінімальне спотворення з перерахунком нового положення центроїдів
WRC	Метод кластеризації з визначенням позиції центроїда на основі вибору вектора-кандидата, який забезпечує мінімальне спотворення без перерахунку нового положення центроїдів

ВСТУП

Актуальність теми. Розвиток Інтернету, цифрових ЗМІ, мультимедійних технологій та нових способів обробки звукової інформації викликав великий інтерес і увагу до шляхів, якими інформаційні технології можуть бути застосовані до такого контенту. З можливістю доступу до великих архівів мультимедіа, практично в будь-якому місці і в будь-який час, необхідно було запропонувати нові способи навігації та взаємодії з цими архівами аудіозаписів.

Суть пошуку на основі аудіоконтенту полягає в тому, щоб автоматично визначати аудіозаписи, що містять частини, подібні до заданого аудіофрагмента. Більшість існуючих систем поєднує в собі такі етапи обробки аудіозаписів: аналіз вхідного аудіосигналу, виділення певних параметрів (ознак), вибір міри їхнього порівняння та визначення аудіозапису або деякої категорії аудіозаписів, найближчих до аудіофрагмента за обраною мірою. У подальшому, говорячи про розподіл аудіозаписів за категоріями (наприклад, музика, мовлення, тиша), будемо використовувати термін «класифікація», тоді як визначення конкретного аудіозапису будемо називати пошуком. Слід зазначити, що в англійській літературі термін «класифікація» використовується в обох випадках, тобто у більш широкому сенсі.

Важливим завданням для систем пошуку на основі аудіоконтенту є масштабованість, оскільки розміри мультимедійних архівів (музики, фільмів, теле- та радіопередач, стенограм судових засідань тощо) сягають сотень тисяч і навіть мільйонів аудіозаписів та зростають постійно. Це, в свою чергу, також вимагає використання компактних параметрів з низькою складністю обчислень та ефективних методів пошуку в архівах аудіозаписів. Відповідно, в процесі розробки методів пошуку в великих архівах аудіозаписів потрібно досягти компромісу між обчислювальною складністю та повнотою пошуку.

Незважаючи на великі успіхи, досягнуті у цій галузі в останні роки, сучасним інформаційно-пошуковим системам властиві певні недоліки, зокрема:

- недостатня продуктивність під час роботи з великими архівами;
- неможливість їх безпосереднього застосування для розв'язання таких задач, як моніторинг трансляції на радіостанціях (телебаченні) для визначення кількості повторень в ефірі рекламного продукту, музичного твору тощо.

Отже, створення інформаційної технології пошуку заданих фрагментів в архівах аудіозаписів є актуальною науково-практичною задачею.

Метою монографії є підвищення повноти та релевантності результатів і швидкості пошуку заданих аудіофрагментів в архіві аудіозаписів за рахунок розроблення нової інформаційної технології.

Запропоновані методи та розроблені на їх основі програми пошуку дозволяють зменшити складність обчислень до 3,6 % або в 28 разів порівняно з точним пошуком на основі kd-дерева.

В монографії висвітлюються результати оригінальних досліджень авторів [1–19]. Автори будуть вдячні за відгуки на монографію та пропозиції щодо розвитку подальших досліджень.

Розділ 1

АНАЛІЗ СУЧАСНИХ МЕТОДІВ І СИСТЕМ ПОШУКУ АУДІОЗАПИСІВ У ВЕЛИКИХ КОРПУСАХ

1.1 Актуальність задачі пошуку на основі аудіоконтенту

Завдяки розвитку технологій Інтернету і комп'ютерам, поширенню онлайн-магазинів та «хмарних» сервісів мультимедіа передавання і зберігання аудіо, зокрема музики, стало одразу простим та загальнодоступним. В даний час величезні обсяги мультимедійного контенту доступні з будь-якої точки світу [20].

1.1.1 Сфери застосування пошуку на основі аудіоконтенту

Розглянемо результати в сферах, для яких формулювання поставленої задачі звучить по-різному, однак для аналізу використовують подібні дані та методи.

Пошук заданого фрагмента в архівах аудіозаписів є важливою задачею аналізу аудіозаписів за контентом, метою якого є визначення подібних записів в архіві [21]. Найважливішою вимогою в цій сфері є здатність ефективно обчислювати відстані до аудіозаписів в корпусі.

Створення ефективного методу оцінювання подібності мультимедійного контенту є важливою та актуальною задачею, про що свідчить велика кількість можливих сфер його застосування.

- Автоматичне генерування метаданих на основі аудіозапису. Методи аналізу контенту на основі метаданих вже реалізовано, проте нові методи аналізу контенту на основі самого аудіозапису можуть бути необхідними, щоб генерувати ці метадані [22–25].

- Сервіси для автоматичного рекомендування слухачам нових, невідомих ще музичних творів, фільмів, наприклад, в потоковому сервісі, он-лайн магазині, або навіть за допомогою невеликого особистого музичного пристрою, щоб автоматично створювати списки відтворення на ходу [26, 27].

- Методи, необхідні для спрощення пошуку подібних аудіозаписів в архівах аудіозаписів експертів (наприклад, Національної ради України з питань телебачення та радіомовлення).

- Методи автоматичного створення карт архівів аудіозаписів [28–30] з метою розробки більш ефективних способів пошуку аудіозаписів.

- Пошук фрагмента аудіозапису.

- Автоматична класифікація невідомих аудіозаписів відповідно до конкретних категорій [31, 32]. Так, наприклад, автоматична класифікація музики за жанрами, аналіз і розподіл відео за категоріями саундтреків [33] є необхідними у професійному виробництві ЗМІ, на радіостанціях, для аудіовізуального управління архівами, розваги і звичайно в Інтернеті.

- Виявлення дублікатів [34] в архівах аудіозаписів.

- Моніторинг трансляції з метою виявлення можливих порушень авторських прав (плагіату) на радіостанціях (телебаченні).

Проте розв'язання такого роду задач є надзвичайно складним в обчислювальному плані, незважаючи на удосконалення обчислювальних ресурсів. Це обумовлює актуальність досліджень у згаданих напрямках.

1.1.2 Оцінювання подібності аудіозаписів

Центральним поняттям інформаційного пошуку аудіозаписів (ПА) є схожість/подібність аудіозаписів (музичних творів зокрема). Обчислення подібності аудіозаписів є однією з основних цілей ПА. Моделювання подібності аудіозаписів необхідне для розв'язання низки задач пошуку, наведених в п. 1.1.1, на основі контенту.

Більшість сучасних систем пошуку аудіозаписів використовує метадані та текстові анотації [35–38]. В загальному випадку метадані можуть бути визначені як дані, що описують інші дані. Генерація таких анотацій аудіоконтенту є трудомістким і тривалим процесом [35, 39]. Стандарт опису мультимедіа MPEG-7 (Moving Picture Expert Group) стандартизував подання таких метаданих, що описують контент, як параметри інструментів [40].

Традиційний і найбільш надійний спосіб визначення подібності аудіозаписів, звичайно, вручну. Вочевидь, що це просто неможливо для великих архівів аудіозаписів, оскільки аналіз вимагатиме дуже багато часу [41]. Хоча за допомогою певної кількості експертів можна класифікувати навіть великі архіви музики за кількома характеристиками (жанром, роком випуску тощо) і використовувати цю інформацію, наприклад, для рекомендації музичних творів, що реалізовано в радіо-сервісі Pandora.com [42]. Проте все ж таки використання експертів це також не швидкий спосіб (в середньому час аналізу треку складає 20–30 хв), а крім того ще й дорогий.

Методи оцінювання подібності аудіозаписів на основі аудіоконтенту використовують тільки звуковий сигнал для побудови моделі по-

дібності. Пошук на основі аудіоконтенту дозволить виконувати пошук аудіозаписів, навіть якщо текстові нотації взагалі відсутні, або підвищити повноту пошуку, якщо метадані є неповними. Саме пошук на основі аудіоконтенту лежить в основі запропонованої інформаційної технології пошуку заданих фрагментів в архівах аудіозаписів.

Спільним для всіх цих методів є те, що вони часто згадуються як «набір кадрів» (можлива аналогія – «набір слів» для класичного пошуку тексту), оскільки параметри для порівняння аудіозаписів обчислюються незалежно на коротких проміжках. Таким чином, основою аналізу аудіозаписів є обчислення векторів параметрів на невеликих фрагментах (кадрах/фреймах), для яких характеристики сигналу залишаються відносно стійкими. Визначимо термін «міра подібності аудіозаписів» як кількісну оцінку близькості двох аудіозаписів на основі будь-якої інформації, яку можна отримати з аудіофайлу, залежно від визначеної мети.

Загальний алгоритм оцінювання подібності аудіозаписів, як правило, визначається двома стадіями: (I) отримання параметрів, (II) вибір міри оцінювання подібності. Щоб використовувати функцію подібності спочатку аудіозаписи необхідно проаналізувати для виділення параметрів. Після отримання параметри зберігаються для використання їх повторно. Зазвичай параметри обчислюють за чотири кроки.

- Декодування: кодований аудіофайл (н-д, MP3) розпаковується і перетворюється на його дискретне подання в часовій області. У часовій області аудіосигнал описується набором рівномірно розташованих дискретних цілих чисел $x(t)$ (відліків) в кожен момент часу t . Кількість відліків за секунду називається частотою дискретизації звуку. Стандартні частоти дискретизації, які використовуються в алгоритмах оцінювання подібності аудіозаписів, 8000 Гц, 16000 Гц, 22050 Гц та 44100 Гц для сигналу моноканалу.

- Подання інтервалу частот: швидке перетворення Фур'є (ШПФ, Short-Time Fourier Transformation (STFT)) обчислює дискретне подання інтервалів частот звуку. У частотній області сигнал подано розподілом його частот $f_b(t)$ в заданій кількості смуг частот b за аналізований проміжок часу T (розмір вікна). Алгоритми оцінювання подібності аудіозаписів, як правило, використовують 1024–2048 відліків у STFT. У більшості випадків обчислюється спектр потужності, який є сигналом з енергією на частотній смузі і часовому інтервалі.

• Застосування нелінійної шкали частот: корисний сигнал зосереджується в діапазоні низьких частот, в той час як шум знаходиться, в основному, на високих частотах. Саме тому алгоритми оцінювання подібності аудіозаписів використовують нелінійні шкали частот. Зазвичай застосовують такі шкали частот:

- Sone-шкалу, що є суб'єктивною мірою гучності J [43]:

$$J = kI^{\frac{1}{3}},$$

де I – фізична інтенсивність (потужність) звуку, фон; k – коефіцієнт, що залежить від частоти;

- Bark-шкалу, що є спектром, пов'язаним з критичними смугами слуху:

$$b = 113a \tan(0,00076 \cdot f) + 3,5a \tan\left(\frac{f}{7500}\right)^2,$$

де f – частота, Гц;

- Mel-шкалу – одиницю висоти звуку. Кількісне оцінювання звуку за висотою базується на статистичній обробці великого числа даних про суб'єктивне сприйняття висоти звукових тонів (запропоновано Стівенсом і Волкманом [44]). Результати досліджень показують, що висота звуку пов'язана, головним чином, з частотою коливань, але залежить також від рівня гучності звуку і його тембру. Звукові коливання частотою 1000 Гц при ефективному звуковому тиску 2,10–3 Па (тобто при рівні гучності 40 фон), що впливають спереду на спостерігача з нормальним слухом, викликають у нього сприйняття висоти звуку, що оцінюється за визначенням в 1000 мелів. Звук частотою 20 Гц при рівні гучності 40 фон має, за визначенням, нульову висоту (0 мелів). Оскільки застосування Mel-шкали дає найкращі результати, для оцінювання подібності аудіозаписів у роботі обрано саме її.

• Щоб оцінити подібність між двома аудіозаписами, після вибору параметрів обчислювальний алгоритм також має визначити міру подібності, перейти до порівняння невідомого аудіозапису з аудіозаписами корпусу та визначення власного аудіозапису, розходження з яким буде мінімальним. Щоб знайти аудіозаписи найбільш близькі до даного, в простому випадку алгоритм пошуку лінійно сканує і порівнює всі аудіозаписи із заданим відповідно до функції подібності.

Для узагальнення сказаного вище на рис. 1.1 показано способи оцінювання подібності аудіозаписів. Виділені блоки застосовано в процесі розробки інформаційної технології пошуку заданого фрагмента в архіві аудіозаписів.

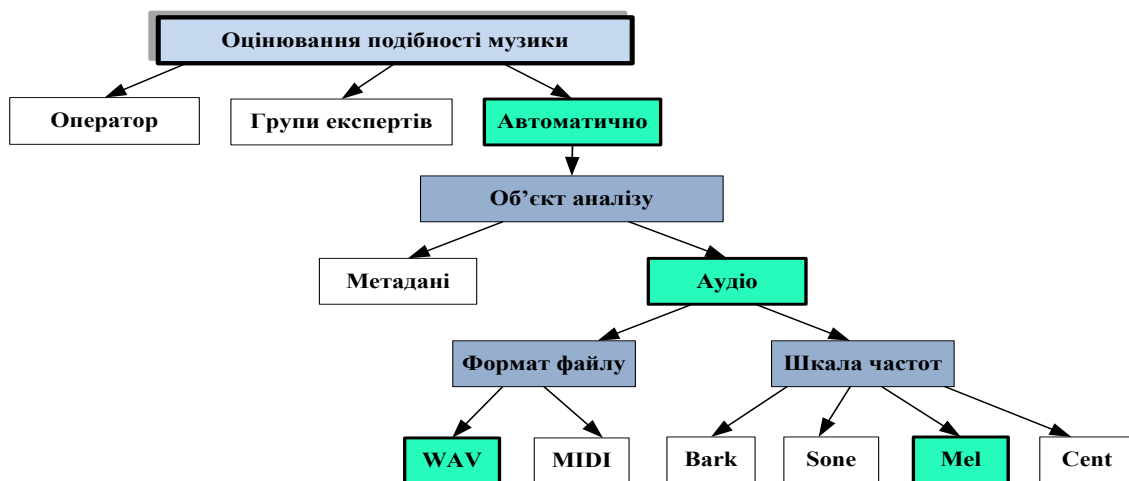


Рисунок 1.1 – Способи оцінювання подібності музики

1.2 Характеристика основних етапів розв'язання задачі автоматичного пошуку аудіозаписів

На основі аналізу літературних джерел [44–51] в процесі виконання автоматичного пошуку аудіозаписів можна виділити такі основні етапи:

- 1) обчислення параметрів опису аудіосигналу;
- 2) способи зменшення розмірності параметрів;
- 3) впорядкування параметрів;
- 4) безпосередньо пошук найближчого сусіда за обраною мірою оцінювання подібності.

Наведемо більш розгорнуту характеристику кожного з цих етапів.

1.2.1 Аналіз параметрів математичної моделі аудіосигналу

Існує велика кількість параметрів аудіосигналу, причому багато з них можуть використовуватись як для мовлення, так і для музики [45, 52].

Музичний звук характеризується чотирма головними властивостями: основний тон, гучність, тривалість і тембр. Саме ці чотири властивості дозволяють слухачу розрізняти музичні твори між собою.

Основний тон можна вважати величиною, що дорівнює оберненому періоду звуків. Основна частота є відповідним фізичним терміном і обчислюється в герцах.

Фізичним двійником гучності є інтенсивність, величина якої пропорційна енергії акустичної хвилі. Коли говорять про тривалість, як правило, мають на увазі тривалість не дуже коротких тонів. Тембр є багатовимірним поняттям, що залежить, в основному, від розподілу енергії спектра та її динаміки в часі [53, 54]. Тембр – це те, що також

називають «кольором» музики, і він щільно пов'язаний з розпізнаванням джерел звуку. Коли два музичні твори мають однакову частоту основного тону, гучність і тривалість тонів, тембр є характеристикою, що дозволяє відрізнити два звуки між собою. Саме тому у даній роботі для опису музичних треків застосовано тембральні параметри.

Характеристики: тембр (короткочасна спектральна інформація), ритм (в т. ч. такт і темп), основний тон ретельно досліджені в [53]. Характеристики високого рівня, що охоплюють оркестровку, текстуру, ритм, динаміку, статистику основного тону, мелодію та акорди, досліджувались в [55]. Згідно з результатами спостережень музичний тембр є домінуючим фактором. З урахуванням цього було створено багато методів оцінювання подібності, що базуються виключно на музичному тембрі. Наразі майже всі найбільш поширені методи визначення подібності музики досі застосовують тембральний компонент аналогічно введеному Logan [46]. Крім музичного тембру ще один важливий аспект, яким, зокрема, визначається поняття музичного жанру, це ритм. Оцінювання подібності музичних творів на основі ритму проводилась, наприклад, Rampralk та ін. [56] або Gouyon та ін. [57].

Більшість параметрів опису аудіосигналу можна поділити за характером інформації, яку вони передають, на такі основні категорії:

- описують ритм – темп (tempo), чіткість імпульсу (pulse clarity) і флуктуації (fluctuation) тощо;
- описують тембр – спектральний центроїд (spectral centroid, SC), спад спектра (roll-off), MFCC тощо;
- показують зміни в часі – частота переходу через нуль (zero-crossing rates, ZCR), амплітуда обвідної (amplitude envelope), короткочасна енергія (STE) тощо;
- несуть інформацію про частоту – коефіцієнти спектра лінійного прогнозування (LPC), лінійні спектральні частоти (LSF) спектральні піки, спектральний потік (spectrum flux, SF) тощо;
- сприйняття звуку людиною – спад спектра (spectrum rolloff, SR), спектральний центроїд, ентропія, хроматичні параметри, частота основного тону (ЧОТ), MPEG-7 [47, 58];
- описують кепстр – спад кепстру (cepstrum flux), мел-частотні кепстральні коефіцієнти (Mel-frequency cepstral coefficients, MFCC), BFCC, коефіцієнти лінійного прогнозування кепстру (LPCC) тощо;
- модуляції на частоті — модуляції енергії на певній частоті 4 кГц, періодичність смуг (band periodicity) тощо.

Розглянемо більш детально параметри за їх типами.

Всі параметри, що їх отримують в часовій області, обчислюють безпосередньо з відліків аудіосигналу, без будь-яких перетворень, відповідно, складність обчислень низька. ZCR – це один з найпростіших параметрів, що визначається як кількість перетинів нуля в часовій області за 1 с. ZCR дозволяє отримати інформацію про основну частоту сигналу, зазвичай застосовується для розрізнення мовлення і музики, хоча також і для інших задач, наприклад, класифікації музики за жанрами тощо.

Енергія сигналу дорівнює квадрату його амплітуди. Потужність сигналу – це енергія, що передається за одиницю часу. Короточасна енергія описує обвідну сигналу і дорівнює середній енергії фрейму. STE широко застосовується для різних задач в сфері пошуку аудіозаписів, наприклад, виділення шуму (тиші) з корисного сигналу.

Параметри, що належать до частотної області, є найбільшою групою. Спектральні параметри набули поширення в літературі. Оскільки порівнювати безпосередньо звукові сигнали в часовій області довго і не дуже ефективно, відліки аудіосигналу ділять на невеликі частини (фрейми), для яких характеристики сигналу залишаються відносно стійкими (стаціонарний випадковий процес), з перекриттям фреймів. Тривалість одного фрейму, як правило, лежить у межах 10–30 мс. Для кожного фрейму виконується спектральний аналіз, на основі якого обчислюється значення вектора параметрів (параметризація). Перехід в частотну область зазвичай виконується за допомогою ШПФ, автокореляції, рідше — косинусного чи вейвлет-перетворень тощо.

LPC описують голосовий тракт людини, відповідно широко застосовуються для розпізнавання мовлення, а також для сегментації аудіосигналу. На практиці частіше застосовують кепстральне подання LPC (LPCC), оскільки вони є більш ефективними. LSF описують ту ж інформацію, що й LPC, але вони є більш стійкими до спотворень. LSF застосовують для класифікації інструментів або звуку на мовлення/музику.

SF показує зміни форми спектра в часі, визначається як сума різниць амплітуд двох сусідніх фреймів. SF застосовують для розрізнення мовлення і музики, інформаційного пошуку аудіозаписів тощо.

SC дорівнює математичному очікуванню магнітуди спектра, тобто визначає точку, де концентрується більшість енергії спектра, і

пов'язаний з основною частотою сигналу. SR – це частота, нижче якої зосереджується щонайменше $N = 85\%$ магнітуди аудіосигналу.

Використання хроматичних ознак обґрунтовано «спіраллю сприйняття музичного тону Шеппарда» [59]. Енергія гамми музичних напівтонів накопичується в дванадцяти класах частот в діапазоні однієї октави [60]. Хроматичний спектр може бути визначений зі спектра магнітуди $X(k)$ за формулою

$$X_{chroma}(k) = \sum_i X(k), 1 \leq i \leq [(N + 1) / 2],$$

де k – індекс частоти; N – кількість коефіцієнтів перетворення Фур'є.

Хроматичні параметри застосовують для аналізу та пошуку музики. Існує кілька варіантів опису музики, подібних до хроматичних ознак [49, 61]. Гістограма частот за Tzanetakis і Cook, яка базується на виявленні та поданні домінантних частот у вигляді гістограми, також тісно пов'язана з хроматичними ознаками [31].

Нещодавно міжнародний стандарт MPEG-7 прийняв нові спектральні декорельовані параметри зниженої розмірності [62]. MPEG-7 параметри складаються зі спектральних векторів невеликої розмірності, отримуваних за допомогою лінійного перетворення спектрограми (Principal Component Analysis (PCA)).

MFCC і їх похідні спочатку отримали широке використання як параметри в системах розпізнавання мовлення та диктора [45, 63, 64]. Вони також стали першим вибором при побудові систем загального аналізу аудіоконтенту або музики. Мел-частотні кепстральні коефіцієнти (MFCC) використовуються для моделювання та розрізнення музичних сигналів. Використання коефіцієнтів MFCC в процесі ІПА постійно зростає, а саме: під час класифікації за жанрами, визначення подібності аудіозаписів [31, 41, 48, 65] тощо. Вони є стійким до шумів, стислим поданням спектра сигналу. MFCC мають важливу властивість: невеликі (відповідно великі) чисельні зміни коефіцієнтів MFCC відповідають невеликим (відповідно великим) змінам сприйняття.

Порівняємо параметри MPEG-7 з MFCC для пошуку аудіозаписів. Хоча і одні й інші несуть в собі інформацію про спектр, MPEG-7 параметри належать до нових в сім'ї параметрів аудіосигналу, тоді як MFCC отримали широке застосування в розпізнаванні мовлення та класифікації аудіосигналів. Для їх отримання замість DCT використовується метод головних компонент (PCA), а іноді також аналіз незалежних компонент (Independent Component Analysis (ICA)). Відповід-

но алгоритм отримання MPEG-7 параметрів для аналізу аудіозаписів займає більше часу і пам'яті. MPEG-7 параметри використовують логарифмічну шкалу, оскільки вона є дуже простою. Для MFCC використовується шкала Mel-частот. Mel-шкала, як уже було зазначено вище, є кращою для пошуку аудіозаписів, ніж розподіл частот в логарифмічному масштабі [58]. Саме параметри MFCC було обрано для опису аудіозаписів.

У той час як коефіцієнти MFCC можуть бути застосовані до великої різноманітності звуків (мовлення, музика, навколишні звуки), хроматичні ознаки є специфічними саме для музики і описують спектральний зміст музичних звуків.

Тобто, хроматичні параметри, як правило, використовуються для аналізу структури музики [60, 66], оцінювання тональності (наприклад, [58, 67]), визначення каверів [68], або виявлення гармонічних змін при аналізі ладів [69]. Bartsch і Wakefield відзначили, що хроматичні ознаки перевершили MFCC у виділенні найяскравіших частин музичного твору [60]. MFCC ж дали найкращу точність класифікації за типом інструмента та взагалі аудіокласифікації, крім того MFCC можуть бути вибрані і за рахунок їх невеликої складності обчислень. В табл. 1.1 в компактному вигляді наведено порівняльну характеристику найпопулярніших параметрів в часовій та частотній областях.

Таблиця 1.1 – Огляд основних параметрів часової та частотної областей

Область	Назва параметра	Враховують сприйняття людиною	Обчислювальна складність	Скаляр/Вектор	Перетворення	Шкала частот	Декорельованість	Сфера застосування			
								Розпізнавання мовлення, диктора	Пошук музики	Аналіз структури музики	Класифікація аудіосигналу
Часу	ZCR	-	H	C				+			
	STE	-	H	C				+			
Частоти	SF	-	H	C				+			+
	Chroma	-	C	B	FFT	Log				+	
	MPEG-7	+/-	B	B	FFT	Log	+ (PCA)		+		
	MFCC	+	C	B	FFT	Mel	+ (DCT)	+	+		+

За результатами проведеного аналізу параметрів, а також відповідно до вимог, що їх наведено у п. 2.1.1, для опису аудіосигналу в архівах аудіозаписів обрано MFCC.

1.2.2 Аналіз методів класифікації та методів швидкого пошуку

Крім великої кількості різноманітних параметрів опису аудіозаписів, існує широкий спектр класифікаторів та методів швидкого пошуку найближчих сусідів, зокрема гаусові моделі (GMM) [31, 50], нейронні мережі (Neural Networks), приховані марківські моделі (Hidden Markov Models (HMM)), k-means, діаграми Вороного (ДВ) [70], дерева прийняття рішень (decision trees) [71], машини опорних векторів (support vector machines (SVM)) [72] самоорганізовані карти (Self Organizing Map, SOM) [73], AdaBoost. Розглянемо детальніше основні з них.

Стандартним підходом обчислення подібності аудіозаписів (що забезпечує високу достовірність результатів) є схожість за тембром на основі параметризації звуку за допомогою коефіцієнтів MFCC та гаусових (статистичних) моделей розсіювання (розподілу).

GMM використовує сімейство гаусових функцій щільності ймовірності для поділу простору параметрів на кластери. Оскільки ймовірності функцій щільності можуть перекриватися, GMM є методом для класифікації з гнучким (soft) вибором належності векторів кластерам [29].

Багатовимірна функція Гаусової щільності ймовірності визначається за формулою

$$N(\mathbf{x} | \mathbf{v}, \Sigma) = \left(\frac{1}{2\pi}\right)^{d/2} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{v})^T \Sigma^{-1}(\mathbf{x} - \mathbf{v})\right),$$

де \mathbf{x} – вектор ознак; \mathbf{v} – вектор математичного очікування. Враховуючи обчислювальну складність зазвичай обчислюють діагональ коваріаційної матриці $\Sigma = \{\sigma_1^2 \dots \sigma_d^2\}$.

GMM є сумішшю M гаусових розподілів, і щільність ймовірності для вектора \mathbf{x} дорівнює зваженій сумі кожного з них.

$$p(\mathbf{x} | \Phi) = \sum_{m=1}^M c_m p(\mathbf{x} | m, \Phi) = \sum_{m=1}^M c_m N(\mathbf{x} | \mathbf{v}_m, \Sigma_m),$$

де $1 \leq m \leq M$, $0 < c_m \leq 1$, $\sum_{m=1}^M c_m = 1$; N – кількість коефіцієнтів перетворення Фур'є;

Φ – набір параметрів, які необхідно обчислити,

$$\Phi = \{\mathbf{v}_m, \Sigma_m; m = 1 \dots M\}.$$

Всі методи оцінювання подібності аудіозаписів, що базуються на багатовимірних гаусових моделях, є неекстремальними і використовують

не стандартну норму вектора (наприклад, R^2), а неметричні міри оцінювання розбіжності – non-metric divergence measures (тобто розбіжність Kullback-Leibler)

$$D_{KL}(p, q) = \sum_{x \in \mathcal{X}} p(x) \ln \frac{p(x)}{q(x)}, p(x) > 0, q(x) > 0, \mathcal{X} \subset R,$$

де p – функція ймовірності випадкової величини X ; q – функція ймовірності випадкової величини Y ; R – множина дійсних чисел.

Тому, незважаючи на високу якість методів оцінювання подібності аудіозаписів, що базуються на GMM, їх не можна використовувати для пошуку у великих архівах аудіозаписів, оскільки для них не можуть бути використані методи швидкого пошуку, що будуть описані нижче, у той час як лінійний пошук може бути легко виконаний [74, 75]. Це є головною проблемою методів оцінювання подібності аудіозаписів, де використовуються багатовимірні Гаусові моделі. В роботах [74, 75] використовувалися багатовимірні Гаусові функції для тембральних параметрів, в [75] Гаусова модель також застосовується для ритмічних параметрів.

Отже, більш ефективні алгоритми індексування, кластеризації або візуалізації просто не призначені для роботи з розподілами Гауса і такими нестандартними метриками, як розбіжність Кульбака–Лейблера. Щоб обійти це обмеження параметри часто штучно векторизують, наприклад, в [29, 76], проте це вимагає великих обчислювальних витрат. Отже, неспроможність застосовувати навіть кластеризацію є великою проблемою для реальної масштабованої системи пошуку подібних аудіозаписів з тією чи іншою кінцевою метою. Для розв'язання такої задачі кластеризація, наприклад, просто необхідна, щоб побудувати швидкий пошуковий індекс. Обчислення повної матриці подібності для таких задач неможливе для архівів з мільйонами аудіозаписів.

Приховані моделі Маркова [29] є дуже потужним статистичним методом характеристики даних з багатовимірного часового ряду, який успішно використовується для розпізнавання мовлення, статистичного моделювання мовлення та машинного перекладу. НММ – це статистична модель Маркова, в якій система, що моделюється, розглядається як марковський процес із неспостережуваними (прихованими) станами. Кожен стан має ймовірнісний розподіл усіх можливих вихідних значень. Оскільки GMM фактично є НММ з одним станом, то застосування НММ вимагає більшої складності обчислень.

Самоорганізовані карти (SOM) є дуже популярним методом штучних нейронних мереж у категорії методів навчання без вчителя, який використовують для кластеризації аудіозаписів і їхньої класифікації [73]. Важливою відмінністю SOM є те, що в них всі нейрони (вузли, центри класів) упорядковані в деяку структуру (зазвичай, двовимірну сітку). За рахунок цього SOM можна вважати одним з методів проектування багатовимірного простору в простір з меншою розмірністю. При реалізації SOM заздалегідь задається конфігурація сітки (прямокутна або шестикутна), а також кількість нейронів у мережі. У випадку, коли розмір карти складає десятки тисяч нейронів, час, необхідний на навчання карти, зазвичай буває занадто великим для вирішення практичних завдань, отже, необхідно досягати допустимого компромісу при виборі кількості вузлів. Вдало вибраний спосіб ініціалізації може істотно прискорити навчання і привести до одержання більш якісних результатів. Схожістю в даній задачі називають відстань між векторами, яка звичайно обчислюється в евклідовому просторі.

Метод кластеризації k -середніх (k -means) є найрозповсюдженішим і найбільш дослідженим серед усіх методів кластеризації. Він мінімізує спотворення, розподіляючи дані між регіонами, що не перетинаються та ідентифікуються за їхніми центрами. Основна ідея його полягає в тому, що на кожній ітерації переобчислюють центр мас для кожного кластера, отриманого на попередньому кроці, потім вектори розбиваються на кластери знову відповідно до того, який з нових центрів виявився ближче за обраною метрикою, зазвичай, евклідовою. Виконання k -середніх завершується, коли не відбувається зміна кластерів. Метод гарантовано збігається за кінцеве число ітерацій.

Поширеність методу k -середніх зумовлена його головними перевагами: простотою, гнучкістю, швидкою збіжністю. Проте практичне застосування класичного методу k -середніх суттєво обмежується його недоліками, зокрема: результати кластеризації за методом k -середніх значною мірою залежать від вибору початкової конфігурації центрів (ініціалізації); робота методу суттєво уповільнюється під час кластеризації великих обсягів даних; він може сходиться до локального мінімуму цільової функції.

З метою подолання цих недоліків методу k -середніх у [77] запропоновано вдосконалений варіант кластеризації k -середніх, названий авторами *greedy global k-means*. В його основі лежить припущення, що глобальний оптимум може бути досягнутий шляхом запуску

ЛІТЕРАТУРА

1. Крупельницький Л. В. Аналого-цифрові пристрої систем, що самокоригуються, для вимірювань і оброблення низькочастотних сигналів : монографія / Л. В. Крупельницький, О. Д. Азаров ; під заг. ред. О. Д. Азарова. – Вінниця : УНІВЕРСУМ-Вінниця, 2005. – 167 с .
2. Ткаченко А. Н. Метод направленного поиска векторов в кодовых книгах / А. Н. Ткаченко, О. Ф. Грийо Тукало // Электронное моделирование. – 2010. – № 2, Т. 32. – С. 77–85.
3. Метод кластеризації на основі послідовного запуску k-середніх з удосконаленим вибором кандидата на нову позицію вставки [Електронний ресурс] / О. М. Ткаченко, О. Ф. Грийо Тукало, О. В. Дзись, С. М. Лаховець // Наукові праці ВНТУ. – 2012. – № 2. – Режим доступу до журн.: <http://praci.vntu.edu.ua/index.php/praci/article/view/323/321>.
4. Ткаченко О. М. Пошук векторів у кодових книгах при ущільненні мовлення на основі бінарного дерева / О. М. Ткаченко, О. Ф. Грийо Тукало // Інформаційні технології та комп'ютерна інженерія. – 2011. – № 1. – С. 38–44.
5. Метод кластеризації на основі послідовного запуску k-середніх з обчисленням відстаней до активних центроїдів / О. М. Ткаченко, О. Ф. Грийо Тукало, Н. О. Біліченко, О. В. Дзись // Реєстрація, зберігання і обробка даних. – 2012. – № 1, Т. 14. – С. 25–34.
6. Ткаченко А. Н. Метод дельта-квантования параметров речевого сигнала с быстрым поиском ближайшего вектора в кодовой книге / А. Н. Ткаченко, О. Ф. Грийо Тукало, А. В. Дзись // Управляющие системы и машины. – 2013. – № 4. – С. 3–11.
7. Ткаченко О. М. Метод швидкого пошуку найближчого сусіда з обчисленням відстані за зваженою евклідовою метрикою / О. М. Ткаченко, О. Ф. Грийо Тукало // Вісник Вінницького політехнічного інституту. – 2013. – № 1. – С. 116–122.
8. Ткаченко О. М. Підхід до оцінювання тривалості фрагмента для пошуку музичного твору за заданим шаблоном / О. М. Ткаченко,

О. Ф. Грійо Тукало // Інформаційні технології та комп'ютерна інженерія. – 2014. – № 1. – С. 31–40.

9. Ткаченко О. М. Метод підвищення швидкості пошуку фрагмента аудіозапису із застосуванням kd-дерев / О. М. Ткаченко, О. Ф. Грійо Тукало // Інформаційні технології та комп'ютерна інженерія. – 2014. – № 3. – С. 57–66.

10. Полігармонійні методи вимірювання частотних характеристик звукових каналів і трактів / О. Д. Азаров, В. А. Гарнага, Л. В. Крупельницький, Д. Ю. Позняк // Інформаційні технології та комп'ютерна інженерія. – 2015. – № 2. – С. 23–29.

11. Ткаченко О. М. Двоетапна стратегія пошуку найближчого сусіда у векторних кодових книгах / О. М. Ткаченко, О. Ф. Грійо Тукало // Методи та засоби кодування, захисту й ущільнення інформації : III Міжнар. наук.-практ. конф., 20–22 квітня 2011 р. : тези доповідей. – Вінниця : ВНТУ, 2011. – С. 188–189. – 2 с.

12. Ткаченко О. М. Метод швидкого пошуку векторів у кодових книгах на основі бінарного дерева / О. М. Ткаченко, О. Ф. Грійо Тукало // Інформаційні технології та комп'ютерна інженерія : III Міжнар. наук.-практ. конф., 29–31 травня 2012 р. : тези доповідей. – Вінниця : ВНТУ, 2012. – С. 224–225. – 2 с.

13. Крупельницький Л. В. Характеристики і структури багатоканальних АЦ-систем, що самокорегуються, для аналізу аудіо сигналів // Методи та засоби кодування, захисту й ущільнення інформації : тези доповідей П'ятої Міжнародної науково-практичної конференції. Україна, Вінниця, 19–21 квітня 2016 р. – Вінниця : ВНТУ, 2016. – С. 129–133.

14. Ткаченко О. М. Ідентифікація аудіозапису за коротким фрагментом на основі приведеної власної відстані / О. М. Ткаченко, О. Ф. Грійо Тукало // Контроль і управління в складних системах: XII Міжнар. конф., 14–16 жовтня 2014 р. : тези доповідей. – Вінниця : ВНТУ, 2014. – С. 94.

15. Ткаченко О. М. Ідентифікація фрагмента музичного твору на основі приведеної власної відстані / О. М. Ткаченко, О. Ф. Грійо Тукало // Оброблення сигналів і зображень та розпізнавання образів :

XII Всеукр. міжнар. конф., 3 –7 листопада 2014 р. – Київ, 2014. – С. 23–26. – 4 с.

16. Спеціалізоване і вимірювальне обладнання власної розробки і виробництва для телерадіомовлення. Каталог НТЦ «Аналого-цифрові системи» ВНТУ / О. Д. Азаров, Л. В. Крупельницький, В. Я. Стейскал, О. А. Білоконь. – Вінниця : ВНТУ, 2015. – 40 с.

17. Крупельницький Л. В. Особливості метрологічної атестації вимірювальних каналів системи офіційного телерадіомоніторингу / Л. В. Крупельницький, М. Ф. Наталюк, С. Л. Мамука // Системи-2008: метрологія, стандартизація, сертифікація : тези доповідей міжнародної науково-технічної конференції. – Львів, 2008. – С. 74–78.

18. Грийо Тукало О. Ф. Быстрый поиск векторов при сжатии речевых сигналов : монография / О. Ф. Грийо Тукало, А. Н. Ткаченко. – Saarbrücken : LAP LAMBERT Academic Publishing. – 2014. – 108 с.

19. AD systems for processing of low frequency signals based on self calibrate ADC and DAC with weight redundancy / O. D. Azarov, L. V. Krupelnitskyi, Vinnytsia National Technical University (Ukraine), P. Komada, T. Ławicki, Lublin University of Technology (Poland), N. Askarova, A. Sagymbekova, Kazakh National Research Technical University (Kazakhstan) // Przegląd Elektrotechniczny, – 2017, – V. R. 93, no. 5. – P. 125–128. – DOI 10.15199/48.2017.05.26.

20. Muller M. Signal processing for music analysis / M. Muller, D. Ellis, A. Klapuri, and G. Richard // IEEE Journal of Selected Topics in Signal Processing. – 2011. – V. 5, no. 6. – P. 1088–1110. – ISSN 1932-4553.

21. Virtanen T. Probabilistic model based similarity measures for audio query-by-example / T. Virtanen and M. Helén // Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA '07), New Paltz, – NY, USA, – October 2007 – P. 82-85.

22. Caplan P. International metadata initiatives: Lessons in bibliographic control / P. Caplan // In Proceedings of the Bicentennial Conference on Bibliographic Control for the New Millennium: Confronting the Challenges of Networked Resources and the Web, Library of Congress, Washington, D. C., November 15-17, 2000. Режим доступу: http://www.loc.gov/catdir/bibcontrol/caplan_paper.Html.

23. Iterative design of metadata creation tools for resource authors / J. Greenberg, A. Crystal, W. Robertson, E. Leadem // Proceedings of the 2003 Dublin Core Conference: Supporting Communities of Discourse and Practice – Metadata Research and Applications, Seattle, Washington, – September 28-October 2, – 2003. Режим доступу: http://www.siderean.com/dc2003/202_Paper82-colorNEW.pdf.

24. Automatic document metadata extraction using support vector machines / H. C. Han, L. Giles, E. Manavoglu, H. Zha, Z. Zhang, E. A. Fox // In Proceedings of the Third ACM/IEEE-CS Joint Conference on Digital Libraries. – New York : ACM Press – 2003. – P. 37–48.

25. Hatala M. System for computer-aided metadata creation / Hatala M., Forth S. // In Proceedings of 12th International Conference of the World Wide Web Consortium (WWW2003), – Budapest, May 20-24, – 2003. Режим доступу: <http://www.sfu.ca/~mhatala/pubs/hatala-forth-www2003.pdf>.

26. Andric A. Automatic playlist generation based on tracking user's listening habits / A. Andric, G. Haus. // Multimedia Tools and Applications. – 2006. – V. 29. – P. 127–151.

27. Playlist generation using start and end songs / A. Flexer, D. Schnitzer, M. Gasser, and G. Widmer // In Proceedings of the 9th International Conference of Music Information Retrieval – ISMIR'08. – 2008. – P. 173–178.

28. Pampalk E. Islands of music: Analysis, organization, and visualization of music archives / E. Pampalk // Journal of the Austrian Society for Artificial Intelligence. – 2003. – V. 22, no. 4. – P. 20–23.

29. Knees P. Exploring music collections in virtual landscapes / P. Knees, M. Schedl, T. Pohle, and G. Widmer // IEEE Multimedia. – 2007. – V. 14, no. 3. – P. 46–54.

30. Islands of Gaussians: The self-organizing map and gaussian music similarity features / D. Schnitzer, A. Flexer, G. Widmer, and M. Gasser // In Proceedings of the 11th International Conference on Music Information Retrieval (ISMIR'10). – 2010. – P. 327-332.

31. Tzanetakis G. Musical genre classification of audio signals / G. Tzanetakis and P. Cook // IEEE Trans. Speech Audio Process. – No. 5. – V. 10. – 2002. – P. 293–301.
32. McKay C. Automatic genre classification using large high-level musical feature sets / C. McKay and I. Fujinaga // In Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR'04). – 2004. – P. 525–530.
33. A flexible framework for key audio effects detection and auditory context inference / R. Cai, L. Lu, A. Hanjalic, H-J. Zhang, L-H. Cai // IEEE Transactions on Speech, Audio and Language Processing. – 2006. – V. 14, no. 3. – P. 1026–1039.
34. Casey M. Fast recognition of remixed music audio / M. Casey and M. Slaney // In IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP 2007. – 2007. – V. 4. – P. 1425–1428.
35. A review of audio fingerprinting / P. Cano, E. Batlle, T. Kalker, and J. Haitsma // Journal of VLSI Signal Processing. – 2005. – V. 41. – P. 271–284.
36. Downie J. S. Music information retrieval / J. S. Downie // Annual Review of Information Science and Technology. – 2003. No. 37. – P. 295–340.
37. Hoos H. H. GUIDO/MIR—an experimental musical information retrieval system based on GUIDO music notation. / H. H. Hoos, K. Renz, and M. Georg // Proceedings of the International Symposium on Music Information Retrieval. – 2001. – P. 41–50.
38. Content-based multimedia information retrieval: State of the art and challenges / Lew M. S. et al. // ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM). – 2006. – T. 2, № 1. – P. 1–19.
39. Hoashi K. Personalization of user profiles for content-based music retrieval based on relevance Feedback / K. Hoashi, K. Matsumoto, and N. Inoue // Proceedings of the ACM International Conference on Multimedia, – 2003. – P. 110–119.

40. Manjunath B. S. Introduction to MPEG-7 multimedia content description interface / B. S. Manjunath, P. Salembier, T. Sikora. – England : John Wiley & Sons, LTD, 2002. – 396 p.
41. Logan B. A music similarity function based on signal analysis / B. Logan and A. Salomon // in Proc. IEEE Int. Conf. Multimedia Expo. – Tokyo, Japan, 2001. – P. 745–748.
42. Pandora internet radio. [Електронний ресурс]. – Режим доступу: <http://www.pandora.com/>
43. Stevens S. S. A scale for the measurement of a psychological magnitude: loudness / S. S. Stevens // Psychological Review. – 1936. – V. 43, no. 5. –P. 405–416.
44. Reiss J. Efficient multidimensional searching routines for music information retrieval / J. Reiss, J. J. Aucouturier, M. Sandler // 2nd Annual International Symposium on Music Information Retrieval. – 2001.
45. Gold B. Speech and audio signal processing / B. Gold and N. Morgan. – New York : Wiley, 2000. – 688 p.
46. Logan B. Mel frequency kestrrel coefficients for music modeling / B. Logan // International Symposium on Music Information Retrieval – Plymouth, – 2000. – Режим доступу: http://ismir2000.ismir.net/papers/logan_paper.pdf.
47. Signal Processing for Music Analysis Meinard Müller, Member, IEEE, Daniel P. W. Ellis, Senior Member, IEEE, Anssi Klapuri, Member, IEEE, and Gaël Richard, Senior Member, IEEE Journal of Selected Topics in Signal Processing. – 2011. – V. 5, no. 6.
48. Pampalk E. Computational models of music similarity and their application to music information retrieval : Ph. D. dissertation / E. Pampalk ; Vienna Univ. of Technology. – Vienna, Austria, 2006. – 153 p.
49. Aucouturier J.-J. The bag-of-frames approach to audio pattern recognition: A sufficient model for urban soundscapes but not for polyphonic music / J.-J. Aucouturier, B. Defreville, and F. Pachet // J. Acoust. Soc. Am. – 2007. – V. 122, no. 2. – P. 881–891.
50. Li T. Toward intelligent music information retrieval / T. Li and M. Ogihara // IEEE Trans. Multimedia. – 2006. – V. 8, no. 3. – P. 564–574.

51. Fukunaga K. Introduction to Statistical Pattern Recognition / K. Fukunaga // Academic Press, 2nd edition. – London, 1990. – 592 p.
52. Quatieri T. F. Discrete-Time Speech Signal Processing / T. F. Quatieri. – New Jersey : Prentice Hall, 2002. – 781 p.
53. Tzanetakis G. Musical genre classification of audio signals / G. Tzanetakis and P. Cook // IEEE Transactions on Speech and Audio Processing. – 2002. – V. 10, no. 5. – P. 293 – 302.
54. Levy L. Lightweight measures for timbral similarity of musical audio / M. Levy and M. Sandler // In Proceedings of the 1st ACM workshop on Audio and music computing multimedia, AMCMM '06, – NY, USA : ACM, 2006. – P. 27–36.
55. McKay C. Automatic genre classification using large high-level musical feature sets / C. McKay and I. Fujinaga // in Proc. Int. Symposium. Music Inform. Retrieval. (ISMIR). – 2004. – P. 525–530.
56. Pampalk E. Audio-based music similarity and retrieval: combining a spectral similarity model with information extracted from fluctuation patterns [Електронний ресурс] / E. Pampalk // Implementation submitted to the 3rd Annual Music Information Retrieval eXchange (MIREX'06). – Victoria, Canada, 2006. – Режим доступу: http://www.ofai.at/~elias.pampalk/publications/pam_mirex06.pdf
57. Evaluating rhythmic descriptors for musical genre classification / F. Gouyon, S. Dixon, E. Pampalk, and G. Widmer // In Audio Engineering Society Conference: 25th International Conference: Metadata for Audio. – London, 2004. – P. 196–204.
58. Comparing MFCC and MPEG-7 audio features for feature extraction, maximum likelihood HMM and entropic prior HMM for sports audio classification / Z. Xiong, R. Radhakrishnan, A. Divakaran, T. Huang // ICASSP. – 2003. – V. 5. – P. 628–631.
59. Shepard R. N. Circularity in judgments of relative pitch / R. N. Shepard // Journal of the Acoustical Society of America. – 1964. – V. 36, no. 12. – P. 2346–2353.
60. Bartsch M. A. Audio thumbnailing of popular music using chroma-based representations / M. A. Bartsch, G. H. Wakefield // IEEE Trans. on Multimedia. – 2005. – V. 7, no. 1. – P. 96–104.

61. Gomez E. Tonal description of polyphonic audio for music content processing / E. Gomez // *INFORMS Journal on Computing, Special Cluster on Computation in Music*. – 2006. – V. 18, no. 3. – P. 194–302.
62. Casey M. MPEG-7 sound-recognition tools / M. Casey // *IEEE Transactions on Circuits and Systems for Video Technology*. – 2001. – V. 11, no. 6. – P. 737–747.
63. Peeters G. Template-based estimation of time-varying tempo / G. Peeters // *EURASIP Journal of Advances in Signal Processing*. – 2007. – No. 1. – P. 158.
64. Ganchev T. Comparative evaluation of various mfcc implementations on the speaker verification task / T. Ganchev, N. Fakotakis, and G. Kokkinakis // *In Proceedings of 9th International Conference on Speech and Computer, SPECOM'05*. – 2005. – P. 191–194.
65. Aucouturier J.-J. Improving timbre similarity: How high's the sky? / J.-J. Aucouturier and F. Pachet // *J. Negative Results Speech Audio Sci*. – 2004. – V. 1, no. 1.
66. Goto M. A chorus-section detection method for musical audio signals and its application to a music listening station / M. Goto // *IEEE Transactions on Audio, Speech, and Language Processing*. – 2006. – V. 14, no. 5. – P. 1783–1794.
67. Peeters G. Musical key estimation of audio signal based on hidden Markov modeling of chroma vectors / G. Peeters // *in Proc. Int. Conf. Digital Audio Effects (DAFx)*. – Montreal, QC, Canada, 2006. – P. 127–131.
68. Jensen J. H. A tempo-insensitive distance measure for cover song identification based on chroma features / J. H. Jensen, M. G. Christensen, D. P. W. Ellis, and S. H. Jensen // *in Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing – Las Vegas, USA, 2008*. – P. 2209–2212.
69. Jehan T. Creating music by listening [Електронний ресурс] / T. Jehan // *PhD diss., Massachusetts Institute of Technology*. – Massachusetts, USA, 2005. – Режим доступу: <http://web.media.mit.edu/~tristan/phd/>.

70. Agrell E. Spectral coding by fast vector quantization / E. Agrell // Proc. IEEE Workshop on Speech Coding for Telecommunications, Sainte-Adèle. – Québec, Canada, 1993. – P. 61–62.

71. West K. Features and classifiers for the automatic classification of musical audio signals / K. West and S. Cox // in Proc. Int. Symp. Music Inform. Retr. (ISMIR). – 2004. – P. 1–6.

72. T. Li. Toward intelligent music information retrieval / T. Li and M. Ogihara // IEEE Trans. Multimedia. – 2006. – V. 8, no. 3. – P. 564–574.

73. Kohonen T. Self-Organizing Maps / T. Kohonen. – Berlin : Springer Sciences, 2001. – 420 p.

74. Mandel M. Song-level features and support vector machines for music classification / M. Mandel and D. Ellis // In Proceedings of the 6th International Conference on Music Information Retrieval, ISMIR'05. – Queen Mary University of London, 2005. – P. 594–599.

75. On rhythm and general music similarity / T. Pohle, D. Schnitzer, M. Schedl, P. Knees, and G. Widmer // In Proceedings of the 10th International Conference on Music Information Retrieval, ISMIR'09, 2009. – P. 525–530.

76. Pampalk E. Content-based organization and visualization of music archives / E. Pampalk, A. Rauber, and D. Merkl // In Proceedings of the tenth ACM international conference on Multimedia, MULTIMEDIA '02. – NY, USA : ACM, 2002. – P. 570–579.

77. Hussein N. A Fast Greedy K-Means Algorithm / Master's Thesis Nr:9668098 ; University of Amsterdam Faculty of Mathematics, Computer Sciences, Physics and Astronomy Euclides Building Plantage muidergracht 24. – November, 2002. – 62 p.

78. Maneewongvatana S. Analysis of approximate nearest neighbor searching with clustered point sets / S. Maneewongvatana and D. Mount // In ALENEX. – Baltimore, 1999. – P. 105–123.

79. Neumayer R. Content-based organization of digital audio collections / R. Neumayer, T. Lidy, and A. Rauber // In Proceedings of the 5th Open Workshop of MUSICNETWORK, 2005. – P. 4–5.

80. Andoni A. Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions / A. Andoni and P. Indyk // In 47th

Annual IEEE Symposium on Foundations of Computer Science. FOCS'06 –NY, USA, 2006. – P. 459–468.

81. Bentley J. L. Multidimensional binary search trees used for associative searching / J. L. Bentley // Communications of the ACM. – 2007. – V. 18, no. 9. – P. 509–517.

82. Yianilos P. N. Data structures and algorithms for nearest neighbor search in general metric spaces / P. N. Yianilos // Proc. 4th ACM-SIAM Sympos. Discrete Algorithms. – Philadelphia, USA, 1993. – P. 311–321.

83. An optimal algorithm for approximate nearest neighbor searching / S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Wu // Journal of the ACM. – 1998. – V. 45. – P. 891–923,

84. Yao A. C. A general approach to d-dimensional geometric queries / A. C. Yao and F. F. Yao // In Proc. 17th Ann. ACM Sympos. Theory Comput. – NY, USA, 1985. – P. 163–168.

85. Clarkson K. L. Fast algorithms for the all nearest neighbors problem / K. L. Clarkson // In Proc. 24th Ann. IEEE Sympos. on the Found. Comput. Sci. – Tucson, AZ, USA, 1983. – P. 226–232.

86. Friedman J. H. An algorithm for finding best matches in logarithmic expected time. / J. H. Friedman, J. L. Bentley, and R. A. Finkel // ACM Transactions on Mathematical Software. – 1977. – No. 3(3). – P. 209–226.

87. Bern M. Approximate closest-point queries in high dimensions / M. Bern // Inform. Process. Lett. – 1993. – V. 45, no. 2. – P. 95–99.

88. Arya S. Approximate nearest neighbor queries in fixed dimensions / S. Arya and D. M. Mount // Proc. 4th ACM-SIAM Sympos. Discrete Algorithms, – Philadelphia, USA, 1993. – P. 271–280.

89. Chan T. Approximate nearest neighbor queries revisited / T. Chan // Proc. 13th Annu. ACM Sympos. Comput. Geom. – NY, USA, 1997. – P. 352–358.

90. Indyk P. Approximate nearest neighbors: Towards removing the curse of dimensionality / P. Indyk and R. Motwani // Proc. 30th Annu. ACM Sympos. Theory Comput. – NY, USA, 1998. – 604–613.

91. Kushilevitz E. Efficient search for approximate nearest neighbor in high dimensional spaces / E. Kushilevitz, R. Ostrovsky, and Y. Rabani // Proc. 30th Annu. ACM Sympos. Theory Comput. – NY, USA, 1998. – P. 614–623.
92. Goto M. A predominant-F0 estimation method for CD recordings: MAP estimation using EM algorithm for adaptive tone models / M. Goto // in Proc. ICASSP. – 2001. – V. 5. – P. 3365–3368.
93. Liu C. C. An approximate string matching algorithm for content-based music data retrieval / C. C. Liu, J. L. Hsu, A. L. P. Chen // IEEE International Conference on Multimedia Computing and Systems. – 1999. – V. 1. – P. 451–456.
94. Duda R. O. Pattern classification and scene analysis / R. O. Duda, P. E. Hart. – New York : John Wiley & Sons, 1973. – 512 p.
95. Audio fingerprinting: concepts and applications. Book chapter / P. Cano, E. Battle, E. Gómez, R. De C. T. Gomes, and M. Bonnet. – Berlin, Heidelberg : Springer-Verlag, 2005. – P. 233–245.
96. Somervuo P. Experiments with linear and nonlinear feature transformations in HMM based phone recognition / P. Somervuo // In proceedings of ICASSP. – 2003. – V. 1. – P. 52–55.
97. Haeb-Umbach R. Linear discriminant analysis for improved large vocabulary continuous speech recognition / R. Haeb-Umbach and H. Ney // In Proceedings of IEEE ICASSP. – San Francisco, CA, 1992. – P. 13–16.
98. Abbasian H. Class-dependent PCA optimization using genetic programming for robust MFCC extraction / H. Abbasian and B. A. NaserSharif // In: 3rd Conference on Information and Knowledge technology (IKT). – Mashhad, Iran, 2007. – P. 1541–1544.
99. Korada S. B. Dimensionality reduction in feature vector using principle component analysis (PCA) for effective speaker recognition / S. B. Korada, Y. Anitha, K. K. V. S Anjana // International Journal of Applied Information Systems (IJ AIS). – New York, USA. – 2013. – V. 5, no. 5. – P. 15–17.
100. Борович А. Психоакустически мотивированный алгоритм фильтрации шума окружающей среды на основе обработки речевого

сигнала в подпространствах / А. Борович, А. А. Петровский // Речевые технологии. – 2013. – № 3–4. – С. 12–29.

101. Grosche P. Audio content-based music retrieval / P. Grosche, M. Müller, J. Serrà // Dagstuhl Follow-Ups Multimodal Music Processing. – Dagstuhl, Germany. – 2012. – V. 3. – P. 157–175.

102. Ke Y. Computer vision for music identification / Y. Ke, D. Hoiem, and R. Sukthankar // In Proc. Computer Vision and Pattern Recognition. – 2005. – V. 2. – P. 1184–1192.

103. Aucouturier J. J. Music similarity measures: What's the use / J. J. Aucouturier and F. Pachet // In 3rd International Conference on Music Information Retrieval. ISMIR'02. – Paris, France, 2002. – P. 157–163.

104. Seyerlehner K. Fusing block-level features for music similarity estimation / K. Seyerlehner, G. Widmer, and T. Pohle // In Proceedings of the 13th International Conference on Digital Audio Effects. DAFx-10. – Graz, Austria, 2010.

105. Tianjing X. Music Identification via Vocabulary Tree with MFCC Peaks / Tianjing Xu, Adams Wei Yu, Xianglong Liu, Bo Lang // In Proc. 1st international ACM workshop on Music information retrieval with user-centered and multimodal strategies (MIRUM). – Scottsdale, USA, 2011. – P. 21–26.

106. Batlle E. Automatic song identification in noisy broadcast audio / E. Batlle, J. Masip, E. Guaus // Proceedings of the International Conference on Signal and Image Processing. – Kauai, USA, 2002.

107. Mohri M. Robust music identification, detection, and analysis / M. Mohri, P. Moreno, and E. Weinstein // International Conference on Music Information Retrieval (ISMIR). – Vienna, Austria, 2007. – P. 197–207.

108. Quantitative analysis of a common audio similarity measure / J. H. Jensen, M. G. Christensen, D. P. W. Ellis, S. H. Jensen // IEEE Transactions on Audio, Speech, and Language Processing. – Aalborg, Denmark. – 2009. – V. 17, no. 4. – P. 693–703.

109. Билобров С. В. Метод идентификации аудиоматериалов с использованием технологии аудиоотпечатков / С. В. Билобров // Штучний інтелект. – 2006. – № 3. – С. 575–583.

110. Mohri M. Robust music identification, detection, and analysis / Mehryar Mohri, Pedro Moreno, and Eugene Weinstein // In Proc. International Conference on Music Information Retrieval (ISMIR), 2007. – P. 123–130.

111. Cano P. An industrial-strength content-based music recommendation system / P. Cano, M. Koppenberger and N. Wack // In Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, SIGIR'05, –NY, USA : ACM, 2005. – P. 673–673.

112. Scalable music recommendation by search / R. Cai, C. Zhang, L. Zhang, and W. Y. Ma // In Proceedings of the 15th international conference on Multimedia, MULTIMEDIA '07, – NY, USA : ACM, 2007. – P. 1065–1074.

113. Casey M. Song intersection by approximate nearest neighbor search / M. Casey and M. Slaney // In Proceedings of the 7th International Conference of Music Information Retrieval. ISMIR'06, 2006. – P. 144–149.

114. McFee B. Large-scale music similarity search with spatial trees / B. McFee, G. R. G. Lanckriet // ISMIR, 2011. – P. 55–60.

115. Exploiting the tradeoff between precision and cpu-time to speed up nearest neighbor search / P. Roy, J. J. Aucouturier, F. Pachet, and A. Beurive // In Proceedings of the 6th International Conference on Music Information Retrieval. ISMIR'05, – London, UK, 2005. – P. 230–237.

116. Schnitzer D. A filter-and-refine indexing method for fast similarity search in millions of music tracks / D. Schnitzer, A. Flexer, G. Widmer // In Proc. International Conference on Music Information Retrieval, ISMIR, 2009. – P. 537–542.

117. Typke R. A Survey of Music Information Retrieval Systems / R. Typke, F. Wiering, R. C. Veltkamp // Proceedings of the Fifth International Conference on Music Information Retrieval (ISMIR 2005), 2005. – P. 153–160.

118. Advances in knowledge discovery and data mining / U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, R. Uthurusamy. – AAAI/MIT Press, 1996. – 611 p.

119. Gersho A. Vector quantization and signal compression. / A. Gersho, R. M. Gray. – Boston : Kluwer Academic. – 1992. – 760 p.
120. Kanungo T. A Local Search Approximation Algorithm for k-Means Clustering / T. Kanungo, D. M. Mount, N. S. Netanyahu, C. Piatko, R. Silverman, A. Y. Wu // Computational Geometry: Theory and Applications. – 2004. – No. 2. – P. 89–112.
121. Moore A. W. Efficient memory based learning for robot control / A. W. Moore // PhD thesis Nr, 1990. – University of Cambridge, Computer Laboratory UCAM-CL-TR-209. – 1990. – 248 p.
122. Daniel P. W. Echoprint - an open music identification service [Електронний ресурс] / P. W. Daniel, E. B. Whitman, A. Porter // 12th International Society for Music Information Retrieval Conference. – Miami, 2011. – Режим доступу: http://static.echonest.com/echoprint_ismir.pdf.
123. Maneewongvatana S. Analysis of approximate nearest neighbor searching with clustered point sets / S. Maneewongvatana and D. M. Mount // DIMACS Series in Discrete Mathematics and Theoretical Computer Science. – Baltimore. – 2002. – V. 59. – P. 105–123.
124. NAudio Overview. [Електронний ресурс]. – Режим доступу: <https://naudio.codeplex.com/>.
125. Libavcodec Documentation. [Електронний ресурс] Режим доступу: <https://www.ffmpeg.org/libavcodec.html>.
126. Козлов А. В. Алгоритм детектирования музыкальных фрагментов в задачах речевой обработки / А. В. Козлов, А. И. Лоханова, К. К. Симончик // Научно-технические ведомости СПбГПУ. – Издательство Политехнического университета. – 2010. – № 4(103). – С. 7–11.
127. Xu C. Automatic music classification and summarization / C. Xu, N. C. Maddage, and X. Shao // IEEE Trans. Speech Aud. Processing. – 2005. – V. 13, no. 3. – P. 441–450.
128. Carey, M. J. Comparison of features for speech, music discrimination / M. J. Carey, E. S. Parris, H. A. Lloyd-Thomas // In Proc. 1999 IEEE International conf. on Acoustics, Speech, and Signal Processing. – 1999. – V. 1. – P. 149–152.

129. Scheirer E. Construction and evaluation of a robust multifeature speech/music discriminator / E. Scheirer, M. Slaney // In Proc. IEEE International conf. on Acoustics, Speech, and Signal Processing. – 1997. – V. 1. – P. 1331–1334.

130. Automatic music detection in television productions / G. Widmer [et al.] // In Proc. of the International conf. on Digital Audio Effects. –2007.

131. Wang A. An industrial strength audio search algorithm [Електронний ресурс] / A. Wang // Conference on Music Information Retrieval. – 2003. – Режим доступу: <http://www.ee.columbia.edu/~dpwe/papers/Wang03-shazam.pdf/>.

132. Audio Tag. info [Електронний ресурс]. – Режим доступу: http://audiotag.info/faq_en.html.

133. Яндекс. Музыка. [Електронний ресурс]. – Режим доступу: <https://play.google.com/store/apps/details?id=ru.yandex.music&hl=ru>.

134. Биков М. М. Аналіз стану проблеми створення ефективних систем пошуку ключових слів / М. М. Биков, В. В. Ковтун, Н. Г. Савінова // Вісник ВПІ. – 2012. – № 1. – С. 179–181.

135. Пилипенко В. В. Распознавание ключевых слов в потоке речи при помощи фонетического стенографа / В. В. Пилипенко // Искусственный интеллект. – 2009. – № 4. – С. 220–224.

136. Пилипенко В. В. Пошук ключових слів в потоці мовлення із записів судових засідань [Електронний ресурс] / В. В. Пилипенко, О. М. Радущкий // Оброблення сигналів і зображень та розпізнавання образів : XII Всеукр. міжнар. конф., 15–19 жовтня 2012 р. – Київ, 2012. – 2 с. – Режим доступу: http://uasoiro.kibermova.com/files/Zbirnyk/2012/3/p_66.pdf.

Наукове видання

**Грійо Тукало Оксана Франсисківна
Ткаченко Олександр Миколайович
Крупельницький Леонід Віталійович**

**ПОШУК ЗАДАНИХ ФРАГМЕНТІВ
В АРХІВІ АУДІО ЗАПИСІВ
ІЗ ЗАСТОСУВАННЯМ КД-ДЕРЕВ**

Монографія

Редактор С. Малішевська

Оригінал-макет підготовлено О. Ткаченком

Підписано до друку 5.02.2020 р.
Формат 29,7×42¼. Папір офсетний.
Гарнітура Times New Roman.
Друк різнографічний. Ум. др. арк. 8,55.
Наклад 300 (1-й запуск 1–75) пр. Зам № В2020-04

Вінницький національний технічний університет,
ІРВЦ ВНТУ,
21021, м. Вінниця, Хмельницьке шосе, 95,
ВНТУ, ГНК, к. 114.
Тел. (0432) 59-85-32.
press.vntu.edu.ua; email: kivc.vntu@gmail.com.

Свідоцтво суб'єкта видавничої справи
серія ДК № 3516 від 01.07.2009 р.

Віддруковано ФОП Барановська Т. П.
21021, м. Вінниця, вул. Порика, 7.
Свідоцтво суб'єкта видавничої справи
серія ДК № 4377 від 31.07.2012 р.