

УДК 004.52

DOI <https://doi.org/10.32782/2663-5941/2024.3.1/13>**Завгородній В.В.**

Державний університет інфраструктури та технологій

Завгородня Г.А.

Державний університет інфраструктури та технологій

Завгородній В.В.

Державний університет інфраструктури та технологій

Глуценко А.І.

Державний університет інфраструктури та технологій

МЕТОД АНАЛІЗУ ПОДІБНОСТІ АУДІОЗАПИСІВ НА ОСНОВІ ХЕШ-ТОКЕНІВ

У роботі розглянуто метод аналізу подібності аудіозаписів на основі хеш-токенів. У світі музики, існує проблема для виконавців, пов'язана з пошуком схожих аудіозаписів за настроєм або стилем, яка виникає через обмежену функціональність традиційних музичних сервісів з алгоритмами рекомендацій. Основна проблема полягає в тому, що існуючі музичні платформи не завжди здатні точно враховувати особисті вподобання та потреби виконавців у пошуку нових музичних ідей. Тому, розробка зручного інструменту для виконавців, який дозволить їм шукати аудіозаписи, схожі за настроєм або стилем, стає актуальною задачею у музичній індустрії.

Основною концепцією у даній статті є поняття «відбитку» аудіозапису. «Відбиток» аудіозапису представляє собою набір хеш-токенів, які ідентифікують конкретний запис за ключем і певну частотну адресацію у спектрограмі. Ці хеш-токени використовуються як ключі для збереження основної інформації про аудіозапис у спеціальних сховищах. Це спрощує швидкий доступ до аудіозаписів та забезпечує зручний формат для зберігання цих даних.

Основні кроки для розробки системи аналізу подібності аудіозаписів включають оцифрування аудіозапису, дискретизацію, перетворення стерео в моно, даунсемплінг, виконання швидкого перетворення Фур'є, створення спектрограми, визначення найвищих частот за одиницю часу, збереження хеш-токенів у базі даних та порівняння цих відбитків між собою.

Для уточнення подібності та видалення аномалій рекомендується використовувати статистичний метод виявлення викидів. Такий крок допомагає точніше визначити список схожих аудіозаписів та зменшити обчислювальне навантаження на подальших етапах аналізу.

Після цього треба упорядкувати список аудіозаписів, що порівнюються з поточним, від найбільш до найменш схожих. Метод аналізу подібності дозволяє швидко і ефективно визначити ступінь схожості між аудіозаписами без необхідності порівняння кожної частини запису.

Ключові слова: відбиток аудіозапису, система рекомендацій, хеш-токен, спектрограма, цільова точка.

Постановка проблеми. На сьогоднішній день, у світі музики, існує проблема для виконавців, пов'язана з пошуком схожих аудіозаписів за настроєм або стилем, яка виникає через обмежену функціональність традиційних музичних сервісів з алгоритмами рекомендацій.

Традиційні музичні платформи, хоч і мають розвинені системи рекомендацій, зазвичай обмежені в тому, що пропонують користувачеві в основному вже популярні та відомі композиції. Це означає, що виконавці, які шукають інші аудіо-

записи для натхнення або для порівняння своєї творчості з іншими, можуть зіткнутися з обмеженнями у доступі до різноманітних музичних творів.

Аналіз останніх досліджень і публікацій. Літературні дослідження з аналізу звукових сигналів і спектрограм були проведені задовго до появи систем аналізу музичних записів. Робота [1–3] є основоположною у світі систем виявлення інформації у музиці (MIR). У ній досліджуються різні аспекти аналізу звукових сигналів та

спектрограм для створення ефективних систем аналізу музичних записів. Проте питання розвитку технологій MIR та використання новітніх методів аналізу у цій роботі розглянуто лише поверхово.

Автори робіт [4; 5] зосередилися на аналізі структури музики з використанням аудіозаписів. Вони описують методи і алгоритми, які використовуються для розпізнавання та аналізу музичних елементів.

У роботі [6–8] автор висвітлює ефективність підходу до класифікації музичних жанрів на основі спектрограм та згорткових нейронних мереж для аналізу музичних записів за їхнім жанром [9]. Автоматичне отримання характеристик музичних патернів з використанням згорткових нейронних мереж досліджують роботи [10–12], проте недостатньо розкритий вплив параметрів нейронних мереж на результати аналізу [13–15].

Дослідження новітніх технологій та алгоритмів для вдосконалення систем виявлення інформації у музиці, зокрема, розробка нових алгоритмів для класифікації музичних жанрів та структур.

Постановка завдання. Потреба у зручному інструменті для виконавців полягає в тому, щоб мати можливість швидко та ефективно знаходити аудіозаписи, які мають схожий стиль з їхніми власними творами.

Основна проблема полягає в тому, що існуючі музичні платформи не завжди здатні точно врахувати особисті вподобання та потреби виконавців у пошуку нових музичних ідей. Тому, розробка зручного інструменту для виконавців, який дозволить їм шукати аудіозаписи, схожі за настроєм або стилем, стає актуальною задачею у музичній індустрії.

Виклад основного матеріалу. Основною концепцією у даній статті є поняття «відбитку» аудіозапису. «Відбиток» аудіозапису представляє собою набір хеш-токенів, які ідентифікують конкретний запис за ключем і певну частотну адресацію у спектрограмі. Ці хеш-токени, зазвичай у форматі md5, використовуються як ключі для збереження основної інформації про аудіозапис у спеціальних сховищах, наприклад, hash-value. Це спрощує швидкий доступ до аудіозаписів та забезпечує зручний формат для зберігання цих даних. Основні кроки для розробки такої системи включають оцифрування аудіозапису, дискретизацію, перетворення стерео в моно, даунсемплінг (зменшення частоти дискретизації), виконання швидкого перетворення Фур'є (FFT), створення спектрограми, визначення найвищих частот за одиницю часу, збереження хеш-токенів (відбит-

ків) у базі даних та порівняння цих відбитків між собою.

Кожен аудіофайл може бути підданий процесу створення відбитка, який включає вилучення хеш-токенів для його представлення. Невідомі відбитки порівнюються з широким набором відбитків із бази даних. Як сама база даних, так і досліджуваний аудіофайл проходять той же процес створення відбитка.

Відповідність кандидатур оцінюється щодо правильності співпадіння. Один з критеріїв правильності – це синхронізація за часом. Група хеш-токенів може бути близькою або віддаленою від аналогічної групи хеш-токенів іншого аудіозапису, що може свідчити про те, що досліджуваний аудіозапис є частиною будь-якого часового діапазону іншого аудіозапису.

Для надійної ідентифікації пісень у присутності шумів та спотворень важливо враховувати лише найсильніші за амплітудою частоти в кожній часовій області. Такі «піки» у спектрограмі з великою ймовірністю витримають спотворення. Набори таких піків на спектрограмі двох подібних аудіозаписів будуть аналогічними або схожими. Спектрограма відображає потужність сигналу відносно часу на різних частотах, може бути представлена як двовимірний графік з третьою змінною – кольором, або тривимірний графік з четвертою змінною – кольором.

Перед зберіганням пікових значень у сховищі необхідно розуміти, як будуть порівнюватися різні аудіозаписи. Наприклад, якщо вирішено порівнювати кожну пікову частоту одного запису з кожною піковою частотою іншого запису, при наявності сотень аудіозаписів у сховищі це може призвести до значної кількості порівнянь.

Замість цього, для порівняння краще шукати кілька точок одночасно – ця група точок називається цільовою зоною. Щоб впевнитися, що як уривок, так і повна пісня, генерують ті самі цільові зони, необхідно встановити відносини між частотно-часовими точками у фільтрованій спектрограмі:

1. Якщо дві точки мають однаковий час, то першою вважається точка з найнижчою частотою.

2. Першими йдуть точки з найменшим часом.

Потрібно створити відносні адреси цільової зони, вибравши опорну точку за межами зони. Наприклад, розмір цільової зони дорівнює 5, а опорна точка – це третя точка перед кожною зоною (рис. 1). Далі необхідно сформулювати адресу наступного формату: [«Частота опорної точки», «Частота цільової точки», «Різниця у часі між опорною точкою та цільовою точкою»].

Тобто, адреса точки номер «4» у межах зони буде [10, 10, 2], а для точки номер «6» у тій же зоні – [10, 30, 2]. Для оптимального зберігання у вигляді ключа в сховищі хешів можна захешувати ці адреси в md5. Ключем може бути кортеж з абсолютного часу опорної точки в пісні та ідентифікатора пісні в базі даних, щоб забезпечити узгодженість часу.

На рисунку 2 показано можливий випадок, коли існують дві цільові зони, що належать двом різним пісням (верхній графік відноситься до першої досліджуваної пісні, нижній – до дру-

гої). Тобто цільові зони просто поміняли місцями у цих аудіозаписах. Якщо не враховувати часову узгодженість, ці цільові зони можуть збільшувати відсоток збігу між двома піснями, тоді як вони можуть звучати по-різному, оскільки ноти в цих цільових зонах не відтворюються в однаковому порядку. Слід оцінювати схожість двох аудіозаписів не лише на основі наявності однакових ключів у сховищі та їх кількості, але й з урахуванням часової узгодженості. Тому в сховище у значенні поля ключа зберігається абсолютний час.

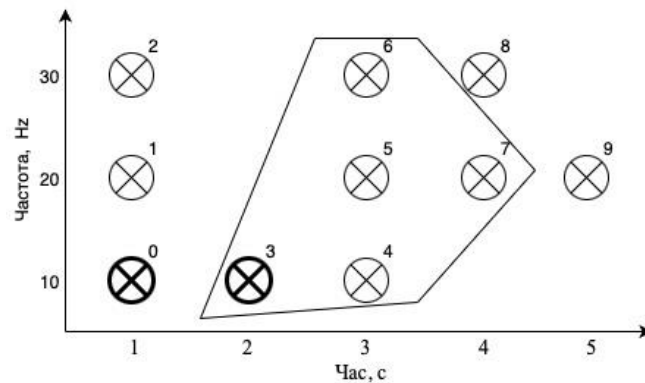


Рис. 1. Зображення цільової зони, а також цільової та опорної точок

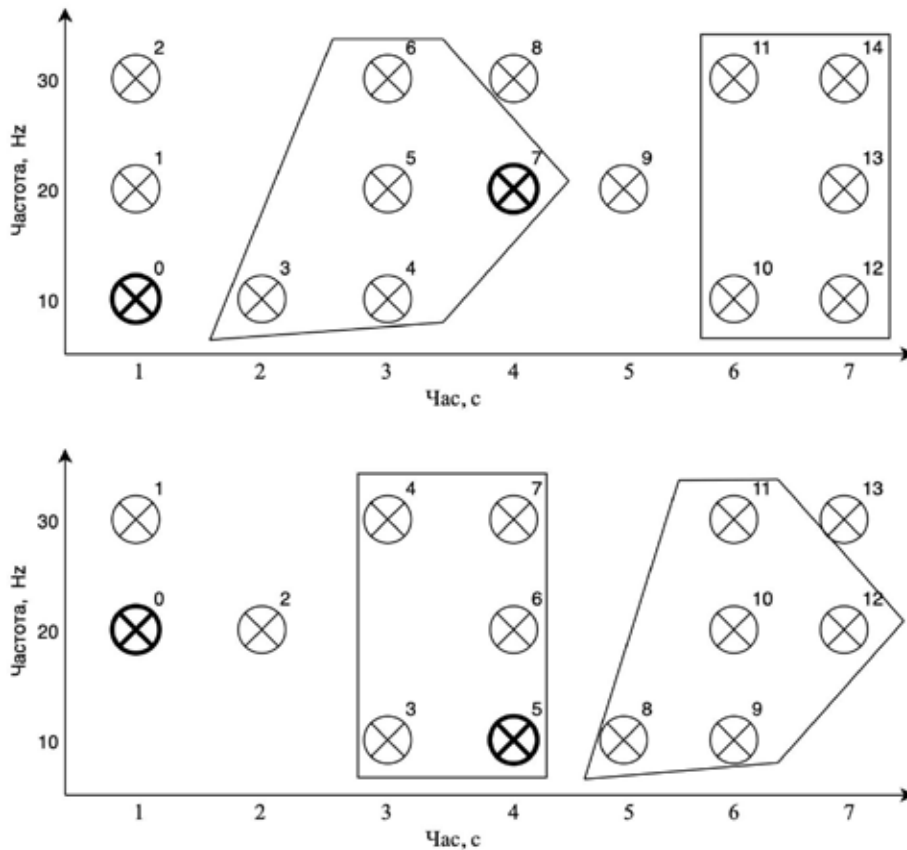


Рис. 2. Подібні цільові зони у різних аудіозаписах

За допомогою абсолютного часу можна визначити різницю (дельту) між поточним аудіозаписом і порівнюваним. Якщо безліч дельт прямує до схожого значення, можна зробити висновок про схожість даних записів. Алгоритм аналізу узгодженості за часом можна провести за такими кроками:

- 1) створення відбитка поточного аудіозапису;
- 2) знаходження запису у сховищі за ключем з використанням отриманих ключів із відбитка;
- 3) у значенні ключа має бути поле з ідентифікатором поточного аудіозапису i , можливо, поля з ідентифікаторами інших аудіозаписів;
- 4) якщо існують поля з ідентифікаторами інших аудіозаписів, то обчислюється дельта між абсолютним часом досліджуваного запису та поточним.

Після збереження даних у хеш-таблиці за алгоритмом аналізу узгодженості за часом, проводиться подальша обробка для визначення схожості аудіозаписів. Така обробка складається з наступних кроків:

1. Оцінка різниці (дельти) між абсолютними часами аудіозаписів проводиться для кожної пари записів у хеш-таблиці. Це допомагає визначити, наскільки схожі або відмінні записи за своїм часовим параметром.

2. За допомогою результатів аналізу дельт за часом групуються аудіозаписи, які мають схожі часові характеристики. Це допомагає створити кластери подібних записів для подальшого аналізу.

3. На основі групування або класифікації аудіозаписів обчислюється ступінь їхньої схожості. Це включає визначення метрик схожості, таких як відстань Манхеттена, косинусна схожість або інші метрики, що враховують особливості аудіозаписів.

4. На основі отриманих даних про схожість проводять класифікацію аудіозаписів за певними критеріями, такими як жанр, настрій, стиль тощо. Це дозволяє створити більш точні та розширені системи рекомендацій або пошуку подібних музичних творів.

5. Отримані дані можуть бути інтегровані з іншими алгоритмами аналізу сигналів або машинного навчання для ще більш точного визначення схожості аудіозаписів та покращення систем обробки музичної інформації.

Система має визначати подібність між аудіозаписами, а не обов'язково шукати лише один найбільш ідентичний запис. Це означає, що сис-

тема повинна виявляти як схожість реміксів треку, що цікавить, так і записів з певними шумами чи тишею.

Ремікси можуть мати лише деякі схожі секунди з оригіналом, тому слід враховувати такі варіації як подібні. Також, аудіозаписи можуть містити періоди тиші, що може впливати на правильність порівняння.

Для уточнення подібності та видалення аномалій рекомендується використовувати статистичний метод виявлення викидів, наприклад, правило трьох сигм. Цей підхід дозволяє відфільтрувати значення дельт часу, які відхиляються від середнього значення на певну кількість сигм. Такий крок допомагає точніше визначити список схожих аудіозаписів та зменшити обчислювальне навантаження на подальших етапах аналізу.

Після цього треба упорядкувати список аудіозаписів, що порівнюються з поточним, від найбільш до найменш схожих. Для цього можуть бути використані два методи: середньоквадратичного відхилення та ранжування за кількістю дельт.

Найбільш схожі записи мають найменшу різницю у значеннях дельт часу, що означає, що значення дисперсії набору дельт є пропорційним ступеню схожості аудіозаписів.

Другий метод полягає у визначенні кількості дельт, які найбільш близькі до певної кількості N дельт. Ці N дельт вибираються як перші N дельт у відсортованому за зменшенням кількості збігів списку дельт. Цей підхід ефективний для невеликих уривків, які мають низьке значення дисперсії порівняно з іншими записами через низьку ентропію дельт, пов'язану з короткою тривалістю запису.

Висновки. Метод аналізу подібності аудіозаписів на основі хеш-токенів базується на використанні унікальних хеш-токенів, які представляють собою відбиток аудіозапису. Основні методи включають: створення відбитка, збереження хеш-токенів, пошук подібних аудіозаписів, визначення ступеня схожості. Цей метод дозволяє швидко і ефективно визначати ступінь схожості між аудіозаписами без необхідності порівняння кожної частини запису.

У майбутній роботі над цим проєктом планується впровадження автоматичного визначення музичного жанру за допомогою MFCC (Mel-frequency cepstral coefficients) і покращення якості визначення подібних аудіозаписів шляхом використання інформації про жанр.

Список літератури:

1. Coviello E., Chan A.B., Lanckriet G. Time Series Models for Semantic Music Annotation. *IEEE Transactions on Audio, Speech and Language Processing*. 2011. Vol. 19. № 5. P. 1343–1359. DOI: <https://doi.org/10.1109/TASL.2010.2090148>
2. Завгородній В.В., Завгородня Г.А., Байраковський В.В., Коваль Р.А. Розробка методу виявлення схожих пісень на основі аналізу їх звукових характеристик. *Вчені записки ТНУ імені В.І. Вернадського. Серія: Технічні науки*. 2023. Том 34 (73). № 3. С. 151–156. DOI: <https://doi.org/10.32782/2663-5941/2023.3.1/24>
3. Paulus J., Müller M., Klapuri A. State of the Art Report: Audio-Based Music Structure Analysis. *Proceedings of the 11th International Society for Music Information Retrieval Conference*. 2010. P. 625–636. URL: http://paulus.kapsi.fi/pubs/paulus_ismir10_star_presentation.pdf
4. Muller M., Ellis D., Klapuri A., Richard G. Signal processing for music analysis. *IEEE Journal of Selected Topics in Signal Processing*. 2011. Vol. 5. № 6. P. 1088–1110.
5. Shaikh T., Jadhav A. Music Genre Classification Using Neural Network. *ITM Web of Conferences*. 2022. Vol. 44. P. 03016. DOI: <https://doi.org/10.1051/itmconf/20224403016>
6. Zavgorodnya A., Zavgorodnii V., Plisenko V, Provatorov N., Kudientsov P. Methods modeling systems for the improvement of their reliability. *International Academy Journal Web of Scholar*. 2019. Vol. 9. Is. 39. P. 3–11. DOI: https://doi.org/10.31435/rsglobal_wos/30092019/6683
7. Mukhin V., Komaga Y., Zavgorodnii V., Zavgorodnya A., Herasymenko O., Mukhin O. Social Risk Assessment Mechanism Based on the Neural Networks. *IEEE International Conference on Advanced Trends in Information Theory (ATIT)*. 2020. P. 179–182. DOI: <https://doi.org/10.1109/ATIT49449.2019.9030519>
8. Lee C.H., Kim D. Automatic Musical Pattern Feature Extraction Using Convolutional Neural Network. *The International MultiConference of Engineers and Computer Scientists*. 2010. Vol. 10.
9. Mukhin V., Zavgorodnii V., Kornaga Ya., Zavgorodnya A., Krylov Ie., Rybalochka A., Kornaga V., Belous R. Devising a Method To Identify an Incoming Object Based on the Combination of Unified Information Spaces. *Eastern-European Journal of Enterprise Technologies*. 2021. Vol. 3 (2). No 111. P. 35–44. DOI: <https://doi.org/10.15587/1729-4061.2021.229568>
10. Yu D., Duan H., Fang J., Zeng B. Predominant Instrument Recognition Based on Deep Neural Network With Auxiliary Classification. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 2020. Vol. 28. P. 852–861. DOI: <https://doi.org/10.1109/TASLP.2020.2971419>
11. Pelchat N., Gelowitz C.M. Neural Network Music Genre Classification. *Canadian Journal of Electrical and Computer Engineering*. 2020. Vol. 43. No. 3. P. 170–173. DOI: <https://doi.org/10.1109/CJECE.2020.2970144>
12. Zavgorodnii V., Zavgorodnya A., Maiko V., Malikov V., Zhuk D. Methods and models for assessment of reliability of structural-complex systems. *World science*. 2018. Vol. 11. Is. 39. P. 5–14.
13. Goto M., Dannenberg R.B. Music Interfaces Based on Automatic Music Signal Analysis: New Ways to Create and Listen to Music. *IEEE Signal Processing Magazine*. 2019. Vol. 36. № 1. P. 74–81. DOI: <https://doi.org/10.1109/MSP.2018.2874360>
14. Kim T., Lee J., Nam J. Comparison and analysis of sampleCNN architectures for audio classification. *IEEE Journal of Selected Topics in Signal Processing*. 2019. Vol. 13. № 2. P. 285–297. DOI: <https://doi.org/10.1109/JSTSP.2019.2909479>
15. Wind W.Y.Ng, Zeng W., Wang T. Multi-Level Local Feature Coding Fusion for Music Genre Recognition. *IEEE Access*. 2020. Vol. 8. P. 152713–152727. DOI: <https://doi.org/10.1109/ACCESS.2020.3017661>

Zavgorodnii V.V., Zavgorodnya A.A., Zavgorodnii V.V., Hlushchenko A.I. METHOD OF ANALYSIS OF AUDIO RECORDINGS BASED ON HASH TOKENS

The paper considers the method of analyzing the similarity of audio recordings based on hash tokens. In the world of music, there is a problem for artists related to finding similar audio recordings by mood or style, which arises due to the limited functionality of traditional music services with recommendation algorithms. The main problem is that the existing music platforms are not always able to accurately take into account the personal preferences and needs of artists in the search for new musical ideas. Therefore, the development of a convenient tool for performers, which will allow them to search for audio recordings similar in mood or style, is becoming an urgent task in the music industry.

The main concept in this article is the concept of “imprint” of an audio recording. The “fingerprint” of an audio recording is a set of hash tokens that identify a specific recording by key and a certain frequency addressing in the spectrogram. These hash tokens are used as keys to store basic information about the audio recording in special repositories. This facilitates quick access to audio recordings and provides a convenient format for storing this data.

The basic steps for developing an audio similarity analysis system include digitizing the audio, sampling, converting stereo to mono, downsampling, performing a fast Fourier transform, creating a spectrogram, determining the highest frequencies per unit time, storing hash tokens in a database, and comparing these fingerprints between by myself

To clarify similarities and remove anomalies, it is recommended to use the statistical method of detecting outliers. Such a step helps to more accurately determine the list of similar audio recordings and reduce the computational load in the subsequent stages of the analysis.

After that, you need to sort the list of audio recordings that are compared with the current one, from the most to the least similar. The similarity analysis method allows you to quickly and efficiently determine the degree of similarity between audio recordings without the need to compare each part of the recording.

Key words: *audio fingerprint, recommendation system, hash token, spectrogram, target point.*