УДК 004.624
DOI https://doi.org/10.32838/2663-5941/2021.1-1/09

Войтко В.В.
Вінницький національний технічний університет

Бевз С.В.
Вінницький національний технічний університет

Бурбело С.М.
Вінницький національний технічний університет

Ставицький П.В.
Вінницький національний технічний університет

АНАЛІЗ ЗАСОБІВ ДЛЯ ПРОГРАМНОЇ РЕАЛІЗАЦІЇ СИСТЕМИ СИНТЕЗУ ТА АНАЛІЗУ МУЗИЧНИХ ЗВУКІВ

У статті розглянуто засоби та підходи до програмної реалізації системи синтезу та аналізу музичних композицій. Це дослідження фокусується на одному з компонентів розроблюваної системи, що відповідає за розпізнавання музичних даних. Для досягнення поставленої цілі цей компонент виконує низку послідовних кроків, що дозволяють виконувати аналіз вхідних аудіоданих, а також створювати базу даних, що використовуватиметься для порівняння та знаходження збігів. Першочерговим завданням компонента розпізнавання музичних композицій є перетворення початкових даних із бінарного формату, що представлений у вигляді масиву байтів, у вигляд, придатний для аналізу. Цей процес називається дискретизацією і дозволяє подати аудіодані у вигляді частотного подання звуку в часі. Додатковим виміром у такому разі є амплітуда звукових хвиль. Наведені три величини можна візуалізувати графічно у вигляді спектрограми, де вісь OY відповідає за частоту звуку за період часу, що представлено на осі OX. Третій вимір є найважливішим у процесі аудіоаналізу, оскільки дозволяє визначити фрагменти звукового спектра, що є стійкими до сторонніх шумів та забезпечують оптимізацію процесу пошуку збігів. Амплітуда звукових хвиль зображується у вигляді кольору на спектрограмі, а найяскравіші точки є локальними амплітудними максимумами або екстремумами, які використовуються в процесі розпізнавання музичних композицій. У статті детальніше розглядається підхід до створення відбитків музичних композицій, на основі яких відбувається пошук збігів. Крім того, описується роль алгоритмів стиснення та хешування в цьому процесі та обґрунтовується вибір алгоритму стиснення з втратами для досягнення заданої мети. Одним із підходів до оптимізації процесу хешування є використання алгоритму хешування з урахуванням розташування в межах спектрограми. Таким чином, доводиться зменшити кількість точок для порівняння і пришвидшити пошук збігів. Крім того, розглядається коефіцієнт Жаккара, який дозволяє визначати розмір дистанцій між точками під час їх розподілу на групи за допомогою алгоритму, описаного вище. Розглянуто подальші кроки та області для аналізу, що є необхідними в процесі програмної реалізації системи синтезу та розпізнавання музичних композицій. Серед таких компонентів є аналіз додаткових характеристик вхідних аудіоданих, таких як жанр, автор тощо. Крім того, визначено місце розглянутих компонентів у загальній розроблюваній системі синтезу та аналізу музичних звуків.

Ключові слова: *аудіоаналіз, розпізнавання музики, аудіовідбиток, спектрограма, стиснення даних, хешування.*

Постановка проблеми. Технології аналізу аудіоданих набувають поширення. Важливим є аналіз технологічного підходу та методу програмної реалізації системи синтезу та розпізнавання музичних композицій. Саме тому метою розробки є вдосконалення алгоритмів програмної реалізації компонентів системи ідентифікації аудіоконтенту для коректного аналізу та синтезу музичних звуків. Об'єктом дослідження є процеси ідентифікації звукових даних, алгоритми та імплементаційні підходи, що є структурними компонентами комбінованого методу синтезу та аналізу музичних звуків. Предметом дослідження є функціональні можливості та компоненти системи аналізу аудіоданих.

Аналіз останніх досліджень і публікацій. Під час розроблення рішень з аналізу та розпізна-

вання звукового контенту слід розглянути дослідження інженерів, що стоять за розробленням сервісу Shazam. Цей продукт є одним із найперших на ринку розпізнавання музичних композицій та надає базис для подальших досліджень [1].

Крім того, у процесі роботи було розглянуто підходи до створення відбитків музичних композицій. Було проаналізовано дослідження процесу хешування аудіоданих, проте особливістю тут є фокус на ідентифікацію початково закодованих таких даних в аудіофайл, як водяні знаки та ідентифікатори для верифікації контенту [2].

У роботі [3] описано складники системи синтезу та аналізу музичних композицій, подальший її розвиток і дослідження. Крім того, важливо розуміти загальне місце досліджуваних компонентів та підходів серед загального набору компонентів і складників системи. Більше інформації про це наведено в окремому дослідженні [4].

Крім того, для розгляду подальших кроків дослідження було розглянуто опис процесу ідентифікації автора музичної композиції за допомогою використання згорткових нейронних мереж [5].

Постановка завдання. Основним завданням є розгляд, розроблення та вдосконалення алгоритмів аналізу музичних композицій та їх застосування в комбінованому методі синтезу та аналізу музичних звуків, що є основою розробленої програмної системи. Так, приділяється увага підходам реалізації компонента розпізнавання та ефективного створення відбитків аудіоданих, що в подальшому використовується для оптимізації процесу пошуку збігів, а також є основою для подальшої розроблення компонента синтезу аудіозвуків.

Виклад основного матеріалу дослідження. Одним із найважливіших складників процесу розпізнавання музичних композицій є створення відбитків.

Процес створення відбитків складається зі стиснення і хешування.

Головною ціллю стиснення є зменшення розміру, який займають вхідні дані. Таким способом удається досягти економії використовуваної пам'яті для зберігання даних, а також підвищення швидкості пошуку самих даних.

Ураховуючи те, що під час розпізнавання аудіокомпозицій важливим є саме швидкий інформаційний пошук, використання алгоритмів стиснення даних може допомогти значній їх оптимізації. Під час програмної реалізації системи можна розглянути два основні чинники, які впливають на швидкість пошуку одиниці даних, що зображені у формулі 1 [5]:

$$\text{ШвидкістьПошуку}(\text{ОД}) \approx \text{Розмір}(\text{ОД}) + \text{Тип}(\text{ОД}), \quad (1)$$

де ОД – окремо взята одиниця даних, із яких складається загальний обсяг даних;

ШвидкістьПошуку(ОД) – швидкість пошуку одиниці даних серед загального набору даних;

Розмір(ОД) – кількість одиниць даних;

Тип(ОД) – тип даних, який визначає кількість байтів, що займає одиниця даних.

Є два типи алгоритмів стиснення: стиснення без втрат та стиснення з утратами.

Перевагою алгоритмів першого типу є те, що, маючи кінцеве оптимізоване значення, завжди є можливість відновити початкові дані без втрат. Проте алгоритми другого типу дозволяють досягти більш ефективної оптимізації, хоча й не дозволяють повною мірою відновлювати вхідні дані.

Для створення відбитків музичних композицій найбільш удалим буде використання алгоритмів стиснення з утратами. Такий підхід дозволяє досягти максимально можливих показників оптимізації вхідних даних. Крім того, враховуючи те, що відбитки створюються лише для локальних амплітудних максимумів [3, с. 2], у відтворенні музичних композицій на базі відбитків немає необхідності. Єдиною вимогою до відбитків є те, що певне значення відбитка має завжди відповідати певному або схожому фрагменту музичної композиції.

Хешування даних дозволяє досягти подальшої оптимізації їх розміру шляхом перетворення в єдине значення фіксованої довжини та розміру.

Таким чином, після стискання та хешування окремий відбиток музичної композиції матиме вигляд одного значення, що можна присвоїти одній змінній. У результаті такого процесу вхідна частотна характеристика аудіосигналу [3, с. 2] буде перетворюватися в масив відбитків, кожен із яких можна подати у вигляді цілого числа або рядка.

Під час створення відбитків для локальних амплітудних екстремумів музичних композицій необхідно враховувати такі характеристики музичних композицій:

– кількість аудіоканалів (моно, стерео);
– рівень дискретизації;
– бітову глибину.

Під час створення відбитків аудіокомпозицій не потрібно обробляти весь набір аудіоданих. Натомість обирається короткий набір локальних

амплітудних максимумів. Такий підхід дозволяє під час оброблення вхідних аудіоданих значно знизити якість аудіофайла зі збереженням усіх характеристик, необхідних для розпізнавання. Досягти цього можна шляхом виконання такої послідовності кроків:

– перетворення стереофайла в моно;
– зниження рівня дискретизації;
– зменшення бітової глибини.

Цей підхід дозволяє значно зменшити розмір вхідних аудіоданих, водночас підвищуючи швидкість розпізнавання.

Для реалізації функціоналу розпізнавання музичних композицій необхідно створити першочергову базу даних відбитків. Таким чином, під час розпізнавання музичної композиції, вона буде проаналізована, а на її основі буде створено набір відбитків.

Потім необхідно виконати пошук серед хеш-значень у базі даних для знаходження збігів. Якщо для цього завдання використовувати лінійний пошук, то швидкість пошуку не буде оптимальною, якщо база даних містить багато значень. Задля вирішення цієї проблеми необхідно виконати оптимізацію на стадії хешування.

Одним із підходів, що дозволяє пришвидшити пошуково-ідентифікаційний процес, є хешування з урахуванням розташування даних. Основною особливістю такого підходу є те, що локальні екстремуми на спектрограмі, що розташовані близько один від одного, належать до одного набору. Як наслідок, декілька хеш-значень, що мають схоже розташування, оптимізуються в одне значення, що описує конкретний набір. Важливим параметром у такому алгоритмі є значення відстані, що дозволяє визначити коефіцієнт схожості локальних максимумів на спектрограмі. Це дозволяє віднести таку точку до того чи іншого набору даних.

Для визначення показника відстані та розбиття початкового набору амплітудних максимумів пропонуємо застосувати коефіцієнт Жаккара.

Використовуючи коефіцієнт Жаккара, можна визначити, наскільки два набори даних схожі між собою. Цей показник вираховується за допомогою формули 2 на множині точок екстремумів [6]:

$$J(A,B) = |A \cap B| / |A \cup B| \quad (2)$$

Для перетворення цього значення в коефіцієнт відстані між наборами амплітудних максимумів необхідно використовувати формулу 3 [6]:

$$dj = 1 - J(A, B) \quad (3)$$

Використання цього коефіцієнта відстані дозволяє скоротити набір точок на спектрограмі, який потрібно обробити, що забезпечує оптимізацію процесу пошуку збігів.

Схематично розподіл амплітудних екстремумів на спектрограмі можна зобразити на прикладі рис. 1. Тут сірими точками позначено амплітудні екстремуми на спектрограмі, а чорним – умовний розподіл таких точок, що розташовані близько одна до одної, на відповідні групи. Далі в процесі аналізу подібними екстремумами можна знехтувати, натомість використовуючи позначення груп обраних точок.

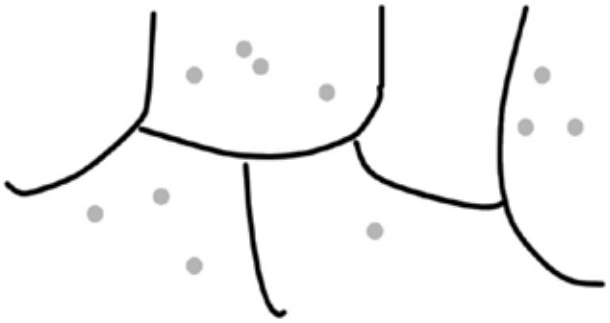

Рис. 1. Схематичний розподіл амплітудних екстремумів спектрограми на групи

Таким чином, із використанням підходу групування екстремумів процес створення бази відбитків зводимо до виконання алгоритму послідовностей етапів аналізу аудіоконтенту (рис. 2).

Рис. 2. Процес створення відбитків вхідних аудіокомпозицій

Крім того, є можливість виконання додаткової класифікації за такими параметрами, як жанр, автор тощо. Для виконання класифікаційної ідентифікації зручно використовувати згорткові нейронні мережі в процесі аналізу особливостей музичних композицій.

Додатковий аналіз та групування музичних композицій за певними характеристиками дозволять досягти розподіленого зберігання і використання фрагментації під час роботи з базою даних. Такий підхід дозволяє пришвидшити процес пошуку даних шляхом одночасного пара-

лельного пошуку в декількох фрагментах даних одночасно.

Під час пошуку збігів алгоритм створення відбитків для вхідного аудіопотоку є схожим. Додатково може бути виконана оптимізація на стадії виконання запиту та пошуку в базі даних.

Серед можливих оптимізацій виділимо розбиття бази даних на логічні такі структурні компоненти за обраними характеристиками вхідних даних, як жанр, автор тощо. Цей підхід називається шардуванням даних. Він дозволяє виконувати паралельний запит до бази даних, що підвищує загальну швидкодію системи ідентифікації аудіоконтенту.

Іншим підходом для оптимізації розпізнавання звукового контенту може бути підвищення рівня паралелізму під час пошуку збігів за рахунок використання ресурсів графічного процесора. У такому разі можна розподілити процес пошуку композицій на тисячі потоків, на відміну від використання центрального процесора. Використання графічного процесора дозволяє досягти оптимізації пошукового процесу за рахунок виконання простих паралельних операцій, що підвищує швидкість й ефективність процесу пошуку збігів музичних композицій.

Наведені підходи та алгоритми є базисом для програмної реалізації системи синтезу та розпізнавання музичних композицій. Поєднання функціоналу розглянутих методів дозволяє сформувати комбінований метод синтезу та аналізу музичних звуків, спрямований на оптимізацію пошукових процесів ідентифікації аудіоконтенту. Запропонований метод передбачає поєднання технологій розпізнавання вхідних аудіоданих, що в подальшому може бути використане для створення власних музичних композицій зі зручним та полегшеним користувацьким досвідом.

Висновки. Розглянуті підходи до реалізації складників системи ідентифікації музичних композицій формують комбінований метод синтезу та аналізу музичних звуків. Подальшого розвитку та розгляду набув компонент аудіорозпізнавання, що базується на створенні відбитків музичних композицій. Результати цього дослідження сприяють оптимізації програмного компонента аудіоаналізу.

Важливим є розгляд та розвиток компонента, що відповідає за класифікацію музичних композицій за жанром, автором чи іншими ідентифікаційними характеристиками. Такий аналіз є можливим за допомогою використання згорткових нейронних мереж. Розроблення класифікаційного компонента в середовищі загальної системи аналізу та синтезу звукових даних сприяє оптимізації пошуково-ідентифікаційних процесів.

Розглянуті компоненти, алгоритми та підходи до їх програмної реалізації є основними структурними блоками системи синтезу та розпізнавання музичних композицій.

Список літератури:

1. Avery Wang. An Industrial Strength Audio Search Algorithm. 2003. URL: https://www.ee.columbia.edu/~dpwe/papers/Wang03-shazam.pdf.

2. Özer, Hamza & Sankur, Bulent & Memon, Nasir & Anarim, Emin. (2005). Perceptual Audio Hashing Functions. *EURASIP Journal on Advances in Signal Processing*. 2005. 10.1155/ASP.2005.1780.

3. Viktoriia V. Voitko, Svitlana V. Bevz, Sergii M. Burbelo, Pavlo V. Stavytskyi, Bogdan Pinaiev, Zbigniew Omiotek, Doszhon Baitussupov, and Aigul Bazarbayeva "Automated system of audio components analysis and synthesis", Proc. SPIE 11045, Optical Fibers and Their Applications 2018, 110450V (15 March 2019); https://doi.org/10.1117/12.2522313

4. Войтко В.В., Бевз С.В., Бурбело С.М., Ставицький П.В. Моделі системи аналізу та розпізнавання музичних композицій. *Інформаційні технології та комп'ютерна інженерія. Міжнародний науково-технічний журнал.* Вінниця : ВНТУ, 2020, № 1. С. 32–38.

5. Zain Nasrullah and Yue Zhao. Music Artist Classification with Convolutional Recurrent Neural Networks. 2019. URL: https://arxiv.org/ pdf/1901.04555.pdf.

6. Sergiu Ciumac. How does Audio Fingerprinting work. An intuitive explanation. 2020. URL: https://emysound.com/blog/open-source/2020/06 /12/how-audio-fingerprinting-works.html#

Voitko V.V., Bevz S.V., Burbelo S.M., Stavytskyi P.V. ANALYSIS OF TOOLS FOR SOFTWARE IMPLEMENTATION OF SYNTHESIS SYSTEM AND ANALYSIS OF MUSICAL SOUNDS

There are considered ways and approaches to the implementation of the music synthesis and analysis synthesis. This particular paper focuses on one of the components of the developed system which is responsible for music data recognition. In order to achieve this goal the given component performs the number of consecutive steps that allow it to analyse the input audio data and generate the database for further music match searches. First task that this component performs is transforming the initial audio data that is represented as a byte array to the format that is suitable for analysis. This process is called sampling and allows to represent the

input audio stream as a relation between sound frequency and time. An additional value that is also included to the analysis is a sound wave amplitude. All these values can be graphically represented as a spectrogram that shows the sound frequency as OY axis in time which is represented as OX. The third value is the most important in the analysis process as it allows to determine those fragments of spectrogram that are noise-resistant and help to optimize the matching step. The sound amplitude is represented with color on the spectrogram and the most saturated points are local maximums that are being used while creating audio footprints. This article focuses mostly on the footprint creation process as it requires to implement the number of optimization in order to increase the recognition speed. There are described compression algorithms and specifically compression algorithms with losses that are more suitable for the music analysis. Moreover, it is important to consider hashing techniques that might help to improve algorithms performance. One of the hashing optimization approaches is a locality sensitive hashing algorithm which allows to partition the initial set of local amplitude maximums into smaller sets by reducing similar fingerprints to a single one that represents them. Moreover, it is important to consider a Jaccard Coefficient that helps to determine size of such sets by specifying constraints to distances between data points on spectrogram. This article also mentions the following steps and areas that must be considered in order to create an implementation of a music synthesis and recognition system. As an example, it is possible to implement a detailed analysis of input audio data and determine music genre, author etc. This will help to improve data persistence and match finding. In addition, the place of the considered components in the general developed system of synthesis and analysis of musical sounds is considered.

Key words: *audio analysis, music recognition, audio footprint, spectrogram, compression, hashing.*