УДК 004.42 **DOI: 10.33042/2522-1809-2023-6-180-2-6**

Н.А. Григоренко, Н.М. Ларіонов, В.М. Бредіхін

Харківський національний університет міського господарства імені О.М. Бекетова, Україна

ДОСЛІДЖЕННЯ ПРОЦЕСУ ТРАНСЛЯЦІЇ ВІЗУАЛЬНОГО МИСТЕЦТВА В МУЗИКУ ТА СТВОРЕННЯ КОЛЕКЦІЙ ДЛЯ ЛЮДЕЙ З ВАДАМИ ЗОРУ

У статті досліджується створення музики шляхом автоматизованої генерації звукової композиції за зображенням. Розроблений метод автоматичної генерації звуків за зображенням ґрунтується на спільному використанні нейронних мереж та світломузичної теорії. Результатом роботи стала модель нейронної мережі з двошаровою пакетною LSTM мережею з 512 прихованими нейронами в кожному осередку LSTM.

Ключові слова: *рекурентна нейронна мережа, світломузична теорія, спектрограма, генерація композицій.*

Постановка проблеми

Компанії навчають нейромережі створювати мелодії, пропускаючи через них тисячі пісень у різних жанрах. Штучний інтелект навчають розпізнавати настрій, темп, жанр мелодій, а потім за налаштуваннями створювати нові треки [1].

Найчастіше створення музики за допомогою нейромережі відбувається на основі пісень, які завантажені до її бібліотеки. Деякі нейромережі пишуть ноти, інші створюють композицію з вокалом та купою музичних інструментів. Також існують нейронні мережі, які створюють мелодії за фотографіями. В цьому випадку зображення кодується в музику, а потім реконструюється на основі згенерованого аудіозапису.

Трансляція візуального мистецтва в музику за допомогою моделей машинного навчання може бути використана для створення великих музейних колекцій доступних для людей з вадами зору завдяки переведенню творів мистецтва з недоступної чуттєвої модальності (зір) у доступну (слух).

Аналіз останніх досліджень і публікацій

Першу візуалізацію музики – Atari Video Music – розробив Роберт Браун у 1976 році. Він хотів створити візуалізацію до стереосистеми Hi-Fi [2].

Незважаючи на те, що це було давно, музична візуалізація досі привертає увагу багатьох вчених зі всього світу. З'явились нові технології, такі як генерація музики штучним інтелектом та генерація музичних композицій на основі зображення з використанням алгоритмів глибокого навчання за допомогою зіставлення візуальних, текстових і аудіофункцій.

Дослідження аудіовізуальних моделей показали, що попередні дослідження були зосереджені на покращенні продуктивності моделі за допомогою мультимодальної інформації, а також на покращенні доступності візуальної інформації через аудіоподання.

Мета статті

Мета статті полягає у аналізі методу автоматичної генерації звуків за зображенням, що ґрунтується на спільному використанні нейронних мереж та світломузичної теорії.

Виклад основного матеріалу дослідження

Для зниження ролі користувача-композитора в генерації звуків частина характеристик музичного твору виходить шляхом аналізу кольорової гами зображення. Таким чином, характер отриманої музичної композиції відповідатиме вхідному зображенню. Ця особливість робить можливим застосування цього підходу для створення музейних колекцій доступних для людей з вадами зору.

Ключовими характеристиками музичного твору є його тональність та темп. Саме ці параметри визначаються шляхом аналізу кольорової гами зображення, як показано в табл. 1.

Таблиця 1

Співвідношення кольорових та музичних характеристик [3]

Колірні характеристики	Музичні характеристики
Відтінок (червоний, синій, жовтий…)	Нота (до, до-дієз, ре, ре-дієз, мі, фа, фа-дієз, сіль, сіль-дієз, ля, ля-дієз, сі)
Колірна група (теплий / холодний)	Музичний лад (мажор / мінор)
Яскравість	Октава ноти
Насиченість	Довжина ноти

З табл. 1 робимо висновок, що тональність твору визначається двома кольоровими характеристиками – відтінок та група кольору, а темп – яскравістю та насиченістю. Алгоритм визначення тональності спирається на аналіз зображення та табл. 1 і складається з чотирьох кроків.

Крок 1. Перетворимо вхідне зображення з простору RGB в HSV. Даний крок дозволяє перетворити зображення до зручнішого вигляду, оскільки HSV-простір вже містить необхідні характеристики: назва кольору (визначається за параметром hue), насиченість (параметр saturation) та яскравість (параметр brightness).

Крок 2. Аналізуючи загалом зображення, визначаємо переважний колір.

Крок 3. Визначаємо назву та групу кольорів переважного кольору.

Крок 4. Згідно з табл. 1 та схемою Ньютона визначаємо тональність твору (ноту та музичний лад).

Для визначення темпу твору необхідно отримати яскравість і насиченість (за параметрами saturation і brightness) переважного кольору і розрахувати темп, згідно з даними параметрами (рис. 1).

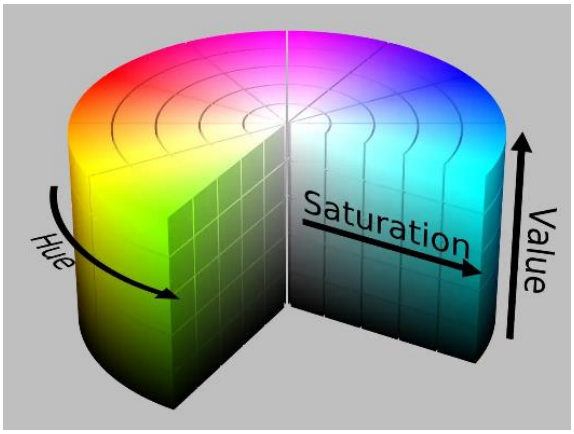

Рис. 1. Циліндр HSV

Результатом цих кроків є графічна анотація перетворення графічного зображення в музикальний ряд з використанням всіх кольорових характеристик, яка передається на вхід нейронної мережі.

Нейронна мережа з математичної точки зору поводиться як звичайна функція, хоч і дуже складно влаштована. Вона має заздалегідь позначену кількість аргументів і позначений формат, в якому вона видає відповідь.

У роботі було використано LSTM мережу, тому що вона має пам'ять про стан осередку і може переносити інформацію про більш довгострокові структури в музиці порівняно з рекурентною нейронною мережею (RNN) та мережею з архітектурою керованого нейрона (GRU), що дозволяло нам передбачати довші послідовності до 1 хвилини, які очікувано будуть звучати узгоджено.

Для реалізації програми автоматизованої генерації музичних композицій за зображенням є сенс використовувати рекурентні нейронні (RNN) мережі з довготривалою пам'яттю (LSTM) [4], які намагаються вирішити проблему звичайних RNN мереж – втрату інформації з часом, використовуючи фільтри та явно задану клітину пам'яті. Метою цих фільтрів є захист інформації. Вхідний фільтр визначає, скільки інформації з попереднього шару зберігатиметься в нейроні. Вихідний фільтр визначає, скільки інформації отримають такі шари [5].

Основна ідея вирішення задачі генерації музики за зображенням представляє собою складання спектрограм за вхідним рядком зображення і конвертації її в аудіокліп.

Аудіо спектрограма – це візуальний спосіб представлення частотного змісту звукового кліпу [6]. Вісь X є час, а вісь Y є частота (рис. 2). Колір кожного пікселя визначає амплітуду звуку залежно від частоти та часу.

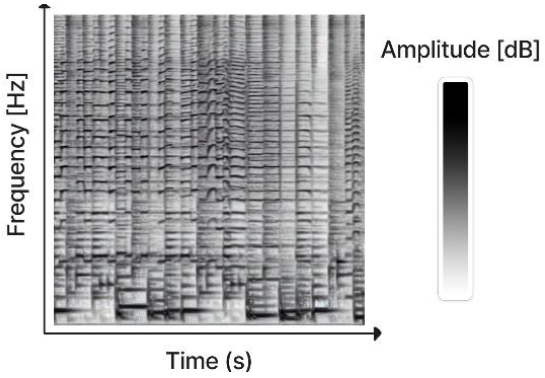

Рис. 2. Спектрограма звукового кліпу

Спектрограма може бути отримана зі звуку з використанням перетворення Фур'є (STFT), яке апроксимує звук як комбінацію синусоїдальних хвиль різної амплітуди та фази (рис. 3).

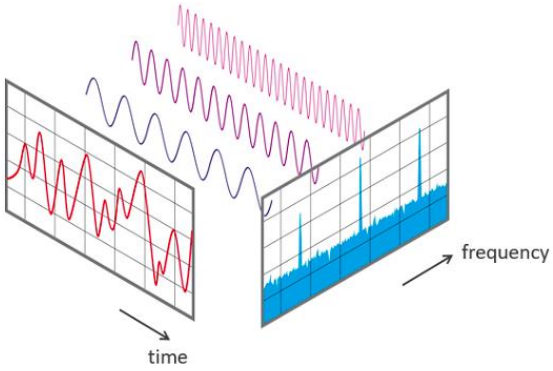

Рис. 3. Спектрограма як комбінація синусоїдальних хвиль різної амплітуди та фази [7]

У процесі дослідження методів синтезу звуків було розглянуто та проаналізовано найбільш популярні методи: адитивний синтез, FM-синтез, фазова модуляція, семплінг, таблично-хвильовий синтез, лінійно-арифметичний синтез, субтрактивний синтез та векторний синтез [8].

FM-синтез добре застосовний для синтезу звуку ударних інструментів, синтез інших музичних інструментів звучить занадто штучно. Головний недолік FM-синтезу – нездатність за його допомогою повноцінно імітувати акустичні інструменти.

Фазова модуляція дає досить гарний звук, але дуже обмежена за своїми можливостями, тому рідко використовується на практиці.

Семплінг застосовується в більшості сучасних синтезаторів, тому що дає найбільш реалістичний звук та досить простий у реалізації [9].

Таблично-хвильовий синтез та лінійно-арифметичний синтез схожі на метод семплінгу, але вони складні в реалізації, тому на практиці перевага віддається семплінгу як найпростішому методу.

Субтрактивний синтез зазвичай використовується спільно з адитивним, має хорошу якість синтезу звуків, проте складний у реалізації.

Векторний синтез використовується для отримання більш багатих і складних тембрів, однак у рамках розгляду системи це не суттєво.

Тому для реалізації системи було обрано саме метод семплінгу. Цей метод дає найбільш реалістичне звучання інструментів, що є важливою її характеристикою.

Для реалізації запропонованих алгоритмів генерації звуків за кольоровою гамою зображення було розроблено базову модель нейронної мережі з двошаровою пакетною LSTM мережею з 512 прихованими одиницями в кожному осередку LSTM. Ми використовували шар впровадження, щоб перетворити кожен токен (усереднене значення кольору) у вектор (рис. 4).

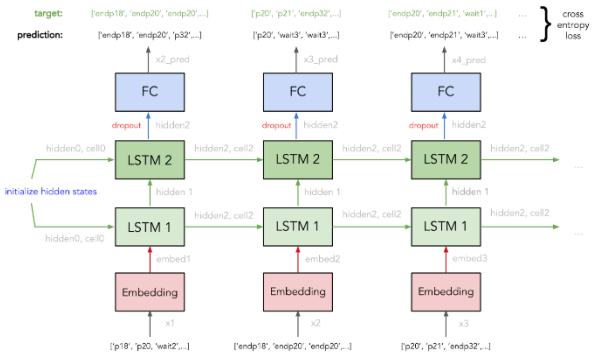

Рис. 4. Архітектура LSTM мережі

Програму генерації музичних композицій з використанням нейронних мереж було навчено на 29 композиціях сучасної музики.

Гіперпараметри, що були налаштовані у LSTM мережі, такі як, наприклад, визначення довжини послідовності для вибірки, не повинні бути занадто короткими, бо їх довжини буде недостатньо для того, щоб створити музичний ланцюжок, який звучить узгоджено, але й занадто довга послідовність займе занадто багато часу для навчання мережі без вивчення додаткової інформації.

Після навчання було складено набір із десяти тестових зображень, що мають різний тип (абстрактні зображення, пейзажі, міста та люди), за якими були отримані та збережені вихідні музичні композиції. Всі музичні композиції були оцінені за такими критеріями:

– відповідність характеру зображення (за п'ятибальною шкалою);
– реалістичність звучання інструменту (фортепіано чи гітара);
– мелодійність композиції;
– якість гармонії (акомпанементу);
– приємність мелодії для сприйняття;
– цілісність композиції;
– реалістичність/штучність композиції.

За результатами оцінки по всім тестовим зображенням було розраховано середні значення, які представлені у табл. 2.

Таблиця 2

Оцінка композицій за критеріями

Критерій	Середнє значення для всіх тестів
Відповідність характеру зображення	4.9
Реалістичність звучання інструмента	3.9
Мелодійність композиції	4.4
Якість гармонії	4.9
Приємність для сприйняття	4.6
Цілісність композиції	4.5
Реалістичність композиції	4.3

Ми використовували метрику BLEU, яка оцінює кількість n-грам (повтори з вихідних даних), що використовуються в згенерованих послідовностях для порівняння нашої базової моделі LSTM із деякими реальними композиціями. Результати: усереднені за тестовими зображеннями послідовностей – 0,25 для LSTM та 0,14 для реальних даних. Це означає, що наші згенеровані зразки хороші за нашими налаштованими метриками. Вища оцінка, яку композиції LSTM дають порівняно з реальними композиціями, підкреслює наявність перенасичення, де згенерована музика містить суттєві повтори з вихідних даних.

Загалом можна зробити висновок, що композиція, згенерована за абстрактним зображенням, приємніша на слух, ніж генерація за пейзажами. Загальне враження від згенерованих композицій позитивне. Серед мінусів слід звернути увагу на однотипність гармонії, іноді рваність та недостатню реалістичність твору.

Висновки

У результаті виконання цього дослідження було запропоновано комбінований підхід до генерації звукових послідовностей. Він використовує рекурентну нейронну мережу для генерації музичного матеріалу та кольорову музичну теорію для визначення параметрів композиції із зображення. У процесі вибору нейронної мережі для генерації музичних композицій було виявлено, що для реалізації програми автоматизованої генерації музичних композицій необхідно використовувати саме рекурентні нейронні мережі з довготривалою пам'яттю – RNN LSTM.

Результати роботи можуть бути використані для створення великих музейних колекцій доступних для людей з вадами зору завдяки переведенню творів мистецтва з недоступної чуттєвої модальності (зір) у доступну (слух).

Література

1. Як штучний інтелект створює музику і змінює креативну індустрію [Електрон. ресурс] / Depositphotos : сайт. – США, 2009–2023. – Оновлюється постійно. – Режим доступу: https://blog.depositphotos.com/ua/yak-shtuchnyj-intelekt-stvoryuye-muzyku.html, вільний (дата звернення: 05.10.2023).

2. GANSynth: Adversarial Neural Audio Synthesis / J. Engel, K. K. Agrawal, S. Chen, I. Gulrajani, C. Donahue, A. Roberts // Proceedings of the 7th International Conference on Learning Representations (ICLR), New Orleans, LA (USA), May 6–9, 2019 yr. – 17 p. – DOI: 10.48550/arXiv.1902.08710.

3. Caivano J. L. Color and Sound: Physical and Psychophysical Relations / J. L. Caivano // Color Research and Application. – 1994. – Vol. 19 (2). – P. 126–133. – DOI: 10.1111/j.1520-6378.1994.tb00072.x.

4. Комарський О. С. Модель рекурентної нейронної мережі для генерації музики / О. С. Комарський, А. Ю. Дорошенко // Проблеми програмування. – 2022. – № 1. – С. 87–93. – DOI: 10.15407/pp.2022.01.87.

5. A Hierarchical Latent Vector Model for Learning Long-Term Structure in Music / A. Roberts, J. Engel, C. Raffel, C. Hawthorne, D. Eck // Proceedings of the 35th International Conference on Machine Learning (ICML), Stockholm (Sweden), July 10–15, 2018 yr. – Proceedings of Machine Learning Research (PMLR), 2018. – Vol. 80. – P. 4364–4373. – Regime of access: http://proceedings.mlr.press/v80/roberts18a/roberts18a.pdf, free (date of the application: 05.10.2023).

6. Яровий М. В. Частотний аналіз в задачах розпізнавання звуку з використанням нейронних мереж / М. В. Яровий, О. С. Назаров // Сучасні аспекти та перспективи розвитку науки : матеріали I міжнар. студ. наук. конф., Кропивницький, Україна, 16 квітня 2021 р. – Кропивницький : Молодіжна наукова ліга, 2021. – Т. 2. – С. 48–50. – Режим доступу: https://ojs.ukrlogos.in.ua/index.php/liga/issue/view/16.04.2021/502, вільний (дата звернення: 05.10.2023).

7. Бондаренко А. І. Виявлення і аналіз акустичних подій в електронній музиці (на прикладі "Мотус" А. Загайкевич) / А. І. Бондаренко // Питання культурології. – 2015. – Вип. 31. – С. 22–28. – Режим доступу: http://nbuv.gov.ua/UJRN/Pkl_2015_31_5, вільний (дата звернення: 05.10.2023).

8. Кущ Є. В. Про деякі аспекти функціонування електромузичного інструментарію у музичній культурі другої половини XX століття / Є. В. Кущ // Наукові записки Тернопільського національного педагогічного університету імені Володимира Гнатюка. Серія: Мистецтвознавство. – Тернопіль: Вид-во ТНПУ ім. В. Гнатюка, 2013. – № 1. – С. 17–23. – Режим доступу: http://dspace.tnpu.edu.ua/bitstream/123456789/3824/1/KUSHCH.pdf, вільний (дата звернення: 05.10.2023).

9. How to Sample Music: Step-by-Step Music Sampling Guide [Electronic resource] / MasterClass : website. – San Francisco, CA (USA), 2014–2023. – Updated continuously. – Regime of access: https://www.masterclass.com/articles/how-to-sample-music, free (date of the application: 05.10.2023).

References

1. Chervinska, N. (2022, August 12). *Generating Music with AI: How it Works*. Depositphotos. Retrieved from https://blog.depositphotos.com/ua/yak-shtuchnyj-intelekt-stvoryuye-muzyku.html

2. Engel, J., Agrawal, K. K., Chen, S., Gulrajani, I., Donahue, C., & Roberts, A. (2019). GANSynth: Adversarial Neural Audio Synthesis. *Proceedings of the 7th International Conference on Learning Representations (ICLR)* (17 p.). DOI: 10.48550/arXiv.1902.08710

3. Caivano, J. L. (1994). Color and Sound: Physical and Psychophysical Relations. *Color Research and Application*, 19(2), 126–132. DOI: 10.1111/j.1520-6378.1994.tb00072.x

4. Komarskyi, O. S., & Doroshenko, A. Yu. (2022). Recurrent neural network model for music generation. *Problems in programming*, 1, 87–93. DOI: 10.15407/pp.2022.01.87 [in Ukrainian]

5. Roberts, A., Engel, J., Raffel, C., Hawthorne, C., & Eck, D. (2018). A Hierarchical Latent Vector Model for Learning Long-Term Structure in Music. *Proceedings of the 35th International Conference on Machine Learning (ICML)* (pp. 4364–4373). Proceedings of Machine Learning Research (PMLR). Retrieved from http://proceedings.mlr.press/v80/roberts18a/roberts18a.pdf

6. Yarovyi, M. V., & Nazarov, O. S. (2021). Frequency analysis in sound recognition tasks using neural networks. *Proceedings of the 1st International Student Scientific Conference 'Modern aspects and prospects for the development of science'*: Vol. 2 (pp. 48–50). Youth Science League. Retrieved from https://ojs.ukrlogos.in.ua/index.php/liga/issue/view/16.04.2021/502 [in Ukrainian]

7. Bondarenko, A. I. (2015). Detection and analysis of acoustic events in electronic music (on the example of "Motus" by A. Zahaikevych). *Issues in Cultural Studies*, 31, 22–28. Retrieved from http://nbuv.gov.ua/UJRN/Pkl_2015_31_5 [in Ukrainian]

8. Kushch, E. V. (2013). About some aspects of functioning of electronic musical instruments in musical culture of the second half of the XX-th century. *The Scientific Issues of Ternopil Volodymyr Hnatiuk National Pedagogical University. Series: Art Studies*, 1, 17–23. Retrieved from http://dspace.tnpu.edu.ua/bitstream/123456789/3824/1/KUSHCH.pdf [in Ukrainian]

9. MasterClass. (2021, June 7). *How to Sample Music: Step-by-Step Music Sampling Guide*. Retrieved from https://www.masterclass.com/articles/how-to-sample-music

Рецензент: д-р фіз.-мат. наук, проф. Н.Д. Сізова, Харківський національний університет міського господарства імені О.М. Бекетова, Україна.

Автор: ГРИГОРЕНКО Наталія Анатоліївна
здобувач вищої освіти 2-го курсу магістратури навчально-наукового інституту енергетичної, інформаційної та транспортної інфраструктури
Харківський національний університет міського господарства імені О.М. Бекетова
E-mail – nataliya.grygorenko@kname.edu.ua

Автор: ЛАРІОНОВ Назарій Миколайович
здобувач вищої освіти 2-го курсу магістратури навчально-наукового інституту енергетичної, інформаційної та транспортної інфраструктури
Харківський національний університет міського господарства імені О.М. Бекетова
E-mail – nazarii.larionov@kname.edu.ua

Автор: БРЕДІХІН Володимир Михайлович
кандидат технічних наук, доцент, доцент кафедри комп'ютерних наук та інформаційних технологій
Харківський національний університет міського господарства імені О.М. Бекетова
E-mail – bredixinv@gmail.com
ID ORCID: https://orcid.org/0000-0002-6063-5046

RESEARCH OF THE PROCESS OF VISUAL ART TRANSMISSION IN MUSIC AND THE CREATION OF COLLECTIONS FOR PEOPLE WITH VISUAL IMPAIRMENTS

N. Hryhorenko, N. Larionov, V. Bredikhin
O.M. Beketov National University of Urban Economy in Kharkiv, Ukraine

This article explores the creation of music through the automated generation of sounds from images. The developed automatic image sound generation method is based on the joint use of neural networks and light-music theory. Translating visual art into music using machine learning models can be used to make extensive museum collections accessible to the visually impaired by translating artworks from an inaccessible sensory modality (sight) to an accessible one (hearing). Studies of other audio-visual models have shown that previous research has focused on improving model performance with multimodal information, as well as improving the accessibility of visual information through audio presentation, so the work process consists of two parts. The result of the work of the first part of the algorithm for determining the tonality of a piece is a graphic annotation of the transformation of the graphic image into a musical series using all colour characteristics, which is transmitted to the input of the neural network. While researching sound synthesis methods, we considered and analysed the most popular ones: additive synthesis, FM synthesis, phase modulation, sampling, table-wave synthesis, linear-arithmetic synthesis, subtractive synthesis, and vector synthesis. Sampling was chosen to implement the system. This method gives the most realistic sound of instruments, which is an important characteristic. The second task of generating music from an image is performed by a recurrent neural network with a two-layer batch LSTM network with 512 hidden units in each LSTM cell, which assembles spectrograms from the input line of the image and converts it into an audio clip. Twenty-nine compositions of modern music were used to train the network. To test the network, we compiled a set of ten test images of different types (abstract images, landscapes, cities, and people) on which the original musical compositions were obtained and stored. In conclusion, it should be noted that the composition generated from abstract images is more pleasant to the ear than the generation from landscapes. In general, the overall impression of the generated compositions is positive.

Keywords: *recurrent neural network, light music theory, spectrogram, generation of compositions.*