DOI 10.31891/2307-5732-2022-305-1-64-67
УДК 004.624

ВОЙТКО В. В.
https://orcid.org/0000-0002-3329-7256
e-mail: defakfki@i.ua
БЕВЗ С. В.
https://orcid.org/0000-0001-5561-9036
e-mail: bevz@vntu.edu.ua
БУРБЕЛО С. М.
https://orcid.org/0000-0002-8554-2292
e-mail: burbelo@vntu.edu.ua
СТАВИЦЬКИЙ П. В.
https://orcid.org/0000-0002-9139-6076
e-mail: morfly3000@gmail.com
Вінницький національний технічний університет

ТЕХНОЛОГІЯ АУДІОГЕНЕРАЦІЇ СИСТЕМИ СИНТЕЗУ ТА АНАЛІЗУ МУЗИЧНИХ КОМПОЗИЦІЙ

Розглянуто технологію та способи аудіогенерації в системі синтезу та аналізу музичних композицій. Серед них є функціонал награвання музичних композицій за допомогою голосу. Крім того, вхідні дані можуть бути використані як шаблон, на основі якого буде згенеровано подібну музичну композицію. Редагування відбувається за допомогою використання двовимірної площини, що визначає залежність висоти ноти від часу і має вигляд секвенсера. Для реалізації аудіосинтезу на основі шаблону використовується архітектура глибинного навчання з варіаційним аудіокодувальником, що дозволяє тренувати нейронну мережу й відтворювати дані, схожі до вхідних. Під час використання такого підходу важливим кроком є перетворення вхідних даних, які подані у вигляді необробленого аудіоформату, в спектрограму, що є зображенням, на основі якого відбувається процес тренування нейронної мережі та аудіогенерація. В результаті операцій створення музичних композицій користувач додатково може використовувати режим ручного редагування для отримання бажаної вихідної музичної композиції.

Ключові слова: аудіосинтез, спектрограма, варіаційний аудіокодувальник.

VIKTORIIA VOITKO, SVITLANA BEVZ, SERGII BURBELO, PAVLO STAVYTSKYI
Vinnytsia National Technical University

AUDIO GENERATION TECHNOLOGY OF A SYSTEM OF SYNTHESIS AND ANALYSIS OF MUSIC COMPOSITIONS

System of audio synthesis and analysis of music compositions is considered. It consists of two primary parts, the audio analysis component, and the music synthesis component. An audio generation component implements various ways of creating audio sequences. One of them is aimed to record melodies played with voice and transform them into sequences played with selected musical instruments. In addition, an audio input created with a human voice can be utilized as a seed, that is used to generate similar music sequences using artificial intelligence. Finally, a manual approach for music generation and editing is available. After automatic mechanisms for composition generation are used, the results of their work are presented on a two-dimensional plane which represents the dependence of music note pitches on time. It is possible to manually adjust the result of audio generation or create new music sequences with this approach. A creation process could be used iteratively to create multiple parallel music sequences that are to be played as a single audio composition. To implement a seed-based audio synthesis, a deep learning architecture based on a variational autoencoder is used to train a neural network that can reproduce input-like data. When using such an approach an additional important step must be considered. All the input data must be converted from a raw audio format to spectrograms which are represented as grayscale images. Moreover, the result of a sound generation is also represented in a spectrogram and therefore, must be converted back to an output audio format that can be played using speakers. This is required as using spectrograms helps to discard redundant data that raw audio format contains and thus significantly reduces resources consumption and increases overall synthesis speed.

Keywords: audio synthesis, spectrogram, variational audio encoder.

Постановка проблеми у загальному вигляді та її зв'язок із важливими науковими чи практичними завданнями

В умовах стрімкого розвитку сучасних інформаційних технологій обробка аудіоконтенту потребує нових методів та засобів аналізу аудіопотоків, розпізнавання і синтезу звукових композицій. Серед функціоналу системи синтезу та аналізу звуків є компонент генерації музичних композицій, головне призначення якого полягає в забезпеченні простого і швидкого інструментарію для створення користувачем власних мелодій. Важливою є реалізація варіативного підходу в технології аудіогенерації, що дозволяє зробити процес створення музичних композицій гнучким і, в той же час, забезпечує детальність процесу їх редагування.

Аналіз досліджень та публікацій

У роботі [1] наведено приклад системи WaveNet, що виконує генерацію аудіо різного призначення. Її особливістю є те, що в якості вхідних даних тут використовується необроблений аудіоформат. У результаті роботи цієї системи генеруються вихідні дані в аудіоформаті. Система WaveNet, у тому числі, використовується для реалізації конвертації тексту в аудіомовлення людини. Іншим прикладом системи, що

реалізує аудіосинтез, є систеа Jukebox [2], яка призначена для генерації музичних послідовностей. Аналогічно з попереднім прикладом, в якості вхідних даних така системи використовує необроблений аудіоформат, а результатом є згенерована в такому ж форматі аудіокомпозиція. У роботі [3] описано проєкт DrumGAN , який бере за основу підхід обробки вхідних даних у вигляді спектрограм, а на виході генерує звуки барабанних інструментів. Наведені дослідження розглядають відомий функціонал аудіогенерації, проте не дозволяють поєднувати і ефективно комбінувати низку підходів для генерації музичних композицій, зокрема награвання мелодій за допомогою голосу, генерацію композицій на основі шаблону, награнного голосом, та ручне точне редагування паралельних музичних послідовностей для створення інтуїтивного та ефективного рішення.

Формулювання цілей статті

Метою роботи є удосконалення технології аудіогенерації системи синтезу та аналізу музичних композицій за рахунок використання комплексного підходу до створення й обробки аудіоконтенту, що дозволяє розширити функціонал системи.

Виклад основного матеріалу

Система синтезу та аналізу музичних звуків реалізує функціонал створення музичних композицій і дозволяє користувачеві використовувати три способи генерації аудіоконтенту (рис. 1).

Рис. 1. Способи генерації музичних композицій

Перший спосіб полягає у ручному створенні музичних композицій. Редагування виконується за допомогою двовимірної площини, де вісь OX визначає часову шкалу, а вісь OY – висоту звуку або ноту [4]. Такий підхід є універсальним способом створення музичних композицій та задає загальний фреймворк, на якому базуються всі інші варіанти аудіогенерації системи.

Другий спосіб звукової генерації полягає у можливості награвання голосом мотиву або мелодії бажаної музичної композиції, що системою буде перетворено на музичну послідовність. Згенерована послідовність подається у вигляді, придатному для ручного корегування. Таким чином, користувач має змогу генерувати композицію, виконуючи декілька ітерацій награвання мелодій та, за потреби, завершаючи процес синтезу композиції ручним її редагуванням. У результаті створення декількох паралельних аудіодоріжок існує можливість вибору музичного інструмента, за допомогою якого кожна з них буде відіграна [5].

Третій спосіб аудіогенерації полягає у використанні награної або введеної музичної послідовності як шаблону для подальшої генерації схожої мелодії. Аналогічно до попереднього способу, після генерації мелодії, у користувача є можливість ручного редагування композиції та вибору конкретного музичного інструмента, за допомогою якого складова мелодії буде відіграна.

Таким чином, у розпорядженні користувача є три інструменти аудіогенерації, що в технології генерації звуку доповнюють один одного шляхом їх комбінованого використання для синтезу бажаних мелодій.

Для реалізації компонента шаблонної аудіогенерації необхідно побудувати нейронну мережу, яку потрібно тренувати на реальних наборах даних звучання різних музичних інструментів та композицій, побудованих за допомогою них. На основі таких даних можна реалізувати можливість генерації аудіо з урахуванням особливостей звучання різних музичних інструментів, а також різних музичних стилів.

Під час реалізації компонента генерації аудіокомпозицій модуль обробки вхідних даних може приймати інформацію у двох різних форматах: необроблений звук і спектрограма.

Перевагою першого підходу є відсутність потреби попередньої обробки чи адаптації вхідного контенту та постобробки результату, оскільки робота проводиться безпосередньо зі звуковими даними у форматі «.wave». Проте, недоліком такої реалізації є те, що кожного разу нейронна мережа повинна обробити значний обсяг даних, що є ресурсо- та часозатратним. Крім того, такий формат даних містить велику кількість надлишкової інформації, що не є корисною під час обробки, тому доречно було б позбавитися зайвих «шумів» для підвищення ефективності роботи системи.

При використанні спектрограм у якості вхідних даних для обробки, необхідно додатково реалізувати компоненти генерації спектрограм з необробленого аудіозвуку. Крім того, при такому підході в результаті система видає спектрограму, яку знову необхідно конвертувати в звукові послідовності. Для

зручності та подальшої оптимізації спектрограми можуть бути використані у чорно-білому форматі, оскільки цього достатньо для якісної обробки даних (рис 2).

Рис. 2. Процес аудіогенерації

Під час награвання мелодії голосом створюється файл у необробленому аудіоформаті, який перетворюється на спектрограму та передається у модуль аудіогенерації. В результаті роботи цього модуля генерується спектрограма, яку, в свою чергу, необхідно конвертувати у звукові послідовності для можливості формування музичної композиції.

Реалізація такого підходу передбачає використання варіаційного автокодувальника [6] в якості архітектури нейронної мережі, що відповідає за аудіосинтез (рис. 3).

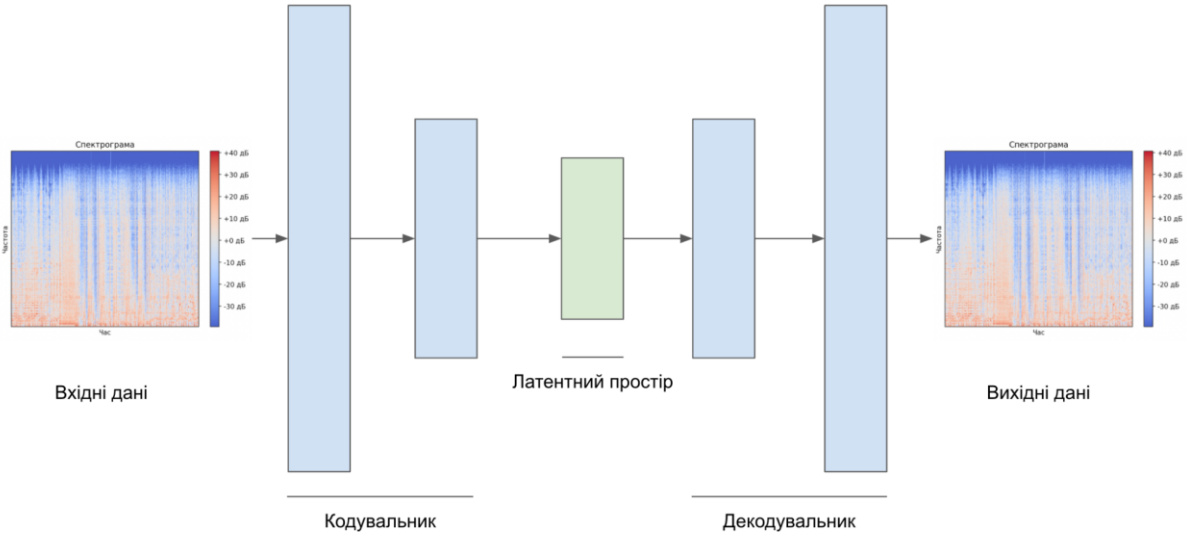

Рис. 3. Модуль аудіогенерації з використанням варіаційного автокодувальника

Цей підхід полягає у використанні нейронної мережі, яка складається з базових компонентів: кодувальника і декодувальника, що є симетричними один одному. Крім того, їх з'єднує вузьке місце, що називається латентним простором. У результаті роботи кодувальника відбувається компресія вхідних даних за рахунок зменшення кількості нейронів на прошарку нейронів. А декодувальник має змогу декодувати дані латентного простору та відтворювати вихідні дані. Перевагою використання варіаційного кодувальника є врахування ним розходження Кульбака-Лейблера, що покращує характеристики відтворення першочергових даних.

Для реалізації такого підходу необхідні першочергові дані для тренування мережі. Тут доцільно використати готові набори даних, серед яких можна виділити NSynth від Magenta, що є високоякісним набором анотованих музичних нот, та MAESTRO від Magenta, що фокусується на звуках фортепіано.

Результуюча музична композиція C може бути побудована за допомогою формули (1):

$$C = \sum_{i=0}^{n} \left(f_{templ}(f_{hum}(X, E), E) \right) + M \tag{1}$$

де n – кількість паралельних музичних послідовностей багатоголосої музичної композиції;

X – необроблена музична послідовність, створена шляхом награвання мелодії голосом користувача;

E – конфігурація згенерованої музичної послідовності, що, в тому числі, визначає музичний інструмент, яким звукова послідовність буде відіграна;

f_{hum} – музична послідовність, згенерована шляхом награвання мотиву мелодії голосом;

F_{templ} – музична послідовність, згенерована шляхом створення стилістично схожої музичної послідовності на основі шаблону як аргумента;

M – набір модифікацій робочої області секвенсера, зроблений користувачем вручну до або після процесу аудіогенерації.

Висновки з даного дослідження і перспективи подальших розвідок у даному напрямі

Удосконалена технологія аудіогенерації системи синтезу та аналізу музичних композицій орієнтована на використання комплексного підходу до створення й обробки аудіоконтенту, дозволяє розширити функціонал аудіосистеми. Запропоновані до використання способи аудіогенерації системи синтезу та аналізу музичних послідовностей дозволяють отримувати вхідний контент різним шляхами, що розширює можливості використання системи. Зокрема, спосіб награвання мелодій музичних композицій голосом у результаті роботи системи забезпечує їх відтворення за допомогою обраного музичного інструмента. Підхід шаблонної аудіогенерації дозволяє за допомогою початкового шаблону генерувати мелодію, подібну за стилем. У результаті ручного створення музичних композицій, користувач має можливість самостійно створювати і редагувати фрагменти композиції, використовуючи двовимірну площину, що характеризує залежність висоти музичної ноти від часу. Поєднання вказаних способів аудіогенерації дозволяє адаптувати процес синтезу та аналізу музичних композицій до умов створення вхідного контенту.

Література

1. Van Den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A.W. and Kavukcuoglu, K., 2016. WaveNet: A generative model for raw audio. SSW, 125, p.2.

2. Dhariwal, P., Jun, H., Payne, C., Kim, J.W., Radford, A. and Sutskever, I., 2020. Jukebox: A generative model for music. arXiv preprint arXiv:2005.00341.

3. Nistal, Javier, Stefan Lattner, and Gael Richard. "DrumGAN: Synthesis of drum sounds with timbral feature conditioning using Generative Adversarial Networks." arXiv preprint arXiv:2008.12073 (2020).

4. 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

5. Viktoriia V. Voitko, Svitlana V. Bevz, Sergii M. Burbelo, Pavlo V. Stavytskyi, Oleksandr M. Khoshaba, Natalia O. Rysynets, Olena Yu. Teplova, Andrzej Smolarz, Saule Smailova, Assel Mussabekova, and Bakhyt Yeraliyeva "Analysis of the development approaches of the system of audio synthesis and recognition with the option of using photonic processors", Proc. SPIE 12040, Photonics Applications in Astronomy, Communications, Industry, and High Energy Physics Experiments 2021, 120400N (3 November 2021); https://doi.org/10.1117/12.2611464

6. Foster, David. Generative deep learning: teaching machines to paint, write, compose, and play. Sebastopol, CA: O'Reilly Media, 2019. Print.

Рецензія/Peer review : 22.01.2022 р. Надрукована/Printed : 27.02.2022 р.