

УДК 004.896

DOI: [10.26102/2310-6018/2022.37.2.026](https://doi.org/10.26102/2310-6018/2022.37.2.026)

Генерация жанровых музыкальных композиций по эмоциональному состоянию человека

Н.А. Никитин[✉], Ю.А. Орлова, В.Л. Розалиев

*Волгоградский Государственный Технический Университет, Волгоград, Российская
Федерация*
set.enter@mail.ru

Резюме. Данная работа направлена на исследование и разработку математического и программного обеспечения для автоматизации и поддержки технического творчества путем автоматизированной генерации музыкальных произведений различных жанров, основанных на эмоциональном состоянии человека (пользователя). В основе работы лежит метод генерации музыкального материала с использованием искусственных нейронных сетей. Для генерации музыки была выбрана рекуррентная нейронная сеть с долгой краткосрочной памятью (LSTM), поскольку именно такой тип сетей позволяет учитывать иерархичность и созависимость музыкальных данных. Работа содержит подробное описание процесса сбора обучающих данных, процесса обучения сети, ее использование для генерации музыки, а также описание архитектуры сети. Кроме этого, дается описание обобщенного метода получения эмоционального состояния человека путем анализа изображения с использованием принципов теста Люшера. Для синтеза звуков по готовому музыкальному материалу в работе используется метод сэмплинга. Именно этот метод позволяет получить звучание музыкальных инструментов, приближенное к реалистичному, также данный метод сравнительно прост в реализации. Также статья включает описание процесса проектирования и разработки программного обеспечения для подтверждения описанных алгоритмов и методов, а именно веб-сайта для генерации музыкальной композиции путем анализа изображения.

Ключевые слова: программная генерация музыки, Spotify API, сэмплинг, рекуррентные нейронные сети, схемы соотнесения цвета и ноты.

Благодарности: Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 20-37-90004.

Для цитирования: Никитин Н.А., Орлова Ю.А., Розалиев В.Л. Генерация жанровых музыкальных композиций по эмоциональному состоянию человека. *Моделирование, оптимизация и информационные технологии*. 2022;10(2). Доступно по: <https://moitvvt.ru/ru/journal/pdf?id=1175> DOI: 10.26102/2310-6018/2022.37.2.026

Generation of genre musical compositions according to the emotional state of a person

N.A. Nikitin[✉], Y.A. Orlova, V.L. Rozaliev

Volgograd State Technical University, Volgograd, Russian Federation
set.enter@mail.ru

Abstract. The aim of this article is research and development of algorithms and software for automation and support of technical creativity process by automated generation of musical compositions of different genres, based on the emotional state of a person. It relies on the method of generating musical material with the aid of artificial neural networks. To generate music, a recurrent neural network with long-short term memory is chosen because this is the type of neural networks that helps to take into account the hierarchy and codependency of musical data. The paper contains a detailed description of training data

collection process, the process of neural network training, its use for generating musical compositions as well as an illustration of the network architecture. In addition, it outlines a generalized method for obtaining the emotional state of a person by analyzing an image by utilizing the principles of the Luscher test. For the synthesis of sounds with the help of the prefabricated musical material, the sampling method is applied. It is this method that makes it possible to emulate the realistic sound of musical instruments, which is also relatively easy to implement. Furthermore, the article includes a description of the software design and development process with a view to confirming the algorithms and methods under review, namely a website for generation musical composition by analyzing an image.

Keywords: automated musical generation, Spotify API, sampling, recurrent neural network, correlation schemes between color and pitches.

Acknowledgments: The reported study was funded by RFBR according to the research project No. 20-37-90004.

For citation: Nikitin N.A., Orlova Y.A., Rozaliev V.L. Generation of genre musical compositions according to the emotional state of a person. *Modeling, Optimization and Information Technology*. 2022;10(2). Available from: <https://moitvvt.ru/ru/journal/pdf?id=1175> DOI: 10.26102/2310-6018/2022.37.2.026 (In Russ.).

Введение

Сегодня процесс создания музыки с использованием программных вычислительных средств становится массовым явлением и постепенно становится важным компонентом современной музыкальной культуры. Персональные компьютеры предоставляют множество новых возможностей в развитии профессионального мышления музыканта во всех областях музыкального творчества, что приводит к широкому внедрению компьютерных технологий в музыку. Это позволяет существенно изменить или дополнить существующий процесс работы музыканта и музыковеда, композитора и даже исполнителя, а также влияет на процесс преподавания и обучения музыке [1].

Первоначально, термин «компьютерная музыка» употреблялся для описания более широкой области наук, направленной на цифровую обработку и запись различных аудиосигналов и звуковых структур. В настоящее время, данный термин всё чаще употребляют для описания любой музыки, так или иначе созданной с использованием компьютерных технологий.

Со времен, когда музыку стали записывать на бумажные носители в виде нотных знаков, стали появляться необычные способы ее создания. Одним из первых методов алгоритмической музыкальной композиции стал способ сочинения музыки, придуманный Вольфгангом Моцартом – музыкальная игра в кости [2]. Считается, что работа Л. Хиллера и Л. Айзексона под названием «Сюита Иллиака для струнного квартета» стала одним из первых музыкальных произведений, написанных с использованием алгоритмов. Хотя данная работа стала одним из первых законченных алгоритмических музыкальных произведений, существуют и более ранние работы, где были попытки использовать различные математические методы для новаторской на тот момент задачи – генерации музыки с использованием компьютеров. В этих исследованиях были попытки адаптировать математический аппарат различных теорий для этой задачи, например, теорию вероятностей и марковские цепи. Однако наиболее интересным выглядит использование порождающих грамматик, поскольку данный метод можно назвать первой попыткой алгоритмически формализовать процесс создания музыки – с помощью грамматики и правил вывода [3]. Несмотря на то, что существует музыкальная теория, достаточно точно формализующая музыку, сам процесс ее создания достаточно тяжело формализовать и автоматизировать.

На текущий момент существует большое количество различных математических методов алгоритмической композиции, однако лидирующие позиции занимают современные методы машинного обучения, а именно нейронные сети [4]. Именно искусственные нейронные сети позволяют выявить все взаимосвязи в музыкальном материале, которые порой не видны даже музыканту. Поэтому в данной работе для генерации музыкального материала была выбрана рекуррентная нейронная сеть с долгой-краткосрочной памятью (LSTM), поскольку именно такой тип сетей позволяет учитывать иерархичность и созависимость музыкальных данных [5].

Важнейшей характеристикой музыкальной композиции является ее соответствие определенному эмоциональному состоянию пользователя-человека. Для этой задачи в работе добавлен восьмицветовой тест Люшера с одним проходом, чтобы определить состояние человека.

Целью работы является исследование возможностей автоматизации технического творчества посредством автоматизированной генерации музыкальной композиции по эмоциональному состоянию человека.

Для достижения поставленной цели, необходимо решить следующие задачи:

- 1) Исследование и разработка моделей и методов автоматизированного определения эмоционального состояния человека;
- 2) Разработка моделей и методов генерации музыкальной композиции по определенному эмоциональному состоянию человека;
- 3) Разработка алгоритмического и информационного обеспечения для создания системы генерации композиций.

Определение эмоционального состояния человека

Автоматическое определение эмоционального состояния человека является достаточно нетривиальной задачей, тесно связанной с психологией. Для наиболее точного определения необходимо активное участие специалиста. Однако существуют различные варианты автоматизированного тестирования, чтобы иметь некое представление об эмоциональном состоянии пользователя. В данной работе было решено оставить пользователю возможность выбора способа определения состояния: первый способ заключается в ручном определении пользователем изображения, которое, по его мнению, наиболее точно характеризует его текущее состояние; второй способ – использование специализированного восьмицветового теста Люшера с одним проходом, который является одним из психологических тестов для определения текущего и желаемого состояния человека, а также возможных конфликтов. Тест Люшера позволяет определить состояние человека путем анализа взаиморасположения предпочтительных цветов. При этом, во внимание берется порядок выбора и расположение так называемых основных цветов (красный, синих, зеленый и желтый). Если все основные цвета расположены на главных позициях (до пятой), то можно говорить, что в той или иной мере выполняются потребности, описанные цветами. Если же какой-либо из основных цветов расположен дальше, то можно говорить о наличии конфликта.

Анализ изображений

В данной работе анализ изображения нужен для получения характеристик результирующей музыкальной композиции, что является одним из вариантов генерации музыки, соответствующей эмоциональному состоянию человека. В первую очередь, нужно преобразовать оригинальное изображение в цветовое пространство, которое позволит получить данные каждого пикселя изображения – насыщенность, яркость и оттенок. Затем, необходимо определить преобладающий цвет изображения. По нему мы

сможем в дальнейшем получить тональность музыкальной композиции. Существуют различные варианты определения преимущественного цвета: подсчетом или с помощью методов кластеризации. В данной работе будем использовать алгоритм кластеризации K-средних, поскольку именно данный метод дает относительно высокую эффективность и качество при сравнительной простоте реализации [6].

Для реализации данного алгоритма, для начала, необходимо прочитать изображение с использованием библиотеки OpenCV и функции imread. Особенностью данной функции является то, что сразу после чтения изображения, мы получаем данные о цветах в сине-зелено-красной цветовой схеме, но для дальнейшей работы нам нужна именно оригинальная схема красный-зеленый-голубой. Поэтому, следующим этапом алгоритма является преобразование каналов с использованием специальной функции cvtcolor. Теперь мы получили трехмерные данные изображения: номера строки и столбца, номер цветового канала. Поскольку нам нет необходимости делить информацию по строкам и столбцам, кроме того, значительно сложнее оперировать трехмерной матрицей, то алгоритм преобразует ее в двумерную. Для программной реализации можно просто использовать готовую функцию кластеризации, например, библиотеку sklearn, передавая ей количество кластеров [7].

Таким образом, мы получили преимущественный цвет изображения, последним этапом обработки изображения является получение по нему тональности композиции и ее музыкальный окрас. Для того, чтобы получить наименование ноты для будущей тональности композиции, необходимо преобразовать полученный оттенок цвета в ноту. Для этого используются специальные схемы соотнесения цветов и нот. Существует несколько таких схем, в данной работе используется схема Ньютона (таблица 1). После получения ноты тональности, необходимо определить музыкальный окрас композиции – мажорная или минорная тональность и композиция. Для этого используется базовый принцип из теста Люшера, а именно разделение цветов на две группы, свидетельствующие об эмоциональном состоянии человека. Таким образом, мы получим итоговую тональность.

Таблица 1 – Схема соотнесения цветов и нот
Table 1 – Correlation scheme between color and musical notes

Цвет	Нота
Зеленый	F
Зелено-голубой	F#
Голубой	G
Сине-фиолетовый	G#
Фиолетовый	A
Желто-зеленый	A#
Розовый	H
Красный	C
Красно-оранжевый	C#
Оранжевый	D
Желто-оранжевый	D#
Желтый	E

Другим вариантом генерации композиции по эмоциональному состоянию человека является определение характеристик композиции по тесту Люшера. Для этого пользователю вначале предлагается пройти тест: пользователь поочередно выбирает наиболее предпочтительные цвета из восьми цветов – зеленый, красный, синий, желтый,

серый, черный, коричневый, пока они не закончатся, таким образом пользователь структурирует цвета в порядке предпочтения. После этого оценивается взаимное расположение основных цветов и согласно этому выбирается музыкальный жанр и окрас по следующему алгоритму:

1) Если все основные цвета в 1-5-й позициях, то считается, что потребности описываемых цветов удовлетворены. Тогда результирующая музыкальная композиция будет мажорной. Для определения музыкального жанра смотрим на первый выбранный цвет:

– синий (1) – потребность в удовлетворении, спокойствии, устойчивой положительной привязанности; результирующий жанр – блюз;

– зеленый (2) – потребность в самоутверждении; результирующий жанр – классика;

– красный (3) – потребность активно действовать и добиваться успеха; результирующий жанр – рок;

– желтый (4) – потребность в перспективе, надеждах на лучшее, мечтах; результирующий жанр – классика.

2) Если какой-либо из основных цветов стоит дальше пятой позиции, то имеет место какой-либо конфликт, тревожность, неудовлетворенность из-за неблагоприятных обстоятельств. Отвергаемый цвет может рассматриваться как источник стресса. Он и выбирается как основной для генерации композиции. Тогда результирующая музыкальная композиция будет минорной. Для определения музыкального жанра смотрим на отвергаемый цвет:

– синий (1) – блюз;

– зеленый (2) – классика;

– красный (3) – рок;

– желтый (4) – классика.

Метод генерации композиции

В предыдущем разделе мы получили характеристики композиции – тональность и жанр. Следующим этапом является генерация композиции с заданными характеристиками. В данной работе предполагается следующий алгоритм получения композиции по изображению:

1) согласно методу соотнесения цветовых и музыкальных характеристик получаем тональность произведения и последовательность первых нот, считанных с изображения;

2) далее по полученной последовательности нот предсказываем продолжение произведения с помощью обученной модели и нейронной сети;

3) по итоговой последовательности нот и тональности, согласно методу соотнесения цветовых и музыкальных характеристик, строим гармоническую часть произведения.

Если же пользователь определяет состояние по тесту Люшера, то вместо получения первых нот по изображению, берется выбранная в тесте последовательность цветов.

На текущий момент нейронные сети являются мощным инструментом для решения различных задач, связанных с обработкой и анализом данных, включая автоматизацию человеческого творчества. Например, нейронная сеть DeepDream от Google может находить различные шаблоны в изображении и улучшать их, а система BachBot может генерировать музыкальные композиции, повторяя стиль Баха [8].

Музыкальная информация достаточно обширна: важны как отношения между предыдущими состояниями композиции, так и между параллельными голосами (партиями). Кроме того, большой проблемой становится обучение сети на различных релевантных данных, чтобы, с одной стороны, иметь достаточную выборку для получения разнообразных композиций, а с другой стороны, не переобучивать сеть на всевозможные жанры и направления, получив в таком случае на выходе микс из разных жанров, т.е. абсолютно нестройную композицию.

Для того, чтобы обойти эту проблему, в данной работе предлагается обучить нейронную сеть на композициях определенных жанров, затем собрать обученную модель для каждого жанра и генерировать музыкальный материал, только используя определенную модель отдельного жанра. Это позволит получить более подходящий выход сети. Предположим, что мы обучили различные модели на разном обучающем наборе данных по жанрам, тогда процесс генерации новой композиции на верхнем уровне выглядит как показано на Рисунке 1.

Сбор и классификация обучающих данных

В методе, описанном в предыдущем разделе, основной задачей является корректное обучение нейронной сети для композиций с различными музыкальными жанрами. И наиболее важным этапом этого процесса является сбор и классификация данных для обучения нейронной сети.

В нашем случае мы будем обучать нейронную сеть midi-файлами – именно данный тип файлов содержит описание последовательности проигрываемых нот с их длительностями, силой звука, инструментами и прочим. Таким образом, именно анализируя midi-файлы, мы можем наиболее точно получить описание музыкальной композиции, которое приближенно к человеческому описанию (музыкальная нотация).

Для данной цели, был взят набор данных «The Lakh MIDI dataset» [9], который содержит коллекцию из 176.581 уникальных midi-файлов, 45.129 из которых соотнесены с метаинформацией из Million Song Dataset [10]. Данный набор данных включает файлы, которые содержат в наименовании название автора и имя композиции в совершенно произвольной форме без указания на жанр, например F.FENDER.Wasted days n wasted nights.mid, i_am_the_walrus.mid, или IBelieveICanFly.mid. Таким образом, следующим этапом подготовки обучающих данных является классификация данных midi-файлов по жанрам.



Рисунок 1 – Верхнеуровневый алгоритм для генерации музыкальных композиций различных жанров

Figure 1 – The top-level method for generating musical compositions of different genres

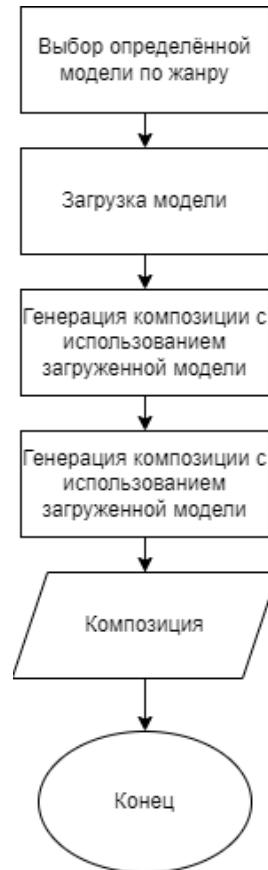


Рисунок 1 (продолжение)
Figure 1 (continued)

Для начала был использован Rest API Spotify, который обеспечил возможность для поиска композиции и получения жанра по автору. Для того, чтобы это сделать, необходимо получить наименование midi-файла и отправить запрос в Spotify с этим наименованием, например, `/v1/search?q=N COLE Almost like being in love&type=track&limit=2`. Данный запрос вернет результат поиска композиции по заданной поисковой строке в формате JSON. Данный ответ содержит некоторые метаданные композиции, например, наименование, изображения и блок с артистом (исполнителем), пример такого блока ответа представлен на Рисунке 2.

```

"artists": [
  {
    "external_urls": {
      "spotify": "https://open.spotify.com/artist/7v4imS0moSyGdXyLgVTIV7"
    },
    "href": "https://api.spotify.com/v1/artists/7v4imS0moSyGdXyLgVTIV7",
    "id": "7v4imS0moSyGdXyLgVTIV7",
    "name": "Nat King Cole",
    "type": "artist",
    "uri": "spotify:artist:7v4imS0moSyGdXyLgVTIV7"
  }
],
    
```

Рисунок 2 – Ответ на запрос поиска композиции с использованием Spotify API
Figure 2 – Query response for a song using the Spotify API

Используя данный ответ, можно получить наименование исполнителя, взяв прямую ссылку из блока href. Данный ответ содержит некоторые метаданные исполнителя, например, наименование и блок жанров, пример такого ответа представлен на Рисунке 3.

```

"genres": [
  "adult standards",
  "lounge",
  "soul",
  "vocal jazz"
]
    
```

Рисунок 3 – Ответ на запрос о получении информации по жанрам с использованием Spotify API
 Figure 3 – Query response for receiving information by genres using the Spotify API

Таким образом, путем отправки двух запросов мы можем получить список жанров, к которым относится композиция. Мы обработали все композиции из набора, часть композиций не нашлась в сервисе, для некоторых не было жанров – после фильтрации всех неактуальных файлов мы получили около 50 тысяч уникальных композиций, которым соответствует 1 213 уникальных жанров. Мы сохранили все результаты в базе данных с довольно простой структурой, показанной на Рисунке 4.

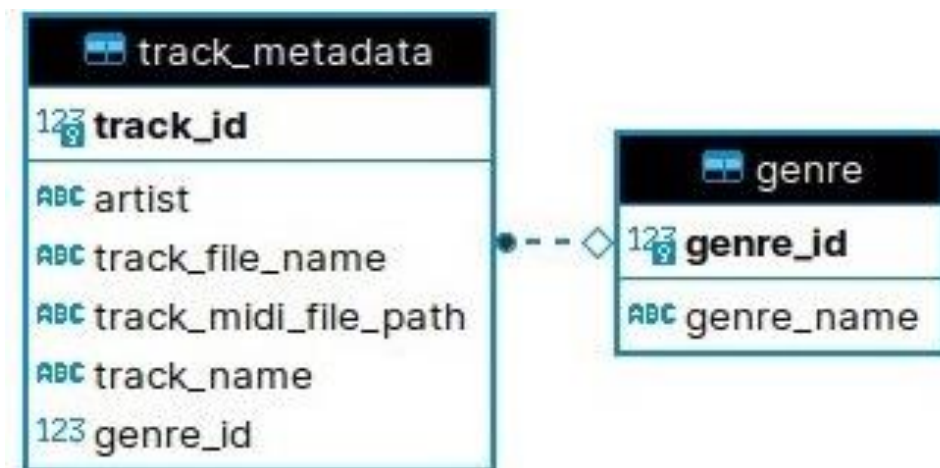


Рисунок 4 – Схема базы метаданных композиций
 Figure 4 – The song metadata database structure

На предыдущем этапе мы получили около 1 тысячи уникальных жанров, большинство из которых слишком узкие, очевидно, что для предсказания композиций это слишком много. Теперь необходимо расширить все жанры до нескольких простых. Подходящего API для этого не нашлось, поэтому данное распределение было сделано вручную. Мы получили около 20 простых жанров, распределение композиций по простым жанрам показано в Таблице 2.

Таблица 2 – Распределение композиций по простым жанрам
Table 2 – The distribution of compositions by simple genres

Наименование жанра	Количество композиций
Blues	354
Children music	219
Relax	109
Choir	58
Classical	1 451
Country	680
Dance	450
Disco	466
Drum	7
Electronic	727
Folk	385
Hip-hop	710
Indie music	312
Jazz	473
Latin music	408
Pop music	10 438
Rap music	280
Reggae	141
Religious	197
Rock	11 464
Soul	407
Soundtrack	882

Нейронная сеть

Для генерации музыкальных данных можно применить множество различных математических методов: марковские цепи, генетический алгоритм, порождающие грамматики и другие – однако в данной работе для этой задачи были выбраны современные методы машинного обучения, а именно нейронные сети, поскольку именно они позволяют сравнительно просто обучить различные модели и сгенерировать композицию, основываясь на таких скрытых взаимосвязях в музыкальных данных, которые не всегда формализуемы и видны человеку.

Существуют разные архитектуры искусственных нейронных сетей, которые обладают своими достоинствами и недостатками, а также используются для различных задач. В данной работе была использована архитектура нейронной сети, называемая рекуррентной, поскольку именно данный тип сетей позволяет выявлять и воспроизводить все взаимосвязи в неструктурированных данных, что делает ее эффективной для генерации музыки. В рекуррентных нейронных сетях функция применяется для каждого входа, при этом результат расчета зависит как от текущего состояния, так и от результата предыдущего вычисления. Эта особенность позволяет сети реализовать некоторое подобие памяти, что критически важно для генерации музыкальных композиций, поскольку каждая нота следующего такта так или иначе взаимосвязана со всеми предыдущими состояниями нот и аккомпанемента. Однако в рекуррентных нейронных сетях существует так называемая проблема долгосрочных зависимостей, которая заключается в неспособности сети вычислить правильное текущее состояние на основе прошлых данных, вычисленных большое количество шагов

назад. Таким образом, вычисление будет произведено только по текущему и нескольким предыдущим состояниям, но без учета контекста в целом [11].

Для решения данной проблемы были изобретены сети с долгой краткосрочной памятью или LSTM. В отличие от обычных рекуррентных нейронных сетей, где повторяющийся модуль представляет собой простую функцию, в сетях LSTM этот модуль содержит четыре взаимодействующих между собой слоя: входной слой (фильтр), выходной слой, забывающий слой и ячейка памяти. Основной целью этих слоев является предотвращение данных от перезаписи и забывания. Такие нейронные сети наиболее хорошо подходят для анализа сложных структур, например, текстовых и музыкальных данных, а также предсказания временных рядов [12].

Нейронная сеть умеет работать только с числовой информацией. Поэтому в данной работе было решено обучать и предсказывать композиции в midi представлении. Плюсом данного формата является то, что данные о начале и окончании звучания конкретной ноты или набора нот представлены в числовом виде, при этом высота и ее название специальным образом зашифровано в числовые значения. Это делает midi формат удобным для работы с нейронной сетью, кроме того, такой подход не требует дополнительных преобразований входных обучающих файлов – их сразу можно подготовить в midi формате. Таким образом, сценарий обучения сети выглядит следующим образом: берем заранее подготовленный набор музыкальных композиций в midi формате для определенного жанра, обучаем сеть и сохраняем модель в отдельный файл для данного жанра. Тогда процесс генерации композиции выглядит следующим образом: по заданному жанру определяем нужную модель для использования, загружаем модель в память и генерируем композицию в midi формате с использованием загруженной модели.

Архитектура системы

Для подтверждения разработанных алгоритмов и методов была разработана программа (веб-сайт) для генерации композиции по цветовой гамме изображений. Условно, систему можно разделить на три больших модуля: основная подсистема для взаимодействия с пользователем и извлечения будущих параметров композиции с изображения, подсистема для обучения нейронной сети и предсказания с ее помощью композиций, подсистема для синтеза звуков по предсказанной композиции. Важным моментом данной системы являются две базы данных: первая база данных содержит сгенерированные модели по разным жанрам, вторая – музыкальные сэмплы для генерации композиции в формате mp3 с реалистичным звучанием инструментов.

Веб-сайт содержит несколько страниц, главная из них – страница генерации композиций. На этой странице необходимо указать изображение и выбрать некоторые параметры – жанр и используемую схему; генерация композиции занимает определенное время, затем система предлагает скачать сгенерированную композицию. В тестовом режиме веб-сайт запущен по адресу mgmusic.site. Веб-сайт умеет работать на данный момент только с тремя музыкальными жанрами – блюз, классика и рок. Исследование возможностей других жанров является областью дальнейших исследований.

Архитектура системы представлена на Рисунке 5.

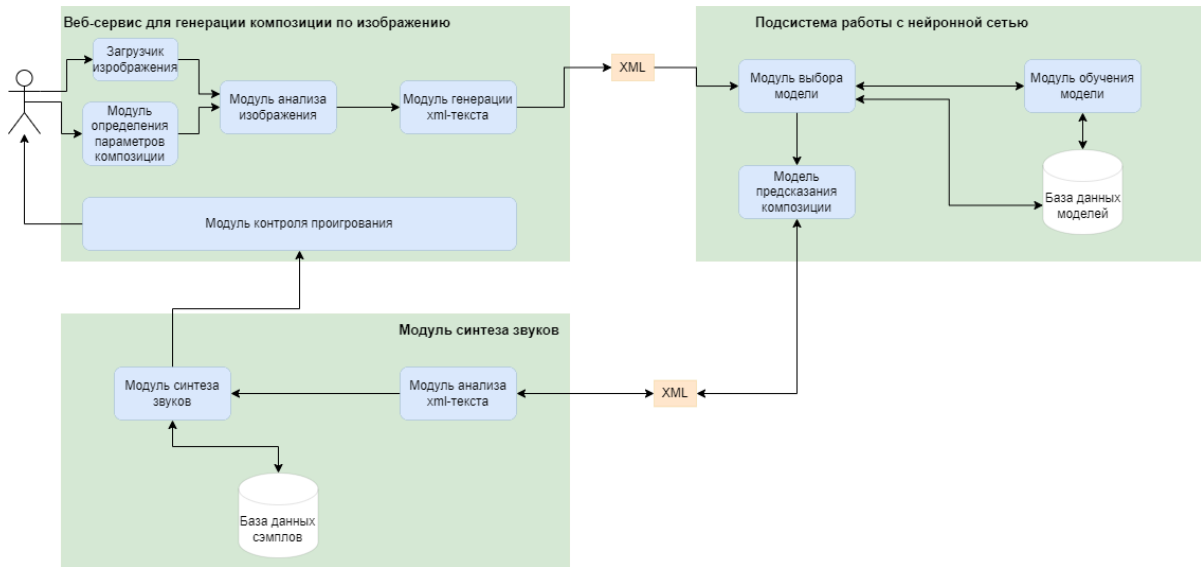


Рисунок 5 – Архитектура системы
Figure 5 – System architecture

Тестирование и валидация

Для подтверждения разработанных методов и алгоритмов, а также программных средств были составлены наборы из изображения и сгенерированной композиции, которые вместе с ссылкой на разработанный веб-сайт были отправлены экспертам-музыкантам, которым предлагалось оценить каждую композицию по следующим критериям:

- соответствие характеру изображения / предполагаемому эмоциональному состоянию;
- реалистичность звучания музыкальных инструментов;
- мелодичность и гармоничность композиции;
- реалистичность и цельность композиции для восприятия.

Оценки экспертов были собраны, обработаны и сведены в Таблицу 3.

Таблица 3 – Экспертные оценки композиций
Table 3 – Expert evaluations of compositions

Критерий	Среднее значение для всех тестов
Соответствие характеру	4.0
Реалистичность звучания инструментов	4.3
Мелодичность и гармоничность	3.9
Цельность восприятия	4.0

Результаты

Проведя анализ оценок экспертов, можно сделать вывод о том, что в целом эксперты считают, что сгенерированные композиции соответствуют характеру изображения или эмоциональному состоянию, определенному посредством теста Люшера. Однако не все остались довольны определенным характером, желая получить композицию другого жанра или с другим эмоциональным окрасом. Критерий «реалистичность звучания инструментов» получил самую высокую оценку: это указывает на эффективность использования метода синтеза сэмплинга в системах, где реалистичность звучания музыкального инструмента играет важную роль.

Критерий «мелодичность и гармоничность композиции» получил самую низкую оценку. Из основных отмечаемых моментов была некоторая равность композиции, то есть некоторые ноты выбиваются из такта и звучат не слитно. Однако общая оценка скорее положительная, отмечается попадание нот в гамму и гармонию.

Общая цельность восприятия была оценена чуть выше, чем предыдущий пункт, однако недостатки мелодичности оставили свой отпечаток.

Заключение

В данной работе был проведен анализ и описание способа автоматизации процесса генерации музыкальных композиций. Можно выделить три части этого процесса:

- анализ изображения для получения характеристик композиции;
- генерация музыкального материала;
- синтез звуков.

Для анализа изображения описан метод, частично основанный на методах преобразования цвета, а также на принципах теста Люшера. Он позволяет по изображению определять музыкальные характеристики, такие как тональность. Кроме того, алгоритм кластеризации K-средних используется для определения преобладающего цвета изображения, а затем преобразования этого цвета в тональность результирующей музыкальной композиции.

Для синтеза музыкального материала описан соответствующий метод. Этот метод основан на светомузыкальных теориях и схеме соотнесения цвета и высоты звука Ньютона. Кроме того, для генерации музыки основным методом являются нейронные сети, которые могут создавать ноты в формате *midí* на основе данных, определенных предыдущим алгоритмом. Именно этот метод совместного использования светомузыкальной теории и нейронной сети представляет особый интерес в данной работе.

И, наконец, описана архитектура системы для генерации музыкальных композиций по предложенным методам и алгоритмам.

Одним из путей дальнейших исследований является уход от изображения и генерация композиции исключительно по результатам теста Люшера, а также проведение тестирования и экспериментов с привлечением экспертов-психологов для оценки соответствия эмоциональному состоянию человека. Кроме того, важной проблемой является появление выбивающихся из тактов нот – нужно доработать методы и укладывать все ноты в определенные такты.

СПИСОК ИСТОЧНИКОВ

1. Chereszniuk I Algorithmic composition and its role in modern musical education. *Art education*. 3:65–68.
2. Ariza C. Two Pioneering Projects from the Early History of Computer-Aided Algorithmic Composition. *Computer Music Journal*. MIT Press. 2012;3:40–56.
3. Никитин Н.А. Автоматизированный синтез музыкальных композиций на основе рекуррентных нейронных сетей. *Искусственный интеллект в решении актуальных социальных и экономических проблем XXI века : сб. ст. по материалам Четвёртой всерос. науч.-практ. конф., проводимой в рамках Пермского естественнонаучного форума «Математика и глобальные вызовы XXI века»*. 2019:80–85.
4. Doornbusch P. Gerhard Nierhaus: Algorithmic Composition: Paradigms of Automated Music Generation. *Computer Music Journal* 2014;4.

5. Graves A., Jaitly N., Mohamed A. Hybrid speech recognition with deep bidirectional LSTM. *Automatic Speech Recognition and Understanding (ASRU)*. IEEE Workshop on IEEE. 2013;273–278.
6. Jain A., Murty M., Flynn P. Data clustering: A review. *ACM Computing Surveys*. 1999;31:264–323.
7. Никитин Н.А. Разработка методов для синтеза музыкальных композиций на основе интуитивного и эмоционального подходов. *Программная инженерия: современные тенденции развития и применения (ПИ–2020): сб. материалов IV всерос. науч.-практ. конф.* 2020:54–61.
8. Feynman L et al. Automatic Stylistic Composition of Bach Chorales with Deep LSTM ISMIR. 2017.
9. Raffel C. Learning-Based Methods for Comparing Sequences, with Applications to Audio-to-MIDI Alignment and Matching. *Doctoral thesis, Columbia University*. 2016;161–163.
10. Bertin-Mahieux Th, Ellis D, Whitman B and Lamere P The Million Song Dataset *Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR)*. 2011:591–596.
11. Bengio Y., Simard P., Frasconi P. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*. 1994;5(2):157–166, DOI: 10.1109/72.279181.
12. Hochreiter S., Schmidhuber J. Long short-term memory. *Neural Computation* 1997;9:1735–1780.

REFERENCES

1. Chereszniuk I Algorithmic composition and its role in modern musical education. *Art education*. 3:65–68.
2. Ariza C. Two Pioneering Projects from the Early History of Computer-Aided Algorithmic Composition. *Computer Music Journal*. MIT Press. 2012;3:40–56.
3. Nikitin N.A. Avtomatizirovannyj sintez muzykal'nyh kompozicij na osnove rekurrentnyh nejronnyh setej. *Iskusstvennyj intellekt v reshenii aktual'nyh social'nyh i ekonomicheskikh problem HKHI veka: sb. st. po materialam CHetvyortoj vseros. nauch.-prakt. konf., provodimoj v ramkah Permskogo estestvennonauchnogo foruma «Matematika i global'nye vyzovy XXI veka»*. 2019:80–85. (In Russ.)
4. Doornbusch P. Gerhard Nierhaus: Algorithmic Composition: Paradigms of Automated Music Generation. *Computer Music Journal* 2014;4.
5. Graves A., Jaitly N., Mohamed A. Hybrid speech recognition with deep bidirectional LSTM. *Automatic Speech Recognition and Understanding (ASRU)*. IEEE Workshop on IEEE. 2013;273–278.
6. Jain A., Murty M., Flynn P. Data clustering: A review. *ACM Computing Surveys*. 1999;31:264–323.
7. Nikitin N.A. Razrabotka metodov dlya sinteza muzykal'nyh kompozicij na osnove intuitivnogo i emocional'nogo podhodov. *Programmnyaya inzheneriya: sovremennye tendencii razvitiya i primeneniya (PI–2020): sb. materialov IV vseros. nauch.-prakt. konf.* 2020:54–61. (In Russ.)
8. Feynman L et al. Automatic Stylistic Composition of Bach Chorales with Deep LSTM ISMIR. 2017.
9. Raffel C. Learning-Based Methods for Comparing Sequences, with Applications to Audio-to-MIDI Alignment and Matching. *Doctoral thesis, Columbia University*. 2016;161–163.

10. Bertin-Mahieux Th, Ellis D, Whitman B and Lamere P The Million Song Dataset *Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR)*. 2011:591–596.
11. Bengio Y., Simard P., Frasconi P. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*. 1994;5(2):157–166, DOI: 10.1109/72.279181.
12. Hochreiter S., Schmidhuber J. Long short-term memory. *Neural Computation* 1997;9:1735–1780.

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Никитин Никита Андреевич, аспирант кафедры «ПОАС» Волгоградского государственного технического университета, Волгоград, Российская Федерация.

e-mail: set.enter@mail.ru

Nikita A. Nikitin, Postgraduate Student of the Software Engineering Department of Volgograd State Technical University, Volgograd, Russian Federation.

Орлова Юлия Александровна, доктор технических наук, профессор Волгоградского государственного технического университета, Волгоград, Российская Федерация.

e-mail: yulia.orlova@gmail.com

ORCID: [0000-0003-4854-7462](https://orcid.org/0000-0003-4854-7462)

Yulia A. Orlova, Doctor of Technical Sciences, Professor of Volgograd State Technical University, Volgograd, Russian Federation.

Розалиев Владимир Леонидович, кандидат технических наук, доцент Волгоградского государственного технического университета, Волгоград, Российская Федерация.

e-mail: vladimir.rozaliev@gmail.com

ORCID: [0000-0002-7372-8364](https://orcid.org/0000-0002-7372-8364)

Vladimir L. Rozaliev, Candidate of Technical Sciences, Associate Professor of Volgograd State Technical University, Volgograd, Russian Federation.

Статья поступила в редакцию 28.04.2022; одобрена после рецензирования 22.06.2022; принята к публикации 29.06.2022.

The article was submitted 28.04.2022; approved after reviewing 22.06.2022; accepted for publication 29.06.2022.