

## **Проблема синтеза музыки и цвета: опыт генерации музыкальной композиции с помощью рекуррентной нейросети на основе анализа изображения**

**Чижик А.В.**, Санкт-Петербургский государственный университет  
chizhik.spinus@gmail.com

### **Аннотация**

Статья посвящена описанию соотношения характеристик звука и цвета. Предлагается метод практического соотнесения в виде модуля, пишущего мелодии на основе анализа изображений. Автором предлагается доработанная цветовая схема И.Ньютона. В основе алгоритма генерации музыкального ряда – RNN.

### **1 Введение**

Социокультурный слой современного общества – результат сформированной под влиянием совокупности знаковых систем онтологической реальности, соприкасаясь с которой индивид становится способным влиться в свою этническую группу (на локальном и макро уровнях), осознав и разделив принятые в ней нормы и ценности. Так, к примеру, все механизмы общественного и массового сознания построены на правильно работающих интерпретациях знаков. Таким образом, семиотическая реальность является фундаментом культурного кода, а ее глобальной подсистемой является язык, который расчерчивает границы ментального и национального единства. Однако необходимо отметить, что существуют не менее важные побочные подсистемы, по которым индивиды ориентируются с той же частотностью. Так одним из эффективных способов познания окружающей действительности для человека является зрение. Зрение выступает в связке со знаковой системой, к которой принадлежит человек, как результат у субъекта появляется суждение. С первобытных времен визуальный способ передачи информации являлся знаковым и символическим кодом, ключом, к расшифровке которого, обладали все представители той или иной культуры. Цвет – один из знаков в семиотической системе, который присутствует в любой культуре; смысловая окраска этого знака разнообразна: он может обозначать часть ритуала, выражать принадлежность к профессии, отражать душевное состояние и

т.д. Каждая культура характеризуется наличием ряда неосознаваемых ассоциативных соответствий между теми или иными образами и цветами, свойственными для того или иного этноса. Декодирование определенного цвета и его оттенков во многом зависит от социальных стереотипов и этнопсихологических особенностей конкретного народа. Это ведет к тому, что каждый язык по-своему описывает различные цвета и их оттенки, что позволяет говорить о символике цветовой палитры в культурной традиции определенного этноса, указывающей на определенные эмотивные понятия. В повседневной жизни мы часто сталкиваемся с ситуацией считывания значения цвета, однако, этот процесс настолько автоматизирован, что не вызывает у человека осознанных когнитивных усилий; причем прочтение цвета как символа производится на разных когнитивных пластах в зависимости от принадлежности к определенной культуре, а также к определенному социальному срезу. А.Ф. Лосев определял символ как «субстанциальное тождество идеи и вещи» [Лосев, 198]. Всякий символ включает в себя образ, но не сводится к нему, поскольку подразумевает присутствие некоего смысла, нераздельно слитого с образом, но ему не тождественного. Образ и смысл образуют два логически связанных между собой и не существующих по-отдельности элемента символа. Символы существуют как символы только внутри интерпретаций. Лосев, описывая типы символов, в частности выделил художественные и мифологические символы, рассмотрение которых помогает описать социокультурное пространство как поле с ярко очерченной символической системой европейской культуры. В качестве одного из центральных символов можно выделить цвет, не ограничивая при этом разнообразие характеристик, которые можно ему дать. Отметим, что в языке любого народа отражаются его цветовое мировосприятие и ментальность, учитывая символику цветовой палитры и ее восприятие этносом. Цветовое мировосприя-

тие и ментальность народа находят свое отражение в лексическом составе языка, который включает ряд слов, относящихся к номинациям цвета или цветообозначениям. Итак, можно заключить, что цветовая палитра является фундаментом системы символов любого масштаба (локального этноса и макросистем), она выходит за рамки исключительно зрительной опоры считывания знаков, предавая легитимную форму языковым тенденциям и психологическим особенностям.

Еще одним важным компонентом периферической анатомо-физиологической системы, обеспечивающей получение и первичный анализ информации из окружающего мира, является слух, что мотивирует выдвинуть систему фиксации звуков с помощью графем в качестве значительной части системы знаков, используемых человечеством (наравне с буквами как основой письменности). Однако следует отметить, что если характеристиками цвета внутри европейской знаковой системы индивиды оперируют достаточно легко, постигая ее в достаточной мере на уровне эмпирического опыта, то музыкальной нотацией, которая передаёт смысл музыкальной логики (звуковысотность и ритм), владеет небольшой круг людей, получивших профессиональное образование.

В этой связи возникла потребность проанализировать, существуют ли параметры, по которым возможно было бы соотнесение системы звуков с системой цвета, что расширило бы сектор индивидов, способных воспользоваться звуком как осознанной знаковой системой. Прикладной целью исследования явилось создание программы, которая способна синтезировать звук по входным параметрам загружаемых изображений, создавая таким образом их звуковую интерпретацию.

## 2 Построение цветотональной системы

Весь ряд музыкальных звуков выстроен в единую систему – звукоряд, в котором все звуки упорядоченно следуют друг за другом от низких до высоких (или наоборот). При этом ряд разделяется на октавы – отрезки, содержащие набор одинаковых по названию нот (до, ре, ми, фа, соль, ля, си). Основные ступени соответствуют звукам, извлекаемым на фортепиано на белых клавишах. Часть звукоряда, которая повторяется на разной высоте, называется октавой. Таким образом, весь звукоряд

можно разделить на октавные участки. Началом октавы принято считать звук «до». Логично считать ключевыми характеристиками музыкального произведения его тональность и темп. Что позволяет начать сопоставление между звуком и цветом, разделив частотный ряд на октавы, т. е. на участки, где длина волны (и частота) изменяются вдвое. Поскольку цвет имеет и другие составляющие (насыщенность, температура и др.) в общую систему необходимо включать и соответствие по этим параметрам.

В рамках данного исследования было принято решение ограничиться разложением на следующие бинарные пары характеристик: яркость/октава, насыщенность цвета/длительность ноты, оттенок/нота, температура цвета/музыкальный лад.

Выделив характеристики, по которым можно состыковать две системы записи информации, необходимо определить идейную основу математической схемы соотнесения цвета и звука. И. Ньютон, занимаясь изучением солнечного света, разложил его с помощью призмы и получил цветовую дорожку [Ньютон, 1927]. В рамках своего исследования ученый сравнил синусы углов преломления выделенных им семи цветов с отношениями семи тонов в октаве. Несмотря на спорное решение выделения цветового спектра звуков (механический по своей сути), он даёт точное установление высоты (температуру цветового ряда). В России основоположниками цветомузыкального искусства считается композитор А.Н. Скрябин и Н.А. Римский-Корсаков. Оба композитора обладали музыкально-цветовой синестезией, но цветомузыку представляли по-разному. С научной точки зрения система, построенная Н.А. Римским-Корсаковым, является более актуальной, так как композитор включал музыкальные образы своих опер в строго определённые цветовые тональности, совпадающие с расчетами И. Ньютона.

Итак, воспользовавшись системой Н.А. Римского-Корсакова, можно соотнести тональность произведения двум цветовым характеристикам – оттенку и цветовой группе, а темп – яркости и насыщенности (рис. 1).

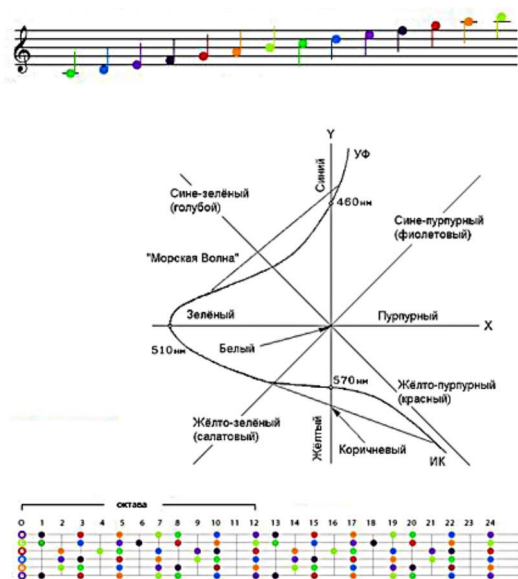


Рис. 1. Цветокоординатная система нелинейной модели и кривая видности глаза. Кривая, описывает положение спектральных цветов (непрерывного спектра полученного разложением призмой белого солнечного света) на координатной плоскости.

Таким образом, очертились параметры, по которым можно проанализировать загружаемое в программу изображение. Первым этапом взаимодействия с поступившим на обработку материалом становится преобразование цветовой модели в удобную для дальнейших операций. Таковой является HSB, в которой координатами цвета являются тон (варьируется в пределах  $0-360^\circ$ ), насыщенность ( $0-100$ ) и яркость ( $0-100$ ). Далее следует определить доминирующие цвета, что задаст тональность музыкального произведения, а числовые эквиваленты яркости и насыщенности установят темп.

### 3 Генерация музыкальной композиции

Музыкальная последовательность является условно бесконечной, в которой важен порядок информации. В связи с этим для генерации музыки имеет смысл использовать рекуррентные нейронные сети, где нейроны обмениваются информацией между собой, получая друг от друга информацию о предыдущем состоянии сети. При этом важно, чтобы у сети была долгая краткосрочная память, что реализуется за счет входного, выходного и забывающего фильтров у нейронов [Будыльский, 2015].

Отметим, что гармония созвучий (для поставленной цели было важным не просто

написать код, который бы выдавал набор звуков, подходящих под числа, полученные при анализе изображений, а создавал приятную для уха музыку) достигается за счет инвариантности по нотам и времени. Инвариантность по нескольким направлениям реализована у сверточных нейронных сетей для распознавания изображений. Если такую сверточную модель дополнить, добавив для каждого пикселя рекуррентную нейросеть с собственной памятью, и заменить пиксели на ноты, то на выходе получится необходимый инвариант (рис. 2). При этом модель типа RNN-RBM обеспечивает сочетание нот (за счет разделения задач).

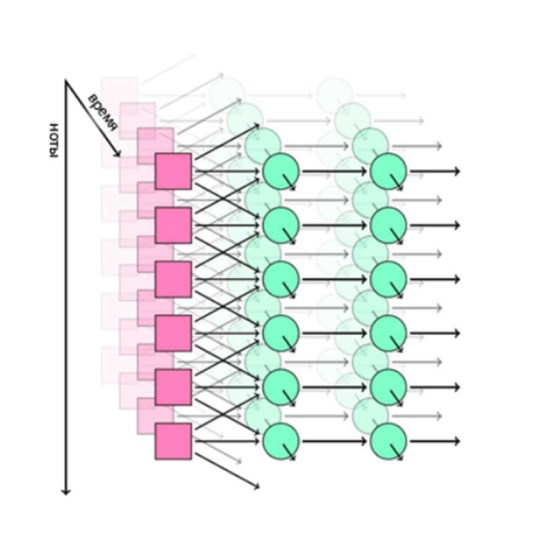


Рис. 2. Инвариантная по времени и нотам система

Ограничение сети можно облегчить за счет введения псевдооси для направления вычислений. Для реализации модели была использована библиотека Theano численного вычисления в Python. Ось по времени в первом слое принимала на вход позицию, высоту звука, значение окружающих нот, предыдущий контекст, ритм [Brinkkemper, 2016]. Самогенерирующие блоки на основе кратко-долговременной памяти срабатывали таким образом, что рекуррентные соединения направлялись по оси времени и по оси нот. После завершающего блока срабатывал простой нереккуррентный слой для выдачи конечного результата с двумя значениями на выходе: вероятность воспроизведения для конкретной ноты и вероятность сочетания ноты с другой.

Программа генерации музыкальных композиций с использованием нейронных сетей была обучена на 11 композициях М.К. Чюрлёниса, композитора и художника, который за-

нимался изучением синтеза искусств и поискам аналогий музыки и изобразительного искусства (для первичного теста работоспособности системы было важно сравнить то, что получится на выходе из программы и то, что писал композитор, исследуя данную проблематику сам). Практическая модель состояла из двух скрытых слоев по времени, каждый из 150 узлов, и двух слоев по оси нот, на 50 и 30 узлов, соответственно. Тренировка проводилась в виртуальной машине g2.xlarge в облаке Amazon Web Services.

После обучения был составлен набор из 8 изображений (серия картин-сонат), для каждого программный модуль создал музыкальные композиции.

#### 4 Заключение

В ходе выполнения работы была определена схема соотнесения цветовых и музыкальных характеристик. Были проанализированы типы архитектур нейронных сетей и сделан выбор в пользу рекуррентных нейронных сетей (RNN) с долгой краткосрочной памятью. В результате проведения эксперимента, была обучена модель (нейронная сеть), способная создать по полученному на входе изображению музыкальное произведение с гармонией, характеризующейся использованием усложненной тоники, ладов с несимметричной структурой, использованием принципа серийности (следствие выборки материала для обучения). Представляется, что созданный модуль может быть использован в доработанном виде на занятиях в музыкальных школах, а также имеет потенциал для применения в рамках прикладных задач аудио дизайна (фоновая музыка для кинематографа и видео-арта). Стоит отметить, что созданный модуль является прототипом реальной системы, так как в рамках этого исследования теория соотнесения цвета и звука изучалась поверхностно и имеет большую погрешность при практической реализации.

#### Список литературы

- Brinkkemper, F. *Analyzing Six Deep Learning Tools for Music Generation*. [Электронный ресурс]. – 2016. – Режим доступа: <http://www.asimovinstitute.org/analyzing-deep-learning-tools-music/> (Дата обращения: 13.03.2019).
- Будыльский Д. В. *GRU и LSTM: современные рекуррентные нейронные сети* // Молодой ученый. №15. 2015. – С. 51-54.
- Лосев А. Ф. *Знак, символ, миф*. М., 1982 – 480 с.
- Аллахвердов В.М. *Психология искусства. Эссе о тайне эмоционального воздействия художественных произведений*. СПб., 2001 – 200 с.
- Ньютон И. *Оптика*. М.-Л., 1927 – 73 с.