

Министерство образования Республики Беларусь  
Учреждение образования  
Белорусский государственный университет  
информатики и радиоэлектроники

УДК 004.032.324:78

Андрадэ  
Александр Исмаэлевич

Система распознавания музыкальной транскрипции при помощи методов  
машинного обучения

**АВТОРЕФЕРАТ**

на соискание степени магистра информатики и вычислительной техники

по специальности 1-40 81 02 «Технологии виртуализации и облачных  
вычислений»

---

Научный руководитель  
Насуро Екатерина Валериевна  
к.т.н., доцент

---

Минск 2019

## ОГЛАВЛЕНИЕ

ОГЛАВЛЕНИЕ .....	2
ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ.....	3
ВВЕДЕНИЕ .....	4
КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ.....	6
ЗАКЛЮЧЕНИЕ .....	7
БИБЛИОГРАФИЧЕСКИЙ СПИСОК .....	8
Список публикаций соискателя.....	8

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

*Цель работы:* состоит в исследовании методов музыкальной транскрипции и реализации метода, основанного на использовании рекуррентной нейронной сети.

*Задачи исследования:* В соответствии с поставленной целью, можно выделить следующие задачи:

1. Выполнить анализ предметной области.
2. Так как задача не решена в общем виде на сегодняшний день, предложить свой метод для решения задачи транскрипции фортепиано.
3. Провести испытания, получить результаты работы метода, если возможно – улучшить их.
4. Оценить, насколько хорошо метод решает задачу, определить его преимущества и недостатки.
5. Предложить улучшения для предложенного метода, которые, возможно, позволят получить более качественный результат.

*Гипотезы:*

1. Частотно-временное представление произведения фортепиано имеет достаточно информации (свойств) для точной классификации каждой ноты в определенный момент времени.
2. Временная составляющая информации (свойств) имеет важное значение, поэтому для решения задачи подходящим инструментом послужат рекуррентные сети типа LSTM.
3. Для работы классификатора достаточно обучить одну нейросеть с независимыми классификаторами, число которых равно колву нот фортепиано.
4. Необходимую информацию (свойства) нейросеть способна выделить из предоставляемого частотно-временное представления.

Данная работа имеет практическую *связь* с такими направлениями, как распознавание речи и цифровая обработка сигнала, поэтому она актуальна.

Результаты исследований опубликованы в виде тезисов на следующих научных конференциях:

1. 54-я научно-техническая конференция аспирантов, магистрантов и студентов БГУИР, которая проходила в мае 2019 в Минске. Тема тезисов: «Музыкальная транскрипция при помощи методов машинного обучения» [1 – А.].
2. Пятая международная научно-практическая конференция «Big Data and Advanced Analytics», 14.03.2019 в Минске. Тезисы: «Средство музыкальной транскрипции при помощи методов машинного обучения» [2 – А.].

## ВВЕДЕНИЕ

Музыкальная транскрипция – это процесс анализа акустического музыкального сигнала и запись музыкальных параметров звуков, возникающих в нем. Традиционно музыка записывается в виде нотной записи, которая позволяет указать высоту тона, время начала и продолжительность каждого воспроизводимого звука.

Музыкальную транскрипцию также можно представить в репрезентационном смысле как преобразование акустического сигнала в символическое представление. С другой стороны, музыкальная нотация - это скорее инструкция для исполнителя, чем представление музыкальной записи.

Она описывает музыку на языке, понимаемом музыкантом и служит для воспроизводства музыкальных звуков. С этой точки зрения, музыкальная транскрипция может быть рассмотрена как обнаружение «рецепта» или реверс-инжиниринг «исходного кода» музыкального сигнала. Прикладная нотация не обязательно должна быть традиционной музыкальной нотацией, скорее любым символическим представлением, которое дает необходимую информацию для исполнения записи используя доступные музыкальные инструменты. В случае электронного синтезатора, используемого для повторного синтеза, MIDI файл является примером такого музыкального представления записи.

Музыкальная нотация не только позволяет воспроизводить музыкальную запись, но также и модифицировать, поправлять и обрабатывать музыку на более высоком уровне абстракции. Другое важное применение музыкальной транскрипции – это структурированное аудио кодирование. MIDI-подобное представление невероятно компактно, но в то же время оно сохраняет идентифицируемость и характеристики музыкального произведения в значительной степени. Еще одно применение музыкальной транскрипции включает *поиск информации* на основе мелодии фрагмента, *музыкальный анализ* импровизируемой этнической музыки и *интерактивные музыкальные системы*, которые генерируют аккомпанемент поющему или играющему солисту.

Автоматическая транскрипция полифонической музыки является субъектом все возрастающих исследований на протяжении последних 30 лет. Проблема во многом аналогична автоматическому распознаванию речи, но не получила сопоставимых академического или коммерческого применения.

Несмотря на большое количество попыток решить проблему, практически применимая, универсальная система музыкальной транскрипции не существует и по сей день. Существующие системы на порядок хуже

специалистов в точности и гибкости. Самые новые системы, однако, достигли определенной степени точности в транскрибировании полифонической музыки ограниченной сложности. Типичные ограничения целевого сигнала таковы:

- число одновременных нот ограничено (или фиксировано);
- интерференция ударных и перкуссионных инструментов обычно запрещена;

Как ни странно, даже транскрипция одноголосого произведения является еще нерешенной проблемой.

*Цель работы:* состоит в разработке метода автоматической транскрипции, который по возможности учитывает опыт и недостатки аналогов для получения лучшего результата распознавания.

С учётом описанных выше недостатков существующих подходов, для достижения поставленной цели перед автором данной работы были поставлены следующие *задачи*:

3. Разработать метод для более точного выделения в спектре звука информации, с целью получения более аккуратного отображения фрагментов звукозаписи в пространство признаков.

4. Выделить новые наиболее информативные признаки, желательно небольшой размерности, например, использовать мел-кепстральные коэффициенты (MFCC) в качестве признаков.

5. Реализовать описанные алгоритмы в виде программного продукта, позволяющего распознавать последовательность нот в поданном на вход звуковом файле.

## **КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ**

В главе «Обзор литературы» описываются физические свойства звука, необходимые понятия из теории музыки, цифровой звук и свойства цифровых звукозаписей. Также рассмотрены несколько аналогов, их сильные и слабые стороны.

В главе «Распознавание нот с помощью машинного обучения» производится формализация поставленной задачи, рассказывается о частотно-временном представлении звукозаписи, объясняется задача классификации применительно к данной работе. Также говорится о предварительной обработке звукозаписи, поясняются различные подходы для получения спектрограммы. В ней также описываются признаки, подходящие для обучения, рассказывается о рекуррентных нейронных сетях и марковских моделях.

В главе «Экспериментальный раздел» описывается структура проекта, подготовка данных для обучения, настраиваемые гиперпараметры сети. Перечислены испробованные архитектуры моделей. Дано сравнение процесса обучения моделей разной архитектуры, подведены результаты обучения.

## ЗАКЛЮЧЕНИЕ

В результате исследования был реализован классификатор нот фортепиано. Для распознавания использована рекуррентная нейронная сеть типа LSTM, с помощью которой решаются специализированные задачи, в которых необходимо распознавать последовательности.

Сеть дала неплохие результаты классификации. Хотя ее уровень не позволяет сказать, что задача распознавания нот решена в общем виде. LSTM слой дал хорошие результаты по временной составляющей классификации. Длительности нот распознавались достаточно хорошо. Аналоги как раз плохо работают с длительностями нот, для улучшения результата они используют дополнительные шаги, такие как скрытые марковские модели. Глубокое обучение же позволяет не беспокоиться о выборе свойств для классификации и нейросеть таким образом решает задачу в общем виде, выделяя значащие признаки самостоятельно.

В качестве будущих улучшений сети можно рассмотреть следующее:

6. Применить в архитектуре вместо слоя LSTM, слой ConvLSTM2D так как данные для обучения представлены в виде двумерного массива и имеют пространственно-временной характер.

7. Использовать новый намного больший датасет - Maestro [31]. Датасет состоит из более чем 200 часов фортепианных выступлений, записанных с точным выравниванием (~ 3 мс) между метками нот и звуковыми сигналами.

8. Улучшение характеристик частотно-временного преобразования, например, увеличение числа частотных интервалов для обучающей выборки, что позволит получить более «четкое» изображение.

## БИБЛИОГРАФИЧЕСКИЙ СПИСОК

### Список публикаций соискателя

[1 – А.] Андрадэ, А. И. Музыкальная транскрипция при помощи методов машинного обучения / А. И. Андрадэ // Компьютерные системы и сети: материалы 54-й научной конференции аспирантов, магистрантов и студентов, Минск, 23 – 27 апреля 2018 г. / Белорусский государственный университет информатики и радиоэлектроники. – Минск, 2018. – С. 15.

[2 – А.] Андрадэ, А. И. Средство музыкальной транскрипции при помощи методов машинного обучения / А. И. Андрадэ, Е. В. Насуро // BIG DATA and Advanced Analytics = BIG DATA и анализ высокого уровня : сборник материалов V Международной научно-практической конференции, Минск, 13–14 марта 2019 г. В 2 ч. Ч. 1 / Белорусский государственный университет информатики и радиоэлектроники; редкол. : В. А. Богуш [и др.]. – Минск, 2019. – С. 376 – 380.