

УДК 004.032.26:581.522-047.44

СРЕДСТВО МУЗЫКАЛЬНОЙ ТРАНСКРИПЦИИ ПРИ ПОМОЩИ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ



А.И. Андрадэ
Магистрант БГУИР



Е.В. Насуро
Доцент кафедры ЭВМ БГУИР,
кандидат технических наук

Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь

E-mail: veimar.94@gmail.com, nasurokv@bsuir.by

А.И. Андрадэ

Окончил Белорусский государственный университет информатики и радиоэлектроники. Магистрант БГУИР. Работает в *simatosoft* в должности инженера-программиста. Разрабатывает систему для транскрибирования записей фортепиано.

Е.В. Насуро

Доцент кафедры электронных вычислительных машин БГУИР, основные направления деятельности: проектирование взаимодействия пользователя и сложных программно-аппаратных продуктов, научное руководство магистрантами.

Аннотация. Музыкальная транскрипция – это процесс анализа акустического музыкального сигнала и запись музыкальных параметров звуков, возникающих в нем. Символьное представление звукового ряда необходимо во многих сферах человеческой деятельности - поиск информации на основе мелодии фрагмента, музыкальный анализ импровизируемой этнической музыки, интерактивная генерация аккомпанемента поющему или играющему солисту, контроль за соблюдением авторских прав. Такие задачи требуют разработки средств интеллектуальной транскрипции музыки, которые позволят получать точные, оперативные данные в автоматическом режиме. В работе проведена оценка существующих методов музыкальной транскрипции, приводится архитектура нейронной сети, предназначенной для автоматической транскрипции, характеристики обучающих данных и результатов транскрибирования. Особый интерес вызывает полифоническая музыка, распознавание которой сопровождается дополнительными сложностями.

Ключевые слова: музыкальная транскрипция, нейронные сети.

Введение. Музыкальная нотация - представление музыкальной записи. Она описывает музыку на языке, понимаемом музыкантом и служит для воспроизводства музыкальных звуков. Прикладная нотация не обязательно должна быть традиционной музыкальной нотацией, скорее любым символьным 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.

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

коммерческого применения.

Несмотря на большое кол-во 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.

Техническая информация. Музыкальный звук характеризуется высотой, тембром, громкостью, длительностью. Обычный человек способен слышать звуковые колебания в диапазоне частот от 16 Гц до 20 кГц. Длительности музыкальных звуков имеют отношение 1:2. Динамические оттенки подчинены шкале громкостей, не имеющей абсолютных значений. Из музыкальных звуков строится музыка.

Итоговое музыкальное звучание получается наложением всех извлекаемых из инструмента звуков (рисунок 1).

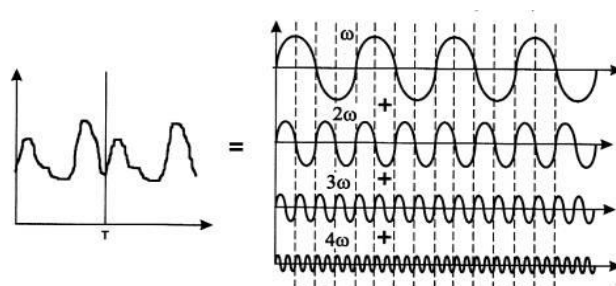


Рисунок 1. Музыкальное звучание

Для сохранения звука в компьютерах используется т.н. дискретизация – сохраняется значение колебания звука через равные промежутки времени. Количество таких измерений в секунду – частота дискретизации. Количество возможных измеренных значений – разрядность квантования (рисунок 2).

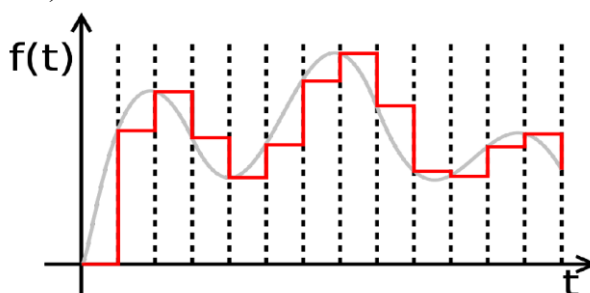


Рисунок 2. Дискретизация музыкального звучания

Используемые методы. Для автоматической музыкальной транскрипции необходимо последовательно использовать несколько методов. Непрерывное вейвлет-преобразование (CWT) позволяет преобразовать звукозапись в частотно-временное представление и получить вектора признаков. Далее организовывается обучение без учителя рекурсивной

нейронной сети (RNN-RBM). Формируется набор классификаторов из 88 штук (по количеству клавиш) – классификация методом опорных векторов (SVM) «один против всех». Пост-процессинг скрытыми марковскими моделями (СММ). Взаимодействие методов показано на рисунке 3.

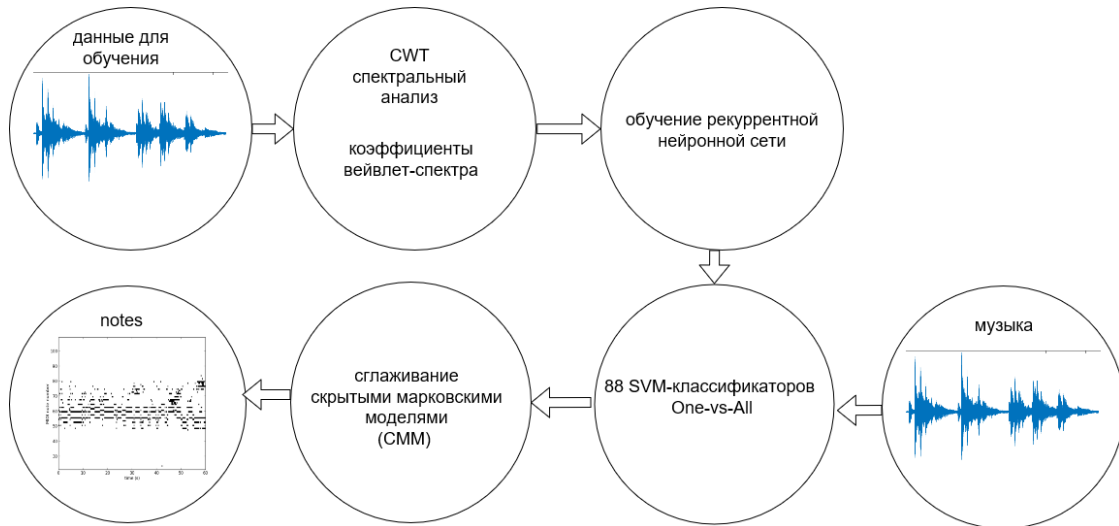


Рисунок 3. Этапы подготовки данных и классификации

Обучающая выборка. Набор данных для обучения находится по следующему адресу <http://www.piano-midi.de/>. Этот ресурс содержит популярные произведения фортепиано в формате midi и в mp3, синтезированном из midi. Для обучения понадобятся оба формата.

Данные, представленные в mp3 формате преобразуем во временно-частотное представление с помощью constant-Q преобразования, получаем спектрограмму (рис. 4). Далее производится нормализация, в результате которой данные становятся числами с плавающей запятой в диапазоне $[0, 1)$.

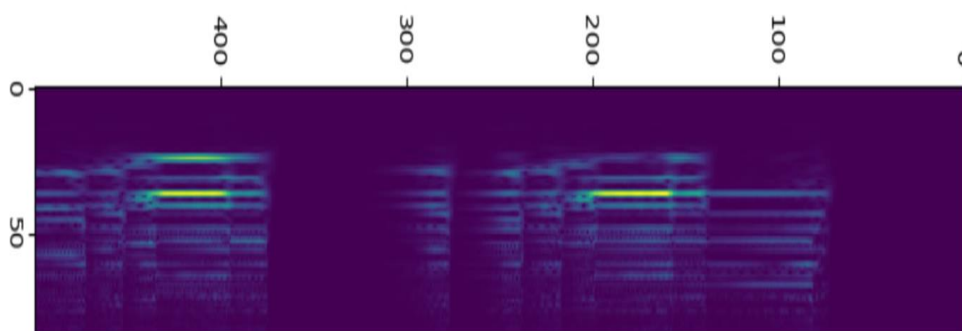


Рисунок 4. Временно-частотное представление отрезка произведения

Каждому фрейму данных спектрограммы ставим в соответствие midi данные, при необходимости выравниваем их по времени начала проигрывания первой ноты.



Рисунок 5. Данные о проигрывании нот, полученные из midi файла

Для обучения рекуррентной нейросети нарезаем композицию (тренировочный сет) на небольшие отрезки, такие, чтобы нейросеть смогла распознать именно последовательность событий проигрывания нот. Например, размер спектрограммы одного из тестовых произведений: 42739 (фреймов) x 84 (частотных бина). Можно разбить на последовательность по 300 (фреймов) x 84 (частотных бина).

Описание и архитектура нейронной сети. Нейронная сеть представляет собой несколько последовательный слоев LSTM (рис. 6). LSTM способны запоминать последовательности информации, поэтому данная архитектура сети подходит для запоминания нот произведения.

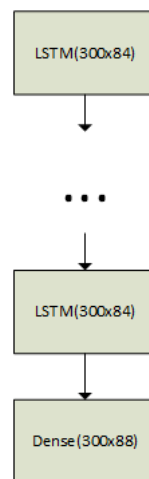


Рисунок 6. Архитектура нейронной сети

Последний слой – полносвязный служит для классификации, его размер 88 нейронов – по количеству нот в фортепиано.

Для кодирования меток (лейблов) используется multi hot encoding, так как в один момент времени могут звучать несколько нот, например аккорд. 1 - обозначает, что в данный момент времени (частотный фрейм) такая нота проигрывается, 0 - обозначает ее отсутствие (рисунок 7).

C1	D1	E1	F1	H1	
1	1	0	0	...	0
1	0	0	1	...	1

Рисунок 7. Пример multi hot encoding

В итоге после обучения нейросети, есть возможность распознавать присутствие нот фортепиано (всего 88) во фрагменте произведения формата 300 (фреймов) x 84 (частотных бина).

Для реализации нейросети использовалась библиотека глубокого машинного обучения keras (<https://keras.io>).

Заключение. Разработанный алгоритм автоматизированной музыкальной транскрипции включает в себя последовательное использование нескольких математических средств и методов. Такой подход позволяет выделять звуки и проводить классификацию звукового ряда, что дает возможность получить символьную нотификацию музыкального произведения.

Литература

- [1] Нижибицкий Е. - Музыкальная транскрипция при помощи методов машинного обучения. Евгений Нижибицкий. ВМК МГУ. 24 сентября 2014 г.
- [2] M. Marolt — A Connectionist Approach to Automatic Transcription of Polyphonic Piano Music (2004)
- [3] Anssi P. Klauri - Automatic Music Transcription as We Know it Today (2004)
- [4] Graham E. Poliner and Daniel P.W. Ellis - A Discriminative Model for Polyphonic Piano Transcription (2006)

MEANS OF MUSICAL TRANSCRIPTION BY MEANS OF MACHINE TRAINING METHODS

A.I. Andrade

*Undergraduate student of
the BSUIR*

K.V. Nasuro

*Assistant professor at BSUIR,
Candidate of Technical Sciences*

Belarusian state university of informatics and radio electronics, Republic of Belarus

Abstract. The musical transcription is a process of the analysis of an acoustic musical signal and recording of musical parameters of the sounds arising in it. Symbolical representation of a sound row is necessary in many spheres of human activity - information search by the part of melody, the musical analysis of the improvised ethnic music, interactive generation of accompaniment for the singing or playing soloist, control of copyright compliance. Such tasks demand development of means of an intellectual transcription of music which will allow to obtain exact, operational data in the automatic mode. The assessment of the existing methods of a musical transcription is carried out, the architecture of the neural network intended for the automatic transcription, training set and results of transcription characteristics are given.

Keywords: musical transcription, neural networks