

РАСПОЗНАВАНИЕ МУЗЫКАЛЬНЫХ ИНСТРУМЕНТОВ НА ОСНОВЕ НЕЙРОСЕТЕВОГО АНАЛИЗА АМПЛИТУДНО-ЧАСТОТНЫХ ХАРАКТЕРИСТИК

Станкевич Ф.В.

Научный руководитель: Белоусов А.А., к.т.н., доцент

Томский политехнический университет, 634050, Россия, г. Томск, пр. Ленина, 30

E-mail: f.stankevich@mail.ru

В данной работе рассматривается проблема распознавания тембра отдельного музыкального инструмента на основе его спектральных особенностей с применением нейронных сетей.

Человеческий мозг способен распознавать различные визуальные и звуковые образы, в том числе и отличать звучание одного музыкального инструмента от другого, даже на фоне звучания других инструментов. Звук представляет собой волну, распространяющуюся в среде и изменяющуюся во времени. Любая волна характеризуется спектром и амплитудой. Каждый музыкальный инструмент имеет свое неповторимое звучание, которое характеризуется набором обертонов содержащихся в звуке данного музыкального инструмента (рисунок 1).

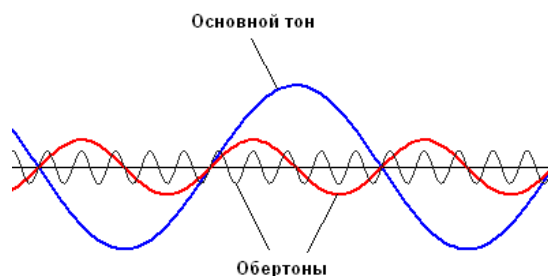


Рисунок 1. Частотный спектр звука музыкального инструмента [1]

В связи с этим для решения задачи распознавания тембра музыкального инструмента было предложено анализировать именно такие обертоны. Так как обертоны представляют собой дополнительные частоты, содержащиеся в звуке, то целесообразно извлекать информацию о таких обертонах из спектральной характеристики звукового сигнала с учетом амплитуды спектральных составляющих и использовать эту информацию для решения задачи распознавания. Однако спектр сигнала представляет собой большой набор данных, а именно все частоты, содержащиеся в сигнале, что является избыточным для поставленной задачи. В связи с этим видится целесообразным снизить размерность входного пространства посредством нелинейного преобразования в пространство признаков меньшей размерности. Для этой цели было предложено использовать мел-кепстральные коэффициенты (mel-frequency cepstral coefficients, MFCC) – набор коэффициентов, характеризующих

спектр сигнала. Данные коэффициенты успешно применяются в задачах распознавания речи.

На рисунке 2 изображен спектр звука трубы ноты *си* первой октавы.

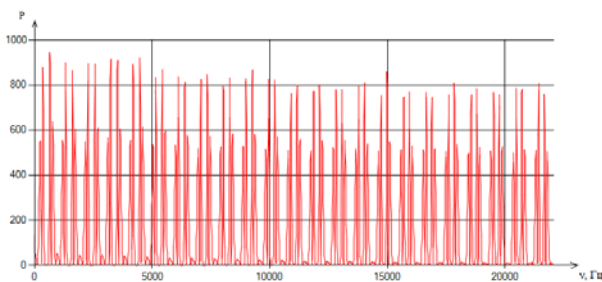


Рисунок 2. Спектр звука

На рисунке 3 представлены мел-кепстральные коэффициенты, полученные из спектра данного сигнала. Коэффициенты нормированы на 1.

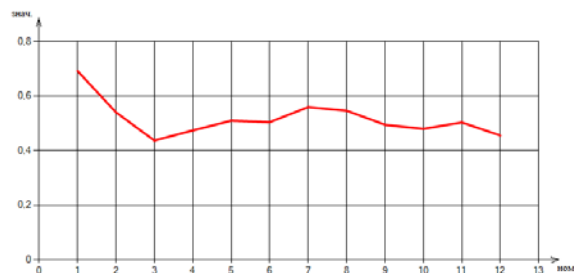


Рисунок 3. Мел-кепстральные коэффициенты

Для решения задачи распознавания тембра было предложено использовать нейронную сеть, так как нейронные сети способны к обобщению и показывают хорошие результаты на данных со сложными закономерностями, что для рассматриваемой задачи является несомненным преимуществом, в сравнении с другими классификаторами (например, с наивным байесовским классификатором).

Для построения нейронной сети была выбрана искусственная нейронная сеть прямого распространения. Иллюстрация такой сети приведена на рисунке 4. В данной задаче используется сеть с одним скрытым слоем, содержащим 12 нейронов.

В соответствии с выбранным подходом, алгоритм распознавания тембра музыкального инструмента можно описать следующий образом:

1. Разбиение входного сигнала на интервалы. Целесообразно разбивать сигнал на небольшие

промежутки от 20 до 100 мс, так как сигнал изменяется во времени и использование больших интервалов может привести к снижению качества распознавания. В данной работе был выбран сравнительно большой интервал ~ 100 мс (4096 отсчетов при частоте дискретизации 44 100 Гц), для ускорения скорости обработки и снижения числа входных данных, так как на текущем этапе данная работа носит исследовательский, нежели прикладной характер.

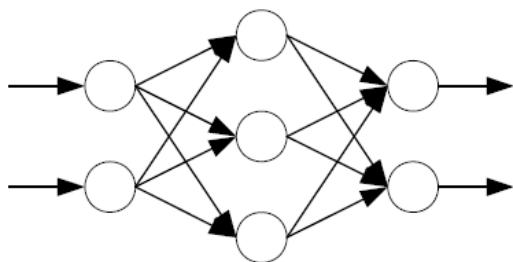


Рисунок 4. Искусственная нейронная сеть прямого распространения

2. Вычисление преобразования Фурье для каждого входного интервала. В данной работе было использовано быстрое преобразование Фурье с применением оконной функции Хэмминга.

3. Вычисление MFCC. Вычисление коэффициентов происходит на основе спектральной плотности мощности сигнала (power spectrum). При этом рассматривается тот интервал, где находится интересующие нас частоты. Для музыкальных инструментов целесообразно рассматривать интервал соответствующий частотному диапазону рояля, а именно от 27,5 Гц (нота *ля* субконтроктавы) до 4186 Гц (нота *до* пятой октавы), однако некоторые обертоны могут лежать за пределами данного интервала частот, поэтому необходимо добавить запас к указанным частотам. Рассматриваемый интервал разбивается на $n + 2$ частей, где n – число вычисляемых коэффициентов, причем интервалы имеет разный размер, то есть плотность выше там, где выше плотность данных. Затем на каждом интервале вычисляется суммарная мощность сигнала, попадающая в треугольную область, основание треугольника лежит в соседних точках разбиения на интервалы, например в точках, n и $n + 1$. Число коэффициентов n может варьироваться. Для рассматриваемой задачи было решено использовать 12 коэффициентов, именно это число применяется в задачах распознавания речи.

Более подробное описание процедуры вычисления MFCC приведено в [2].

4. Обучение нейронной сети на основе полученных коэффициентов. Для обучения нейронной сети был использован метод градиентного спуска, а именно метод обратного распространения ошибки (backpropagation). Подробное описание метода может быть найдено в [3].

Для выполнения данной работы был использован два набора инструментов. Первый набор состоял из трех инструментов — фортепиано, труба и скрипка. Во второй набор были дополнительно добавлены еще два инструмента — флейта и кларнет.

Для обучения нейронной сети для каждого инструмента были выбраны основные ноты (ноты гаммы *до*-мажор) из трех октав: малая, первая и вторая. В качестве критерия остановки обучения нейронной сети использовался набор дополнительных нот (диезы и бемоли) из указанных октав, которые не участвовали в процесс обучения, а именно ноты *ре* бемоль и *си* бемоль.

Для оценки сети после обучения использовались ноты *ми* бемоль, *соль* бемоль и *ля* бемоль, которые так же не принимали участие в процессе обучения. В результате удалось достичь точности около 90% для трех инструментов и около 80% для пяти инструментов.

В ходе исследования были решены следующие задачи: изучены характеристики звуков музыкальных инструментов, определен значимый набор признаков, выбран классификатор, подготовлены данные и произведено распознавание тембра музыкальных инструментов. Данное решения было реализовано в виде программного модуля на языке C#.

Список использованных источников

1. Ф.В. Станкевич Использование мел-кепстральных коэффициентов для нейросетового распознавания тембра музыкального инструмента. Материалы X международной научно-практической конференции «Ключевые аспекты научной деятельности - 2014». Наука и образование, Польша, 2014 г.

2. Mel Frequency Cepstral Coefficient (MFCC) tutorial. Practical Cryptography. URL: <http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs> Access Date: 01.01.2014 г.

3. В.Г. Спицын, Ю.Р. Цой. Интеллектуальные системы. Томский политехнический университет, 2012 г.