

МУЗЫКАЛЬНАЯ ТРАНСКРИПЦИЯ ПРИ ПОМОЩИ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Андрадзе А.И.

Насуро Е. В. - к.т.н., доцент

Транскрибирование – распознавание нот в цифровых музыкальных аудиозаписях. Исследуется полифоническая запись фортепиано, т.е. запись, на которой воспроизводится множество нот одновременно. Необходимо определить время воспроизведения каждой отдельно сыгранной ноты и ее длительность. На данный момент нету единого оптимального подхода для решения этой задачи, а существует некоторое их кол-во, каждый из которых имеет как свои индивидуальные сильные и слабые стороны.

Звук — распространение в виде упругих волн механических колебаний в твёрдой, жидкой или газообразной среде. Обычный человек способен слышать звуковые колебания в диапазоне частот от 16 Гц до 20 кГц. В музыке звук обычно получается колебаниями частей инструмента, исключение — духовые инструменты. Звукам, извлекаемым из фортепиано, в итоге соответствуют частоты от приблизительно 16 Гц до 15800 Гц. Фундаментальная частота - самая низкая частота ноты, это также частота, соответствующая высоте, которая идентифицируется человеком слушатель. Все гармоники кратны фундаментальной частоте.

Чтобы получить представление о частотах необходимо перевести временное представление композиции в частотно-временное.

Рассмотренные существующие решения осуществляют переход к частотно-временному представлению при помощи STFT (Short-time Fourier transform), получают спектрограмму звукозаписи. Это позволяет работать с отдельными частотными компонентами звука. Ось x спектрограммы – время, y – частота, цвет – интенсивность определенной частоты в данный момент времени. Для получения спектрограммы сигнал разделяется на временные сегменты с перекрытием, выполняется на каждом сегменте FFT и далее совмещаются все FFT сегментов на одном графике. Недостаток STFT в сложности нахождения компромисса между временным и частотным разрешениями. Вместе с тем, при использовании алгоритмов быстрого преобразования Фурье невозможно произвольным образом выбирать частоты его компонент. Это создает неудобства при дальнейшей обработке, поскольку невозможно точно определить количество звуковой энергии, приходящейся на частоты, соответствующие ступеням звукоряда. При фиксированном размере окна FFT, информация, используемая для расчета FFT ограничена. Если выбрать большое окно, то получим очень хорошее частотное разрешение, но это приведет к перекрытию нот. С другой стороны, выбирая малое окно, достигаем хорошего разделения нот, но частотное представление нот хуже, что, возможно, уменьшит способность разделять ноты.

В преобразовании постоянного качества (constant-Q преобразование) в отличие от преобразования Фурье, размер анализируемого фрагмента и размер оконной функции зависят от номера соответствующей частотной компоненты. Достоинством этого преобразования является легкость дальнейшей работы со спектром, поскольку его компоненты напрямую соответствуют ступеням звукоряда. Недостатками являются большая сложность вычислений и зависимость от правильного определения частоты настройки.

Гребёнка фильтров (filter bank). В цифровой обработке сигналов любое преобразование сигнала называют фильтром. В данном случае под фильтром понимается полосовой фильтр – преобразование, сохраняющее в звуке только частоты, находящиеся в некоторой полосе частот. Быстрое преобразование Фурье эквивалентно вполне определенной гребёнке достаточно грубых фильтров. Вместо них можно использовать любые другие фильтры, у каждого из которых центр полосы пропускания соответствует частоте одной из ступеней звукоряда, а ширина полосы пропускания достаточно мала, чтобы не охватывать частоты соседних ступеней. Эти фильтры можно подобрать так, чтобы они были менее грубыми, то есть более точно определяли количество звуковой энергии, приходящейся на их полосы пропускания. Недостатком данного метода является большая вычислительная сложность в сравнении с алгоритмом быстрого преобразования Фурье.

Предлагаемое улучшение – использовать вейвлет-преобразование.

Вейвлет-преобразование Вейвлет-преобразование (WT) – еще одна техника для перевода сигнала из временного представления в частотно-временное. WT является альтернативой STFT и позволяет преодолеть проблемы связанные с неопределенностью частотного и временного разрешений. STFT представляет одинаковое временное разрешение для всех частот, в то время как DWT предоставляет высокое временное разрешение и низкое частотное для высоких частот и высокое частотное разрешение и низкое временное для низких частот. Человеческое ухо демонстрирует аналогичные частотно-временные характеристики разрешения.

Список использованных источников:

1. George Tzanetakis, Georg Essl, Perry Cook, Audio Analysis using the Discrete Wavelet Transform
2. Ross Kidson, Musical Note Identification Violin Score Reproduction from Audio Recordings
3. E. Poliner, P. W. Ellis — A Discriminative Model for Polyphonic Piano Transcription (2006)
4. M. Marolt — A Connectionist Approach to Automatic Transcription of Polyphonic Piano Music (2004)