

ОТЗЫВ

официального оппонента

на диссертацию Глазырина Николая Юрьевича, представленную на соискание ученой степени кандидата физико-математических наук по специальности 05.13.18 – математическое моделирование, численные методы и комплексы программ на тему: «Алгоритмическое распознавание аккордов в цифровом звуке»

Актуальность темы диссертационной работы

На современном этапе развития компьютерной техники стало возможным внедрение методов точных наук во многие гуманитарные области, в частности, в музыку. В практической музыкальной деятельности все шире применяются современные технологии цифровой обработки сигналов. Многократно возросшая производительность компьютеров, в том числе персональных, позволяет теперь ставить задачи и разрабатывать алгоритмы обработки акустических сигналов, которые еще сравнительно недавно были нереализуемы из-за слишком больших вычислительных затрат.

Представленная работа направлена на решение задач, связанных с анализом цифровых фонограмм с целью восстановления текста композиции, что дает возможность последующей ее идентификации, обнаружение заданной композиции в фонограммах, выявления фонограмм с разными версиями аранжировки композиции, а также ее последующего воспроизведения на музыкальном инструменте. Такая система распознавания музыкального текста, работающая в реальном времени, может использоваться также в учебном процессе — она позволит контролировать процесс исполнения и немедленно фиксировать ошибки.

Данная диссертация посвящена определению последовательности аккордов в звуке. Для достижения поставленной цели решались следующие задачи. Разрабатывался метод более точного выделения в звуке компонент, соответствующих музыкальным инструментам; исследовалась применимость некоторых методов на основе моделей нейронных сетей; была предпринята попытка улучшить алгоритм определения аккорда на основе сопоставления с их шаблонами. Разработанные алгоритмы реализованы в виде комплекса программ языках Java и Python, позволяющих распознавать последовательность аккордов в фонограмме; качество распознавания аккордов сравнивалось с зарубежными аналогами путем участия автора в официальном соревновании программ в международном научном объединении MIR (на коллекции из более чем 300 звукозаписей).

Научная новизна исследования и значимость полученных результатов

В представленной работе получены новые результаты, касающиеся распознавания последовательности аккордов в звукозаписи. В частности, был предложен и реализован новый метод распознавания последовательности аккордов в звукозаписи, не

использующий алгоритмы машинного обучения; его важным качеством является отсутствие опасности его переобучения под конкретную музыкальную коллекцию. Разработан также новый метод представления результатов анализа звукозаписи в виде последовательности векторов признаков с применением многослойной нейронной сети.

Обоснованность научных положений и достоверность результатов

Выносимые на защиту научные положения соответствуют результатам, полученным в диссертации. Достоверность результатов подтверждается как теоретическим обоснованием, так и экспериментальной проверкой.

Практическая ценность проведенных исследований

Полученные при выполнении диссертационной работы результаты могут применяться для анализа звукозаписей с целью их звуковысотной расшифровки, т. е. восстановления исходного нотного текста — как с целью последующего исполнения, так и с целью поиска схожих музыкальных композиций, а также в процессе обучения музыке.

Содержание работы

Диссертационная работа (всего — 88 страниц) состоит из введения, пяти глав, заключения и списка литературы. Список литературы содержит 100 наименований.

Во введении обосновывается актуальность темы диссертации, формулируется цель и основные задачи, показаны научная новизна и практическая значимость работы, а также сформулированы научные положения, выносимые на защиту.

В первой главе приводятся теоретические сведения о звуке и его свойствах, о музыкальном строе и составе аккордов, необходимые для более конкретной постановки задач и для дальнейших построений.

Во второй главе приводится весьма обширный обзор научной литературы по теме исследований, обсуждаются и классифицируются известные методы решения аналогичных задач и подходы, используемые в разных работах для каждого из этапов преобразования анализируемой фонограммы в условную запись последовательности аккордов.

В третьей главе описывается реализованный в работе метод распознавания аккордов, не использующий алгоритмы машинного обучения. Метод может быть условно разделен на четыре этапа преобразований, на каждом из которых использованы новые алгоритмы. В частности, отдельно рассмотрена операция определения ритма, которая позволяет делать измерение спектральных составляющих в фазе стационарного звучания инструмента, что повышает точность распознавания. Для определения ритма в данной работе использовались внешние библиотеки. Для выбранных на основе ритмической структуры моментов времени вычислялось т. н. Фурье-преобразование «постоянного качества», в котором использовалась оконная функция Хэмминга. В результате для

каждого момента времени вычисляются спектральные столбцы, используемые далее для определения аккорда; учитывается также возможность отличия настройки инструментов от точной «стандартной» (ля первой октавы — 440 Гц).

Далее к спектрограмме применяется серия преобразований для лучшего выделения компонент, которые несут информацию, важную для идентификации аккорда. В частности, используется двухмерная фильтрация спектрограммы, рассматриваемой как изображение, для подавления вертикальных полос, являющимися следами звука ударных инструментов. Используется также наличие повторов в исполнении: формируется матрица самоподобия, позволяющая повысить качество распознавания за счет этого фактора.

В четвёртой главе рассматривается метод получения тональных признаков при помощи многослойной нейронной сети. При таком подходе не используется большая часть из описанных выше преобразований спектра. Вместо этого каждый из столбцов спектрограммы подаётся на вход нейронной сети, на выходах которой получается вектор признаков. Обучение сети производится методом мини-пакетного стохастического градиентного спуска, при этом в качестве целевых векторов используются 12-мерные бинарные шаблоны аккордов. Здесь отмечается наличие проблемы, связанной со статистически неравномерным использованием в исполнении разных аккордов и рассматриваются возможности преодоления этой ситуации при обучении.

Пятая глава посвящена экспериментам, нацеленным на подбор оптимальных параметров методов и определение степени влияния каждого из шагов предлагаемых методов. Описываются используемые метрики качества распознавания аккордов, в соответствии с которыми возможно сравнивать разные методы или разные варианты одного метода. Для проверки наличия статистически значимой разницы между разными вариантами метода предлагается использовать непараметрический критерий Фридмана. Приводятся некоторые количественные характеристики тестовой коллекции. Классифицируются основные типы ошибок, что позволяет более точно определить влияние различных параметров метода на качество распознавания аккордов. Далее поочередно рассматривается эффект от выбора различных значений параметров и различных алгоритмов на промежуточных этапах. Вычисляются значения метрик на всей коллекции, приводятся диаграммы ошибок.

По результатам экспериментов сделаны выводы, позволяющие оптимальным образом сформировать совокупность параметров методов. В частности, подтверждена целесообразность реализованных в работе методов повышение разрешения спектрограммы по времени и задержки относительно моментов начала метрических долей.

Эффект от определения частоты настройки, напротив, оказался не очень существенным, как и использование свёртки для очистки спектрограммы. Использование факта самоподобия в исполнении, как показали эксперименты, позволяет существенно улучшить

результат.

Использование нейронной сети для получения хроматических признаков оказалось не более выигрышным, чем применение методов непосредственного анализа спектра (без обучения), описанного в третьей главе. Работа на основе такой сети оказывается также в целом более «медленной».

Было показано, что реализованный в рамках работы метод позволяет добиться качества распознавания, сравнимого с наилучшими существующими в мире аналогами, а именно, правильно распознаются аккорды в среднем для 77,5% от продолжительности композиции (при использовании только мажорных и минорных трезвучий). Этот результат получен для используемого в международном сообществе ISMIR тестового набора из 318 композиций (записи известных эстрадных групп).

В заключении диссертации сформулированы основные научные и практические результаты диссертационной работы.

Основные замечания по работе

По диссертационной работе имеется также ряд замечаний. Часть из них относится к области музыки, где хотелось бы видеть более четкое использование музыкальной терминологии. Другая часть связана с акустикой музыкального звука, где в изложении также встречаются неясности.

В заголовке работы значится «распознавание аккордов», однако на самом деле распознается только тип аккорда и тональность, но не высотное положение по октавам. Таким образом, нотный текст не восстанавливается полностью, что с точки зрения применимости, например, в процессе обучения создает заметные ограничения.

Кроме того, вся разработка ориентирована на 12-полутоновую европейскую звуковысотную систему, в связи с чем, например, в автореферате утверждается: «...музыкальный звук является совокупностью отдельных звуков, соответствующих нотам». Однако понятие музыки охватывает гораздо больший круг явлений и, в частности, традиционные культуры с совершенно другим музыкальным строем. Вопрос о применимости разрабатываемого аппарата к таким культурам, по-видимому, был бы интересен музыкантам-исследователям, однако он даже не затрагивается в тексте.

Далее (С. 17) со ссылкой на источник [63] утверждается, что «...музыка основана на повторениях», а на (С. 36) — «Музыка нравится человеку в том числе из-за повторений одного и того же мотива в разных вариациях, с некоторыми изменениями» — такие упрощенные формулировки вызывают возражения и к тому же выходят за рамки темы. Используемое далее (С. 37) для сглаживания спектрограммы свойство самоподобия фонограммы опирается, видимо, не на восприятие человека, а на распространенное в музыке явление повтора, но обычно — частичного, вариативного. (Как учитывается в этом случае вариативность, остается не слишком понятным.) Кстати, положительный (судя по тексту) результат такого сглаживания спектрограммы (рис. 3.5) визуально

неочевиден а пояснения по поводу требуемого (или желаемого) качества спектрограммы отсутствуют.

На С. 16 сказано, что «В звучании таких инструментов (с ясно выраженной высотой звучания — А.Х.) можно выделить отдельную ноту. При этом наряду с частотой, соответствующей этой основной ноте, звучат другие частоты. Их звучание менее выражено, но они могут соответствовать другим ступеням музыкальной системы». Из последующих пояснений видно, что здесь имеется в виду наличие в тональном звуке колебания основного тона с амплитудой, заведомо большей, чем колебания других гармоник — однако следует отметить, что это далеко не всегда выполняется.

Например, в звуке скрипки 1-я гармоника исходно слабее 2-й; кроме того, при работе с фонограммой никогда нельзя быть уверенным, что при ее записи не были подавлены низкочастотные составляющие спектра, включая основной тон каких-то звуков — и тогда ожидаемое соотношение нарушится. В целом с ростом номеров гармоники амплитуды составляющих убывают, но у всех инструментов есть в спектре «поднятия»-форманты, которые локально нарушают эту обобщенную тенденцию.

При обсуждении соотношений амплитуд (в составляющих «шаблонных векторов») возникают также терминологические казусы. Так, на С. 25 и С. 65 сказано: «Поскольку первый обертон любой ступени звукоряда соответствует звуку с таким же названием,...». «Обертон» — это частичный тон звука, лежащий *выше* основного тона; соответственно «первый обертон» — это *вторая* гармоника, октавный частичный тон. В этом месте автор ссылается на источник [84], в котором обсуждается выбор нужного числа учитываемых обертонов, но упомянутые авторы однозначно употребляют термин «гармоника», что вполне правильно.

На С. 19 обсуждается возможность экономии количества вычислений путем понижения частоты дискретизации «со стандартной для компакт-дисков 44100 Гц до 11025 Гц», причем предлагается делать это «заменой каждых 4 подряд идущих отсчетов...» на первый из них, т. е. просто «выбрасывать» из фонограммы по три отсчета подряд, оставляя каждый четвертый. Указано, что «при этом теряется информация о частотах выше 5.5 кГц», что в данном случае не критично (речь идет об определении звуковысотного состава звука), что верно.

Однако следует заметить, что при такой операции теряется соответствие теореме Котельникова — необходимо *предварительное подавление* всех высокочастотных (выше 5,5 кГц) составляющих спектра, которое обязательно выполняется в любой фирменной программе (но в тексте оно не упоминается). Эта дополнительная фильтрация тоже требует времени, и его следовало бы учесть при расчете временного выигрыша от снижения частоты дискретизации.

При описании способа повышения разрешения спектрограммы по времени (С. 32, С. 55) для ее последующего сглаживания предлагается следующий метод: «между

каждыми двумя соседними моментами начала метрических долей... вставляются (Т-1) промежуточных значений».

Однако остается неясным, что именно вставляется: искусственно созданные спектральные столбцы или отсчеты сигнала? Алгоритм формирования этих дополнительных отсчетов не описан, как и не разъяснен вопрос о том, почему такое внесение в исходные данные искусственно созданных фрагментов не искажает последующий результат (ведь они попадут в качестве исходных данных и в операцию сглаживания!).

При описании результатов численных экспериментов в начале (С. 45) приводится перечень параметров, при которых производились расчеты, однако используются только их буквенные обозначения, введенные существенно раньше, так что вспомнить их все в этом месте текста уже затруднительно (можно было бы снабдить их пояснениями).

Сами результаты представлены графическими иллюстрациями и таблицами, причем комментариев к графическим изображениям, как правило, нет. Что отложено по осям на рис.5.2. и последующих, в частности, по вертикальной оси (там числа от 8000 до 20000)? Число ошибок (из какого общего)? Что такое графики а) и б)?

Аналогичный вопрос возникает на С. 69 (результаты испытаний на данных MIREX): — Что по горизонтальной оси? Используемые там обозначения встречаются, кажется, впервые. Что означают 2 кривые на всех графиках? — Хотелось бы иметь здесь гораздо более полные пояснения — как и сопоставление с работой программ других авторов (в заключении сказано, что полученные результаты не хуже других, однако прямое сравнение отсутствует). В тексте имеется также указание на «доступность через Интернет» разработанных программ, однако ссылки отсутствуют.

Общая оценка диссертации

Несмотря на отмеченные недостатки, в целом работа выполнена на хорошем научно-техническом уровне. Многие из сделанных замечаний носят терминологический характер и могут быть исправлены путем редактирования при дальнейших публикациях, но на суть и результаты работы они не влияют. Основные результаты диссертационной работы получены самостоятельно, практические результаты и теоретические положения опубликованы в 6 научных работах (в том числе 2 — в изданиях, рекомендованных ВАК); они соответствуют содержанию диссертации. Результаты разработки прошли также апробацию на ряде научных конференций.

Диссертация написана в целом хорошим языком, материал изложен логически стройно, аргументировано, содержание автореферата соответствует основным положениям диссертации.

Заключение

По качеству решения поставленных задач, структуре, содержанию и полноте представления материала диссертация представляет собой завершенную научную работу, имеющую научную и практическую ценность, соответствует требованиям п. 9 Положения о присуждении ученых степеней ВАК РФ, предъявляемым к кандидатским диссертациям, а ее автор, Глазырин Николай Юрьевич, заслуживает присуждения ученой степени кандидата физико-математических наук по специальности 05.13.18 — математическое моделирование, численные методы и комплексы программ.

29 января 2015 года

Официальный оппонент

кандидат технических наук, доцент, заведующий кафедрой
музыкально-информационных технологий

 Харуто Александр Витальевич

Федеральное государственное бюджетное образовательное учреждение высшего
профессионального образования (университет)

«Московская государственная консерватория имени П. И. Чайковского»

Адрес организации:

125009, Москва, ул. Большая Никитская, д. 13/6

Подпись официального оппонента

Харуто Александра Витальевича заверяю

Начальник отдела кадров

Московской государственной консерватории имени П. И. Чайковского



 Жмуров А.В.