

УДК 004.424.4

**А.Ю.Белобородов**

# Распознавание аудиообразов с применением обертонового ряда

*Рассмотрены возможности использования методов распознавания аудиообразов для решения задач поиска. Предложен алгоритм распознавания нот по спектру сигнала.*

**Ключевые слова:** поиск, распознавание аудио, спектр звука, спектральный анализ, обертоновый ряд.

Появление различных музыкально-поисковых сервисов в сети Интернет свидетельствует о востребованности систем подобного рода. Широкое распространение получили сервисы, позволяющие искать аудиозаписи по имени файла, описанию, или же тега, ассоциируемым с музыкальной композицией. Такой вариант решения поисковой задачи приводит к появлению ряда новых проблем. Результатом поиска должно быть отображение списка релевантных запросу музыкальных композиций. Поиск аудиозаписи по текстовой информации требует от пользователя знания всех полей, однозначно идентифицирующих композицию (композитор, исполнитель, альбом и прочее). В базе данных, в свою очередь должны содержаться достоверные данные о каждом файле. На практике же, не редко встречаются и обратные задачи. Например, узнать исполнителя имеющегося аудиофрагмента. В подобных случаях на помощь приходят системы распознавания и анализа аудиоданных. Не смотря на то, что подавляющая масса аудиоданных хранится в оцифрованном виде, проблема распознавания и по сей день остаётся актуальной и обсуждаемой, о чём свидетельствуют многочисленные научные публикации по данной тематике.

## Цель работы

Целью данной работы является разработка алгоритмов распознавания аудиоданных по спектру сигнала для организации системы поиска аудиофайлов по нотам. Поиск аудиофайлов по содержанию

Методы, применяемые для поиска аудиофайлов (аудиоданных в оцифрованном виде), зависят в первую очередь от способа их представления, т.е., по сути, формата файлов.

**Национальный аэрокосмический университет  
им. Н.Е.Жуковского «ХАИ»**

**кафедра компьютерных  
систем и сетей**

**Научный руководитель  
Волковой А.В.  
(к.т.н., доцент)**

Ярким примером такой зависимости может служить MIDI-формат.

MIDI (Musical Instrument Digital Interface) – цифровой интерфейс музыкальных инструментов. Особенность MIDI-формата состоит в том, что аудиоинформация хранится в виде нот. Следовательно, если файл содержит в себе ноты, то можно разработать алгоритм поиска по нотам, что и было выполнено в предыдущей работе [1]. Соответственно, распознавания в данном случае не происходило.

Распознавание образов – это отнесение исходных данных к определенному классу с помощью выделения существенных признаков, характеризующих эти данные, из общей массы несущественных данных [2].

Поиск файлов по содержанию включает в себя последовательное сравнение исходного фрагмента и фрагмента файла. Сравнение может быть выполнено различными способами: побайтно, при помощи расчёта корреляции сигналов, с применением хеш-функции или другими методами. При этом можно условно выделить уровни, на котором происходит сравнение. Побайтный способ можно принять за самый низкий уровень, т.к. сравнение производится для последовательностей байт файла и ключевого образца. Поиск будет успешен только тогда, когда ключевой образец является фрагментом рассматриваемого файла. Подобно работает способ, основанный на применении хеш-функций. Сопоставление фрагментов файла и ключевого образца при помощи расчёта корреляции сигналов назовём вторым уровнем сравнения. Этот способ является более предпочтительным, чем первый, т.к. качество оцифровки сигнала менее сказывается на результа-

тах поиска. Если в побайтовом сравнении файл и искомый образец будут иметь различный битрейт (показатель скорости передачи и обработки цифровых данных) или кодируемое число уровней [3], то результат поиска может существенно отличаться от желаемого. Третьим, или в нашем случае верхним, уровнем будет сравнение аудиофайла и исходного фрагмента по нотам. Такой подход позволит использовать те же алгоритмы поиска, которые были применены для MIDI-файлов [1]. Основной трудностью здесь является преобразование, которое необходимо выполнить. Фактически, это преобразование дискретного аудиосигнала в MIDI-данные. Данная операция включает в себя распознавание нот из аудиоданных.

#### Обертоновый ряд и спектр звука

Гармонический ряд (или обертоновый ряд) – звукоряд, возникающий в результате спонтанного деления струны на равные части во время колебания [8].

Спектр звука – частотное представление состава звука. Спектр звука представляют обычно на координатной плоскости, где по оси абсцисс отложена частота, по оси ординат – амплитуда или интенсивность каждой гармонической составляющей звука.

Количество частей, на которое спонтанно делится струна, в общем случае, не ограничено. Источник звука, колеблющийся целиком, производит основную частоту, наиболее слышимый звук, кажущийся единственным. Вторые части (половины) производят звук с частотой вдвое большей, чем основная; третьи части – втрое, четвёртые – вчетверо большей, чем основная, и т. д. В результате таких сложных одновременных колебаний и возникает звук сложного состава [0]. На рисунке 1 приведён

пример спектра звука, получаемого при исполнении на трубе звука «до» второй октавы (525.478 Гц). Необходимо отметить, что под «колеблющейся струной» здесь и далее понимается источник звука музыкального инструмента. Для скрипки, рояля – это струна, для голоса – голосовые связки и т.д.

#### Применение обертонового ряда в задаче распознавания аудиоданных

Существуют различные подходы реализации спектрального анализа. Для распознавания невозможно использовать все обертоны, т.к. их бесконечное множество. Источник звука может спонтанно делиться на множество частей, что и приведёт к появлению множества гармоник в его спектре. Поэтому на практике используют не все обертоны, а только их часть. Рисунок 1 свидетельствует о том, что каждый обертон имеет свою определённую амплитуду (характеристика амплитуд гармонических составляющих формирует тембр), а амплитуда верхних обертонов рассматриваемого спектра весьма мала по сравнению с амплитудой основного тона (самая нижняя гармоника, на рисунке 1 – слева).

Для распознавания могут использоваться все подряд следующие обертоны, либо определённая выборка.

Распознавание по алгоритму опорных обертонов (далее – АОО) предусматривает суммирование значений амплитуд для «опорных» обертонов. Т.к. обертоны 1, 2, 4, 8 и 16 имеют одинаковые названия (на рисунке 1 это нота «до» первой, второй, третьей и четвёртой октав; отмечены красными треугольниками), то их можно принять за основу. Такую операцию необходимо провести над всеми звуками, которые необходимо искать в спектре.

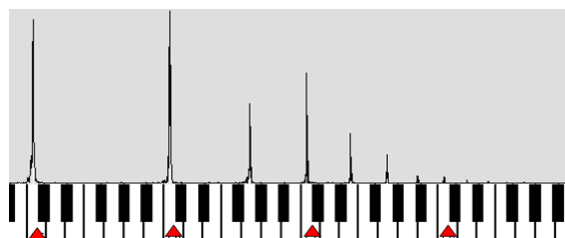


Рис. 1 Спектр звука и "опорные" обертоны

Описание алгоритма.

- 1) Установить пороговое значение амплитуды.
- 2) Найти в спектре гармоники, значение амплитуды которых больше, чем установленное пороговое в п.1. Ноты соответствующие

найденным гармоникам будут проверяться в спектре.

- 3) Выполнить расчёт для каждой ноты из п.2: вычислить сумму уровней амплитуд для всех одноимённых тонов (опорных обертонов).

4) Найти сумму с максимальным значением. Нота, для которого эта сумма была вычислена и будет основной нотой данного спектра.  
Примеры, иллюстрирующие работу алгоритма при звучании духового оркестра отображены на

рисунках 2 и 3 (оркестр играет «до мажорное» трезвучие с нотой «соль» в мелодическом голосе). Вертикальными линиями отмечено положение «опорных» обертонов.

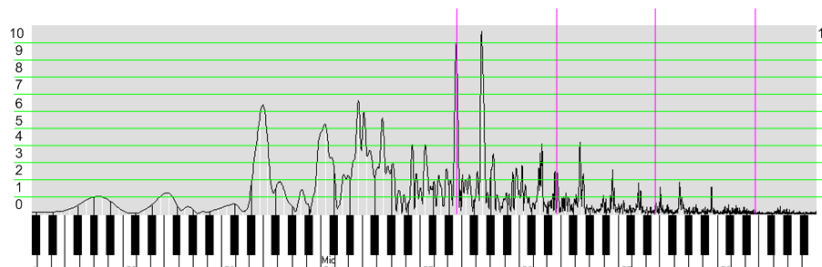


Рис. 2 – Анализ спектра по АОО от ноты «ми второй октавы»

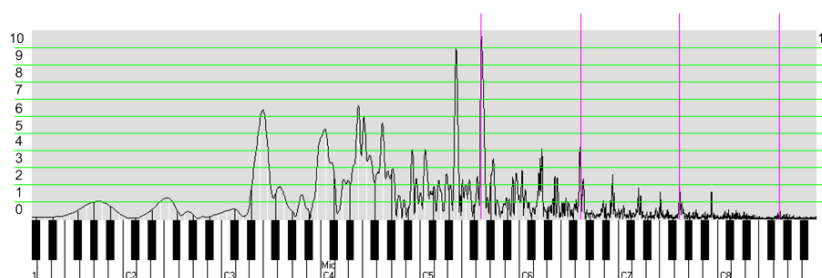


Рис. 3 – Анализ спектра по АОО от ноты «соль малой октавы»

C5 соответствует «до второй октавы».

1) Пороговое значение амплитуды равно 7.

2) Гармоники, амплитуда, которых выше, чем 7: «ми второй октавы» и «соль второй октавы».

3) Сумма по «ми второй октавы» =  $8 + 2 + 0 + 0 = 10$ ;

Сумма по «соль второй октавы» =  $10 + 4 + 1 + 0 = 16$ ;

4) Максимальное значение равно 16, которое было обнаружено для ноты «соль второй октавы».

Второй алгоритм действует подобно первому, за исключением того, что в распознавании используется 16 подряд следующих обертонов.

### Проблемы распознавания

При анализе нестационарных сигналов применяется кратковременное преобразование Фурье. Полный временной сигнал разделяется на подинтервалы – временные окна, и преобразование проводится для каждого окна в отдельности [4]. Соответственно, относительно длительности окна и ноты, которая попала во временной интервал, можно различать: короткое, большое и эквивалентное окна Фурье.

Выбор короткого окна Фурье приводит к тому, что ноты, длиннее, чем выбранный временной интервал, будут раздроблены на более мелкие. Этот случай иллюстрирует рисунок 4. Выбор большого окна Фурье тоже приводит к ошибкам распознавания. Данный случай иллюстрируют рисунки 5-8.



Рис. 4 - Дробление ноты при выборе короткого окна

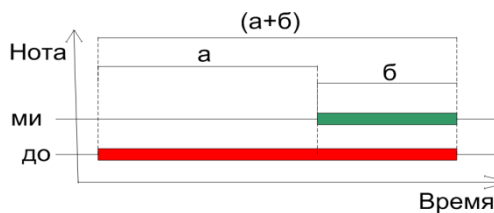


Рис. 5 – Большое окно Фурье. Исходные данные

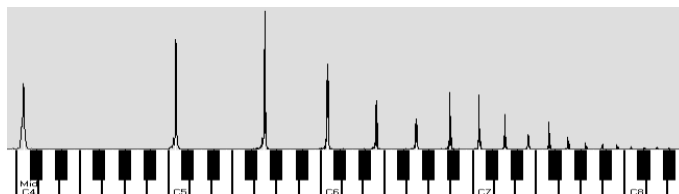


Рис. 6 –Спектр участка "а" (нота «до»)

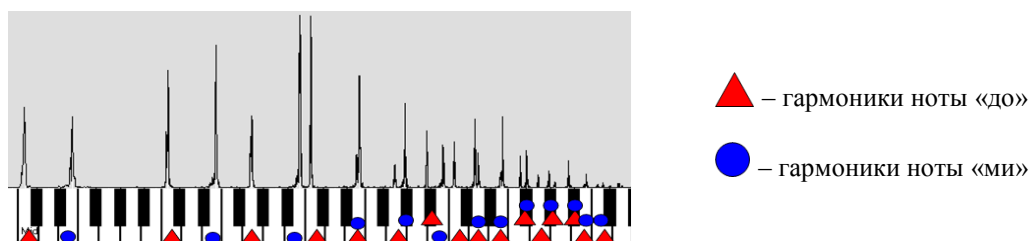


Рис. 7 –Спектр участка "б"

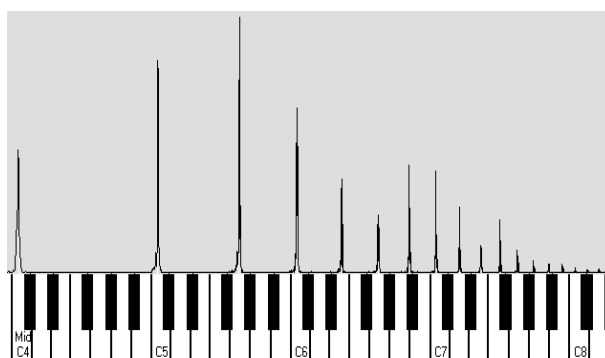


Рис. 8 –Спектр участка "(а+б)"

Рис. 8 иллюстрирует потерю гармоник ноты «ми», что вызвано использованием большого окна Фурье.

Из вышеизложенного следует сделать вывод, что окно Фурье должно иметь переменный размер при распознавании и совпадать с началом и концом звучания каждой ноты, которую распознаём. Данный вопрос требует отдельного исследования.

### Пути решения проблем распознавания

Применение методов распознавания нот, основанных на оконном преобразовании Фурье, как мы уже видели выше, имеет проблемы, связанные с особенностью анализа нестационарных сигналов. Для решения таких проблем предлагаются следующие подходы:

1) Применение окон Фурье переменной длины. Для определения размера окна, возможно, потребуется предварительный анализ сигнала.

2) Использование ограничения по частоте или тембру при использовании окон.

3) Применение дискретных вейвлет-преобразований, которые позволяют проверить наличие гармоник на заданном отрезке времени [6].

4) Использование «словарей музыкальных фраз». Подобно как программы распознавания речи и текста опираются на свой внутренний словарь [5], так же можно поступить и с музыкальными композициями. Некоторые ритмо-мелодические рисунки невозможно исполнить по техническим причинам, если речь идёт о записи реального, не синтезированного электронным способом, инструмента. В ряде случаев мелодия движется плавно (естественно, в данном случае подразумевается музыкальные композиции определённого жанра). Правила общепринятой гармонии и аранжировки музыкальных композиций накладывает определённые ограничения на движение мелодии и сопровождающих мелодию голосов. Эти особенности могут быть использованы для разработки алгоритма фильтрации случайных мелодических скачков, который уменьшит число ошибочных нот распознавания. Разработка такого алгоритма является отдельной задачей.

5) Применение нейронной сети для нахождения границ переходов между нотами или аккордами [7], а также и как механизм распознавания.

#### **Выводы**

Таким образом, в ходе проведенных исследований, были разработаны алгоритмы распознавания нот в аудиоданных на основании обертонового ряда. Также проанализированы основные причины, влияющие на качество распознавания, и способы улучшения качества распознавания. Данные результаты позволят разработать систему поиска файлов различных аудиоформатов. Основным достоинством изложенного подхода является то, что результат поиска по нотам не зависит от тембра, темпа и тональности записанного аудиофрагмента в файле и ключевого образца, что достигается

благодаря применению разработанных алгоритмов поиска в MIDI-файлах [1].

Распознавание и анализ аудиоданных не ограничивается только задачей поиска музыкальных композиций. Спектральный анализ используется для определения причин крушения самолёта при флаттере. По гармоникам, входящим в состав механического колебания, можно понять, какие части самолёта колебались сильнее всего и вызвали (могут вызвать) разрушение конструкции.

#### **Литература**

1. Белобородов А.Ю. Система поиска MIDI-файлов по заданной мелодии/Тезисы докладов Международной научно-практической конференции аспирантов и студентов «Инженерия программного обеспечения 2009», Национальный авиационный институт, Киев, 2009. – с. 38.

2. Статья [Электронный ресурс]: Теория распознавания образов. – Электрон. дан. – Режим доступа:

[http://ru.wikipedia.org/wiki/Теория\\_распознавания\\_образов](http://ru.wikipedia.org/wiki/Теория_распознавания_образов)

3. Статья [Электронный ресурс]: Цифровой звук. – Электрон. дан. – Режим доступа:

<http://www.3dnews.ru/multimedia/sound/>

4. Лекции по цифровой обработке сигналов [Электронный ресурс]: Тема 3d: Оконное преобразование Фурье – Электрон. дан. – Режим доступа:

<http://prodav.narod.ru/dsp/app/dsp.zip>

5. Bernd Planer. An Introduction to Speech Recognition. – Germany, 2005 – 68 pages.

6. Бохан К.О., Кучук Г.А. Методи цифрової обробки сигналів: Навчальний посібник для студентів комп'ютерних спеціальностей вищих навчальних закладів. – Харків: Національний аерокосмічний університет ім. М.Є. Жуковського „Харківський авіаційний інститут”, 2008 – 84 с.

7. Грэхем Иан. Объектно-ориентированные методы. Принципы и практика. 3-е издание. – Москва: Издательский дом «Вильямс», 2004 – 880 с.

8. Юцевич Ю. Словарь музыкальных терминов. – Киев: Музычна Україна, 1988 – 262 с.

Способин И. В. Элементарная теория музыки. – Москва: Музгиз, 1959 – 202 с.