

Общедоступный сервис поиска музыки по ритмическому сходству

© Николай Бузикашвили

Отдел когнитивных и компьютерных технологий, Институт системного анализа
buzik@cs.isa.ru

Аннотация

Изложен проект системы музыкального поиска, основанного на ритмическом сходстве запроса и хранящихся в базе музыкальных произведений. Помимо собственно музыкально-звукового поиска, рассмотрена трактовка поиска музыкальной информации как многомерного поиска в мультимодальных документах.

1 Введение

Сегодня в электронном виде накоплены миллионы экземпляров музыкальных записей [3]. Хотя большинство этих записей не имеют отношения к профессиональной музыке, тем не менее, в открытых базах, в принципе, можно найти любые музыкальные произведения, включая разнообразные исполнительские интерпретации и ноты. В то же время де-факто это изобилие остается недоступным обычному пользователю. Причина состоит в отсутствии систем поиска музыки, пригодных для эффективного использования. Вопреки сложившемуся мнению, существующие системы не столь плохи, однако чтобы воспользоваться ими, пользователь должен обладать, как минимум, профессиональными навыками музицирования по памяти, что резко ограничивает круг реальных пользователей таких систем. “Музыкальный запрос”, в лучшем случае, нужно наиграть или напеть, а в худшем, ввести в своеобразной нотной записи, что очень непросто и для профессионального музыканта.

В статье обсуждается проект музыкально-поискового сервиса, основным описателем музыки в котором является ритм, а основным способом задания запроса — настукивание ритма, в частности, на клавиатуре.

2 Задача поиска музыкальных произведений

За последние несколько десятилетий формальный подход к музыке претерпел существенные изменения, причем, к удивлению (возмущению) самих их авторов, эти изменения утилитарны и целиком лежат в русле “индустриализации”. Сегодня важными считаются совсем не те задачи, которые были важны или интересны в период становления этого направления “искусственного интеллекта”. Например, если классическая работа [1] посвящена задачам синтеза музыки, прежде всего гармонического строения, то выполненная в тех же стенах работа [14] посвящена задачам распознавания и нотификации звучащей музыки. На смену популярной в общем контексте 60-х парадигме “машина сочиняет”, в рамках которой задача нотификации того, что изначально выражено в нотах, достаточно нелепа, пришла гораздо более практически-ориентированная парадигма.

Последние десятилетие в число основных задач, как в области поиска информации, так и в области математической музыкологии, входит поиск музыкальной (звучащей) информации. Однако, при всем обилии исследовательских работ, поиск музыкальной информации сегодня так и не стал сервисом специализированной промышленной поисковой системы или одним из сервисов массовых поисковых машин. Причина, на наш взгляд, состоит прежде всего не в несовершенстве механизмов поиска, а в том, что в силу самого своего построения, а именно по типу поддерживаемых запросов, музыкальный поиск оказался жестко ориентирован на музыкантов-профессионалов. Тем самым от его возможностей оказалась отсечена многократно большая аудитория потенциальных пользователей.

3 Музыка в сети

В электронном виде фиксируются либо (1) компьютерно-нотная запись музыкального произведения (MIDI-файлы и им подобные), либо (2) исходная аналоговая нотификация (отсканированные партитуры) или оцифрованная аналоговая аудиозапись (“WAV”-файлы, включая MP3). В случае (2) для обработки музыки с помощью

компьютера требуется сначала выполнить процедуру распознавания изображения партитуры [5] или распознавания звуков [14].

Музыкальную информацию, хранящуюся в электронных библиотеках, музыкальных репозиториях (специализированных хранилищах, например, в архиве лаборатории звукозаписи и акустики Московской консерватории) и рассеянную по Интернету, можно разделить на две группы по тому, содержит ли она явную “компьютерную нотификацию”. MIDI-файл (например, введенный с музыкальной клавиатуры) содержит однозначно интерпретируемую компьютером нотификацию¹. Напротив, в аудиозаписи, снятой, с динамиков персонального компьютера, на котором произведение исполнялось с MIDI-клавиатуры, нотификация утрачена, и для ее восстановления потребуется не всегда надежная процедура распознавания звучащей музыки. Точно также изображение партитуры должно быть распознано (распознавание “символов” нотной записи).

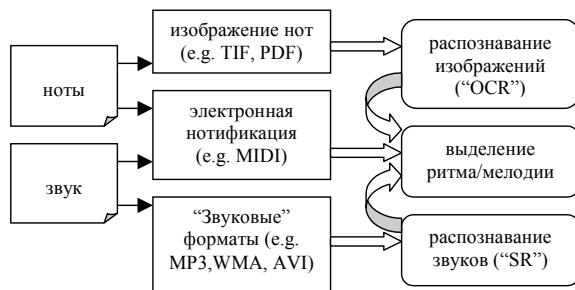


Рис.1. Источники музыкальной информации, типы их электронного представления и способы извлечения информации о ритме/мелодии

Из достаточно размытого корпуса “музыкальной информации” выделим звукозаписи — любое представление музыки, которое *актуально* допускает автоматическое звуковоспроизведение, и будем называть их *музыкой*. Соответственно, поиск звукозаписей музыкальных произведений будем называть не *поиском музыкальной информации*, а *поиском музыки (ПМ)*. (Обычно упоминаемая в контексте поиска музыки задача различения музыки и речи здесь не рассматривается даже вскользь и, более того, выглядит надуманной как в концептуальном плане (скрип гвоздя по стеклу,

¹ Согласно экстремистской точке зрения, пропагандируемой, например, А.Танганом, MIDI-нотификация безнадежно уступает традиционной и влечет необратимую утрату “музыкальной информации”. Хотя распечатка MIDI-файла в форме традиционной нотной записи имеет мало общего с исходной нотной записью произведения, а MIDI-язык ориентирован на клавишные, тем не менее, всякий (в т.ч. профессиональный музыкант), прослушивающий MIDI-файл, легко распознает произведение. Различия примерно те же, что между двумя разными исполнениями одного произведения.

горловое пение, стихи или ритмизованная проза — что здесь более “музыка”, а что “речь”?), так и в чисто практическом (вряд ли выбранный наугад MP3-файл окажется не “музыкой”).

“Проектирование и оценка существующих систем музыкального поиска в значительной мере основана на ни в коей мере не представительных (anecdotal) сведениях о потребностях пользователей, интуитивных воззрениях на их поведение и априорных соображениях о типичных сценариях использования” [9]. Системы ПМ изначально создавались “для себя” и не претендовали на массовое распространение. Оценка качества такой системы с позиций “человека с улицы” невозможна, поскольку он не в силах воспользоваться этой системой. В то же время априорное представление о наличии у человека с улицы требуемых музыкальных навыков совершенно адекватно: их нет.

4 Почему ритм?

Однако редко у кого хотя бы изредка не возникало желание послушать когда-то слышанное произведение. Практически каждый является потенциальным пользователем ПМ. Кто сегодня может реально использовать ПМ? То есть каковы возможности систем ПМ, прежде всего в плане пользовательского интерфейса?

Сегодня используются следующие формы ввода запросов: напевая (query-by-singing и более популярное название — query-by-humming) [8], наигрывая (query-by-example, хотя, конечно, напев — тоже задание образца) [6,7], набирая нотную запись (query-by-note) [11]. Понятно, что последний способ наиболее трудоемок даже для профессионала и хорошо покрывается первыми двумя способами (наигрышем и напевом). Пример запроса в квазинотной записи:

```
%F-4$bB@3/8#`8C.6.3$B`C&/,8A`D6(-)D/,8G`8.C,6B/8F
```

То есть предполагается, что пользователь не только может запомнить фрагмент произведения и выразить его в нотах, но еще и сделать это в столь специфической нотации. Такой запрос вводится с обычной клавиатуры ПК.

Для ввода наигранного или “напетого” запроса используются музыкальная клавиатура и микрофон. Чтобы воспользоваться такими возможностями поиска нужно 1) достаточно хорошо помнить искомое, 2) уметь выразить то, что помним.

Поиска собственно музыкальной информации может производиться по любому набору музыкальных признаков, который:

- 1) присутствует в любом музыкальном произведении;
- 2) позволяет различать музыкальные произведения между собой и находить их сходство, например, с аналогичным набором, представляющим поисковый запрос;
- 3) достаточно легко вводится:

3.1) почти любой человек без специальной подготовки способен это сделать;

3.2) может быть введен с помощью уже имеющейся периферии компьютера (так, едва ли ради единственно ситуативно возникшего запроса человек отправится в магазин или к знакомому за микрофоном).

Признаки-кандидаты: нотная запись, звуковысотное задание, ладогармоническое представление (мелодия), ритм. Только ритм удовлетворяет всем перечисленным выше требованиям.

Ввод настукиванием наиболее доступен как по возможностям пользователя — отсутствие чувства ритма более редко чем отсутствие слуха и голоса, так и по требованиям к компьютеру — их нет, поскольку настукивать можно не только с микрофоном, но и на клавиатуре, а при желании, и с помощью кнопки мышки или джойстика. Предоставляемые при этом пользователю две градации ненулевой громкости кажутся реально ему доступными (нетрудно стукнуть не по одной, а сразу по двум клавишам), но и достаточными в плане задания ритма.

Что касается недостатков ритмического описания музыки, то “почти единственный” и хорошо известный недостаток состоит в том, что один и тот же ритм может воспроизводиться в нескольких разных произведениях [2]. Однако недостаток ли это: даже оставив в стороне тему заимствований в музыке, обнаружить ритмическое сходство разных произведений — как правило, просто интересно для пользователя и едва ли станет препоной для его поисковой активности. (Более серьезный и связанный с “почти единственным” недостаток — то, что ритмический рисунок многих произведений беден).

5 Ритмическое сходство

Сегодня существуют программы, позволяющие выделять ритм или мелодию из аудиозаписи (см., например, работы WIDISoft под руководством С. Андреевского [16]). Сегодня существуют программы распознавания изображений, позволяющие переводить обычную нотную запись в электронную форму [5]. При том, что надежность этих программ невысока, последующее выделение ритма можно построить устойчивым к несистематическим ошибкам распознавания (ведь речь идет о повторяющейся компоненте произведения).

В случае ПМ, мы имеем дело с несколько иной задачей, чем при решении задачи распознавания (нотификации). Действительно, пусть имеется коллекция записей музыкальных произведений $\{R\}$ и некоторая “слуховая” мера несходства $t(*,*)$, используя которую человек говорит о сходстве произведений (замечим, что человеку нетрудно проранжировать по сходству сходные произведения, но крайне затруднительно — несхожие). Задача распознавания (“WAV”→”MIDI”) состоит в построении такого отображения $f()$, что $t(R, f(R))=0$.

При поиске, в дополнение к $\{R\}$, $f()$ и $t(*,*)$, мы имеем дело с напетым (наигранным или, как в нашем

случае, настуканным) запросом Q и его “машинным образом” $g(Q)$, а также с “машинной” мерой сходства $s(*,*)$ двух “машинных представлений”. Задача поиска состоит в том, чтобы ранжировка $\text{rank}(t(\{R\}, Q))$, построенная человеком по сходству $\{R\}$ с Q , была в своей начальной части такой же, что ранжировка $\text{rank}(s(\{f(R)\}, g(Q)))$, построенная машиной по сходству машинных образов произведений $\{f(R)\}$ с машинным образом запроса $g(Q)$. Эта задача отличается от задачи нотификации.

Для определения ритмического сходства используются различные метрики. Многие из них используют интервальное представление длительности ноты (удара), а именно, интервал между началом одной (входящей в ритм) ноты и началом следующей. В качестве мер ритмического сходства применяют евклидово и хэммингово расстояние между векторами длительностей, обобщения расстояния Хэмминга, а также хронометрическое расстояние [2,11,13,15].

6 Индексация музыки

Какая информация должна попадать в индекс о каждом музыкальном файле? Это, прежде всего, информация о собственно музыке, т.е. несколько неповторных наборов ритмов, встречающихся в произведении. Остальная информация текстовая и включает тип файла, сведения об авторстве, произведении и исполнителе. Эти данные обычно четко идентифицируются в источнике (в том числе, когда источником служит нотная запись), а в индексе хранятся в отдельных полях, что позволяет выполнять многомерный (атрибутный) поиск. Кроме того, в индексе следует предусмотреть поля нечетко идентифицированной и дополнительной информации.

К счастью, файлы, содержащие музыку, — строго структурированные и такие поля как “автор”, “название произведения” и “исполнитель” присутствуют практически во всех форматах. Тот факт, что в подавляющем большинстве музыкальных файлов эти поля пустые, является отражением вторичности таких файлов (например, переписанная в MP3 “для себя” композиция с лицензионного CD) и свидетельством непрофессионализма их создателей. Пустое поле — явное указание на то, что данным файлом нужно воспользоваться в последнюю очередь.

7 Музыкальные запросы и многомерный поиск музыки

Запрос может покрывать сразу несколько полей описания документа. Например, пользователь может настукивать какой-то ритм и, сверх того, указать, в поле автор “Дунаевский”. То есть запрос изначально является многомерным (фасетным) — для его задания используются сразу несколько полей.

Хотя ситуация многомерности типична для “мультимедиа-поиска”, применительно к текстовому поиску, она, с одной стороны, была значительно

потеснена одномерным поиском в Вебе, а с другой стороны, при всей своей модельной прозрачности, осталась недостаточно изучена теоретически. Речь идет о принятом в автоматизированных библиотечных каталогах (online public access catalogues) описании документа, как минимум, тремя раздельно задаваемыми полями: автор, название, рубрика. Понятно, что такой трехмерный поиск “трехмерных” документов гораздо удобнее поиска в одном, объединенном окне документов, для которых поля не были изначально разграничены.

С одной стороны, появляется возможность косвенного задания полей (например, найти автора введенного ритма — насколько поле “ритм” специфицирует музыкальное произведение, настолько же оно специфицирует и значение в поле “автор”). Тем самым, возникает и возможность формулировать запросы типа “дай биографию того, кто написал то, что я наступал”.

С другой стороны, что более традиционно для многомерного поиска, приблизительно специфицировав несколько полей, мы можем получить достаточно точный результат запроса, поскольку не точные значения в этих полях, грубо говоря, “по пересечению” лучше определяют объект поиска. Так, если приблизительно задан ритм и приблизительно автор, то шансы найти нужное произведение выше, чем при приблизительно задании только одного поля.

Спецификой музыкальных запросов, отчасти являющейся следствием того, что пользователь не обязан быть музыкантом-профессионалом, является то, что вместо значений “квартет де-дур” или “Иоганн Христиан Бах” в запросах будут использоваться приблизительная запись и *родовидовые* понятия. Что касается приблизительной записи (Кристиан, Крестьян), то сегодня достаточно популярный поисковый спеллинг состоит в подстановке тех написаний, которые были бы близки на слух к заданному. Однако, ясно, что при задании запроса непрофессионалом, Христиан имеет шансы превратиться в несозвучного Себасти(ь)ана. То есть надежность запроса достаточно невысока и о точном соответствии речь может идти скорее как об исключении, чем о правиле.

Более интересно, что для характеристики музыки часто используются “родовые” понятия, частным случаем которых являются жанр и стиль [10]. Вне зависимости от нашей способности исполнить запрос на “веселую”, “типа ска с барабаном”, именно такие запросы удобны или единственно возможны не только для непрофессионалов, но и для полупрофессионалов, скажем, при подборе музыки к драматическому спектаклю или фильму. В таком случае, система музыкального поиска должна включать и активно использовать тезаурусы, соответствующие как стандартному музыковедческому членению на жанры и эпохи, так и обиходно употребляемой классификации.

Мы намеренно не обсуждаем здесь первого кандидата на упоминание — очень популярный в

теории поиска информации механизм обратной связи (relevance feedback), в котором пользователь указывает в текущем списке результатов поиска те объекты, которые наиболее похожи или, напротив, непохожи на нужный, а система анализирует признаки и формирует новый список, учитывая эти неявно заданные признаки. В противоположность теории, реальные пользователи используют этот механизм обычно раз в жизни, после чего уже к нему не возвращаются. Теория поиска изображений сегодня практически полностью построена на механизме обратной связи. Однако реализация напоминает развлечение: пользователь не находит нужного результата, но остается доволен самим процессом (при поиске текстовой информации удовольствия он не получает). Хотя, в отличие от поиска изображений, при поиске музыки, кажется, более ясно, по каким признакам искать сходство, тем не менее, и здесь говорить о возможной эффективности механизма положительной обратной связи преждевременно.

8 Заключение

В статье очерчен проект системы музыкального поиска, основанного на ритмическом сходстве запроса и хранящегося в базе музыкального произведения. Такой подход делает поиск музыки общедоступным. И именно с собственно поиска музыки предполагается начать построение системы музыкального поиска. Последовательность действий на этом этапе: (1) построение мер ритмического сходства и их прогонка на “лабораторных” ритмах, почти напрямую введенных как ритм (например, с клавиатуры), а не выделенных из полноценного аудио, (2) выделение ритма (ритмов) из реального аудиофайла (по нарастанию сложности формата и музыкального материала — сначала MIDI, затем WAV и более “продвинутые”, сначала моно-, затем полифония), (3) экономная (бесповторная) индексация ритмов и поиск.

Такой порядок диктуется стремлением снизить размер возможно непродуктивных затрат. Выделение ритма из реальных аудиофайлов требует объемной, в том числе чисто вспомогательной работы по созданию инструмента. Эта работа осмысленна лишь в случае достижимости практического результата. В свете того, что “распознавание музыки”, в полном соответствии с давней искусственно-интеллектуальной традицией, более богато [возможно, богатыми] идеями, а также как бы работающими и как бы существующими системами, эффективность этих затрат не гарантирована. В то же время, выполнение подэтапа (1) не очень затратно, но очень полезно как в теоретическом, так и в практическом плане, вне зависимости от утилитарности результатов.

Выше мы охарактеризовали музыкальный поиск как многомерный. Использование этой многомерности — предмет второго этапа практических работ, судьба которого полностью

зависит от успеха реализации первого. Используя многомерную спецификацию объекта поиска можно улучшить качество музыкального поиска, но при условии приемлемости качества выделения ритма и ранжирования ритмического сходства (подобно тому, как использование лингвистических знаний резко улучшает качество распознавания линейных текстов, но лишь в том случае, если надежность распознавания символов выше некоторого порога).

Наконец, мы сознательно не касались дополнительной преграды для ПМ, а именно юридических и коммерческих проблем, связанных с защитой авторских прав. Однако нам кажется, что в этой самостоятельной области возможны самые, на первый взгляд, неожиданные, но коммерчески эффективные построения взаимодействия тройки “владелец музыки” — “владелец поискового сервиса” — “потребитель”.

9 Благодарности

Автор признателен А.С. Тангяну за помощь в знакомстве с формальной музыкологией и распознаванием музыки, а также за непредвзятую характеристику реально достигнутых в этой области результатов. Автор благодарен В.Л. Арлазарову и И.В. Сегаловичу за обсуждение задач поиска музыки и изложенное ими (достаточно скептическое) видение существующей сегодня ситуации и возможных перспектив музыкального поиска.

Автор также действительно признателен анонимному рецензенту, рекомендовавшему автору “обратиться к классическим трудам (начиная, с работы Ранганатана по фасетным классификациям)”.

Литература

- [1] Зарипов Р.Х.. Кибернетика и музыка. – М.: «Наука» – 1971. – 235 с.
- [2] Dixon S., Gouyon F., Wildmer G. Towards characterization of music via rhythmic patterns. In *Proceedings of ISMIR 2004*.
- [3] Downie J. Music information retrieval. In *Annual Review of Information Science and Technology*, volume 37. NJ: Medford. pages 295–340, 2003
- [4] Downie J. Toward the scientific evaluation of Music Information Retrieval systems. In *Proceedings of ISMIR 2003*.
- [5] George S. *Visual Perception of Music Notation: On-Line and Off-Line Recognition*. IRM Press, 2004, 372 p.
- [6] Haitsma J, Kalker T. A highly robust audio fingerprinting system. In *Proceedings of ISMIR 2002*.
- [7] Harb H., Chen L. T A query by example music retrieval algorithm. In *Proceedings of 4th European Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2003)*, pages 122–128, 2003.

- [8] Haus G., Pollastri E. An audio front end for query-by-humming systems. In *Proceedings of ISMIR 2001*, pages 65–72, 2001.
- [9] Lee J.H., Downie J. Survey of music information needs, uses, and seeking behaviors: Preliminary findings. In *Proceedings of ISMIR 2004*.
- [10] Leitich S., Rauber A. Information Retrieval in Digital Libraries of Music. In *Труды 6-й конференции RCDL 2004*.
- [11] Pickens J. A comparison of language modeling and probabilistic text information retrieval approaches to monophonic music retrieval. In *Proceedings of ISMIR 2000*. – 2000.
- [12] Tangian, A. How do we think? Modeling interactions of memory and thinking. In *Cognitive Processing*, 2(1), pages 117-151, 2001.
- [13] Tangian, A. Constructing Rhythmic Canons. In *Perspectives in New Music*, 41(2), pages 66-94, 2003.
- [14] Tanguiane, A.S. *Artificial Perception and Music*. LNAI 746. Springer-Verlag, 1993, 210 p.
- [15] Toussaint G. A comparison of rhythmic similarity measures. In *Proceedings of ISMIR 2004*.
- [16] <http://www.widisoft.com>

A Rhythm-Based Easy of Access Music Retrieval Service

Nick Buzikashvili

A project of rhythm-based music information retrieval service is proposed. It provides an easy of access way to formulate musical queries and to search music. Besides proper MIR, ‘multi-dimensional’ (multimodal) searching is considered.