

И.А. БАЛТИЙСКИЙ, С.И. НИКОЛЕНКО
**ОБЗОР ГРАФИЧЕСКИХ ВЕРОЯТНОСТНЫХ
МОДЕЛЕЙ ГАРМОНИИ ДЛЯ АНАЛИЗА
МУЗЫКАЛЬНЫХ ПРОИЗВЕДЕНИЙ**

Балтийский И.А., Николенко С.И. Обзор графических вероятностных моделей гармонии для анализа музыкальных произведений.

Аннотация. Цель статьи — познакомить читателя с современным состоянием дел в области автоматического анализа музыкальной гармонии. Мотивацией для исследований в этой области может являться создание автоматических систем рекомендации музыки, ориентированных на содержание (наподобие Pandora, но без ручного труда экспертов-музыковедов). Основное внимание уделено графическим вероятностным моделям как одному из наиболее перспективных подходов, но описываются и альтернативные методы. Рассмотрены работы, использующие марковские цепи, скрытые марковские модели, многоуровневые графические модели. Приведены как работы, моделирующие только гармонию — последовательности аккордов, в некоторых случаях и тональность, — так и работы, включающие в себя информацию о структуре анализируемого произведения (ритмической, голосовой).

Ключевые слова: информационный поиск музыки, рекомендательные системы, гармония, оценка схожести, графические модели, вероятностные модели, байесовский вывод, марковские цепи, скрытые марковские модели.

Baltiyskiy I.A., Nikolenko S.I. Probabilistic graphical models of harmony in automated music analysis: a survey.

Abstract. This paper presents the current state of art in the field of automated musical harmony analysis. Research in this field can be motivated by the real-world problems of creating completely automated content-based music recommendation systems (similar to Pandora, but without the manual work of expert musicologists). The paper is mainly focused on probabilistic graphical models as one of the most promising approaches, although we also give background in alternative methods. We consider works that use Markov chain models, hidden Markov models, and multi-level graphical models. Along with the models that capture only harmonic information—chord progressions, in some cases also the key,—we also list several models that combine harmonic structure with rhythmic or stream structure.

Keywords: music information retrieval, recommendation systems, harmony, similarity task, graphical models, probabilistic models, Bayesian inference, Markov chains, hidden Markov models.

1. Введение. Компьютерная музыка появилась почти одновременно с электрическими компьютерами — один из первых, CSIRAC, в 1951 году уже умел проигрывать мелодии [22]. Постепенно появился интерес к компьютеру не только как к средству синтеза музыки, но и как к средству ее анализа — извлечения из записанной музыкальной композиции информации сверх необходимой для ее воспроизведения. Сегодня анализ музыки стал одним из центральных направлений ис-

следований в области информационного поиска музыки (англ. *music information retrieval*).

В этой работе нас будет интересовать анализ музыкальной гармонии с использованием вероятностных моделей.

Разумеется, анализ музыки не ограничивается этой областью. Так, многие ранние исследования занимались задачей транскрипции музыки — переведения записанного аудиосигнала в нотную запись — и анализа основного тона [19, 54–55, 61–62]. Более современные обзоры работ в области транскрипции можно найти в [40–41]; в этой задаче также используются вероятностные модели [18, 64, 73]. В задаче сегрегации голосов требуется разделить полифоническую музыку на отдельные голоса [11, 39]. Похожие задачи — задача разделения отдельных музыкальных инструментов [84] и выделения одного голоса [25]. Целый пласт исследований посвящен анализу ритмической составляющей музыки — определению темпа, метрической структуры (поиску отображения событий во времени в такты и доли) [21, 63, 72]. Выделим работы А.Т. Кемгила, использующие графические вероятностные модели, в том числе, фильтры Калмана [13–16]. Многие методы автоматического анализа музыки, в том числе и байесовские методы, находят применение в задачах классификации музыки [17]. Ещё одно интересное практическое применение — восстановление повреждённых записей [24].

Анализ композиций методами искусственного интеллекта — не новинка, подобные методы применяются уже достаточно давно [66]. Байесовский подход, который мы рассматриваем в этой работе, появился в применении к задаче анализа музыкальных произведений в 1990-х гг. [38]

Отметим связь исследований в области автоматического анализа музыки с исследованиями, посвящёнными анализу человеческого восприятия музыки [46–48, 68–69, 71, 78, 82]. Как можно было ожидать, интересующая нас область испытывает взаимное влияние теории предметной области — теории музыки, в частности, теории гармонии [45, 81].

Цель настоящей работы — описание современных вероятностных моделей, используемых для анализа гармонии. Обзор организован следующим образом. Раздел 2 описывает практическую мотивацию таких исследований — задачу построения музыкальных рекомендательных систем; как мы покажем, автоматический анализ гармонии актуален с точки зрения таких задач. В разделе 3 мы обсуждаем постановку задачи моделирования гармонии, помещаем её в контекст практически

важных задач и кратко обсуждаем несколько подходов, имеющих иную, не вероятностную природу. В разделе 4 мы кратко напоминаем основные понятия теории байесовского вывода и графических вероятностных моделей. Раздел 5 — основной; в нём мы приводим обзор нескольких вероятностных моделей, которые использовались для моделирования гармонии. Раздел 6 подводит итоги.

2. Практическая задача — музыкальные рекомендательные системы. Развитие информационных систем — снижение стоимости носителей информации, увеличение пропускной способности сетей, рост популярности сети Интернет — изменило характер распространения и потребления продуктов культуры, в том числе музыки. Существенное отличие магазина, продающего произведения культуры на физических носителях, от интернет-магазина, продающего произведения в электронном виде, — в ограниченном количестве предлагаемых наименований. Природа ограничения — экономическая: место для содержания физических носителей ограничено, и его стоимость нельзя не учитывать. Так, место на полке, занимаемое одним CD-диском в музыкальном магазине, не окупится, если он будет продаваться реже, чем раз в квартал [3]. С другой стороны, затраты на хранение информации в электронном виде существенно ниже, а географический охват намного шире, что накладывает значительно меньшие ограничения на ассортимент [3].

С подачи К. Андерсона [2] в широкое употребление вошел термин «длинный хвост» (англ. *long tail*), фиксирующий сложившийся еще в середине XX века факт существенно неравномерного распределения внимания потребителей культуры к ее произведениям: так, например, самыми популярными песнями становятся только единицы, в то время как подавляющее большинство песен известно весьма ограниченному кругу людей (они и составляют «длинный хвост»). В условиях продажи продуктов культуры на физических носителях, то есть в условиях диктуемого ограниченным физическим пространством размера каталога, «длинный хвост» вообще не попадает в продажу. В условиях интернет-магазинов затраты на содержание одного произведения ничтожны, и «длинный хвост» становится доступен. «Длинный хвост» приносит существенный доход: так, типичный магазин Barnes & Noble, продающий бумажные книги, содержит около 130 000 наименований, в то время как более половины продаж электронных книг в магазине Amazon приходится на книги, *не* входящие в список 130 000 самых популярных. Подобная ситуация наблюдается и в музыке [2].

Итак, наличие «длинного хвоста» в каталоге современного интернет-магазина экономически выгодно продавцу и полезно покупателю, который более не ограничен в выборе рамками самых продаваемых названий. Как, однако, донести информацию о произведении до людей, которые потенциально могут им заинтересоваться? Для этого требуется механизм, подсказывающий покупателю новые наименования — механизм рекомендаций. Практика показывает, что крупные интернет-магазины имеют свои рекомендательные системы, — например, Amazon.com [26] (имеются в виду рекомендации для книг), Netflix [30] (специализируется на фильмах), iTunes [29] (специализируется на музыке).

Таким образом, рекомендация музыки в условиях обилия музыкальных произведений — это актуальная проблема, значит, и задача создания автоматических систем рекомендации музыки актуальна. В дальнейшем нас будет интересовать такой сценарий использования рекомендательной системы: пользователь задает известную ему композицию, в ответ ему предлагается прослушать набор композиций, в некотором смысле схожих с заданной.

В каком именно смысле композиции должны быть схожи? Ответ на этот вопрос неочевиден, поскольку он должен зависеть от цели пользователя, которая может меняться со временем — в один момент требуется энергичная, имеющая фиксированный простой ритм музыка для спорта, в другой — не отвлекающая внимания музыка для сосредоточенной работы, в третий — серьезная музыка, наоборот, требующая внимания; добавим к этому, что пользователь не обязан, вообще говоря, осознавать и формализовывать эту цель для того, чтобы воспользоваться системой. На сегодняшний день ни одна из популярных рекомендательных систем не позволяет в явном виде задавать критерий схожести с помощью заложенных в ней принципов выбора композиций.

По характеру этих принципов системы рекомендаций делятся на два типа. Первый тип систем использует анализ предпочтений пользователей, имеющих схожие вкусы с текущим пользователем. Такой подход носит название *совместной фильтрации* (также *коллаборативная фильтрация*, англ. *collaborative filtering*). Яркий пример — сайт *last.fm* [27], определяющий вкус пользователя как неявно, за счет фиксирования всех прослушанных им композиций, так и явно, за счет специального флага «Круто!» (англ. *Like*), которым можно отмечать понравившиеся композиции.

Второй тип систем использует объективный музыкальный анализ заданной пользователем композиции. Такой подход называют *ориентированным на содержание* (встречается также *контентно-ориентированный*, англ. *content-based*). Яркий пример — сайт *Пандора* (англ. *Pandora*) [12, 28, 37]. Задаваемая пользователем композиция должна присутствовать в базе данных системы, поскольку для поиска схожих композиций используются результаты проведенного ранее анализа (называемые в терминологии Пандоры *геном* композиции). *Гены*, составляющие геном, различны для каждого стиля; так, например, для джаза есть ген «эмоциональная вовлеченность соло на трубе», для которого у конкретной композиции может стоять оценка от 1 до 5. Оценки выставляются командой профессиональных музыковедов, которые заранее договариваются о наборе генов для каждого стиля и выработывают общие методики оценивания.

У каждого из типов систем есть свои преимущества и недостатки. Системы, ориентированные на содержание, лучше используют «длинный хвост», поскольку системы совместной фильтрации подвержены положительной обратной связи — чем более популярен автор, тем больше вероятность, что его композиции будут рекомендованы. Тот же эффект снижает возможность качественно новой, неожиданной рекомендации; например, для любителя рок-музыки 60-х вряд ли будет открытием рекомендация группы *Rolling Stones* — скорее всего, он осведомлен о ней вне зависимости от того, сколько прослушиваний их песен зафиксировала система. Более того, системы, ориентированные на содержание, допускают возможность задания критерия схожести за счет правильного выбора учитываемых генов. С другой стороны, системы совместной фильтрации учитывают влияние на вкус пользователя вкусов других людей — друзей, критиков, музыкальных обозревателей, вкуса масс; важно иметь в виду, что музыка — явление в том числе и социальное. Разумно заключить, что право на существование имеют оба подхода.

Практический интерес для разработки рекомендательных систем, ориентированных на содержание, представляет задача автоматизации труда коллектива музыковедов: нельзя ли автоматически оценивать схожесть музыкальных композиций исходя только из звуковой информации? На данный момент доступных коммерческих решений этой задачи нет. Можно решать задачу по частям, пользуясь тем, что музыка имеет несколько граней — тембр, гармония, мелодия, ритм, — и реализуя методику сравнения специально для каждой из граней с тем,

чтобы скомбинировать получившиеся методы в единый. Пример такого комбинирования рассмотрен в работе [1].

Как мы видим, анализ музыки имеет интересное практическое применение в задаче оценки схожести, причем допустимо разбивать задачу соответственно граням музыки, в частности, выделяя анализ гармонии. Дальнейшее изложение в этой статье исходит из задачи анализа гармонии.

3. Современные подходы к моделированию гармонии музыкальных произведений. Анализ гармонии применяется не только в задаче оценки схожести. Традиционно больше внимания уделялось получению представления, удобного для человека-исследователя (как правило, профессионального музыковеда), либо для поиска на основе запросов, записанных в виде последовательностей аккордов, гармонических функций либо полифонических приемов (например, «найти все параллельные квинты» или «найти задержание»); такой поиск тоже ориентирован в основном на профессиональную среду. В качестве нескольких примеров можно привести [8, 51, 56].

На самом деле, оценка схожести до недавнего времени нечасто решалась с позиций гармонии; это отмечено и в обзорной работе [70], и в [7]. Большинство работ, исследующих схожесть композиций либо жанровую классификацию, фокусировалось на тембральных характеристиках [50], однако за последнее десятилетие накопилось достаточно эмпирических свидетельств в пользу того, что этого недостаточно ни для оценки схожести, ни для многих других задач информационного поиска музыки [5, 50].

Ниже мы опишем несколько гармонических подходов к задаче оценки схожести и смежным к ней, не основанным на использовании вероятностных моделей.

В [6] применяется один из самых старых методов анализа гармонии, основанный на построении генеративных грамматик. Традиция грамматических подходов восходит к лингвистическому анализу [45], что неудивительно, поскольку музыкальная форма, подобно языкам, имеет иерархическую структуру, которой подчиняется и гармония. В этом методе расстояние между последовательностями аккордов вводится через расстояние между поддеревьями дерева разбора, построенного над ними.

Другой метод, также работающий на последовательности аккордов, рассмотрен в [7]. Предложив функцию расстояния между аккордами, авторы вычисляют «профиль» — последовательность расстояний от тонического трезвучия в тональности произведения до текуще-

го аккорда последовательности. Расстоянием между последовательностями аккордов служит расстояние между «профилями», вычисляемое как площадь между их графиками.

Интересно отметить работу [4] в смежной задаче — жанровой классификации, — использующую индуктивное логическое программирование (ИЛП). ИЛП в данном случае используется как метод машинного обучения, позволяющий указывать фоновое знание; задача ИЛП формулируется так: основываясь на фоновом знании B , вывести гипотезу H , объясняющую положительные примеры E^+ ($B \wedge H$ покрывают все примеры из E^+) и не противоречащую отрицательным примерам E^- (ни один из отрицательных примеров не покрыт $B \wedge H$). $B, H, E = E^- \wedge E^+$ выражаются на языке логики первого порядка. В [4], как и в вышеупомянутых трудах, используется представление музыки в виде последовательности аккордов. На этапе обучения авторы задают в качестве положительных примеров все подпоследовательности аккордов длины 4, в качестве фонового знания выступают правила гармонии (закодированы понятия функции аккорда, последовательности основных и басовых тонов аккордов и т.д.) Полученный набор правил, составляющий гипотезу, характеризует жанр. Тогда произведение, относительно которого требуется сделать вывод о жанровой принадлежности, относится к тому жанру, для которого выполняется больше всего правил из соответствующей гипотезы. Этот подход можно было бы применить к задаче оценки схожести музыкальных композиций, однако вопрос о том, как распространить его с последовательностей аккордов на аудиоданные, нетривиален и пока не решен.

К сожалению, сфера применения вышеупомянутых методов ограничена вследствие того, что необходимо обладать данными о последовательности аккордов, созданными вручную, либо пользоваться автоматическими анализаторами аккордов, точность которых пока что не идеальна.

В последующих частях обзора мы рассмотрим несколько работ, использующих вероятностные модели.

4. Графические вероятностные модели. Графические вероятностные модели — это простой и наглядный способ представлять структуру вероятностной модели в виде графа. Более того, этот способ ещё и (часто) транслируется в конкретные алгоритмы байесовского вывода на этой модели. Напомним кратко основные положения теории байесовского вывода; более подробное изложение см. в [9, 53, 89–93].

Теорема Байеса $p(x|y) = (p(y|x)p(x))/p(y)$ обычно применяется в машинном обучении в виде

$$p(h|D) = (p(D|h)p(h))/p(D),$$

где h — гипотеза, D — данные. Чтобы, например, найти самую правдоподобную гипотезу, нужно найти

$$\arg \max_h p(h|D) = \arg \max_h p(D|h)p(h),$$

где $p(h)$ — априорное распределение, $p(D|h)$ — правдоподобие данных при условии той или иной гипотезы. Как правило, в модели есть не только наблюдаемые переменные (собственно данные D), но и скрытые переменные Z , которые выражают те или иные важные параметры или процессы, но не измеряются при наблюдениях. Тогда задача превращается в

$$\arg \max_h p(h|D) = \arg \max_h \int_Z p(D, Z|h)p(h) dZ,$$

и для её решения приходится применять более сложные алгоритмы вывода.

Многие из них сводятся к тому, что функция правдоподобия $p(D|h)$ или лог-правдоподобие $\log p(D|h)$ приводится к такому виду или приближается такой функцией, у которой можно подсчитать градиент (обычно в замкнутой форме). Затем происходит обычный градиентный подъём, который приводит к локальному максимуму правдоподобия. К таким алгоритмам относится классический EM-алгоритм [20, 57], его вариации и обобщения, например, алгоритмы миноризации-максимизации [31–32] и целый ряд приближённых алгоритмов, основанных на вариационных приближениях [10, 33, 36]. Другой класс подходов основан на алгоритме передачи сообщений, которым можно провести маргинализацию в графической модели без циклов [59, 85–86], и его приближённых вариантах, например Expectation Propagation [52–53]. Третий класс алгоритмов — алгоритмы приближённого вывода, основанные на сэмплировании [67]; они проводят сэмплирование либо по Гиббсу [23], либо по одному из методов, основанных на алгоритме Метрополиса–Гастингса [34–35] (например, MCMC, от англ. Markov Chain Monte Carlo). Мы не будем сейчас подробно разбирать эти алгоритмы, а заинтересованному читателю рекомендуем книги [9, 49, 92] и другие источники.

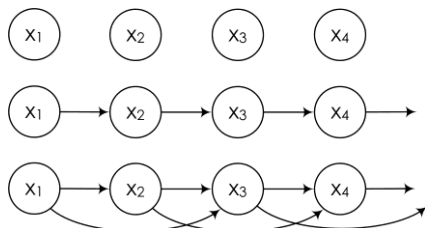


Рис. 1. Марковские цепи порядка 0, 1 и 2.

5. Вероятностные модели музыкальной гармонии. В этом разделе мы рассматриваем несколько конкретных моделей, которые в последние годы применялись для анализа музыкальной гармонии.

5.1. Модель Пикенса–Кроуфорда — марковские цепи. Метод, рассмотренный в [60], решает интересующую нас задачу схожести и поиска полифонических музыкальных композиций. Дж. Пикенс и Т. Кроуфорд концентрируются на анализе полифонической музыки — задаче, которую на тот момент ещё мало кто решал. Интересно, что авторы отказались от вычисления последовательности аккордов, вместо этого рассматривая векторы вероятностей события «данному временному отрезку музыкальной композиции соответствует аккорд c ». Такое же представление использовали и мы в нашем анализе модели «bag of chords», который представлен в параллельной статье в этом же томе [88]; в разделе 2 работы [88] базовое представление описано более подробно.

Каждая композиция представляется в [60] вероятностями цепочек переходов между аккордами — авторы строят марковские модели 0, 1, 2 и 3 порядка. Марковская модель порядка n представляется в виде матрицы размера $24^n \times 24^n$; строки матрицы представляют собой пространство предыдущих состояний (модель порядка n учитывает n предыдущих состояний). Графические модели марковских цепей разных порядков представлены на рис. 1.

Описанная выше модель не учитывает возможность транспонирования: в случае, когда искомая композиция отличается от заданной только сдвигом высот всех нот на некоторую фиксированную величину, она не будет найдена. Авторы предлагают устойчивый к транспозиции вариант модели, использующий вместо цепочек длин 2, 3 и 4, состоящих из аккордов, цепочки длин 1, 2 и 3, состоящих из расстояний между аккордами на квинтовом круге. Сравниваются инвариантные модели и модели без этой оптимизации.

Для задачи поиска авторы используют расстояние Кульбака-Лейблера, подсчитывая расстояние между распределениями марковской цепи d , обученной для каждой композиции в базе данных, и марковской цепью запроса q . Расстояние Кульбака-Лейблера вычисляется как

$$D(q\|d) = \sum_{q_i \in q, d_i \in d} \sum_{x \in X} q_i(x) \log \frac{q_i(x)}{d_i(x)}.$$

Интересны полученные результаты. Они свидетельствуют, что в целом увеличение длины марковской цепи в модели песни улучшает работу поиска, но среди марковских цепей фиксированной длины без факторизации по транспозициям лучше всех оказалась самая простая модель — марковская цепь нулевого порядка.

Отметим, что модель в [60] не рассматривает аккорды в контексте тональности. Полученное на основе такой модели расстояние будет слабее расстояния, учитывающего тональность — расстояние между гармонически разными композициями будет меньше, поскольку слушатель воспринимает переходы между аккордами в первую очередь с точки зрения их функции в тональности.

5.2. Методы, основанные на скрытых марковских моделях.

Наиболее интересны для нас модели, способные вычислять не только последовательности аккордов, но и тональность для каждого момента времени композиции, а значит, и функцию аккордов в тональности. Рассмотрим работу [65]. В ней строится скрытая марковская модель, в которой скрытыми переменными являются гармонические метки — тройки (тональность, лад, функция), где под тональностью понимается первая ступень тональности (от *do* до *si*), под ладом — мажор либо минор, под функцией — ступень основного тона аккорда в данной тональности (I-VII). Гармонические метки рассматриваются для каждого временного отрезка между двумя ближайшими событиями «начало ноты» либо «конец ноты». Наблюдаемые значения — вектор класса высот (от *do* до *si*), содержащий все ноты, присутствующие в композиции на данном временном отрезке, — считаются наблюдениями случайной величины, распределение которой зависит только от текущей гармонической метки. Соответственно, параметрами модели являются вероятности переходов между соседними гармоническими метками и условные вероятности наблюдаемых данных. Идея использования гармонических меток для условных вероятностей наблюдения музыкальных данных была рассмотрена и в применении к нашей модели в [88]; там же подробно вводится понятие вектора класса высот.

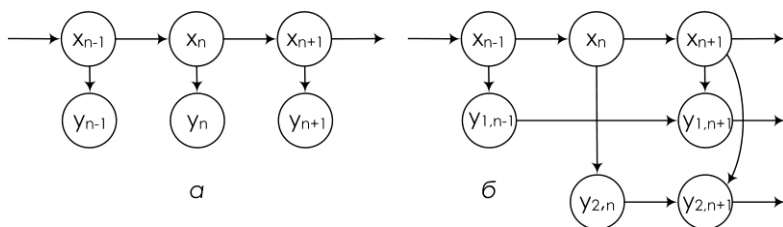


Рис. 2. а — скрытая марковская модель;
 б — скрытая марковская модель с двумя голосами.

Рис. 2,а представляет графическую модель обычной скрытой марковской модели; x_i — события скрытой марковской цепи, y_i — наблюдаемые. Основной причиной для использования скрытых марковских моделей в данном случае является то, что их можно обучать на *неразмеченных* данных — любой MIDI-файл, в котором размечены наблюдаемые ноты, может быть добавлен в систему для обучения. Авторы также предлагают (но не обучают) модель, которая при помощи варианта скрытых марковских моделей с несколькими наблюдаемыми раскладывает полифоническую музыкальную композицию на отдельные голоса; очередная наблюдаемая в одном голосе y_{ik} зависит от текущего состояния всей мелодии x_i и предыдущей наблюдаемой в этом голосе $y_{i,k-1}$. Пример графа такой модели представлен на рис. 2,б. Заметим, что в работе не приведены результаты практических исследований; нам не удалось найти их и в последующих трудах авторов.

В [44] рассмотрена скрытая марковская модель первого порядка, которая также способна вычислять последовательности аккордов и текущую тональность. Приведенные в статье результаты свидетельствуют, что включение в модель информации о тональности дает преимущество по сравнению с методами, не учитывающими тональность. Учёт тональности происходит аналогично методам распознавания речи с небольшим словарём: для K тональностей обучаются K разных скрытых марковских моделей $\lambda_1, \dots, \lambda_K$ (на данных, размеченных по принадлежности к той или иной тональности). Когда в систему поступает запрос на извлечение аккордов, система сначала выбирает наиболее правдоподобную тональность $key = \arg \max_k p(D|\lambda_k)$, а затем привлекает аккорды, используя модель λ_{key} . Таким образом, система, построенная в работе [44], может быть использована и для классификации композиций по тональностям.

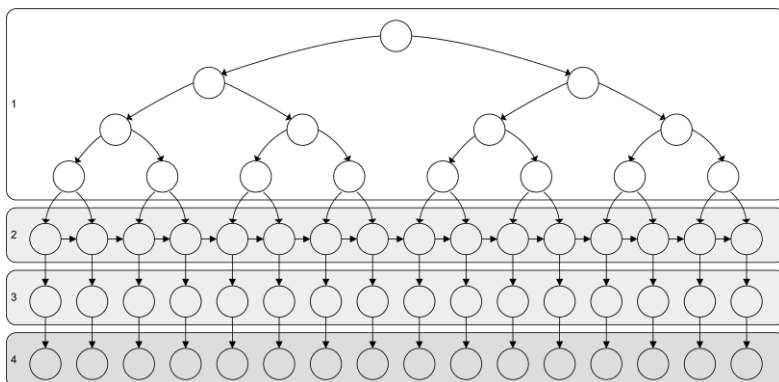


Рис. 3. Вероятностная графическая модель с несколькими скрытыми уровнями для последовательностей аккордов [58].

5.3. Более сложные графические модели. Более тонкая вероятностная модель для анализа гармонии предложена в [58]. В отличие от скрытых марковских моделей в [44; 65], графическая вероятностная модель в [58] включает в себя не только локальную зависимость аккорда от его соседей в последовательности, но и его глобальную зависимость от позиции в последовательности аккордов. Несмотря на явную выгоду такого подхода, тональный контекст в этой модели не учитывается; модель достаточно сложна — она состоит из четырех уровней переменных, три из которых являются скрытыми (см. рис. 3). Первый скрытый уровень имеет иерархическую древовидную структуру и моделирует контекстуальные зависимости, связанные с размером. Второй скрытый уровень дополняет глобальный контекст локальным, сглаживая последовательность аккордов. Третий и четвертый уровни соответствуют аккордам: четвертый уровень наблюдается, а третий (через фиксированную матрицу похожести) «размазывает» наблюдаемые аккорды на те, которые могли бы породить подобное наблюдение.

Можно пойти еще дальше и моделировать гармонию композиции не раздельно, а вместе с остальными гранями музыки. По этому пути пошел Дэвид Темперли, предложивший в серии работ [75–76; 80] и книге [77] объединенную модель, работающую с музыкой, представленной в виде MIDI. Темперли, с одной стороны, синтезировал анализ ритма, гармонии и разбиения на голоса, а с другой стороны, объединил анализ музыкальной структуры и анализ ее фактуры (набор звучащих в некоторый момент времени нот).

Есть основания полагать, что объединенная модель решает каждую из задач, решаемых отдельными ее компонентами, лучше. Во-

первых, смена гармонии обычно приходится на долю, что Темперли подтверждает статистикой по корпусу классической музыки [42]; значит, информация о метре произведения может уточнить информацию о гармонии. Верно и обратное: есть свидетельства [78], что при ритмическом анализе следует учитывать моменты смены гармонии, поскольку последние редко происходят в моменты, принадлежащие ритмической сетке более низкого уровня (более частой), чем доля.

Далее, интересные последствия влечет включение информации о голосах. Термин «голос» в данном случае следует воспринимать в контексте классической музыки — в частности, хорошим примером голоса может служить отдельная партия хора. Длинный промежуток времени между началом последовательных нот (англ. *IOI*, от *inter-onset-interval*) свидетельствует в пользу сильных долей (одна приходится на его начало, другая — на начало следующей ноты). Ясно, что следует рассматривать промежуток времени между последовательными нотами *внутри* одного голоса — следовательно, модель разбиения на голоса связана с моделью ритма.

Наконец, объединение анализа текущего набора звучащих нот и структурного анализа также может дать свои плоды. Поставим задачу транскрипции — определения наиболее вероятного набора нот, звучащих в некоторый момент времени в аудиосигнале:

$$\arg \max_N p(N|s) = \arg \max_N (p(s|N)p(N)),$$

где N — транскрипция, s — аудиосигнал.

Как определить $p(N)$? Для осмысленного определения нам потребуется структурный анализ композиции — наиболее вероятный набор нот подчиняется ритмической структуре, гармонической структуре и ограничениям, накладываемым голосами.

Подробное описание построенной Темперли модели выходит за рамки данного обзора. Модель позволяет два варианта использования: транскрипция (рассмотрена выше) и анализ (извлечение всех видов структурной информации из MIDI). Для последнего строится модель, общий вид которой представлен на рис. 4. Обратим внимание на то, что генеративный процесс для гармонической структуры является марковским: рассматриваются только вероятности переходов между основными тонами аккордов в текущий и последующий момент времени. Сначала порождается метрическая структура, т.е. последовательность долей. Гармоническая структура зависит от метрической и состоит из последовательности аккордов, которым разрешается звучать только на долю.

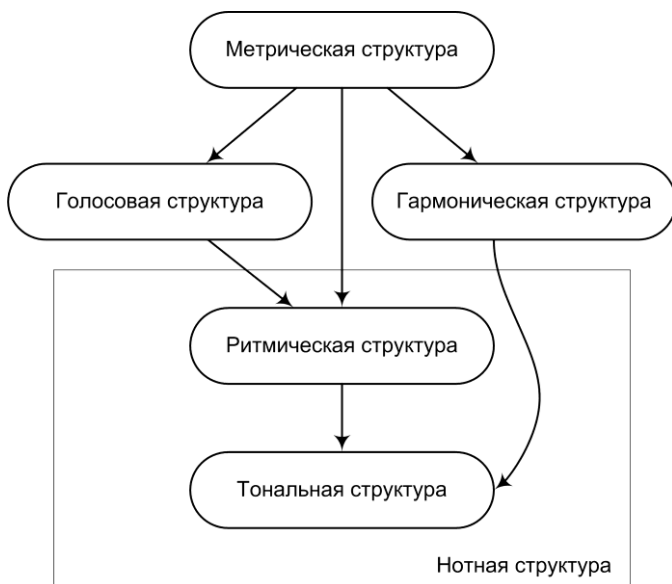


Рис. 4. Общая структура генеративного процесса в модели Темперли [75].

При генерации голосовой структуры порождается множество отрывков непрерывно звучащей музыки, ограниченных долями. Затем порождается собственно нотная структура, состоящая из ритмической (последовательность звучащих нот) и тональной (собственно высота каждой ноты). Модель Темперли легла в основу анализатора гармонии *Melisma*, разработанного Темперли и Слитором [79].

Отметим, что описанная модель является достаточно сложной и содержит много параметров; по признанию автора, их выбор зачастую велся методом проб и ошибок. К сожалению, ни в работе [43], ни в работе [83] не удалось воспроизвести результатов, указанных в [75].

5.4. Другие вероятностные модели. В работе [74] для повышения точности определения аккордов в композиции, представленной в виде аудио, используются оценка высоты басового тона и вероятности переходов между всеми парами аккордов. Рассматриваются 4 типа аккордов — мажорный, минорный, задержанный (трезвучие, содержащее задержание, как правило, на четвертой ступени) и уменьшенный; наблюдаемый в аудиоданных хрома-вектор моделируется смесью четырех гауссиан, по одной на каждый тип аккорда. Интересно, что авторы используют дополнительные данные, извлеченные из аудио (басовый тон), но преимущества для задачи оценки схожести неоче-

видны. Как и в [58], модель не учитывает роли аккордов в контексте тональности.

Отметим также работу [87], в которой также рассматривается вероятностная модель, включающая в себя последовательность аккордов, но не включающая в себя информацию о тональности. Особенность работы — использование априорного знания о типичных гармонизациях аккордов на определенных инструментах для улучшения распознавания аккордов.

6. Заключение. Подведем итоги. Выше были представлены некоторые варианты моделирования гармонии в музыкальных произведениях. Как можно заметить, большинство из них предназначено для другой задачи, нежели оценка сходства композиций; как было указано в начале раздела, это отражает действительное положение дел. Иными словами, авторы преимущественно не указывают возможность сравнения моделей, построенных для разных композиций; разумеется, построенные последовательности гармонических функций могут служить входом для рассмотренных выше алгоритмов (наподобие [6] или [7]), однако прямое сравнение моделей выглядит более привлекательно вследствие неизбежных ошибок распознавания аккордов.

Использование вероятностных моделей — достаточно распространенный подход. Несмотря на то, что немногие подходы включают в модель информацию о тональности, существуют результаты, свидетельствующие о пользе такой информации [44]. В нашей работе, опубликованной в этом же томе [88], аккорды будут анализироваться в контексте текущей тональности. Во многих работах встречаются оригинальные идеи по дополнению модели каким-либо видом априорной информации либо вспомогательной информацией, извлеченной из аудио.

Литература

1. *Allan H., Mullensiefen D., Wiggins G.* Methodological Considerations In Studies Of Musical Similarity // Proceedings of the International Conference on Music Information Retrieval (ISMIR 2007) (Vienna, Austria, September 23–27, 2007) URL: http://ismir2007.ismir.net/proceedings/ISMIR2007_p473_allan.pdf
2. *Anderson C.* The Long Tail // Журнал Wired: [сайт]. URL: <http://www.wired.com/wired/archive/12.10/tail.html>
3. *Anderson C.* The Long Tail: Why the Future of Business Is Selling Less of More. New York: Hyperion, 2006. 256 pp.
4. *Anglade A., Dixon S.* Characterisation of Harmony With Inductive Logic Programming // Proceedings of the Ninth International Conference on Music Information Retrieval (ISMIR 2008) (Philadelphia, Pennsylvania, September 14–18, 2008) Pp. 63–68.

5. *Aucouturier J.-J., Pachet F.* Improving timbre similarity: How high's the sky? // Journal of Negative Results in Speech and Audio Sciences. 2004. № 1(1). URL: <http://ayesha.lti.cs.cmu.edu/ojs/index.php/jnrsas/article/view/2/2>
6. *Bas de Haas W., Rohrmeier M., Veltkamp R., Wiering F.* Modeling Harmonic Similarity Using a Generative Grammar of Tonal Harmony // Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009) (Kobe, Japan, October 26–30, 2009) URL: <http://ismir2009.ismir.net/proceedings/OS7-2.pdf>
7. *Bas de Haas W., Veltkamp R., Wiering F.* Tonal Pitch Step Distance: a Similarity Measure for Chord Progressions // Proceedings of the Ninth International Conference on Music Information Retrieval (ISMIR 2008) (Philadelphia, Pennsylvania, September 14–18, 2008) Pp. 51–56.
8. *Bergeron M., Conklin D.* Structured Polyphonic Patterns // Proceedings of the Ninth International Conference on Music Information Retrieval (ISMIR 2008) (Philadelphia, Pennsylvania, September 14–18, 2008) Pp. 69–74.
9. *Bishop C.* Pattern Recognition and Machine Learning. Springer, 2006. 738 pp.
10. *Blei D., Jordan M.* Variational Inference for Dirichlet Process Mixtures // Journal of Bayesian Analysis. 2006. № 1(1). Pp. 121–144.
11. *Cambouropoulos E.* Voice and stream: Perceptual and computational modeling of voice separation // Music Perception. 2006. № 26. Pp. 75–94.
12. *Castelluccio M.* The Music Genome Project // Strategic Finance. 2006. № 88 (6). Pp. 57–58.
13. *Cemgil A.T.* Bayesian Music Transcription. Ph.D. Thesis, Radboud University of Nijmegen, Netherlands, 2004. 119 pp.
14. *Cemgil A.T., Kappen B.* Monte Carlo methods for tempo tracking and rhythm quantization. // Journal of Artificial Intelligence Research. 2003. 18. Pp. 45–81.
15. *Cemgil A.T., Kappen B., Barber D.* A generative model for music transcription // IEEE Transactions on Speech and Audio Processing. 2005. 14 (2). Pp. 679–694.
16. *Cemgil A.T., Kappen B., Desain P., Honing H.* On tempo tracking: Tempogram representation and Kalman filtering // Journal of New Music Research. 2000. 29. Pp. 259–273.
17. *Chai W., Vercoe B.* Folk music classification using hidden Markov models // In Proceedings of the International Conference on Artificial Intelligence (Las Vegas, USA, 2001). Las Vegas: CSREA Press, 2001. URL: <http://pubs.media.mit.edu/?section=docdetail&id=210134&collection=Media+Lab&filtercollection=Media+Lab>
18. *Davy M.* Multiple fundamental frequency estimation based on generative models // In A.P. Klapuri & M. Davy (Eds.), Signal Processing Methods for Music Transcription (pp. 203–227). New York: Springer, 2006. 452 pp.
19. *de la Cuadra P., Master A., Sapp C.* Efficient pitch detection techniques for interactive music // In Proceedings of the 2001 International Computer Music Conference (La Habana, Cuba, 2001) URL: <http://quod.lib.umich.edu/cgi/p/pod/dod-idx?c=icmc;idno=bbp2372.2001.061>
20. *Dempster A.P., Laird N.M., Rubin D.B.* Maximum Likelihood from Incomplete Data via the EM Algorithm // Journal of the Royal Statistical Society. Series B (Methodological). 1977. № 39 (1). Pp. 1–38.
21. *Dixon S., Cambouropoulos E.* Beat tracking with musical knowledge // In Horn, W. (Ed.), Proceedings of the 14th European Conference on Artificial Intelligence (ECAI 2000) (Berlin, Germany, August 20–25, 2000). Berlin, IOS Press, 2000. Pp. 626–630.
22. *Doornbusch P.* Computer Sound Synthesis in 1951 — The Music of CSIRAC // Computer Music Journal. 2004. Volume 28 Issue 1. URL: <http://www.mitpressjournals.org/doi/pdf/10.1162/014892604322970616>

23. *Gelfand A.E., Smith A.F.M.* Sampling-Based Approaches to Calculating Marginal Densities // Journal of the American Statistical Association. 1990. № 85 (410). Pp. 398–409.
24. *Godsill S.J., Rayner P.J.W.* Digital Audio Restoration — A Statistical Model-Based Approach. Springer-Verlag, 1998. 328 pp.
25. *Grubb L.* A Probabilistic Method for Tracking a Vocalist. Ph.D. thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, 1998. 258 pp.
26. Amazon.com: Online Shopping for Electronics, Apparel, Computers, Books, DVDs & more. URL: <http://amazon.com>
27. Last.fm - Listen to internet radio and the largest music catalogue online. URL: <http://last.fm>
28. Pandora Radio — Listen to Free Internet Audio, Find New Music. URL: <http://pandora.com>
29. Apple - iTunes - Everything you need to be entertained. URL: <http://www.apple.com/itunes/>
30. Netflix - Watch TV Shows Online, Watch Movies Online. URL: <https://www.netflix.com/>
31. *Hunter D.R., Lange K.* A Tutorial on MM Algorithms // The American Statistician. 2004. № 58. Pp. 30–37.
32. *Hunter D.R.* MM algorithms for generalized Bradley-Terry models // The Annals of Statistics. 2004. № 32 (1). Pp. 384–406.
33. *Jaakkola T.S.* Tutorial on variational approximation methods / In Opper, M. and Saad, D. (eds.), Advanced mean field methods: theory and practice. MIT Press, 2000. Pp. 129–159.
34. *Jackman S.* Estimation and Inference via Bayesian Simulation: An Introduction to Markov Chain Monte Carlo / American Journal of Political Science. 2000. № 44(2). Pp. 375–404.
35. *Jaśra A., Holmes C.C., Stephens D.A.* Markov Chain Monte Carlo Methods and the Label Switching Problem in Bayesian Mixture Models / Statistical Science. 2005. № 20(1). P. 50.
36. *Jordan M.L., Ghahramani Z., Jaakkola T.S., Saul L.K.* An Introduction to Variational Methods for Graphical Models / Machine Learning. 1998. № 37. Pp. 183–233.
37. *Joyce J.* Pandora and the Music Genome Project / Scientific Computing. 2006. № 23 (10). Pp. 14, 40–41.
38. *Kashino K., Nakadai K., Kinoshita T., Tanaka H.* Application of Bayesian probability networks to musical scene analysis / In D.F. Rosenthal & H.G. Okuno (Eds.), Computational Auditory Scene Analysis (pp. 115–137). Mahwah, NJ: Lawrence Erlbaum, 1998. 416 pp.
39. *Kirlin P., Utgoff P.* VOISE: Learning to segregate voices in explicit and implicit polyphony // In J. Reiss & G. Wiggins (Eds.), Proceedings of the Sixth International Conference on Music Information Retrieval (ISMIR 2005) (London, UK, September 11–15, 2005). London, UK: University of London, 2005. Pp. 552–557.
40. *Klapuri A.P.* Automatic music transcription as we know it today / Journal of New Music Research. 2004. № 33. Pp. 269–282.
41. *Klapuri A.P., Davy M.* Signal Processing Methods for Music Transcription. New York: Springer, 2006. 452 pp.
42. *Kostka S., Payne D.* Workbook for Tonal Harmony. New York: McGraw Hill, 1995.
43. *Kröger P., Passos A., Sampaio Marcos, de Cidra Givaldo.* Rameau: A System For Automatic Harmonic Analysis // In Proceedings of ICMC 2008 (Belfast, UK). URL: <http://quod.lib.umich.edu/i/icmc/bbp2372.2008.077?view=toc>
44. *Lee K., Slaney M.* A Unified System for Chord Transcription and Key Extraction Using Hidden Markov Models // Proceedings of the Fourth International Conference on Music Information Retrieval (ISMIR 2003) (Baltimore, Maryland (USA), 26–30 October 2003). Pp. 245–250.
45. *Lerdahl F., Jackendoff R.* A Generative Theory of Tonal Music. Cambridge: MIT Press, 1983. 368 pp.

46. *Longuet-Higgins H.C.* Mental Processes: Studies in Cognitive Science. MIT Press, Cambridge, 1987. 424 pp.
47. *Longuet-Higgins H.C.* The perception of melodies // *Nature*. 1976. № 263. Pp. 646–653.
48. *Longuet-Higgins H.C., Lee C.* Perception of musical rhythms // *Perception*. 1982. № 11(2). Pp. 115–128.
49. *Mackay D.J.* Information Theory, Inference, and Learning Algorithms. Cambridge University Press, 2003. 640 pp.
50. *Marques G., Lopes M., Sordo M., Langlois T., Gouyon F.* Additional Evidence That Common Low-Level Features of Individual Audio Frames Are Not Representative of Music Genre // Proceedings of the Eleventh International Society for Music Information Retrieval Conference (ISMIR 2010) (Utrecht, Netherlands, August 9–13, 2010). URL: <http://www.mtg.upf.edu/system/files/publications/20.pdf>
51. *Mauch M., Dixon D., Harte Ch., Casey M., Fields B.* Discovering Chord Idioms Through Beatles and Real Book Songs // Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007) (Vienna, Austria, September 23-27) Pp. 255–258.
52. *Minka T.* A family of algorithms for approximate Bayesian inference. PhD thesis, Massachusetts Institute of Technology, 2001. 75 pp.
53. *Minka T.* Expectation propagation for approximate Bayesian inference // In Breese, Jack S. and Koller, Daphne (eds.), Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence (Seattle, WA, August 2–5, 2001). Massachusetts: Morgan Kaufmann, 2001. Pp. 362–369.
54. *Mont-Reynaud B.* Problem-Solving Strategies in a Music Transcription System // In Proceedings of the 1985 International Joint Conference on Artificial Intelligence (Los Angeles, California, August 18–23, 1985). Pp. 916–918.
55. *Moorer J.* On the transcription of musical sound by computer // *Computer Music Journal*. 1977. № 1(4). Pp. 32–38.
56. *Mouton R., Pachet F.* The Symbolic vs. Numeric Controversy in Automatic Analysis of Music // In: Widmer G. (ed.); Artificial Intelligence and Music (IJCAI–95 Workshop Program Working Notes). 1995. Pp. 32–40.
57. *Neal R., Hinton G. Michael I. Jordan. ed.* A view of the EM algorithm that justifies incremental, sparse, and other variants // *Learning in Graphical Models*. Cambridge, MA: MIT-Press, 1999. Pp. 355–368.
58. *Paiement J-F., Eck D., Bengio S.* A Probabilistic Model for Chord Progressions // In J. Reiss & G. Wiggins (Eds.), Proceedings of the Sixth International Conference on Music Information Retrieval (ISMIR 2005) (London, UK, September 11–15, 2005). London, UK: University of London, 2005. Pp. 312–319.
59. *Pearl J.* Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference (2nd ed.) San Francisco, CA: Morgan Kaufmann, 1988. 552 pp.
60. *Pickens J., Crawford T.* Harmonic Models for Polyphonic Music Retrieval // Proceedings of International Conference on Information and Knowledge Management (CIKM 2002) (McLean, VA, November 4–9, 2002). Pp. 430–437.
61. *Pressing J., Lawrence P.* Transcribe: A comprehensive autotranscription program // In Proceedings of the International Computer Music Conference (Tokyo, Japan, 1993). Pp. 343–345.
62. *Rabiner L.R., Chen M.J., Rosenberg A.E., McGonega A.* A comparative study of several pitch detection algorithms // *IEEE Transactions on Acoustics, Speech and Signal Processing*. 1976. № 23. Pp. 552–557.
63. *Raphael C.* A mixed graphical model for rhythmic parsing // In Breese, Jack S. and Koller, Daphne (eds.), Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence (Seattle, WA, August 2–5, 2001). Massachusetts: Morgan Kaufmann, 2001. Pp. 462–471.

64. *Raphael C.* Automatic transcription of piano music // In 3rd International Conference on Music Information Retrieval (Paris, France, October 13–17, 2002). Pp. 15–19.
65. *Raphael C., Stoddard J.* Harmonic Analysis With Probabilistic Graphical Models // In Proceedings of the Fourth International Conference on Music Information Retrieval (ISMIR 2003) (Baltimore, Maryland (USA), 26-30 October 2003). URL: <http://ismir2003.ismir.net/papers/Raphael.PDF>
66. *Roads C.* Research in Music and Artificial Intelligence // ACM Computing Surveys. 1985. № 17(2). Pp. 163–190.
67. *Robert C.P., Casella G.* Monte Carlo Statistical Methods (second edition). New York: Springer-Verlag, 2004. 645 pp.
68. *Rosenthal D.* Emulation of human rhythm perception // Computer Music Journal. 1992. № 16(1). Pp. 64–76.
69. *Rosenthal D.* Machine Rhythm: Computer Emulation of Human Rhythm Perception. Ph.D. Thesis, Department of Computer Science, Massachusetts Institute of Technology, 1992. 139 pp.
70. *Scaringella N., Zoia G., Mlynek D.* Automatic genre classification of music content: a survey // IEEE Signal Processing Magazine. 2006. vol. 23, no. 2. Pp. 133–141.
71. *Scheirer E.D.* Music-Listening Systems. Ph.D. thesis, Massachusetts Institute of Technology, 2000. 248 pp.
72. *Scheirer E.D.* Tempo and beat analysis of acoustic musical signals // Journal of Acoustical Society of America. 1998. № 103:1. Pp. 588–601.
73. *Sterian A.* Model-Based Segmentation of Time-Frequency Images for Musical Transcription. Ph.D. thesis, University of Michigan, Ann Arbor, 1999. 149 pp.
74. *Sumi K., Itoyama K., Yoshii K., Komatani K., Ogata T., Okuno H.* Automatic Chord Recognition Based on Probabilistic Integration of Chord Transition and Bass Pitch Estimation // Proceedings of the Ninth International Conference on Music Information Retrieval (ISMIR 2008) (Philadelphia, Pennsylvania, September 14–18, 2008) Pp. 39–44.
75. *Temperley D.* A Unified Probabilistic Model of Polyphonic Music Analysis // Journal of New Music Research. 2009. № 38. Pp. 3–18.
76. *Temperley D.* Bayesian Models of Musical Structure and Cognition // Musicae Scientiae. 2004. № 8 (2). Pp. 175–205.
77. *Temperley D.* Music and Probability. MIT Press, 2007. 256 pp.
78. *Temperley D.* The Cognition of Basic Musical Structures. Cambridge, MA: MIT Press, 2001. 422 pp.
79. *Temperley D., Sleator D.* Modeling meter and harmony: a preference-rule approach // Computer Music Journal. 1999. № 23(1). Pp. 10–27.
80. *Temperley, D.* An algorithm for harmonic analysis / Music Perception. 1997. № 15. Pp. 31–68.
81. *Terhardt E.* Pitch, consonance and harmony // Journal of the Acoustical Society of America. 1974. № 55(5). Pp. 1061–1069.
82. *Todd N.P.M.* The auditory “primal sketch”: A multiscale model of rhythmic grouping // Journal of new music Research. 1994. vol.23, № 1. Pp. 25–70.
83. *Tsui, V.W.S.* Harmonic analysis using neural networks. Master’s thesis, University of Toronto, 2002. 96 pp.
84. *Walmsley P.J.* Signal Separation of Musical Instruments. Ph.D. thesis, University of Cambridge, 2000. 237 pp.
85. *Weiss Y., Freeman W.T.* Correctness of Belief Propagation in Gaussian Graphical Models of Arbitrary Topology // Neural Computation. 2001. № 13 (10). Pp. 2173–2200.

86. *Yedidia J.S., Freeman W.T.* Understanding Belief Propagation and Its Generalizations / In Lakemeyer, Gerhard; Nebel, Bernhard. Exploring Artificial Intelligence in the New Millennium. Massachusetts: Morgan Kaufmann, 2003. Pp. 239–236.
87. *Zhang X., Gerhard D.* Chord Recognition using Instrument Voicing Constraints // Proceedings of the Ninth International Conference on Music Information Retrieval (ISMIR 2008) (Philadelphia, Pennsylvania, September 14–18, 2008) Pp. 33–38.
88. *Балтийский И.А., Николенко С.И.* Графическая вероятностная модель для оценки схожести гармонии музыкальных произведений // Труды СПИИРАН. СПб.: Наука, 2011.
89. *Тулупьев А.Л., Николенко С.И., Сироткин А.В.* Байесовские сети доверия: логико-вероятностный вывод в ациклических направленных графах. СПб.: Изд-во С.-Петерб. ун-та, 2009, 400 с.
90. *Тулупьев А.Л.* Алгебраические байесовские сети: реализация логико-вероятностного вывода в комплексе java-программ // Труды СПИИРАН. СПб.: Наука, 2009. Вып. 8. С. 191–232.
91. *Тулупьев А.Л.* Задача локального автоматического обучения в алгебраических байесовских сетях: логико-вероятностный подход // Труды СПИИРАН. 2008. Вып. 7. СПб.: Наука, 2008. С. 11–25.
92. *Тулупьев А.Л., Николенко С.И., Сироткин А.В.* Байесовские сети: логико-вероятностный подход. СПб.: Наука, 2006. 608 с.
93. *Фильченков А.А., Тулупьев А.Л.* Структурный анализ систем минимальных графов смежности // Труды СПИИРАН. 2009. Вып. 11. С. 104–127.

Балтийский Игорь Андреевич — магистр прикладной математики и информатики, студент Санкт-Петербургского государственного института информационных технологий, механики и оптики (ИТМО). Область научных интересов: информационный поиск музыки, вероятностные модели, теория музыки, цифровая обработка сигналов. Число научных публикаций — 2. iosank@gmail.com; СПбГУ ИТМО, ул. Саблинская, д. 14, г. Санкт-Петербург, 197101; +7(921)792-73-22. Научный руководитель — С.И. Николенко.

Baltiyskiy Igor Andreevich — M. Sc. in Applied Mathematics and Informatics, student, Saint-Petersburg State Institute of Fine Mechanics and Optics (Technical University IFMO). Research interests: music information retrieval, probabilistic models, music theory, digital signal processing. Number of publications — 2. iosank@gmail.com; Technical University IFMO, Russia, 197101, Sablinskaya street, 14; +7(921)792-73-22. Supervisor: S.I. Nikolenko.

Николенко Сергей Игоревич — к.ф.-м.н.; научный сотрудник лаборатории математической логики УРАН Санкт-Петербургского отделения Математического института им. В.А. Стеклова РАН, старший научный сотрудник проблемной лаборатории алгоритмической биологии УРАН Санкт-Петербургского Академического Университета — научно-образовательного центра нанотехнологий РАН. Число научных публикаций — 60; sergey@logic.pdmi.ras.ru, <http://logic.pdmi.ras.ru/~sergey>; наб. р. Фонтанки, 27, Санкт-Петербург, 191023, Россия; р.т. +7(812)3124058, факс +7(812)3105377.

Nikolenko Sergey Igorevich — Ph.D.; Researcher at the Laboratory of Mathematical Logic, Steklov Mathematical Institute, St. Petersburg, Senior Researcher at the Laboratory of Algorithmic Biology, St. Petersburg Academic University. Number of publications — 60; sergey@logic.pdmi.ras.ru, <http://logic.pdmi.ras.ru/~sergey>; nab. r. Fontanki, 27, St. Petersburg, 191023, Russia; tel. +7(812)3124058, fax +7(812)3105377.

Поддержка исследований. Работа второго автора была поддержана грантом Правительства РФ для государственной поддержки научных исследований, проводимых под руководством ведущих ученых (грант № 11.G34.31.0018), грантом Президента РФ для молодых кандидатов наук (грант МК-4089.2010.1), грантом Президента для ведущих научных школ НШ-5282.2010.1, ФЦП «Научные и научно-педагогические кадры инновационной России» и Российским Фондом Фундаментальных Исследований.

РЕФЕРАТ

Балтийский И.А., Николенко С.И. **Обзор графических вероятностных моделей гармонии для анализа музыкальных произведений.**

Цель данной работы — описание современных вероятностных моделей, используемых для анализа гармонии. Прежде чем перейти к самому описанию, мы даем краткий обзор задач, рассматриваемых в анализе музыки, а также приводим мотивацию для исследований в описываемой области. Мотивацией служит задача схожести, на основе решения которой может быть построена полностью автоматическая система рекомендации музыки, ориентированная на содержание. Это актуальная практическая проблема — достаточно упомянуть, что единственный широко известный коммерческий проект, реализующий рекомендацию на основе содержания, вынужден использовать ручной труд экспертов-музыковедов.

Далее приводится краткий обзор методов анализа гармонии, применяемых для задач, далеких от задачи схожести, а также методов оценки схожести композиций по гармонии, не использующих вероятностные модели. Вероятностные модели составляют предмет обсуждения всей дальнейшей части работы, поэтому ее предваряет краткое введение в математическую суть графических вероятностных моделей.

Рассматривается модель, использующая марковские цепи над векторами вероятностей появления в данный момент времени аккорда из некоторого заранее заданного набора. Обсуждаются достоинства использованного представления, указываются и недостатки — отсутствие контекста тональности в модели аккордов. Далее описывается несколько скрытых марковских моделей, в которых, в частности, появляется учет текущей тональности.

Приводятся и более сложные модели: модель, включающая в себя дополнительную информацию о ритмической структуре произведения, а также модель, которая, помимо ритмической структуры, использует сегрегацию голосов. Приводятся аргументы и ссылки на эмпирические данные, подтверждающие, что такие объединенные модели позволяют лучше решать задачу анализа гармонии музыкальных произведений.

Наконец, описаны несколько встреченных в литературе оригинальных идей по улучшению работы вероятностных моделей в смежной задаче извлечения последовательности аккордов. Суть идей сводится к привлечению дополнительного знания предметной области — в одном случае, особенностей гармонизации аккордов, присущих определенным инструментам, а в другом — использование извлеченной из аудиосигнала высоты басового тона.

SUMMARY

Baltiyskiy I.A., Nikolenko S.I. **Probabilistic graphical models of harmony in automated music analysis: a survey.**

This paper surveys probabilistic models used for musical harmony analysis. We begin with a short survey of the central problems of musical analysis and provide motivation for the research in this field. Our motivation stems from the music similarity task which may serve as a basis for a completely automated content-based music recommendation system. This is a very relevant and important practical problem: the only widely popular commercial project providing content-based recommendation has had to resort to manual work of expert musicologists.

Further on we present a short survey of harmony analysis methods that are applied to problems other than similarity evaluation; we also present several methods for measuring musical harmony similarity that are not based on probabilistic models. We consider probabilistic models throughout the rest of the paper, so next we give a short introduction into the mathematical background behind the probabilistic (graphical) models.

We consider a model that uses Markov chains on the vectors of chord probabilities (chords are taken from a fixed set.) We discuss the advantages of this representation and point out some disadvantages—for one, the lack of mode or key context in the model of chords. Next, we describe several hidden Markov models, which, among other things, consider the current key as well.

More complex models are described, too: a model that includes additional information about the rhythmic structure of the piece and a model that employs both rhythm and voice segregation information. We give some arguments and empirical evidence to support the idea that such unified models tend to perform better than harmonic, rhythmic, and stream (voice) models that constitute them.

Finally, we describe several original ideas encountered in the literature that aim to improve probabilistic models in the related task of chord sequence extraction. The essence of these ideas is to leverage additional domain knowledge—in one case, authors use chord voicing constraints inherent to some musical instruments, in another—bass tone estimation (in addition to chroma vectors).