

МЕТОД АВТОМАТИЧЕСКОГО ОПРЕДЕЛЕНИЯ ХАРАКТЕРИСТИК МУЗЫКАЛЬНОЙ КОМПОЗИЦИИ

А. В. Курьян, А. И. Парамонов

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

Аннотация. Предлагается новый метод автоматического определения характеристик музыкальных композиций. Метод основан на применении модифицированного алгоритма дискретного преобразования Фурье с оконной функцией Блэкмана-Наттала. Дополнительно при определении темпа выполняется анализ корреляции полученного частотного спектра со спектром эталонных нот, которые используются для настройки музыкальных инструментов. Один из результатов работы авторского метода – это нотное представление музыкальной композиции. Приближение нотного представления анализируется с использованием статистических данных о среднем процентном соотношении нот разных относительных длительностей в музыкальных композициях. Нотное представление используется как входные данные для алгоритма Крумхансля-Шмуклера с целью определения тональности композиции. Проведена серия компьютерных экспериментов, которые показали эффективности предложенного метода в сравнении с аналогами. Для проведения эксперимента создано программное средство, которое выполняло сбор данных и визуализацию работы метода. Приведено описание прикладного программного продукта, который разработан для тестирования метода на широкой аудитории. Приложение реализовано в форме настольного клиента известного стримингового аудио веб-сервиса. Приложение позволяет получить значение темпа и тональности для композиций, которые скачиваются на устройство пользователя. Эти характеристики используются для подбора набора композиций с похожим темпом по предпочтениям пользователей.

Ключевые слова: характеристики звука, BPM, тональность, частотно-временной анализ, дискретное преобразование Фурье, алгоритм Крумхансля-Шмуклера.

Введение

Цифровая музыкальная индустрия при внедрении новых технологий сталкивается с целым рядом задач. Одной из таких задач является выявление и применение характеристик музыкальных фрагментов. В настоящее время методы анализа музыкальных композиций с точки зрения музыкальной нотации являются одним из быстро развивающихся направлений в сфере информационных технологий. Подобные методы могут использоваться для получения значений таких характеристик как темп, преобладающая высота мелодии, тональность, некоторое приближение нотного представления композиции и прочее. Существует даже ежегодный конкурс MIREX (проводится Школой информационных наук Университета Иллинойса), на котором различные алгоритмы тестируются на их способность к определению темпа композиции.

Темп, или удары в минуту (Beats per minute, BPM) в музыке – это показатель, определяющий скорость исполнения или воспроизведения композиции [1]. С точки зрения музыкальной нотации представляет собой количество четвертных долей ноты, потенциально способных разместиться в одной минуте. Для большинства существующих композиций значение данного параметра лежит в диапазоне от 30 до 240. Также существует соответствие между диапазонами значений BPM и традиционными обозначениями темпа, используемыми музыкантами, такими как граве, лярго, анданте, модерато и т. д. [2].

Большинство известных сегодня методов определения BPM для нахождения пиков энергии, которые принимаются за единичные биты, применяют анализ зависимости энергии сигнала от времени. Однако если принимать во внимание понятие BPM с точки зрения музыкальной нотации, потенциально существует возможность добиться большей точности.

Тональность композиции состоит из двух компонентов: лада и тоники. Лад определяет характер звучания мелодии: яркий, радостный в случае мажорного лада и печальный, драматичный в случае минорного. Тоника определяет доминирующий тон и чаще всего является нотой, завершающей фрагмент композиции соответствующей тональности.

Для определения тональности широко применим алгоритм Крумхансля-Шмуклера [3], точность работы которого напрямую зависит от корректности входных данных, представляющих собой суммарную длительность каждой отдельной ноты. Таким образом, повышение эффективности решения задачи определения тональности сводится к задаче определения длительностей отдельных нот, что может выступать одним из этапов алгоритма определения ВРМ.

Постановка задачи

В работе исследуется проблема выявления характеристик музыкальной композиции и предлагается метод определения ВРМ, точность работы которого превышает аналоги. С целью подтверждения работы методов следует провести серию экспериментов, в рамках которых предполагается разработка программного обеспечения на основе предложенного метода. Исходными данными для экспериментов выступают музыкальные композиции, которые получены из аудио сервисов в открытом доступе.

Метод определения характеристик музыкальной композиции

С точки зрения частотного представления сигнала, каждой ноте соответствует одна основная или фундаментальная частота, а также несколько обертонов, количество, амплитуда и смещение которых по частоте во многом зависят от музыкального инструмента. В общем случае частоты всех обертонов примерно кратны частоте основного тона [4], что позволяет с некоторой точностью смоделировать частотный спектр для любой из 12 нот в выбранной октаве. Более точного представления можно достичь, если использовать имеющиеся рассчитанные значения частот для каждого музыкального инструмента, что негативно скажется на универсальности метода при работе с композициями разных жанров.

При разработке авторского метода принято решение, что первые пять обертонов соответствуют эталонным частотам фортепиано, как инструмента, имеющего наибольший частотный диапазон. Следует отметить, что для первых пяти обертонов фортепиано отклонение от теоретической частоты не превышает 13,7 Гц. Это позволяет использовать те же значения частот и для определения нот других музыкальных инструментов с некоторой большей погрешностью.

Для разложения исходного сигнала на спектр частот используется алгоритм оконного дискретного преобразования Фурье (ДПФ). Данный алгоритм позволяет представить сигнал на заданном интервале в виде суммы синусоидальных составляющих вида:

$$f(x) = A * \cos(2\pi f_0 x + \varphi_0),$$

где A – амплитуда; f_0 – частота; φ_0 – начальная фаза гармоник.

В отличие от обычного дискретного преобразования Фурье оконное ДПФ вычисляется из произведения исходного сигнала на ограниченном интервале и некоторой оконной функции, предназначенной для уменьшения эффекта растекания ДПФ [5]. В зависимости от конкретной задачи могут использоваться разные виды оконных функций, улучшающие значения одних параметров и ухудшающие значения других, менее существенных в данной задаче. В авторском методе используется оконная функция Блэкмана-Наттала, которая обладает более низким уровнем боковых лепестков спектральной плотности мощности по сравнению с другими оконными функциями. Данная функция определяется формулой:

$$w(m) = a_0 - a_1 \cos(2\pi m/N) + a_2 \cos(4\pi m/N) - a_3 \cos(6\pi m/N),$$

где N – размер окна ДПФ; $a_0 = 0,3635819$; $a_1 = 0,4891775$; $a_2 = 0,1365995$; $a_3 = 0,0106411$.

Однако главной проблемой ДПФ является принцип неопределенности Гейзенберга, в основе которого лежит тот факт, что невозможно точно определить, какая частота присутствует в сигнале в данный момент времени (можно говорить только про диапазон частот), и невозможно определить, в какой точно момент времени частота присутствует в сигнале (можно говорить лишь про период времени) [6]. В связи с этим возникает проблема разрешающей способности.

На рисунке 1а представлен результат оконного ДПФ с узким окном с масштабом 0,01. За

исходный пример взят сигнал общей длительностью 1 с, являющийся стационарным каждые 250 мс. На первом отрезке он имеет частоту 300Гц, на втором – 200Гц, на третьем – 100Гц и на четвертом – 50Гц. Как можно видеть, полученное преобразование имеет высокую точность относительно времени и низкую точность относительно частоты (каждый максимум занимает некоторый диапазон частот).

На рисунке 1б приведен результат преобразования с более широким окном с масштабом 0,00001. В этом случае очевидна более высокая точность относительно частоты, но при этом более низкая точность относительно времени.

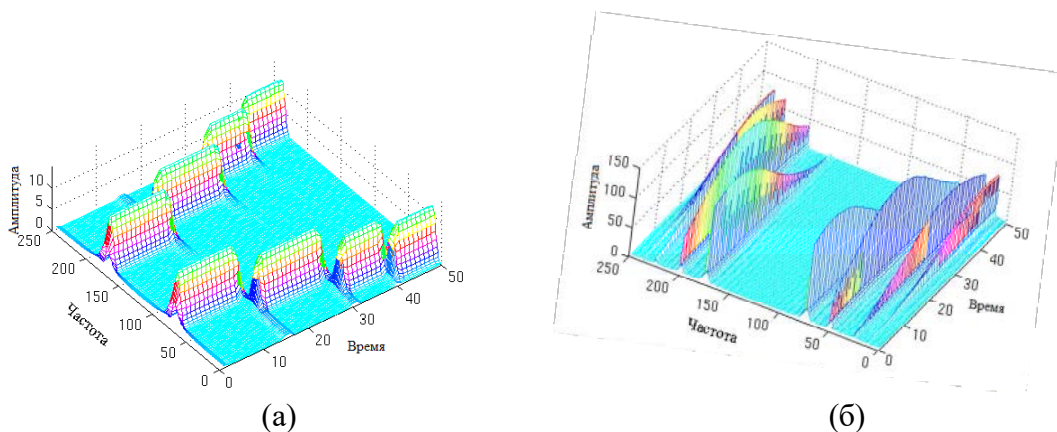


Рис. 1. Результат оконного ДПФ:
а) с окном в масштабе 0,01; б) с окном в масштабе 0,00001

Поскольку для определения ВРМ необходимо как высокое разрешение по частоте (для определения наличия тех или иных нот), так и высокое разрешение по времени (для определения абсолютной длительности нот), то в авторском методе используется разложение сигнала на частотном спектре в два этапа.

На первом этапе осуществляется оконное преобразование Фурье с окном шириной 16384 отсчета, что при частоте дискретизации в 44,1 КГц, являющейся наиболее распространенной в современных музыкальных композициях, обеспечивает разрешение в 2,69 Гц по частоте и 371,5 мс по времени. На втором этапе производится ДПФ с узким окном шириной 512 отсчетов, обеспечивающим разрешение по времени в 11,6 мс и 86,13 Гц по частоте. После чего спектр, полученный преобразованием с окном меньшего размера, состоящий из меньшего количества отсчетов, аппроксимируется многочленом Лагранжа четвертого порядка до размерности спектра, полученного ДПФ с окном большего размера. В конце вычисляется поточечное произведение двух спектров с целью получения максимальных значений в тех точках, где максимальную амплитуду имеют как спектр с большим разрешением по частоте, так и спектр с большим разрешением по времени. Результат вычисления произведения принимается за итоговый спектр, который затем анализируется на возможность наличия тех или иных нот. Графическое представление одного шага преобразования представлено на рисунке 2. На рисунке приведен частотный спектр исходного сигнала, полученный путем ДПФ окном большего размера (синий график), окном меньшего размера, аппроксимированный до размерности большего спектра (зеленый график), и результат поточечного умножения обоих спектров (красный график). Следует отметить, что для наглядности масштаб итогового графика отличается в большую сторону в 4,5 раза по сравнению с масштаб графиков отдельных спектров. Заметно, что благодаря использованию произведения двух спектров пики, соответствующие частотам, локализованным в области ДПФ с окном меньшего размера, становятся более выраженными, чем те, которые находятся вне этой области, но в пределах участка, ограниченного окном большего размера.

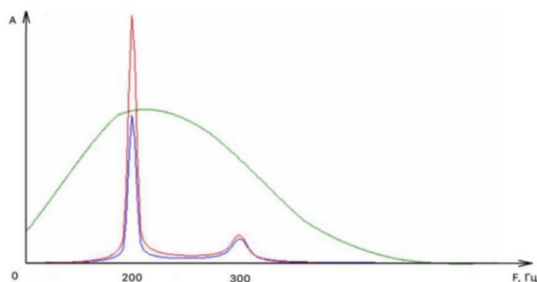


Рис. 2. Результат вычислений на одном шаге ДПФ

На следующем этапе производится определение вероятности наличия тех или иных нот в спектре. Для этого на каждом шаге ДПФ рассчитывается корреляционный момент полученного спектра и смоделированного спектра каждой из нот (12 нот в девяти октавах – всего 108 нот). После чего строится гистограмма, по оси абсцисс которой отложены ноты, а по оси ординат – результат вычисления корреляционного момента. Пример одной из таких гистограмм представлен на рисунке 3.

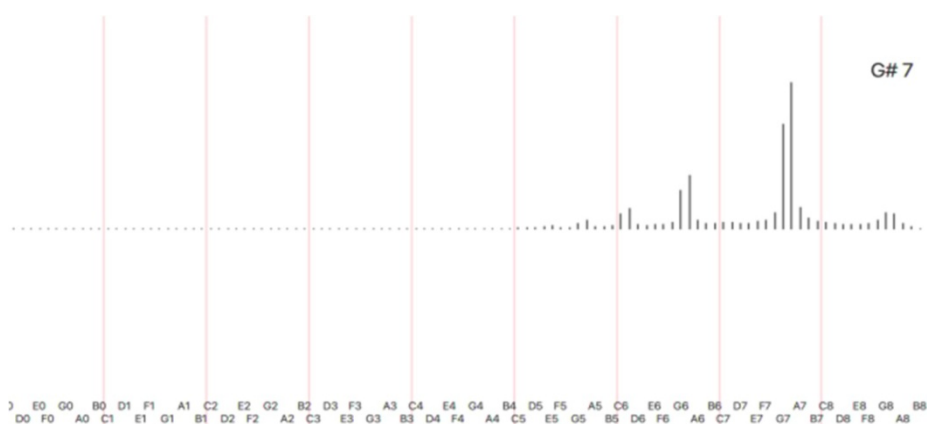


Рис. 3. Гистограмма вероятности наличия нот в спектре

Для получения ограниченного количества нот с наибольшими значениями корреляционного момента на данном интервале в начале рассчитывается среднее значение амплитуд для всех частот в спектре. Затем полученное значение умножается на фиксированный коэффициент, наилучшим значением которого эмпирическим способом было подобрано число 26. Ноты, значения корреляционного момента для которых превышают результат произведения, принимаются за искомые на интервале, ограниченном шириной окна ДПФ меньшего размера.

В качестве результата предыдущих двух этапов получаем некоторое приближение нотной записи в виде расположения нот разных частот на временной шкале. Выполнив сумму длительности всех участков, на которых присутствует каждая из нот, можно получить суммарную длительность отдельных нот в композиции, что является исходными данными для алгоритма Крумхансля-Шмуклера. Данный алгоритм основан на использовании фиксированных последовательностей, называемых профилями, состоящих из 12 вещественных значений, соответствующих каждой из 12 нот [3]. Для значений каждого профиля вычисляется коэффициент корреляции со значениями суммарных длительностей, полученных на предыдущем этапе. После этого последовательность значений профиля циклически сдвигается на одну позицию, и вычисление повторяется. В алгоритме используются два профиля: мажорный и минорный, следовательно, всего требуется вычислить 24 коэффициента корреляции. Каждый из них соответствует вероятности наличия тональности, лад которой определяется типом профиля, а тоника - нотой, значение которой было первым в сдвинутой последовательности значений профиля. За искомую тональность принимается та, для которой коэффициент корреляции оказался наибольшим.

Однако полученных таким образом данных недостаточно для определения BPM. Это связано с тем, что одной и той же длительности в секундах при разных значениях BPM могут

соответствовать несколько относительных длительностей. Следовательно, имея только абсолютную длительность нот, невозможно однозначно определить их относительную длительность. Решение данной проблемы рассматривается в исследовании [7]. В своей работе ученые из университета Коимбры предлагают вывести среднее процентное соотношение относительных длительностей нот в музыкальных композициях.

В соответствии с этим исследованием для двух классов ритма мелодии (веселая/агрессивная и грустная/спокойная) характерно наличие двух вариантов соотношений восьмых, четвертных, половинных и целых нот. Имея это соотношение, можно последовательно сопоставлять абсолютные длительности нот, полученные в результате предыдущих этапов метода, к относительным длительностям, для которых имеются процентные соотношения. На каждом шаге минимальной имеющейся относительной длительности сопоставляется очередное в порядке возрастания абсолютное значение в секундах. Абсолютные длительности, соответствующие последующим относительным длительностям, получаются путем умножения минимальной абсолютной длительности на степень числа два. При этом ноты с абсолютными длительностями, значения которых оказались в интервале между двумя полученными эталонными значениями, включаются в число нот с относительной длительностью, соответствующей ближайшему абсолютному значению. Затем рассчитываются процентные соотношения количества нот всех имеющихся относительных длительностей к общему количеству нот и оцениваются отклонения полученных значений от средних значений, имеющих в исследовании. Значение абсолютной длительности четвертной ноты на том шаге, где было получено минимальное отклонение, принимается за итоговое. Данная последовательность действий повторяется с обеими вариантами соотношений, после чего выбирается тот, для которого отклонение было минимальным. Таким образом, помимо определения темпа, можно также провести классификацию композиций по ритму на веселые/агрессивные и грустные/спокойные.

Искомое значение ВРМ вычисляется отношением:

где $duration$ – длительность четвертной ноты, выраженная в окнах дискретного преобразования Фурье; $windowSize$ – размер окна ДПФ в отсчетах сигнала; $sampleRate$ – частота дискретизации исходного сигнала.

Для большей точности на протяжении всей композиции рассчитываются восемь отдельных значений ВРМ в интервалах частот, ограниченных частотами нот “Ля” в соседних октавах. Таким образом, делается попытка исключить одновременный учет как самих нот, так и их обертонов. В качестве итогового значения принимается минимальное значение ВРМ среди всех интервалов.

Компьютерный эксперимент

С целью оценки эффективности предложенного метода было разработано экспериментальное программное средство на языке программирования Kotlin, которое в режиме реального времени отображает характеристики обрабатываемой музыкальной композиции. Программное средство имеет графический интерфейс, что позволило визуально оценивать отдельные этапы обработки композиции. Интерфейс отображает такие данные как: график зависимости амплитуды от времени; диаграмму вероятности наличия нот в спектре для текущего участка и соотношение длительностей, выраженных в окнах ДПФ к количеству нот таких длительностей, обнаруженных на текущий момент; график зависимости амплитуды от частоты для текущего анализируемого участка. Пример работы при обработке музыкального фрагмента представлен на рисунке 4.

Для проведения эксперимента по сравнению точности авторского метода с точностью известных подходов были подготовлены две выборки по десять композиций с заранее установленными значениями ВРМ. Первая выборка была сформирована исключительно из композиций, которые исполняются на фортепиано. Для этого были использованы материалы каналов AtinPiano [8] и Patrik Pietschmann [9] видеохостинга YouTube. Вторая выборка сформирована из композиций, которые генерировались программой-синтезатором для

нескольких музыкальных инструментов и получены на канале Samuel Kim Music [10] видеохостинга YouTube. В обоих случаях выборки содержат композиции, которые имеют как можно больший диапазон эталонных значений BPM (от 60 до 240), но при этом отличаются по жанру и ритму. А в случае второй выборки разные композиции исполнены на разных музыкальных инструментах. Для сравнительного анализа в качестве ближайших аналогов были выбраны онлайн сервисы GetSongbpm [11], Tunebat [12], Vocalremover [13] и библиотека языка программирования Scala – scala-audio-file (с использованием класса SoundEnergyBPMDetector) [14]. С помощью каждого аналога были рассчитаны значения BPM для каждой из 20 композиций, а затем рассчитано отклонение полученного значения от эталонного и получено среднее отклонение для каждой выборки. Результат проведенного эксперимента сравнения в графическом виде представлен на рисунке 5.

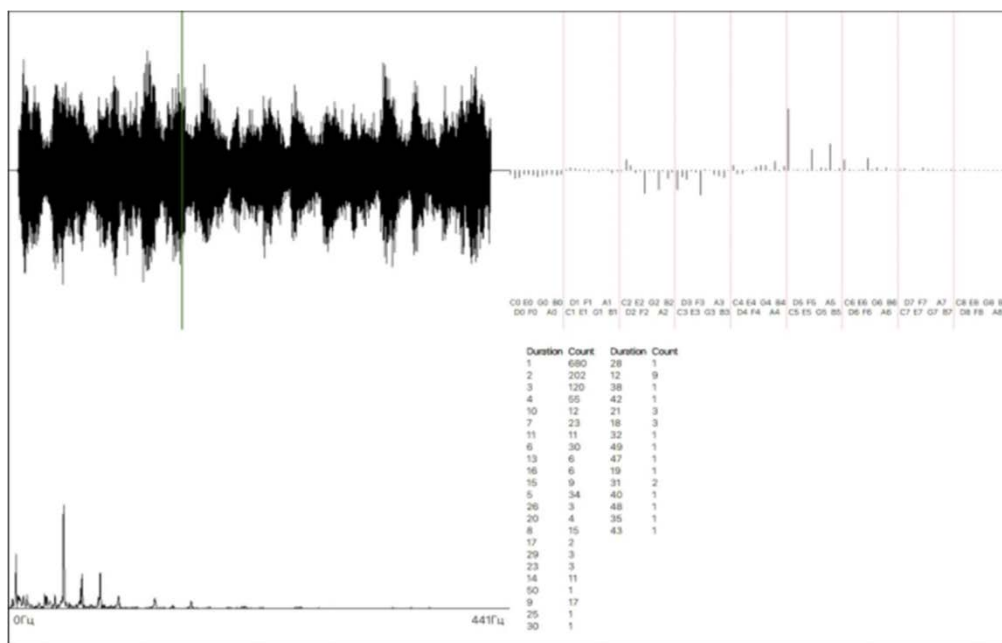


Рис. 4. Пример функционирования экспериментального программного средства

Как видно из графиков отклонений, разрабатываемый авторский метод показал лучшие результаты в данном эксперименте. Последующая серия экспериментов проводится с большей выборкой музыкальных композиций с различным темпом звучания.

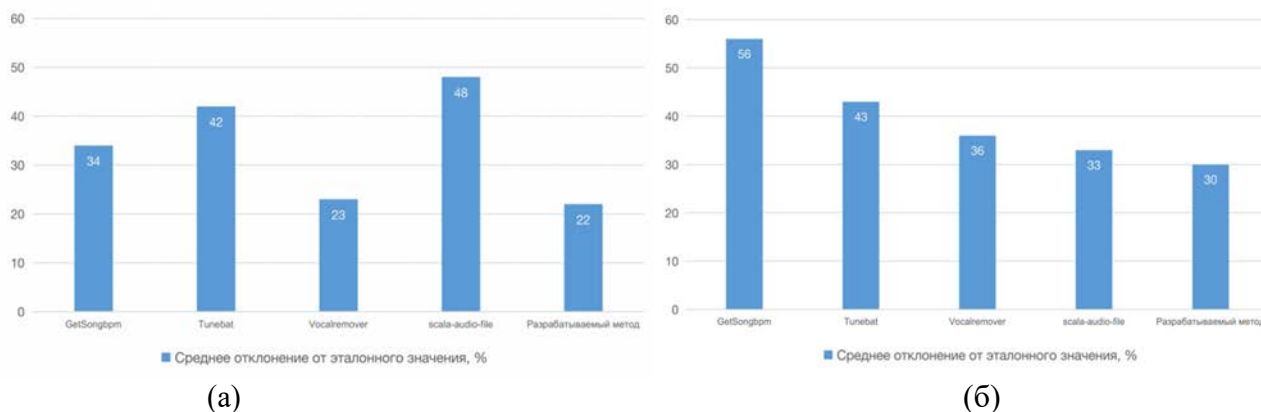


Рис. 5. Результаты тестирования для композиций:
 а) на фортепиано; б) сгенерированных синтезатором музыкальных инструментов

С целью дополнительного сбора данных и разнообразия целевой выборки для дополнительных экспериментов было разработано прикладное программное обеспечение для

пользователей музыкальных стриминговых веб-сервисов. Разработанное приложение представляет собой настольный клиент. Кроме стандартных базовых функций аудио-сервиса, которые доступны в оригинале, приложение позволяет дополнительно создавать выборки музыкальных композиций для пользователей на основании схожести их темпа звучания. Это реализуется путем расчёта значения BPM и тональности для ранее скачанных в приложении композиций. Расчет характеристик выполняется в фоновом процессе, что не вызывает задержек интерфейса. В результате расчета характеристик для всех композиций пользователя можно строить плейлисты по интересующим темпам звучания напрямую либо по схожести с выбранной композицией.

Выводы

Предложен новый метод определения значения BPM, который основан на применении модифицированного алгоритма дискретного преобразования Фурье с оконной функцией Блэкмана-Натгалла и дополнительного анализа корреляции частотного спектра.

Точность предложенного метода была проверена в ходе эксперимента при сравнении с аналогами эмпирическим путем. На основе полученного сравнения делается вывод, что разработанный метод обеспечивает большую по сравнению с аналогами точность для композиций, включающих партию фортепиано или полностью состоящих из таковой, и сравнимую с аналогами точность для других композиций.

Дополнительно в ходе компьютерного эксперимента разработано приложение, с помощью которого пользователи существующих стриминговых веб-сервисов музыки могут удовлетворить свою потребность в получении характеристик интересующих их композиций с точки зрения музыкальной нотации.

Литература

- [1] Beats per minute. The Free Dictionary by Farlex. <https://www.thefreedictionary.com/beats+per+minute> (Дата доступа: 23.10.2021).
- [2] Таблица темпов метронома. Метроном. <http://www.metronomid.ru/metronome-tempo-table> (Дата доступа: 23.10.2021).
- [3] Key-finding algorithm. Rnhart.net. <http://rnhart.net/articles/key-finding> (Дата доступа: 23.10.2021).
- [4] Цифровые Синтезаторы Музыкальных Звуков. Audio Electronics. https://astersart.net/desktop/synthesators/syntz_ms.htm (Дата доступа: 23.10.2021).
- [5] Функции оконного сглаживания. DSPLIB. <https://ru.dsplib.org/content/windows/windows.html> (Дата доступа: 23.10.2021).
- [6] The Wavelet Tutorial. Robi Polikar. <http://users.rowan.edu/~polikar/WTtutorial.html> (Дата доступа: 23.10.2021).
- [7] M. Seıça, A. Rodrigues, A. Cardoso, P. Martins, “Computer Generation and Perception Evaluation of Music-Emotion Associations”, представлено на CMMR, Марсель, Франция, Октябрь 2019.
- [8] AtinPiano. YouTube.com. <https://www.youtube.com/channel/UCkPRBxO9SbNE-2h0XXKz-jg> (Дата доступа: 23.10.2021).
- [9] Patrik Pietschmann. YouTube.com. <https://www.youtube.com/channel/UCs7O9sOUQiBGBxaaAgulwig> (Дата доступа: 23.10.2021).
- [10] Samuel Kim Music. YouTube.com. <https://www.youtube.com/user/samuelkimmusic> (Дата доступа: 23.10.2021).
- [11] GetSongbpm, <https://getsongbpm.com/tools/audio> (Дата доступа: 23.10.2021).
- [12] Tunebat. <https://tunebat.com/Analyzer> (Дата доступа: 23.10.2021).
- [13] Vocalremover. <https://vocalremover.org/bpm-finder> (Дата доступа: 23.10.2021).
- [14] Scala-audio-file. Mizccard. Дата доступа: 23.10.2021. [Онлайн]. Режим доступа: <https://github.com/mziccard/scala-audio-file>.

References

- [1] Beats per minute. The Free Dictionary by Farlex. <https://www.thefreedictionary.com/beats+per+minute> (accessed: Oct. 23, 2021).
- [2] Metronome tempo table. Metronome. <http://www.metronomid.ru/metronome-tempo-table> (accessed: Oct. 23, 2021).
- [3] Key-finding algorithm. Rnhart.net. <http://rnhart.net/articles/key-finding> (accessed: Oct. 23, 2021).
- [4] Digital Synthesizers of Musical Sounds. Audio Electronics. https://astersart.net/desktop/synthesators/syntz_ms.htm (accessed: Oct. 23, 2021).
- [5] Window smoothing functions. DSPLIB. <https://ru.dsplib.org/content/windows/windows.html> (accessed: Oct. 23, 2021).
- [6] The Wavelet Tutorial. Robi Polikar. <http://users.rowan.edu/~polikar/WTtutorial.html> (accessed: Oct. 23, 2021).
- [7] M. Seïça, A. Rodrigues, A. Cardoso, P. Martins, "Computer Generation and Perception Evaluation of Music-Emotion Associations", presented at CMMR, Marseille, France, Oct. 2019.
- [8] AtinPiano. YouTube.com. <https://www.youtube.com/channel/UCkPRBxO9SbNE-2h0XXKz-jg> (accessed: Oct. 23, 2021).
- [9] Patrik Pietschmann. YouTube.com. <https://www.youtube.com/channel/UCs7O9sOUQiGBxaaAguIwig> (accessed: Oct. 23, 2021).
- [10] Samuel Kim Music. YouTube.com. <https://www.youtube.com/user/samuelkimmusic> (accessed: Oct. 23, 2021).
- [11] GetSongbpm, <https://getsongbpm.com/tools/audio> (accessed: Oct. 23, 2021).
- [12] Tunebat. <https://tunebat.com/Analyzer> (accessed: Oct. 23, 2021).
- [13] Vocalremover. <https://vocalremover.org/bpm-finder> (accessed: Oct. 23, 2021).
- [14] Scala-audio-file. Mizccard. Accessed: Oct. 10, 2021. [Online]. Available: <https://github.com/mziccard/scala-audio-file>.

A method for automatically detecting the musical composition characteristics

Andrey Kuryan, Anton Paramonov
Belarusian State University of Informatics and Radioelectronics
(Minsk, Belarus)

Annotation. A new method for automatic detection of the musical compositions characteristics is proposed. Method is based on the application of a modified discrete Fourier transform algorithm with a Blackman-Nuttall window function. Additionally, when determining the tempo, the correlations analysis of the obtained frequency spectrum with the spectrum of benchmark notes, which are used to tune musical instruments, is carried out. One of the work results of author's method is the compositions musical notation. The approximation of musical notation is analyzed using statistical data on the average percentage of notes of different relative durations in musical compositions. The musical notation is used as input to the Krumhansl-Schmukler algorithm to determine the key to the composition. A series of computer experiments have been carried out, which have shown the effectiveness of the proposed method in comparison with analogs. For the experiment, a software tool was created that was collecting data and visualizing the method functioning. The description of the applied software product, which was developed for testing the method on a wide audience, is given. The application is developed in the desktop client form of well-known audio streaming web services. The application allows the user to get the tempo and key values for compositions that are downloaded

to his device. These characteristics are used to select a set of compositions with a similar tempo according to user preferences.

Key words: sound characteristics, BPM, tonality, time-frequency analysis, discrete Fourier transform, Krumhansl-Schmuckler algorithm.

Сведения об авторах:

Курьян Андрей Витальевич – бакалавр кафедры Программное обеспечение информационных технологий Белорусского государственного университета информатики и радиоэлектроники. E-

Парамонов Антон Иванович - кандидат технических наук, доцент, заведующий кафедрой Информационных систем и технологий Института информационных технологий Белорусского государственного университета информатики и радиоэлектроники. ORCID <https://orcid.org/0000-0001-6616-2481>. SPIN-код: 4280-3133. E-mail: a.paramonov@bsuir.by