

**ФИРСОВА С. А., МЕЛЕШКИН А. А.**

## **РАЗРАБОТКА СИСТЕМЫ ДЛЯ АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ АККОРДОВОЙ ПОСЛЕДОВАТЕЛЬНОСТИ В ЦИФРОВЫХ АУДИОФАЙЛАХ**

**Аннотация.** В статье рассматривается алгоритм распознавания аккордовой последовательности в аудиофайлах с помощью разделения частот нот на частотные классы. Описываются принципы работы и реализация разработанной авторами системы, основанной на данном алгоритме.

**Ключевые слова:** быстрое преобразование Фурье, распознавание аккордов, гармоника, спектрограмма, теория музыки, музыкальный строй, хромограмма, алгоритмическое распознавание аккордов, частотно-временное представление, частотные классы.

**FIRSOVA S. A., MELESHKIN A. A.**

## **DEVELOPMENT OF A SYSTEM FOR AUTOMATIC CHORD SEQUENCE RECOGNITION IN DIGITAL AUDIO FILES**

**Abstract.** The article considers an algorithm for recognizing a chord sequence in audio files by dividing the frequencies of notes into frequency classes. The principles of operation and implementation of the system developed by the authors based on this algorithm are described.

**Keywords:** fast Fourier transform, chord detection, harmonics, spectrogram, music theory, musical scale, chromogram, algorithmic chord recognition, time-frequency domain, pitch class profile.

**Введение.** В настоящее время вся музыкальная индустрия интенсивно использует информационные технологии для создания музыки. Для начинающих музыкальных продюсеров открылось множество способов испытывать свои художественные идеи дешево и быстро. Больше не нужно нанимать целый оркестр, чтобы добавить звучание дополнительных инструментов в создаваемый трек. Существует большое количество библиотек, содержащие сэмплы любых музыкальных инструментов, которые можно воспроизвести на midi-клавиатуре. Использование цифрового звука также позволяет анализировать звуковую дорожку путем вывода спектрограммы и при этом позволяет изменять интенсивность некоторых частот. Музыкальные инженеры имеют в своем распоряжении утилиты для настройки записанного вокала. Они могут увидеть всю последовательность нот, которую спел вокалист, а также корректировать их. Инструмент, который будет распознавать аккорды может сильно упростить анализ неизвестной ранее музыкальной записи звуковому инженеру или аранжировщику для дальнейшей работы с треком.

Информационные технологии не только упростили работу для профессиональных музыкантов. Обучение музыке дома может дать все более полное представление о музыкальной теории и может стать более эффективным. Существуют приложения для тренировки музыкального слуха, изучения нотации, демонстрации высотности звука вокалиста в реальном времени. Приложение, распознающее аккорды в музыке, поможет избавить новичков от поиска нотации в интернете. Особенно это актуально, когда песня недавно вышла, и профессиональные музыканты еще не успели записать и выложить нотацию к ней.

**Анализ цифрового звука для распознавания аккордов.** Для начала рассмотрим основные понятия теории музыки. Мы представляем ноту как восприятие человеком звуковой волны, колеблющейся на определенной частоте. Например, в современной западной музыке инструменты принято настраивать по ноте – «Ля» первой октавы», которая имеет стандартизованную частоту 440 Гц.

Аккорд – это сочетание одновременно звучащих трёх и более звуков разной высоты.

Чтобы получить из цифрового аудиофайла информацию о содержащихся в нем частотных классах, нужно перевести этот файл в частотное представление. Для этого может использоваться быстрое преобразование Фурье. На рис. 1 представлено преобразование Фурье для ноты «Ля малой октавы» [1]:

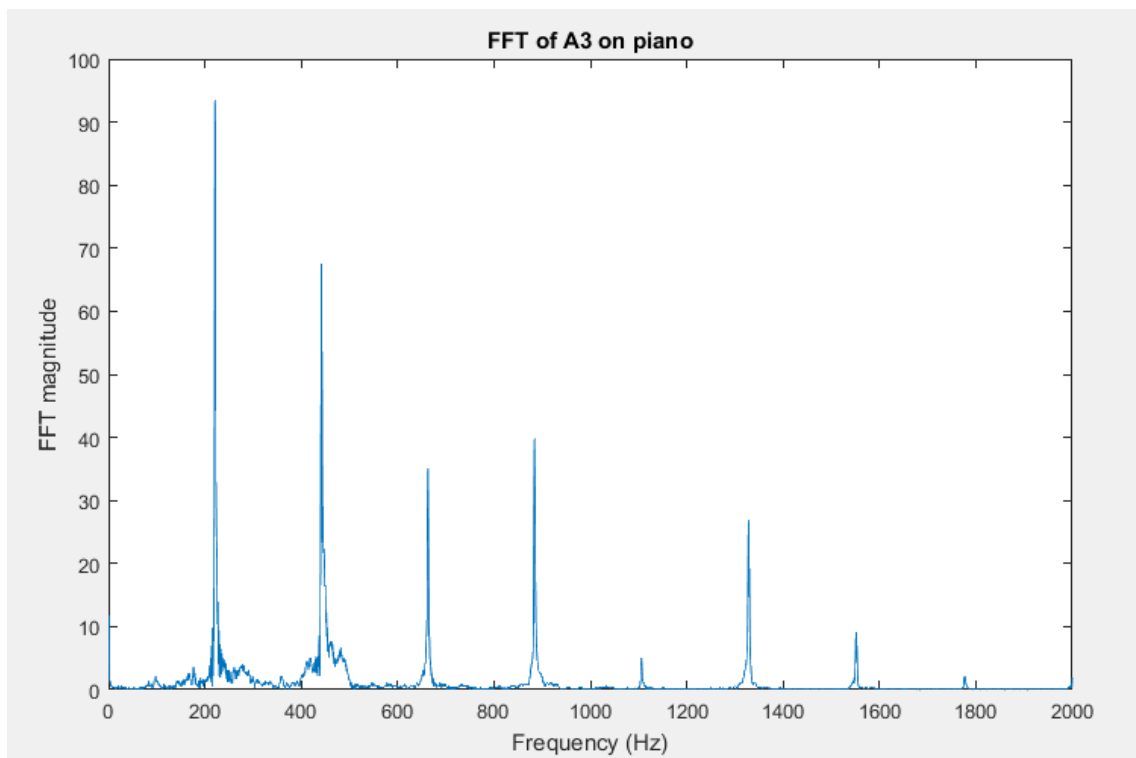


Рис. 1. Преобразование Фурье для записи, на которой звучит нота Ля, сыгранная на фортепиано.

Данная нота имеет основную частоту 220 Гц. Но тогда откуда берутся остальные пики? Одна из основных проблем распознавания сложных аккордов заключается в том, что музыкальные ноты также содержат частоты, кратные основной частоте, известной как обертоны или гармоники. Например, «Ля малой октавы» при частоте 220 Гц имеет гармоники на 440 Гц, 660 Гц, 880 Гц, 1100 Гц и так далее. Чем больше гармоник содержит инструмент, тем он кажется «богаче» на слух. Большинство частот образуют интервал «октава» с основным тоном, которые не мешают распознаванию аккорда, но также в обертоновом ряду присутствуют «квинта», «большая терция» и «малая септима» так, что первые 6 гармоник образуют мажорное трезвучие, а первые 8 – малый мажорный септаккорд. При этом в аккорде звучит более одной ноты, и каждая имеет свою последовательность обертонов.

Даже когда мы имеем идеальную запись без шумов и дополнительных инструментов, нельзя напрямую выбрать все частоты нот и провести соответствие со словарем аккордов.

Так как современная стандартная настройка инструментов неидеальна, можно попытаться находить частоты, которые отклоняются от равномерно-темперированного строя, и не учитывать их. Однако в записи присутствуют такие инструменты, как гитара, которую часто невозможно идеально настроить, и тогда нужные ноты будут теряться.

Другой вариант – нахождение основного тона, как тона с самой низкой частотой, и игнорирование тонов, которые могли бы являться его обертонами, так как мы можем их рассчитать. Этот вариант тоже не подходит, так как в современной музыке основной тон аккорда дублируется на 1 или 2 октавы ниже. И в данном случае будут удаляться уже основные тона аккорда, так что точность распознавание будет понижаться.

В данной разработке был использован третий вариант. Было принято решение распознавать лишь мажорные и минорные трезвучия путем накопления интенсивности каждого частотного класса. Частотный класс представляет собой множество всех частот ноты на каждой октаве. Например «До первой октавы» и «До второй октавы» относятся к одному частотному классу.

**Разработка алгоритма распознавания аккордов.** Идея разделять частоты нот на частотные классы для распознавания аккордов была предложена Т. Фуджишимой в 1999 г. [2]. В этой работе рассматривалось распознавание с использованием скрытых марковских моделей для учета последовательности, в которой аккорды звучат. Для этого способа нужно большой объем музыки с полной нотацией аккордов и времени, в течение которого они звучат. Также способ не сильно повышал точность распознавания, если использовались отличные от обучающей выборки группы или жанры музыки.

Наиболее близкий к примененному в данной работе алгоритм был рассмотрен в работе Кристофа Хауснера [3].

Разработанная авторами система предназначена не для использования в сценариях, работающих в режиме реального времени, то есть примененный алгоритм сначала считывает данные всей песни, а затем выводит все аккорды. Обнаружение аккордов в реальном времени считается сложной задачей, поскольку аккорды необходимо оценивать «на лету» без информации о нескольких следующих моментах времени.

Набор аккордов, оцениваемых данной системой, ограничен 24 мажорными и минорными аккордами (C, Cm, C#, C#m, D, ..., B, Bm) и символом NC (no chord – если считается, что аккорд отсутствует). Такой довольно ограниченный набор объясняется несколькими причинами. Во-первых, это, безусловно, самые распространенные аккорды, присутствующие в популярной музыке. Во-вторых, точность обнаружения значительно падает, когда распознаются слишком много аккордов. Кроме того, часто можно функционально заменить более сложные аккорды простым мажорным или минорным аккордом, имеющим общий основной тон (например, Gm7 на Gm, Dmaj7 на D и т. д.).

Рассмотрим стадии преобразования сигнала (см. рис. 2).

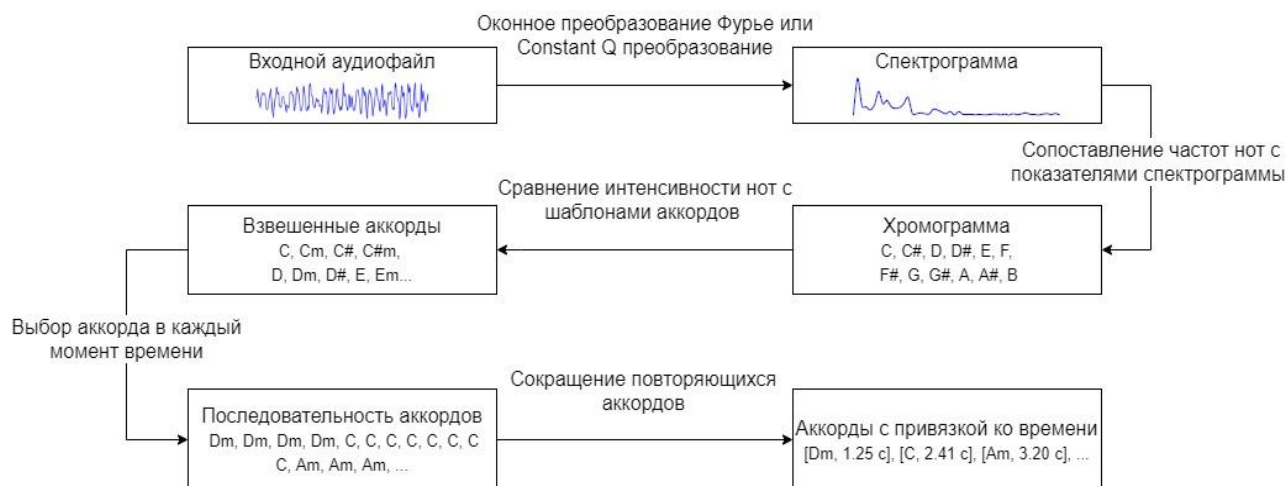


Рис. 2. Стадии преобразования сигнала для распознавания аккордов.

Алгоритм принимает на вход несжатые аудиофайлы в формате\*.wav. На выходе алгоритма получается список предполагаемых аккордов и их продолжительность. Необработанные входные данные очень избыточны, поэтому, учитывая, что нас интересуют только аккорды, и нет способа извлечь эту информацию непосредственно из аудиофайла, было принято решение последовательно преобразовывать эти данные и затем распознавать в них частоты, отвечающие за конкретные аккорды.

После поступления аудиофайла он с помощью оконного преобразования Фурье либо преобразования постоянной добротности преобразуется в данные спектрограммы [4].

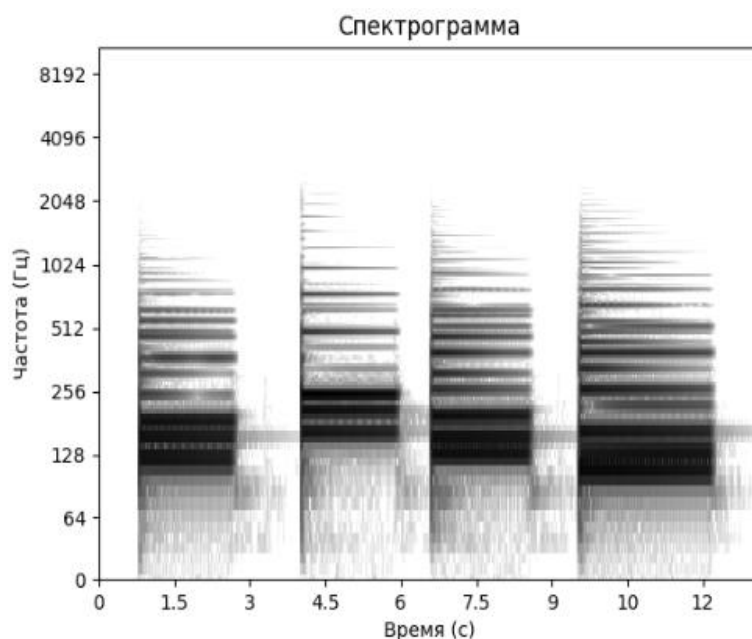
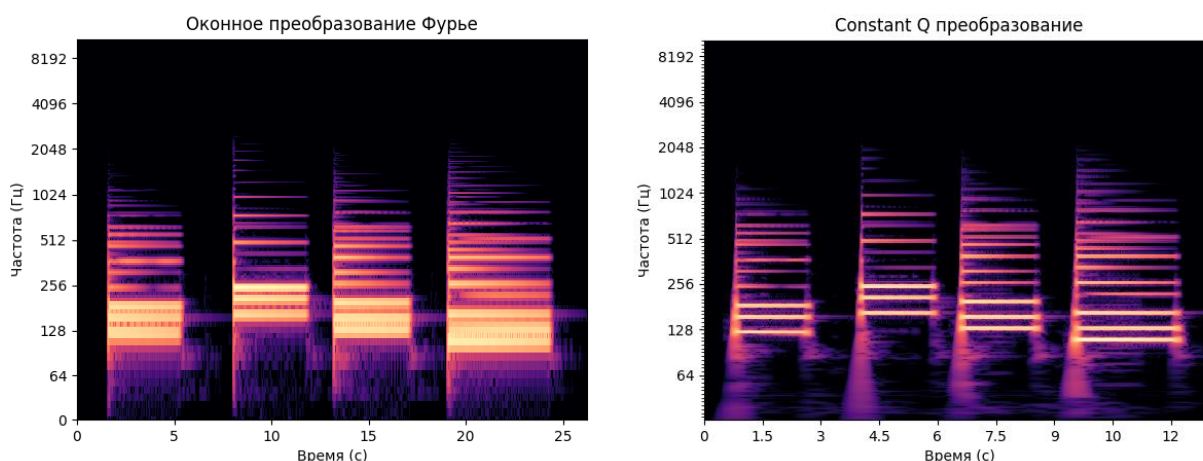


Рис. 3. Построенная диаграмма спектрограммы.

В данном случае спектрограммой называется двумерная структура данных, одна ось которой представляет измерение времени, а другая – измерение частоты. Значение каждого элемента структуры, показывает интенсивность частоты (см. рис. 3).

В связи с тем, что частоты нот расположены в логарифмическом ряду, при обычном оконном преобразовании Фурье разрешение на низких частотах ниже, чем на высоких. Поэтому было принято решение использовать преобразование Constant Q (см. рис. 4) [5].



а) б)

Рис. 4. Сравнение оконного преобразования Фурье (а) и Constant Q преобразования (б).

Следующим шагом алгоритма является вычисление так называемой хромограммы.

Как и спектрограмма, хромограмма представляет собой двумерную структуру, однако вместо частотного спектра в ней содержится информация об интенсивности 12 классов частот. Каждый класс соответствует 12 нотам хроматической гаммы: C, C#, D, D#, E, F, F#, G, G#, A, A#, B. При этом не игнорируется номер октавы. Например, частоты нот C1 и C4 соответствуют одному и тому же классу, и их интенсивности усредняются (см. рис. 5).

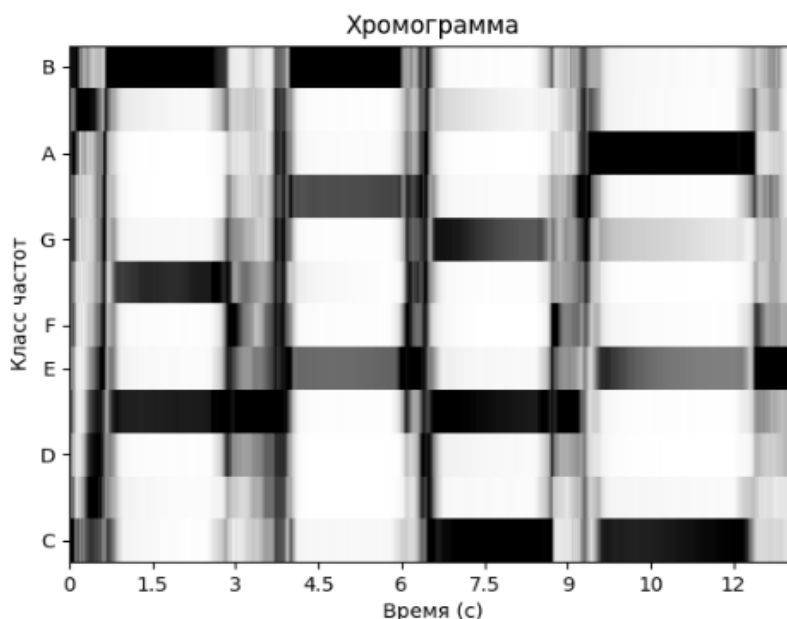


Рис. 5. Построенная диаграмма хромограммы.

После получения классов хромограммы есть возможность получить аккорды, которые состоят из данных классов. В хромограмме присутствует множество обертонов, а также тонов, не относящихся к гармоническому инструменту, например вокал. Поэтому, во-первых, необходимо рассмотреть, насколько хорошо данные хромограммы в каждый момент времени подходят под шаблон каждого аккорда, а во-вторых, рассматривать небольшое количество аккордов. В данном случае будут использоваться простые минорные и мажорные аккорды.

В итоге список шаблонов, будет состоять из 24 аккордов и обозначения «NC» – отвечающие за случай, когда ни один аккорд не подходит: C, Cm, C#, C#m, D, Dm, D#, D#m, E, Em, F, Fm, F#, F#m, G, Gm, G#, G#m, A, Am, A#, A#m, B, Bm, NC.

Шаблон для каждого аккорда будет состоять из 12 весовых коэффициентов, каждый из которых равен 1 либо 0, в зависимости от того входит ли данная нота в состав аккорда. Для обнаружения отрывков без аккордов (N.C.) используется шаблон, в котором весовые коэффициенты всех частотных классов равны 1 (см. табл. 1).

Структуры аккордовых шаблонов занесены в таблицу 1.

Таблица 1

**Структуры аккордовых шаблонов**

Название аккорда	Классы частот											
	C	C#	D	D#	E	F	F#	G	G#	A	A#	B
C	1	0	0	0	1	0	0	1	0	0	0	0
Cm	1	0	0	1	0	0	0	1	0	0	0	0
C#	0	1	0	0	0	1	0	0	1	0	0	0
C#m	0	1	0	0	1	0	0	0	1	0	0	0
D	0	0	1	0	0	0	1	0	0	1	0	0
Dm	0	0	1	0	0	1	0	0	0	1	0	0
D#	0	0	0	1	0	0	0	1	0	0	1	0
D#m	0	0	0	1	0	0	1	0	0	0	1	0
E	0	0	0	0	1	0	0	0	1	0	0	1
Em	0	0	0	0	1	0	0	1	0	0	0	1
F	1	0	0	0	0	1	0	0	0	1	0	0
Fm	1	0	0	0	0	1	0	0	1	0	0	0
F#	0	1	0	0	0	0	1	0	0	0	1	0
F#m	0	1	0	0	0	0	1	0	0	1	0	0
G	0	0	1	0	0	0	0	1	0	0	0	1
Gm	0	0	1	0	0	0	0	1	0	0	1	0
G#	1	0	0	1	0	0	0	0	1	0	0	0
G#m	0	0	0	1	0	0	0	0	1	0	0	1
A	0	1	0	0	1	0	0	0	0	1	0	0
Am	1	0	0	0	1	0	0	0	0	1	0	0
A#	0	0	1	0	0	1	0	0	0	0	1	0
A#m	0	1	0	0	0	1	0	0	0	0	1	0
B	0	0	0	1	0	0	1	0	0	0	0	1
Bm	0	0	1	0	0	0	1	0	0	0	0	1
NC	1	1	1	1	1	1	1	1	1	1	1	1

Обозначим через  $a$  – вектор интенсивностей частотных классов в единичный момент времени,  $b$  – вектор весовых коэффициентов для определенного шаблона аккорда. Чтобы оценить, насколько вектор  $a$  соответствует вектору  $b$ , нужно найти коэффициенты,

показывающие степень совпадения прозвучавшего аккорда и шаблона. В работе Хауснера [3] было предложено использовать косинусное сходство, которое определяется следующей формулой:

$$\cos(\theta) = \frac{(a,b)}{|a| \cdot |b|} = \frac{\sum_{i=1}^n a_i \cdot b_i}{\sqrt{\sum_{i=1}^n (a_i)^2} \cdot \sqrt{\sum_{i=1}^n (b_i)^2}},$$

где  $n = 12$ ,  $a_i$  – интенсивность  $i$ -го частотного класса в единичный момент времени,  $b_i$  – весовой коэффициент  $i$ -го частотного класса для определенного шаблона аккорда.

Преимущество данного выражения состоит в том, что его значения изменяются в диапазоне от 0 до 1.

После применения косинусного сходства для вектора хромограммы в каждый момент времени будет получена двумерная структура, в которой будут отображаться коэффициенты, показывающие насколько совпадает прозвучавший аккорд с шаблонами эталонных аккордов в каждый момент времени.

После этого нужно найти, для какого аккорда коэффициент будет максимальным во все моменты времени. Затем сократить повторяющиеся в последовательности аккорды, записывая время начала звучания аккорда. В композиции с несколькими партиями неизбежно будут появляться лишние ноты, и такая комбинация частот будет распознаваться как другой аккорд. Однако в большинстве случаев такие аккорды будут длиться достаточно короткий промежуток времени.

Алгоритм может эффективно применяться для музыки с относительно простым аккомпанементом и не слишком быстрым темпом, поэтому целесообразно отбрасывать аккорды, которые длятся менее половины секунды. Также необходимо повторно применить функцию, которая уберет дубликаты записей аккордов так, чтобы длительность их звучания сохранялась.

Максимально точно распознавание проходит в тех случаях, когда все ноты аккорда звучат одновременно. При усложнении звука и наполнении его другими партиями точность алгоритма будет падать. Поэтому перед применением алгоритма была использована обученная модель Spleeter от компании Deezer [6]. Она позволяет разделять музыкальный трек на стэмы – группы музыкальных инструментов.

**Реализация алгоритма.** Для реализации системы автоматического распознавания аккордовой последовательности в цифровых аудиофайлах на основе вышеизложенного алгоритма был выбран язык программирования Python. В нем присутствуют такие библиотеки как NumPy, Matplotlib, помогающие в работе с математическими операциями и построением графиков. В пакете Librosa существует много удобных функций и инструментов для работы с звуком, например, функция, выполняющая оконное преобразование Фурье, а также структура



для постройки спектрограммы, автоматически сопоставляющая ноты и частоты.

Первым этапом работы системы было получение спектрограммы сыгранного аккорда (см. рис. 6).

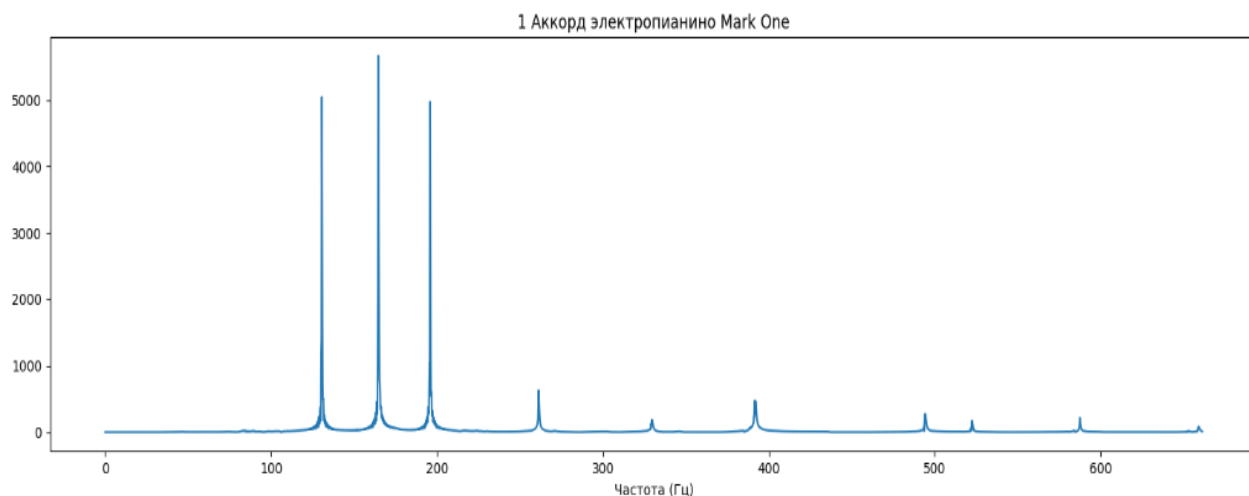


Рис. 6. Спектрограмма аккорда, сыгранного на электропианино.

Можно заметить, что на графике легко можно выделить три основных тона аккорда, а гармоники выражены не сильно. Это связано с тем, что для тестирования работы алгоритма было выбрано электропианино, имеющее более мягкий и джазовый звук.

Затем был построен график хромограммы, где отображаются не частоты, а интенсивности отдельных частотных классов (см. рис. 7).

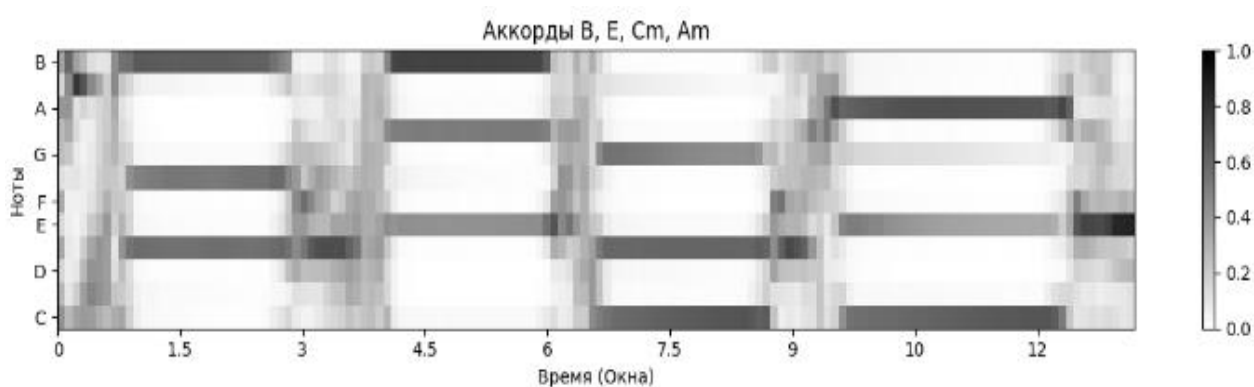


Рис. 7. Хромограмма последовательности аккордов, сыгранного на электропианино.

Благодаря пакету Librosa нет необходимости вручную прописывать коррекцию настройки инструмента и отображать интенсивности частот в интенсивности частотных классов. Поэтому далее можно найти сходство интенсивности нот в каждый промежуток времени с шаблонами аккордов. При этом получится последовательность аккордов (см. рис. 8).



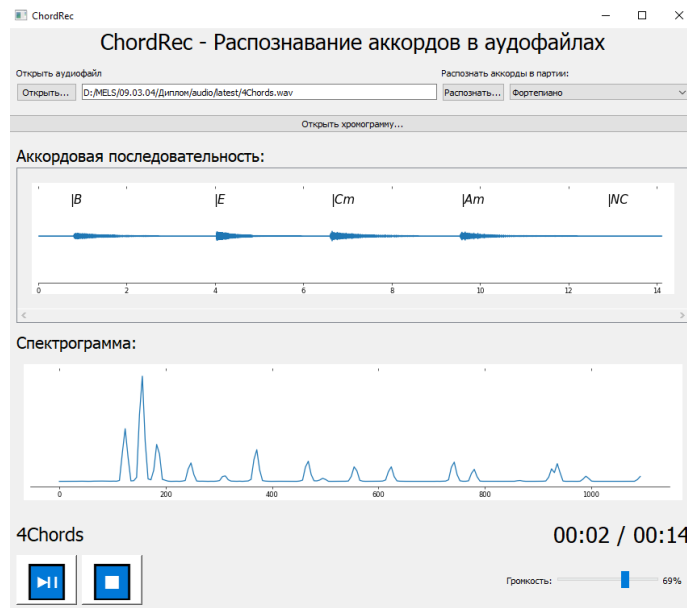


Рис. 10. Демонстрация работы главного окна приложения.

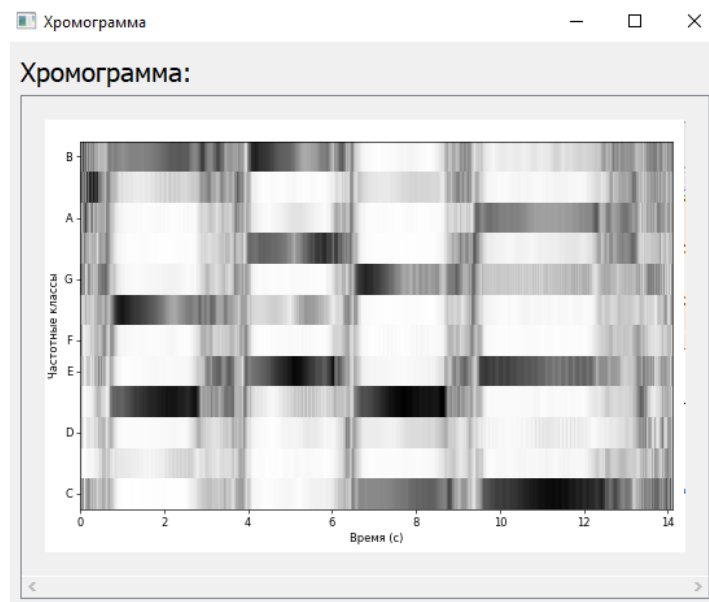


Рис. 11. Демонстрация работы отображения хромограммы.

Для повышения точности распознавания аккордов в системе присутствует функция разделения аудиофайлов на отдельные составляющие.

Пакет Spleeter предоставляет три варианта разделения музыкального аудиофайла. При этом осмысленно определять аккорды лишь в некоторых компонентах. В группе «вокал / аккомпанемент» – в аккомпанементе, в группе «вокал / барабаны/ бас /все остальное» – во всем остальном, а в группе «вокал / барабаны / бас / фортепиано / все остальное» – в фортепиано.

В разрабатываемой системе это можно сделать, выбрав в выпадающем меню

необходимую группу инструментов и нажав кнопку «Распознать...» (см. рис. 12).

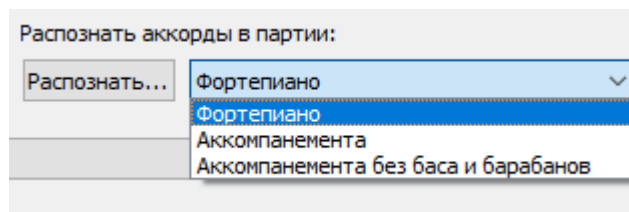


Рис. 12. Выпадающее окно для распознавания аккордов в определенных партиях аккомпанемента.

При выборе группы инструментов программа предложит пользователю выбрать папку для сохранения стэмпов. В папке сохранятся компоненты аудиофайла в соответствии с выбранным вариантом разделения. Сразу после этого нужный аудиофайл автоматически загрузится в систему.

**Заключение.** Рассмотренная в статье система умеет достаточно точно распознавать аккорды в аудиофайлах, в которых присутствует только звук фортепиано, и все ноты звучат одновременно. В системе присутствует пакет для предварительной фильтрации сигнала путем выделения сигнала определенного музыкального инструмента. Тем не менее данный подход не позволит распознавать аккорды сложнее, чем стандартные трезвучия. Проблема заключается в большом количестве обертонов в звуках разных инструментов. В дальнейшем можно попытаться определять гармоника для разных музыкальных инструментов при помощи машинного обучения [7,8].

#### СПИСОК ЛИТЕРАТУРЫ

1. Sisk P., Hughes B., Paul R., Aggarwal A., Patankar M. Chordination Detection Algorithm // University of Michigan [Электронный ресурс]. – Режим доступа: <http://www-personal.umich.edu/~rbpaul/> (дата обращения: 11.03.2021)
2. Fujishima T. Realtime Chord Recognition of Musical Sound: a System Using Common Lisp Music // Proc. ICMC. – 1999. – pp. 464–467 [Электронный ресурс]. – Режим доступа: <https://ci.nii.ac.jp/naid/10013545881/en/> (дата обращения: 14.12.2020)
3. Hausner C. Design and Evaluation of a Simple Chord Detection Algorithm // University of Passau, Faculty of Computer Science and Mathematics. – 2014. – 59 p [Электронный ресурс]. – Режим доступа: [https://www.fim.uni-passau.de/fileadmin/dokumente/fakultaeten/fim/lehrstuhl/sauer/geyer/BA\\_MA\\_Arbeiten/BA-HausnerChristoph-201409.pdf](https://www.fim.uni-passau.de/fileadmin/dokumente/fakultaeten/fim/lehrstuhl/sauer/geyer/BA_MA_Arbeiten/BA-HausnerChristoph-201409.pdf) (дата обращения: 11.10.2020).
4. Lenssen N., Needell D. An Introduction to Fourier Analysis with Applications to Music // Journal of Humanistic Mathematics. – 2014. – Vol. 4. – pp. 72–91 [Электронный

ресурс]. – Режим доступа:[https://www.researchgate.net/publication/314918556\\_An\\_Introduction\\_to\\_Fourier\\_Analysis\\_with\\_Applications\\_to\\_Music](https://www.researchgate.net/publication/314918556_An_Introduction_to_Fourier_Analysis_with_Applications_to_Music) (дата обращения: 14.12.2020).

5. Brown Judith C. Calculation of a constant Q spectral transform // *Journal of the Acoustical Society of America*. – 1991. – Vol. 89(1). – pp. 425–434 [Электронный ресурс]. – Режим доступа: <http://academics.wellesley.edu/Physics/brown/pubs/cq1stPaper.pdf> (дата обращения: 12.03.2021).
6. Deezer Research – Spleeter [Электронный ресурс]. – Режим доступа: <https://research.deezer.com/projects/spleeter.html> (дата обращения: 25.04.2021).
7. Глазырин Н. Ю. Алгоритмическое распознавание аккордов в цифровом звуке: автореф. дисс. ... канд. физ.-мат. наук. – Екатеринбург, 2015. – 22 с [Электронный ресурс]. – Режим доступа: <https://elar.urfu.ru/handle/10995/27846> (дата обращения: 12.12.2020).
8. Дудырев Е. О. Исследование методов распознавания аккордов, основанных на машинном обучении [Электронный ресурс] // Всероссийский форум научной молодежи «Богатство России»: Сборник докладов, Москва, 4–6 дек. 2017 г. – 2018. – С. 56–57. – Режим доступа: <https://elibrary.ru/item.asp?id=32874106> (дата обращения: 06.04.2021).