

## АНАЛИЗ МЕТОДОВ РАСПОЗНАВАНИЯ И ПРЕОБРАЗОВАНИЯ АУДИОИНФОРМАЦИИ В НОТЫ

---

Предлагаются методы решения задачи распознавания и преобразования аудиозаписи в ноты с помощью создания универсального программного решения. В качестве метода решения задачи распознавания аудиоинформации рассматривается возможность применения численных методов (преобразование Фурье) и методов искусственного интеллекта.

### 1. Введение

Идентификацией объекта является определение его характеристик на основе опытного исследования, она самая трудоемкая и самая ответственная операция при анализе объектов. К настоящему времени теория идентификации превратилась в обширное по содержанию и богатое по методам учение, тем не менее, еще многие актуальные проблемы ждут в ней своего решения. Одной из самых актуальных задач в настоящее время является задача расширения класса объектов, поддающихся эффективной идентификации [1].

Проблема поиска музыки становится актуальной в современном мире. Большое количество музыки в различных форматах существует в электронном виде, и одной из трудноразрешимых проблем является преобразование этих файлов в нотный ряд.

*Цель* данного исследования – разработка методов распознавания аудиоинформации и дальнейшего ее преобразования в ноты. Для достижения поставленной цели необходимо разработать программное обеспечение, позволяющее решать следующие задачи [3]:

- а) поиск аудиоинформации по фрагменту аудио;
- б) поиск аудиоинформации по описанию, её каталогизация и категоризация аудио в поисковых целях;
- в) поиск аудиоинформации по «напетой» мелодии;
- г) распознавание аудиоинформации и получение нот.

Для реализации первого пункта на данный момент уже существуют определённые программные средства. Их работа основана на том, что в базе данных хранится большое количество аудиозаписей. Однако каждое из этих программных приложений имеет свои, весьма существенные недостатки. Предлагается использовать новый метод, который одновременно предоставит возможность решения всех описанных выше проблем: аудиозапись распознаётся и переводится в ноты (технически это близко к хранению аудио в миди-формате). Это предоставит возможность сравнивать аудиозапись не только с её фрагментом для поиска, но также и различных вариаций, вплоть до напетого кусочка в микрофон. Таким образом, решится глобальная проблема многих музыкантов: автоматическое получение нот из записей.

Практическая значимость решения задачи распознавания аудиоинформации и преобразования в нотный ряд заключается в:

- автоматизации ручного процесса снятия нот, т.е. перевода из аудио в ноты;
- создании потенциальной базы для разработки “музыкального поисковика”.

### 2. Постановка задачи

Для решения задачи распознавания и преобразования аудиоинформации в ноты необходимо создать универсальный программный продукт, базирующийся на численных методах (преобразование Фурье) и использующий методы искусственного интеллекта (многослойный перцептрон) [2].

Такой программный продукт должен выполнить следующие действия:

1. Считать звуковой файл.
2. На основе совокупности методов проанализировать его.
3. Получить на выходе нотное представление записи.
4. Дальнейшие действия зависят от конкретного применения.

При более подробном рассмотрении процесса распознавания и преобразования аудиоинформации можно выделить следующие этапы исследования:

1. Открытие звукового файла и считывание информации (частота дискретизации, длительность).
2. Спектральный анализ некоего «минимального» фрагмента звука при помощи преобразования Фурье.
3. По наиболее сильным гармоникам определяем, какие ноты присутствуют в аккорде (метод обертонового ряда)[3].
4. По гармоникам, присутствующим в спектре каждой ноты, можем определить, какие тембры присутствуют.
5. Распознавание в динамике – на каждом интервале. Если картина не изменилась для какой-то ноты, мы соответственно предполагаем, что нота длится дольше, и не разрываем её.

6. Оформление результатов.

### 3. Анализ существующих методов и программных продуктов, реализующих проблему распознавания и преобразования аудиоинформации

На данный момент существует несколько программных решений подобной направленности. Вот некоторые из них:

*Amazing Midi* – программа для распознавания аудио и преобразования в midi (рис. 1).

Вначале пользователю предлагается выбрать аудиофайл с образцом тембра музыкального инструмента, который звучит в записи, входной и выходной файлы. Далее находится поле, в котором будут отображаться распознанные ноты.

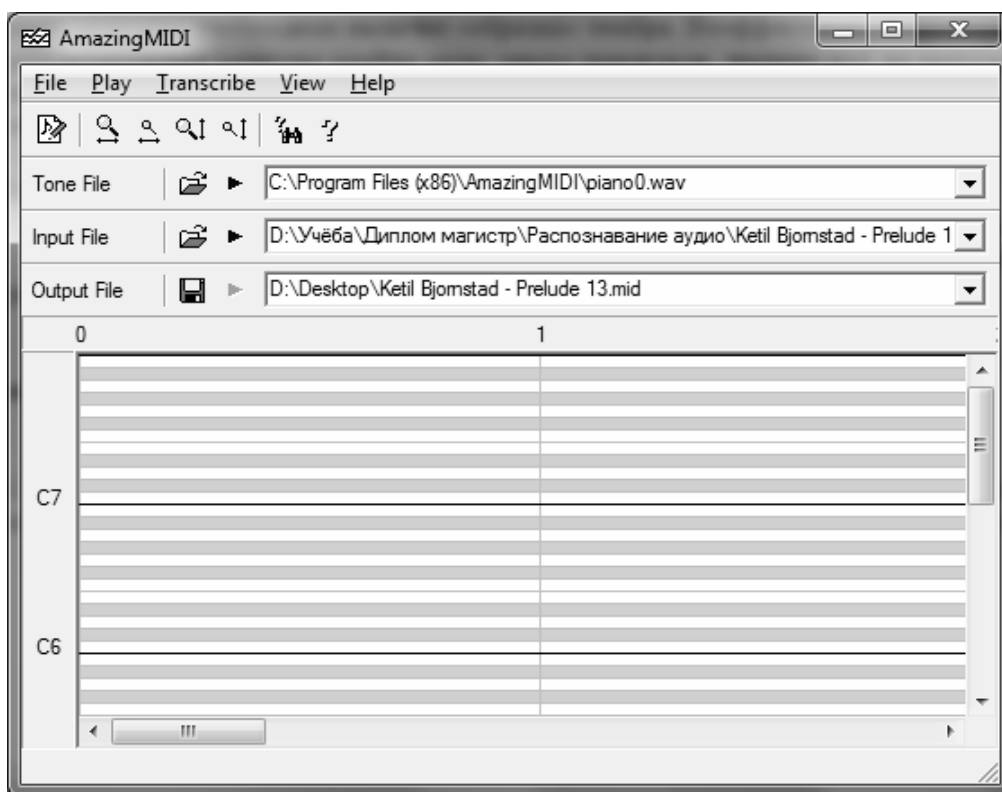


Рис. 1. Главное окно программы AmazingMidi

Недостатки программы: только один инструмент; не позволяет распознавать голос. Для распознавания необходимо наличие «образца» тембра. Неэффективна, так как результирующее качество крайне низко: много призвуков, лишних нот, на выходе, по сути, выдает «кашу» из нот, представляющую нечто близкое к исходной мелодии.

*Digital Ear* – в целом, решение, близкое к предыдущему (рис. 2). Возможно наличие нескольких инструментов в составе.

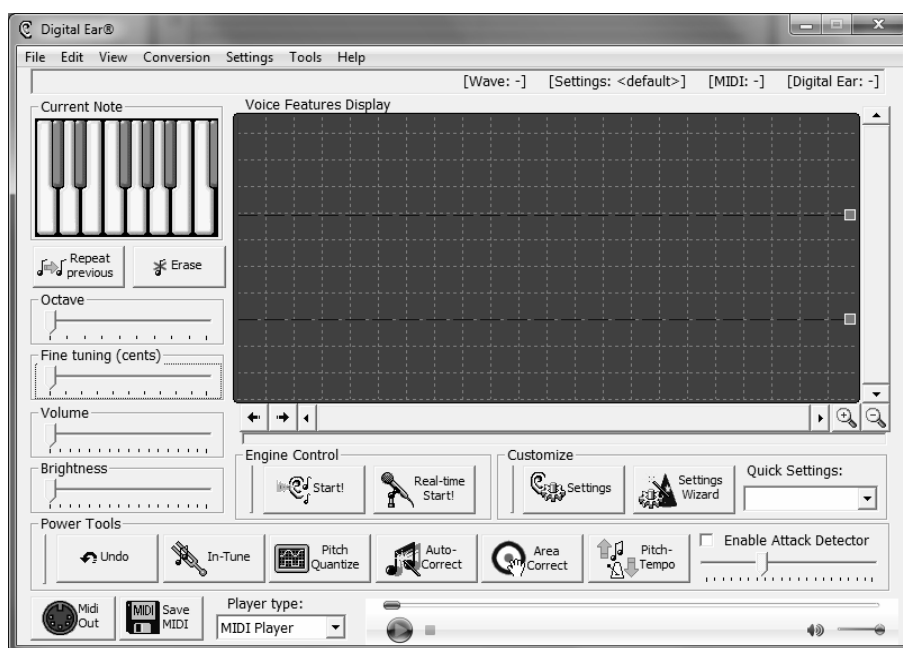


Рис. 2. Главное окно программы «Digital Ear»

В меню «File» пользователь может открыть распознаваемый файл. Далее программа предлагает пользователю выполнить анализ аудиозаписи (попытаться определить диапазон нот, отклонение от стандартного строя), некоторые из них потом можно сменить. После этого можно приступить к распознаванию. В поле «Voice Features Display» будут отображаться ноты.

Если программа «Amazing Midi» захватывала огромное количество призывков, делая результат распознавания бесполезным, то эта программа, наоборот, распознаёт только несколько ключевых нот на эпизоде, что опять-таки делает её неэффективной.

*D'Accord iChords 2.0* – программное решение для распознавания аккордов.

Данная программа устанавливается как подключаемый модуль к проигрывателю на компьютере и позволяет распознавать аккорды, входящие в состав композиции. Этот вариант малоинтересен, так как он распознаёт только гармоническую, но не мелодическую, составляющую.

*IntelliScore Ensemble WAV to MIDI Converter*.

На рис. 3 – главное окно этой программы. Это самое эффективное программное решение – по сути, единственное из существующих, которое позволяет получить хоть какой-нибудь полезный и работоспособный результат. При запуске программы открывается мастер создания проекта, в котором предлагается выбрать звуковой файл и выполнить ряд настроек.

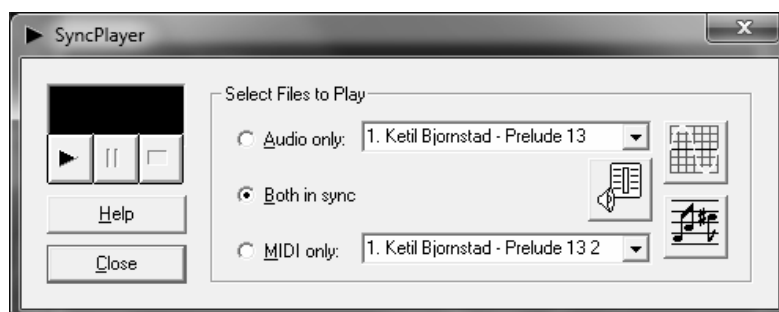


Рис. 3. Главное окно программы IntelliScore Ensemble WAV to MIDI Converter

Недостатки: значительная неточность распознавания вокальной музыки, необходимость вручную обозначать темп, нажимая клавишу на сильные доли, неровность ритма в конеч-

ном результате, платность программы (демо-версия распознаёт первые 15 секунд композиции, что достаточно для оценки качества).

Кроме того, модуль также имеется в секвенсоре *Cakewalk Pro Audio*. Позволяет распознавать одноголосные мелодии.

#### 4. Решение задачи распознавания и преобразования аудиоинформации в ноты

Предложенный метод базируется на следующих действиях:

1. *Преобразование Фурье*, применяемое в целях получения энергии спектральных составляющих сигнала на каждом участке. Поскольку аудиозапись имеет дискретное представление музыкальных данных, используем дискретный вариант записи закона. Энергия  $i$ -го сигнала составляется из двух компонент:

$$G_j : a_j = \frac{2\Delta t}{T} \sum_{i=0}^{f_d} S(i\Delta t) \cos(j \frac{2\pi}{t} i\Delta t) ;$$
$$b_j = \frac{2\Delta t}{T} \sum_{i=0}^{f_d} S(i\Delta t) \sin(j \frac{2\pi}{t} i\Delta t),$$

где  $G_j$  – энергия сигнала;  $f_d$  – частота дискретизации;  $\Delta t$  – промежуток времени между отсчётами;  $T$  – период измерения.

Сама энергия тогда составит:  $|G_j| = a_j^2 + b_j^2$ .

2. *Выделение нот*. Для каждой ноты характерно наличие в спектре нескольких частот. Одна, самая главная, частота имеет наибольшую энергию. Например, как известно, для ноты «ля» первой октавы это – 440 Гц. Остальные частоты определяются в равномерно темперированном строе следующим образом:  $f(i) = f_0 \cdot 2^{i/12}$ , где  $f_0$  – частота ноты, от которой считаем,  $i$  – количество полутонов, которые отстаем от этой частоты.

Например, можно вычислить частоту звука на тон (2 полутона) ниже от «ля» – ноты «соль»:  $i = -2$ ,  $f(-2) = 440 \text{ Гц} \cdot 2^{-2/12} \approx 391,995 \text{ Гц}$

Кроме того, в спектре каждая нота имеет более слабые, обертоновые, спектральные составляющие.

3. *Анализ присутствующих музыкальных инструментов*. Самый сложный этап решения проблемы. Здесь предлагается прибегнуть к методам искусственного интеллекта, а именно – к обычному многослойному перцептрон.

Каждый инструмент имеет характерный тембр, который определяется количеством и энергией гармоник, входящих в обертоновые частоты данной ноты. Построив многослойный перцептрон, можно обучить его, подавая на вход образцы тембров разных инструментов разных нот. Параметрами, подаваемыми на вход, в простейшем случае будут  $n$  первых спектральных составляющих ноты (так как обертоновый ряд в теории бесконечный). На выходе перцептрон будет выдавать ноту.

4. *Корректировка*. В процессе распознавания неизбежно столкнёмся со следующими проблемами: как распознать корректно ноту, если накладываются несколько нот разных инструментов разной динамики. Возможно, перцептрон будет выдавать некорректные значения. Для этого предлагается использовать ещё один, корректирующий, перцептрон. Обучение его будет заключаться в анализе различных сочетаний нот и аккордов разных тембров.

Проблемой здесь также является определение необходимой сложности используемых нейронных сетей.

#### Выводы

Результаты анализа задачи распознавания и преобразования аудиоинформации в ноты показали, что на данный момент она актуальна в связи с отсутствием программных продуктов, осуществляющих качественное ее решение. Дальнейшее исследование включает практическую реализацию программного продукта, базирующегося на численных методах (преобразовании Фурье) и методах искусственного интеллекта (многослойный перцептрон), который позволит максимально качественно распознавать аудиоинформацию и преобразовывать ее в необходимый формат.

**Список литературы:** 1. Бодянский Е.В., Руденко О.Г. Искусственные нейронные сети. Харьков: ТЕЛЕ-ТЕХ, 2004. 369 с. 2. Уоссермен Ф. Нейрокомпьютерная техника. 1992. С. 144-162. 3. Белобородов А.Ю. Распознавание аудиообразов с применением обертонового ряда. 2010. С.72-77.

*Поступила в редколлегию 20.11.2011*

**Булавин Дмитрий Алексеевич**, канд. техн. наук, доцент кафедры теоретической и прикладной системотехники факультета компьютерных наук Харьковского национального университета им. В.Н. Каразина. Научные интересы: искусственный интеллект, теория принятия решений. Увлечения: футбол, теннис. Адрес: Украина, 61166, Харьков, ул. Ленина, 3, кв. 20, тел. 702-30-79, e-mail: dimetroid@yandex.ru.

**Харитонов Илья Александрович**, студент факультета компьютерных наук Харьковского национального университета им. В.Н. Каразина. Научные интересы: искусственный интеллект, компьютерные сети. Увлечения: музыка. Адрес: Украина, 61000, Харьков, ул. Вальтера, 14, e-mail: iliyaisd@gmail.com