

УДК 004.424.4:78.06

**Р. А. Корж**

**Криворожский национальный университет**

# К ПРОБЛЕМЕ АВТОМАТИЧЕСКОЙ ИДЕНТИФИКАЦИИ МУЗЫКАЛЬНЫХ ПРОИЗВЕДЕНИЙ

*Целью данной работы является критическая оценка существующих методов и описание нового подхода для решения имеющихся на сегодняшний день актуальных проблем в области автоматической идентификации музыкальных произведений.*

*Метою даної роботи є критична оцінка існуючих методів та опис нового підходу до рішення існуючих на сьогоднішній день актуальних проблем в області автоматичної ідентифікації музикальних творів.*

*The aim of this article is both critical appraisal of existing methods and description of a new approach to solve today's actual problems in the field of automatic music transcription.*

**Ключевые слова:** автоматическая идентификация музыкальных произведений, объектно-ориентированный формат, частотно-временная плоскость, преобразование Фурье, вейвлет-преобразование

## Введение

В настоящее время, большое количество разработок посвящено проблемам *автоматической идентификации музыкальных произведений*, стартом для которых стал 1975 год, когда Джеймс Мюрер спроектировал первую автоматизированную систему управления (АСУ), которая была способна разделять дуэтные партии музыкальных произведений [1].

Позднее, все больше и больше исследователей в области информационных технологий было включено в разработку программных средств для решения данной проблемы. АСУ становились как более эффективными и результативными, так и более сложными [2].

Общий подход к проблеме заключается в следующем. Сначала, аудио сигнал раскладывается в частотный ряд с помощью различных математических методов. Затем, локальные частотные экстремумы группируются в наборы гармоник по различным критериям. И, наконец, выполняется определение тембральных характеристик музыкальных инструментов, используя различные подходы [3].

В связи с трудностями определения всех звуковых объектов музыкальных произведений (во многих случаях это даже теоретически невозможно), выделяют *полную музыкальную идентификацию* (определение звуковых объектов в максимально-возможном количестве) и *частичную музыкальную идентификацию* (определение исключительно самых сильных звуковых объектов) [1].

## Критический анализ

Согласно [4], существующие АСУ предназначены для очень узких классов музыкальных произведений, из этого следует, что они не способны преобразовать реальную звуковую информацию в объектно-ориентированный формат. Рассмотрим более детально причины данного ограничения.

Как известно, звуковая информация, в общем случае, представляется нестационарным сигналом, то есть статистические свойства варьируют и зависят от времени [5].

Преобразование Фурье, в равной степени с Z-преобразованием и преобразованием Лапласа являются методами *спектрального анализа* и применяются в случае стационарных сигналов, в которых статистические свойства не зависят от времени. Поэтому, использование этих преобразований ко всему сигналу является нерационально и неэффективно [6].

Но, если взять во внимание тот факт, что нестационарный сигнал на всем временном интервале является квазистационарным на очень коротких временных отрезках, то это дает возможность применять Фурье и подобные ему преобразования к нестационарным аудио сигналам, предварительно разбив его на множество очень коротких отрезков [7]. Именно таким приемом широко пользуются разработчики большинства АСУ [1]. Казалось бы, проблема решена, но на самом деле это не так.

Дело в том, что окна, с помощью которых просматриваются сигналы, имеют фиксированный размер. Это приводит к тому, что частотное разрешение на низких и высоких частотах одинаковое (рис. 1а), к тому же наиболее распространенные окна – это окна Гаусса (рис. 2) и им подобные, на границах которых солидно сглаживается амплитуда сигнала. А в случае музыкальных преобразований:

- необходимо, чтобы с увеличением частоты увеличивалось частотное разрешение;
- на границах окон могут быть размещены звуковые объекты, которые, в равной степени с остальными, должны быть определены.

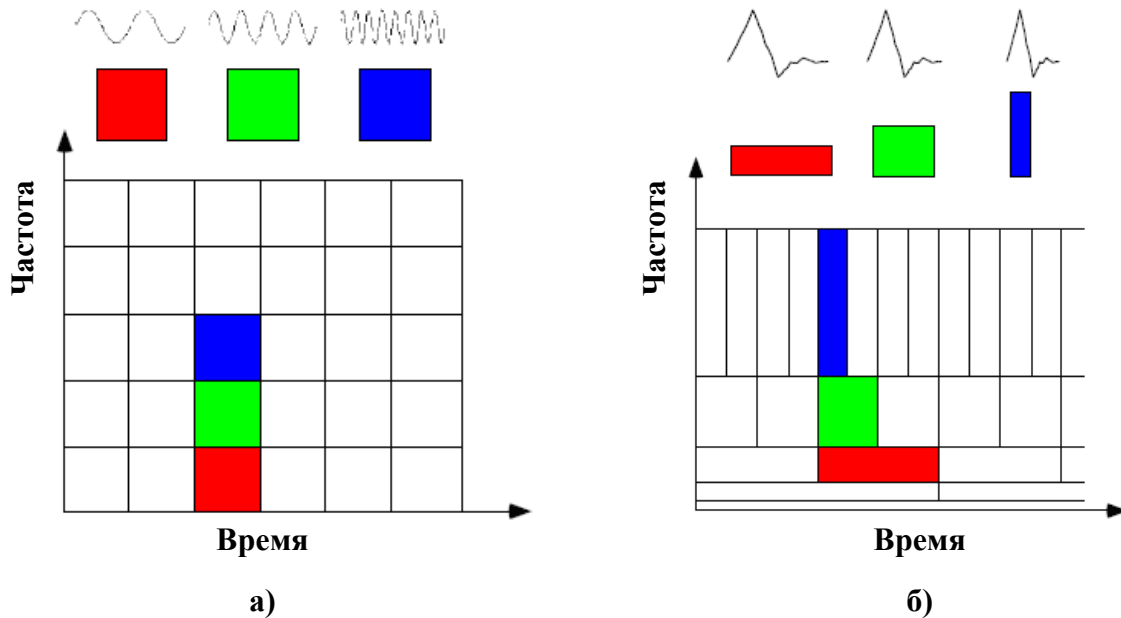


Рис. 1. Частотно-временное разрешения:  
а) преобразование Фурье;  
б) вейвлет-преобразование

Таким образом, эта тройка преобразований не удовлетворяет перечисленным выше требованиям, следовательно, не подходит для анализа музыкальных произведений.

Эти две проблемы решаются с помощью вейвлет-преобразования (рис. 1б). Здесь базисные функции ограничены как по времени, так и по частоте (в отличие от Фурье-преобразования, базисами которого являются синусоиды – функции, имеющие идеальную частотную локализацию и не имеющую временного ограничения) и являются аналогом окна.

Дискретным вейвлет-преобразованием (ДВП) последовательности  $x(nT)$  называется следующий ряд [8]:

$$W(a,b) = \frac{1}{\sqrt{a}} \sum_{n=0}^{\infty} x(nT) w\left(\frac{nT-b}{a}\right), \quad (1)$$

где

$W(a,b)$  – вейвлет-изображение (вейвлет-образ) последовательности  $x(nT)$ , результат дискретного вейвлет-преобразования;

$x(nT)$  – оригинал – вещественная или комплексная последовательность;

$w\left(\frac{nT-b}{a}\right)$  – базисная вейвлет-функция

(рис. 3), которая удовлетворяет следующим условиям:

1. принадлежность к пространству  $L^2(R)$ :

$$\int_{-\infty}^{+\infty} |w(t)|^2 dt < \infty \quad (2)$$

2. нулевое среднее:

$$\int_{-\infty}^{+\infty} w(t) dt = 0 \quad (3)$$

3. частотная локализация:

$$w(\omega) \rightarrow 0, \text{ при } \omega \rightarrow \infty \quad (4)$$

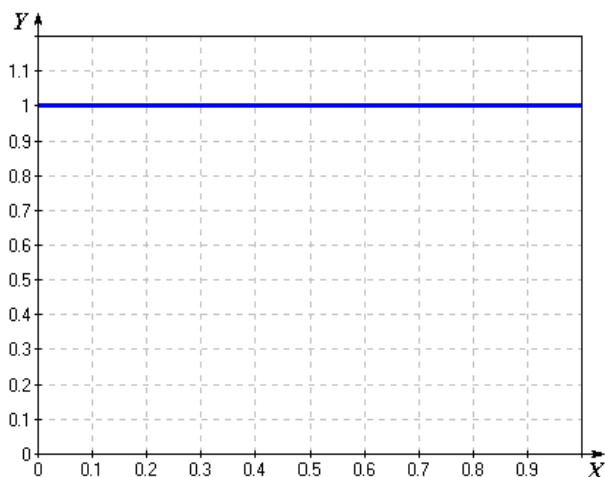
4. временная локализация:

$$w(t) \rightarrow 0, \text{ при } t \rightarrow \infty \quad (5)$$

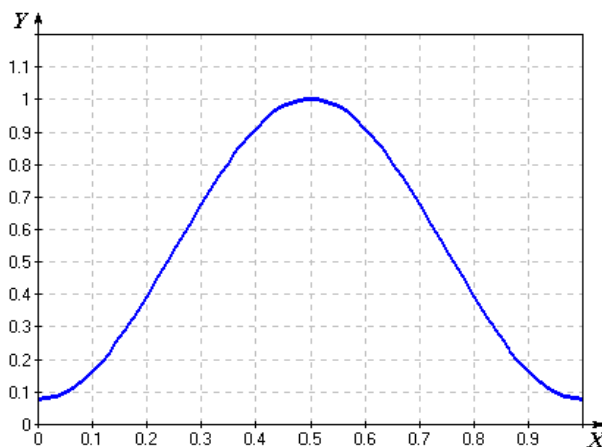
$a$  – параметр масштаба ( $a > 0$ ), благодаря которому вейвлет-функция  $w(t)$ : сужается, при  $a \rightarrow 0$  и расширяется, при  $a \rightarrow \infty$ ;

$b$  – параметр смещения, благодаря которому вейвлет-функция  $w(t)$ : перемещается вправо, при  $b > 0$  и перемещается влево, при  $b < 0$ .

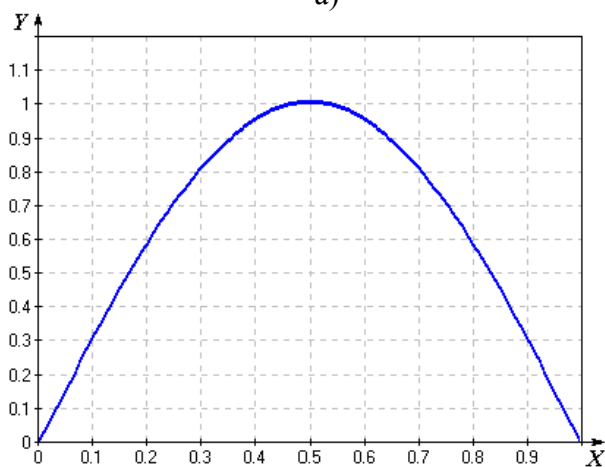
Благодаря масштабированию и смещению, вейвлеты позволяют проводить анализ с высоким разрешением по времени, но с низким разрешением по частоте и, наоборот, с низким разрешением по времени, но зато с высоким разрешением по частоте (рис. 1б). А этот принцип удовлетворяет главному требованию музыкального анализа.



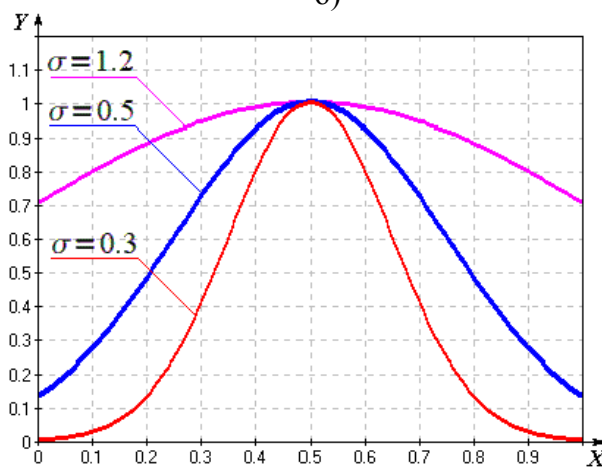
а)



б)



в)



г)

Рис. 2. Оконные функции:

- а) прямоугольное окно,
- б) окно Хемминга,
- в) синус-окно,
- г) окна Гаусса

Теперь перейдем к фундаментальным проблемам звукового анализа музыкальных произведений. В их число входят:

1. Наложение тонов оказывает большое влияние на достоверность результата, что отражается либо в слитии, либо в разделении одновременно-звучащих звуковых объектов;

2. Тембральное наложение, присущее для партий различных музыкальных инструментов;

3. Спектральное разрешение – чем выше частота, тем выше разрешение требуется для более точного определения звуковых характеристик;

4. *Размерное деление*, особенно если музыкальное произведение, достаточно выровненное по амплитуде.

Данное деление является достаточно условным, так как содержит в себе только ос-

новную часть проблем, с которыми сталкиваются исследователи в области обработки звуковых сигналов и музыкальных информационных потоков.

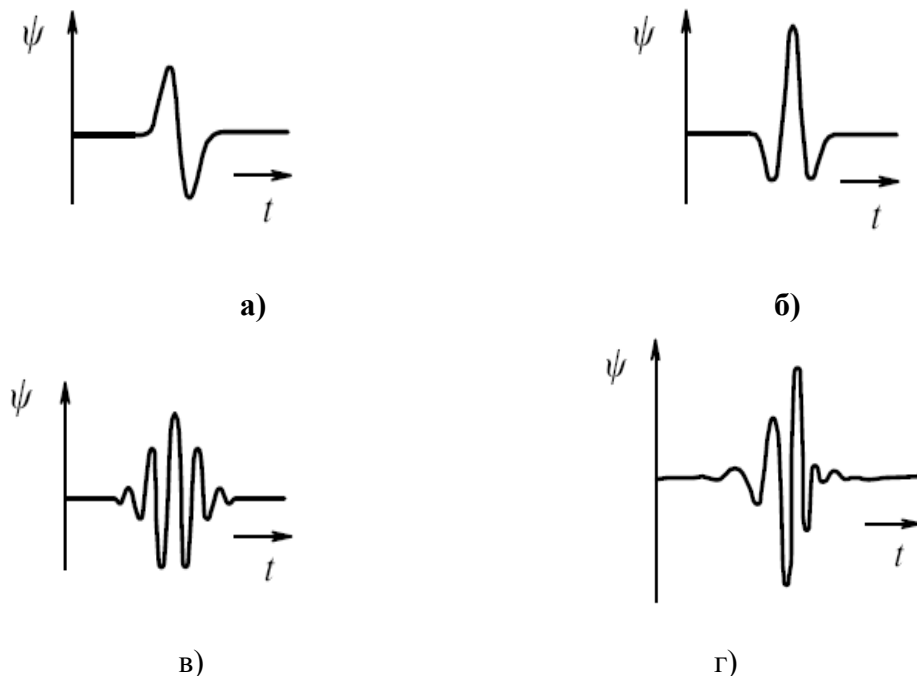


Рис. 3. Примеры вейвлет-функций:

- а) вейвлет-Гаусса 1-го порядка (волна),
- б) «мексиканская шляпа» (вейвлет Гаусса 2-го порядка),
- в) вейвлет Морлета,
- г) вейвлет из семейства Добеши

Совершенство нот одного инструмента образует *партию*. Партия, исполненная на одном музыкальном инструменте, называется *мелодией*. Совершенство партий различных музыкальных инструментов образует *партию*. Партия музыкального произведения содержит полное описание нот всех музыкальных инструментов, другими словами, она представляет собой музыкальное произведение в *объектном* виде.

Учитывая эти проблемы, можно выделить следующие общие ограничения, которыми обладает большинство АСУ:

1. *Степень музыкальной полифонии* (максимальное количество одновременно звучащих звуковых объектов в пределах одного музыкального произведения);
2. *Степень объектной полифонии* (максимальное количество одновременно звучащих звуковых объектов в пределах одного инструмента);
3. *Степень инструментальной полифонии* (максимальное количество одновременно зву-

чащих музыкальных инструментов в пределах одного музыкального произведения).

В связи с трудностями определения всех звуковых объектов музыкальных произведений (во многих случаях это даже теоретически невозможно), выделяют *полную музыкальную идентификацию* (определение звуковых объектов в максимально-возможном количестве) и *частичную музыкальную идентификацию* (определение исключительно самых сильных звуковых объектов) [1].

#### Описание предложенного метода

На рис. 4 представлен алгоритм работы нового подхода, который разработан с целью достичь максимально-качественных результатов преобразования звуковой информации в объектно-ориентированный формат путем адаптации к конкретному звуковому сигналу.

Перед тем, как аранжировщик подбирает на слух ноты, он сначала делает предварительное прослушивание, на протяжении которого определяет степень инструментальной полифо-

нии. Другими словами, он составляет список инструментов, которые звучат на протяжении всего музыкального произведения. И только

потом он начинает подбирать партии для каждого из них [9].

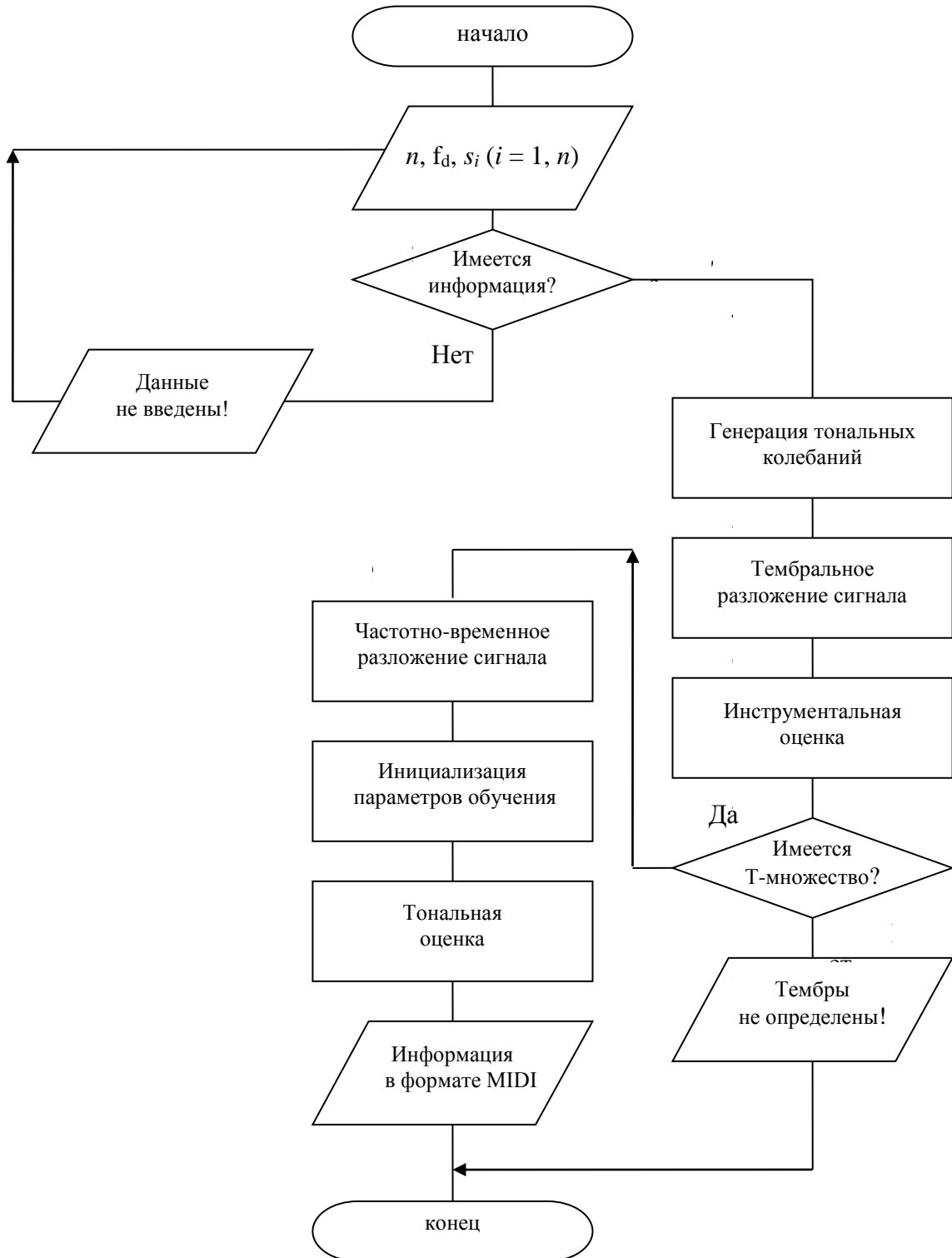


Рис. 4. Алгоритм работы предложенного метода

Данный подход аналогичен работе аранжировщика. В качестве доказательства, приведем детальное описание каждого блока (рис. 4).

В блок *исходных данных* поступают значения звукового сигнала  $s_i$  ( $i = 1, 2, \dots, n$ ), которые соответствуют мгновенным значениям

амплитуды в каждый момент времени и частота дискретизации  $f_d$  (рис. 5).

Затем, выполняется проверка на результат чтения исходных данных. Если чтение не было выполнено или было выполнено не полностью,

а частично, то выводится соответствующее сообщение (блок 3), которое информирует пользователя о том, что необходимо повторно выполнить ввод исходных данных. Если же все прошло успешно, то переходим к блоку 4.

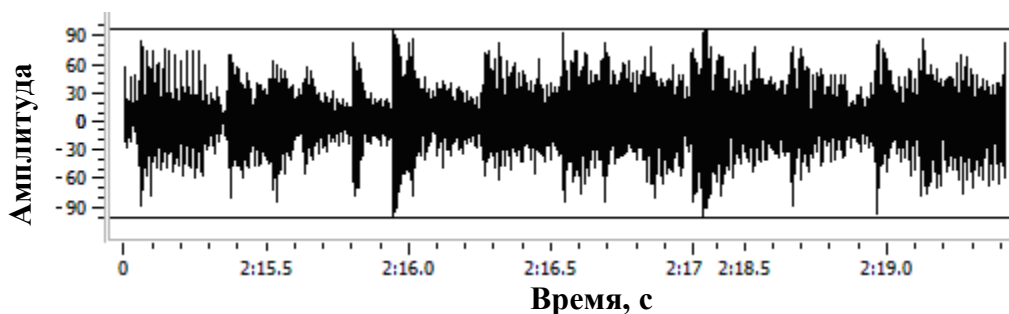


Рис. 5. Входная звуковая информация

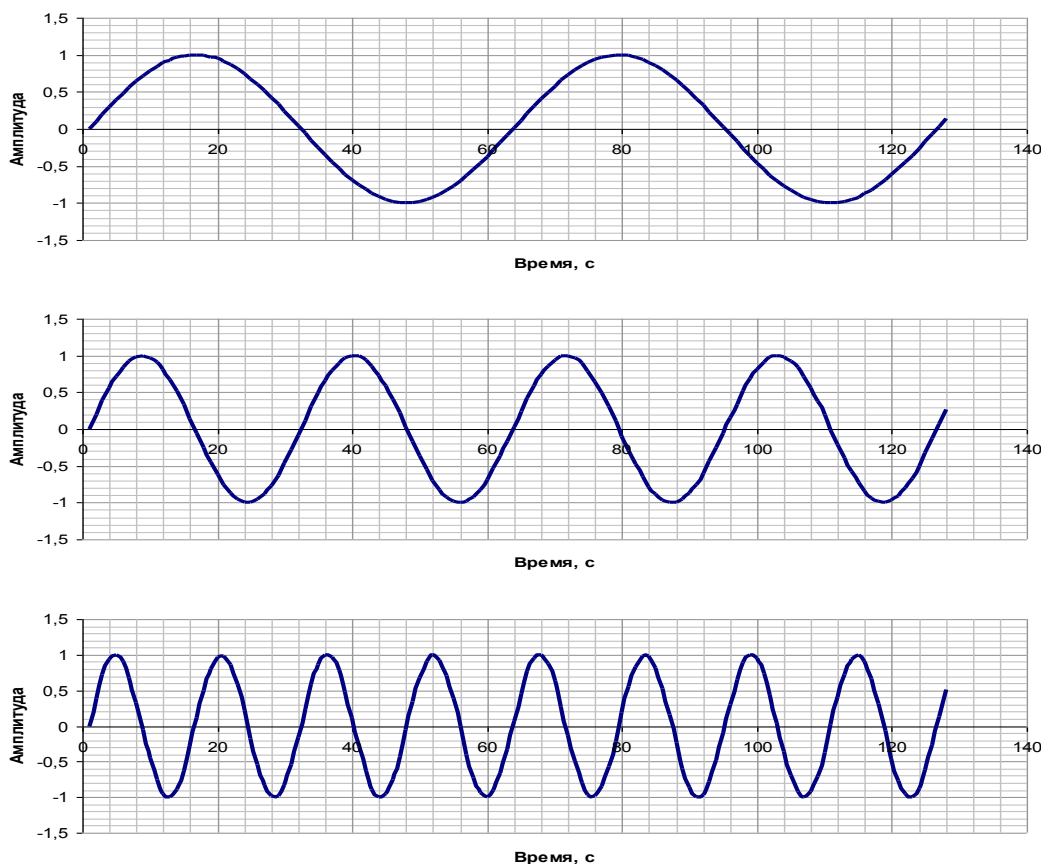


Рис. 6. Основные гармоники некоторых звуков фортепиано

В блоке *генерации тональных колебаний* формируются гармоники, соответствующие частотному диапазону звуков фортепиано (6). Выбор данного инструмента объясняется тем, что он имеет самый широкий звуковой диапазон среди остальных музыкальных инструментов (табл. 1). Примеры таких гармоник показаны на рисунке 6 [10].

$$g_i(t) = \sin(mnt),$$

б)

где

$g_i(t)$  –  $i$ -тая гармоника;

$m$  – расстояние между двумя полутонами ( $m = \sqrt[12]{2} \approx 1,06$ );

$n$  – номер гармоники.

В блоке 5 выполняется *тембральное разложение сигнала* с помощью дискретного вейвлет-преобразования (ДВП) согласно формуле:

$$T = \sum_{n=0}^{\infty} s[n]w(t), \quad (7)$$

где

$T$  – результирующий вейвлет-спектр;  
 $s[n]$  – входной аудио сигнал;  
 $w(t)$  – базисная вейвлет-функция.

В качестве базисных функций используются гармоники (6), определенные в предыдущем блоке. Это позволит локализовать составляющие тембров различных музыкальных инструментов, так как результирующий  $T$ -спектр будет отображать степень "сходства" для каж-

дой гармонике на всем временном интервале анализируемого сигнала.

В блоке *инструментальной оценки* определяются моменты времени с минимальным гармоническим составом и составляется  $T$ -множество – множество элементарных тембральных образов звукового сигнала из  $N_{unc}$  элементов (рис. 7).

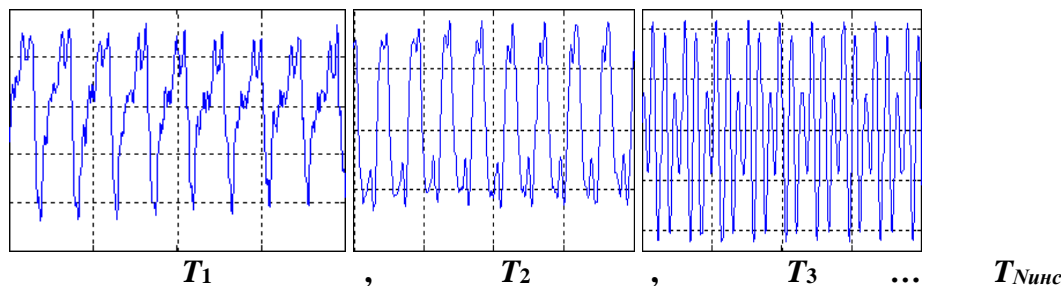


Рис. 7.  $T$ -множество

Таблица 1. Частотные диапазоны некоторых музыкальных инструментов

| Инструмент                | Граничные частоты, Гц |             |
|---------------------------|-----------------------|-------------|
|                           | нижняя                | верхняя     |
| <b>Рояль (фортепиано)</b> | <b>27</b>             | <b>4200</b> |
| Контрабас                 | 40                    | 300         |
| Туба                      | 45                    | 320         |
| Валторны                  | 60                    | 740         |
| Виолончель                | 65                    | 880         |
| Тромбон                   | 80                    | 500         |
| Кларнет                   | 140                   | 1980        |
| Труба                     | 160                   | 990         |
| Скрипки                   | 210                   | 2800        |
| Гобой                     | 230                   | 1480        |
| Флейта                    | 240                   | 2300        |
| Пикколо-флейта            | 560                   | 2500        |

Затем, выполняется проверка  $T$ -множества. Если оно не было определено, то выводится соответствующее сообщение (блок 8), которое информирует пользователя о том, что отсутствуют тембральные элементы. Если же все прошло успешно, то переходим к блоку 9.

В блоке *частотно-временное разложение сигнала* выполняется дискретное вейвлет-преобразование для каждого элементарного  $T$ -образа согласно формуле:

$$W = \sum_{i=1}^{N_{ein}} \sum_{n=0}^{\infty} s[n] w_i(t), \quad (8)$$

где

$W$  – результирующее множество вейвлет-спектров ( $W$ -множество);

$s[n]$  – входной аудио сигнал;

$w_i(t)$  – базисная вейвлет-функция.

В качестве базисных вейвлет-функций  $w_i(t)$  выступают элементы из  $T$ -множества:

$$w_i(t) = T_i, \quad i = 1, 2, \dots, N_{unc}; \quad (9)$$

Масштабирование происходит по рекуррентному закону исходя из формулы распределения частот в музыкальных инструментах [11].

$$a_{j+1} = a_j + 2^{1/12}; \quad (10)$$

В результате формируется  $W$ -множество – множество тембральных вейвлет-разложений из  $N_{unc}$  элементов.  $W$ -множество содержит разложения исходного аудио сигнала по тембральным составляющим и локализует звуковые объекты с их временными и частотными характеристиками для каждой партии музыкальных инструментов.

Блок 10 *инициализирует параметры обучения* для определения звуковых образов в рамках  $W$ -множества. Затем, выполняется распозна-

навание звуковых объектов с помощью специальных обучающих алгоритмов.

Информация, полученная в блоке 11, синтезируется и преобразовывается в MIDI-формат (Music Instrument Digital Interface).

### Заключение

Используя данный алгоритм, можно достичь хороших результатов в спектральном разрешении и при наложении тонов, благодаря использованию алгоритмов вейвлет-преобразований. Тембральное наложение можно будет разделить с помощью набора базисных вейвлет-функций, а размерное деление выполнить тогда, когда уже будет вычислен набор звуковых объектов для каждой инструментальной партии.

### Выводы

Таким образом, была проведена критическая оценка существующих методов в области автоматической идентификации музыкальных произведений, с указанием фундаментальных проблем извлечения звуковых объектов. Вместе с этим, сделано описание нового адаптивного метода, который может разрешить ряд актуальных на сегодняшний день проблем обработки звуковых сигналов.

Главное его достоинство в том, что преобразование звуковой информации будет выполняться, непосредственно, исходя из тембрального анализа конкретного звукового сигнала. Если в предыдущих разработках тембральные образы различных музыкальных инструментов были заранее определены, то здесь они будут формироваться на этапе тембральной музыкальной оценки. Данный подход повысит качество при идентификации музыкальных произведений различных стилей и синтеза звуковой информации.

### Список использованной литературы

1. Klapuri A. Signal Processing Methods for Music Transcription / A. Klapuri, M. Davy. — Springer, New York, 2006.
2. Ellis, D. Extracting Information from Music Audio / D. Ellis // LabROSA, Dept. of Electrical Engineering Columbia University, NY, March 15, 2006.
3. Фадеев, А. С. Идентификация музыкальных объектов на основе непрерывного вейвлет-преобразования / А. С. Фадеев // Диссертация. — Томский политехнический университет. — 2008.
4. Martin, K. D. Toward automatic sound source recognition: Identifying musical instruments / K. D. Martin // Proc. of the 1998 NATO Advanced Study Institute on Computational Hearing, II Ciocco, Italy, July, 1998.
5. Лукин, А. Введение в цифровую обработку сигналов (математические основы), Лаборатория компьютерной графики и мультимедиа, МГУ, 2007.
6. Солонина, А. И. Цифровая обработка сигналов: учебное пособие / А. И. Солонина, Д. А. Улахович, С. М. Арбузов, Е. Б. Соловьева. — Санкт-Петербург, "БХВ-Петербург", 2005, 768 с.
7. Every, M. Separation of musical sources and structure from single-channel polyphonic recordings / M. Every // PhD thesis. — Department of Electronics. — University of York. — 2006.
8. Яковлев А. Н. Введение в вейвлет-преобразования: учебное пособие / А. Н. Яковлев. — НГТУ, 2003. — 104 с.
9. Андреева, А. В. Развитие звуковысотного слуха младших школьников при обучении игре на скрипке / А. В. Андреева // Диссертация. — Нижний Новгород. — 2009.
10. Попов, О. Б. Цифровая обработка сигналов в трактах звукового вещания: учебное пособие / О. Б. Попов, С. Г. Рихтер. — "Горячая линия - Телеком". — М. — 2007.
11. Способин, И. В. Элементарная теория музыки. — Государственное музыкальное издательство. — М. — 1963.

### Сведения об авторе



**Корж Роман Анатольевич** – аспирант, кафедра моделирования и программного обеспечения, научный руководитель – д.т.н. профессор Азарян А.А., Криворожский национальный университет. Научные интересы – информационные технологии, распознавания образов.  
E-mail: [mk3089@mail.ru](mailto:mk3089@mail.ru)

Статья поступила в редакцию 10.10.2011 г.