

УДК 004.424.4:78.06

**Р.А. Корж**

**Криворожский национальный университет**

# **КРИТИЧЕСКИЙ ОБЗОР МЕТОДОВ АВТОМАТИЧЕСКОЙ ИДЕНТИФИКАЦИИ МУЗЫКАЛЬНЫХ ПРОИЗВЕДЕНИЙ**

*Целью данной работы является знакомство с одним из фундаментальных направлений в области обработки звуковых сигналов и обзор автоматизированных систем управления (АСУ) с кратким описанием, которые функционируют в данной предметной области.*

*Метою даної роботи є знайомство з одним із фундаментальних напрямів в області обробки звукових сигналів та огляд автоматизованих систем управління (АСУ) із коротким описом, які функціонують у даній предметній області.*

*The aim of this article is both introduction to the one of the fundamental branches of audio signal processing and overview of automatic control systems (ACS) with brief description, which operate in this subject field.*

**Ключевые слова:** автоматическая идентификация музыкальных произведений, объектно-ориентированный формат, система экспертных знаний, преобразование звуковой информации

## **Введение**

Проблема преобразования звуковой информации в объектно-ориентированный формат стала востребованной с ростом и распространением ЭВМ в различных отраслях промышленности в конце 80-х годов XX столетия. На западе ее название сформировалось как *автоматическая идентификация музыкальных произведений*. Было изложено множество подходов и разработано большое количество автоматизированных систем управления (АСУ), которые использовали последние научные достижения на то время. Но большинство из них так и не нашло практического применения в связи с жесткими ограничениями, которые накладывались на анализируемый сигнал [1].

В 1975 году *Джеймсом Мюрером* была реализована первая АСУ, которая могла отделять музыкальные партии в произведениях со второй степенью инструментальной полифонии. Но, из-за ряда строгих ограничений, которые накладывались на входной аудио сигнал (частотные диапазоны инструментов не должны пересекаться, партии инструментов не должны накладываться друг на друга и другие), она не нашла своего практического применения [2].

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 [3].

Звуковая информация представляется в виде совокупности звуковых объектов. Звуковые объекты одного инструмента образует *партию*. Партия, исполненная на отдельном музыкальном инструменте, называется *мелодией*. Совокупность партий различных музыкальных инструментов образует *партитуру*. Партитура музыкального произведения содержит полное описание нот всех музыкальных инструментов, другими словами, она представляет собой музыкальное произведение в *объектном* виде. Музыкальное произведение также характеризуется *темпом* (скоростью движения) и степенями *музыкальной* (максимальное количество одновременно звучащих звуковых объектов в пределах одного музыкального произведения), *объектной* (максимальное количество одновременно звучащих звуковых объектов в пределах одного инструмента) и *инструментальной* (максимальное количество одновременно звучащих музыкальных инструментов в пределах одного музыкального произведения) полифонии [4].

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].

Процесс идентификации музыкальных произведений подразделяется на три направления [1, 5]:

1. *Тональная оценка* (определение нот и аккордов);

2. *Ритмическая оценка* (определение ритма, такта и долей размера);

3. *Инструментальная оценка* (определение музыкальных инструментов).

Согласно [1, 3, 5], выполнена классификация музыкальных произведений по их объективному содержанию (табл. 1). Характеристика музыкальных произведений проводилась согласно двум параметрам: количество инструментов и количество нот.

Таблица 1. Объектная классификация музыкальных произведений

Инструменты	Один	Много
Один	очень редко встречается	практически не встречается
Много	встречается, но не часто	очень часто встречается

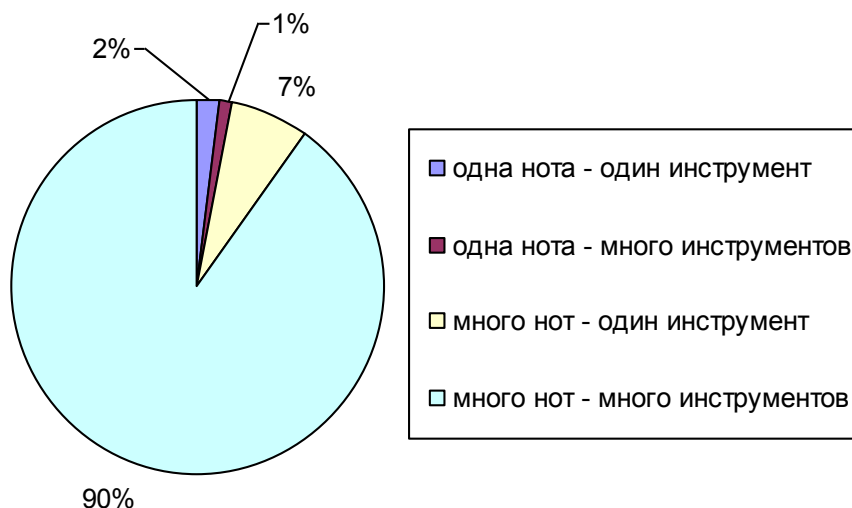


Рис. 1. Объектная классификация музыкальных произведений

Из рис. 1 видно, что очень востребованной является разработка АСУ, которая ориентирована на работу с классом музыкальных произведений «много инструментов – много нот».

#### Обзор существующих методов

Одной из самых неудачных попыток автоматизировать процесс преобразования является разработка Роберта Мейхера [6].

На входной аудио сигнал накладываются жесткие ограничения:

1. Степень инструментальной полифонии должна равняться двум.

2. С целью определения приблизительных частот, соответствующих частичным тонам, предполагается, что звуковые волны – квазигармонические (почти гармонические).

3. Не допускаются тембральные наложения в рассматриваемых временных отрезках.

4. Не допускается реверберация и прочие шумовые окраски.

Поэтому, данный подход обладает следующими недостатками:

1. Данный метод имеет жесткие ограничения и предназначен для очень узкого класса музыкальных произведений.

2. Данный метод не позволяет определять аккорды, а также и другие музыкальные объекты.

3. Шум и прочие искажения аудио сигнала резко снижают качественную оценку инструментального разделения.

4. Данный метод не позволяет отделять партии и распознавать тембральные характеристики инструментов.

Сравнительно лучшие оценки получили разработки Эмilia-Бодо-Давида [7] и Кашино-Мураса [8]. Они являются представителями

небольшого класса АСУ, которые проводят свой анализ исключительно во временной области звукового сигнала. Другими словами, они сопоставляют временные отрезки входного сигнала с огромным набором образцов-шаблонов из специальных баз данных. Такой подход является весьма грубым и неточным, что подтвердилось на практике и обладает следующими недостатками:

1. Метод грубого сопоставления спектров игнорирует слабовыраженные частотные компоненты сигнала, которые соответствуют основным тонам.
2. Низкое качество распознавания присутствует при анализе музыкальных произведений с высоким темпом звучания.
3. С увеличением степени музыкальной полифонии, увеличивается количество допускаемых ошибок распознавания.

4. Имеются ошибки распознавания в басовых и тремоло партиях.

5. Данный метод не позволяет отделять партии и распознавать тембральные характеристики инструментов.

Другим сформировавшимся направлением среди исследователей стало использование алгоритмов дискретного преобразования Фурье (ДПФ), которые раскладывают дискретную временную последовательность в гармонический ряд (ряд Фурье) и, затем, проводят спектральную гармоническую оценку, используя различные методы статистического анализа. Данный подход требует предварительного временного разбиения и дает достаточно низкие показатели в случае наложения гармоник.

Ярким примером служит разработка *Такуи Фуджишимы*, алгоритм которой изображен на рис. 2 [9].

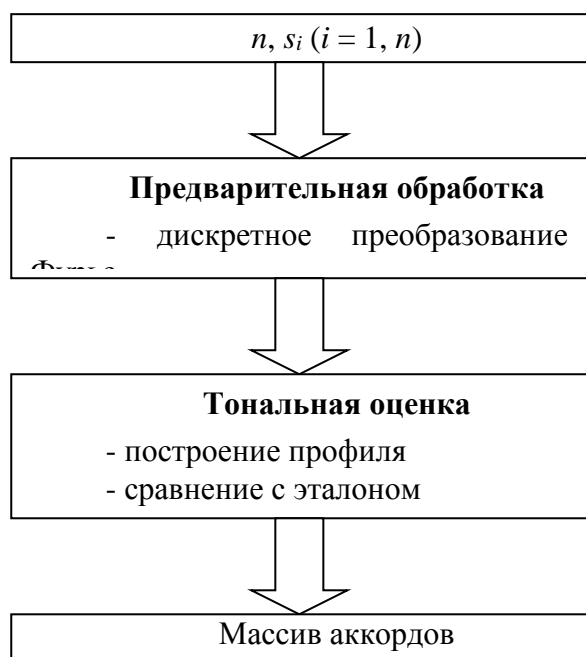


Рис. 2. Фрагмент алгоритма работы системы Такуи Фуджишимы

Входной сигнал  $s[n]$  разбивается на равномерные временные отрезки  $x[n]$ .

В блоке *предварительной обработки* происходит разложение каждого  $x_i$  в гармонический ряд  $X[n]$  с помощью дискретного преобразования Фурье (ДПФ):

$$X[n] = \sum_{i=1}^N s_i e^{-i\omega N} \quad 1)$$

Блок *тональной оценки* состоит из двух этапов. На первом этапе, для каждого  $X_i$  формируется профиль тональной классификации  $p_i$ , который показывает распределение интенсив-

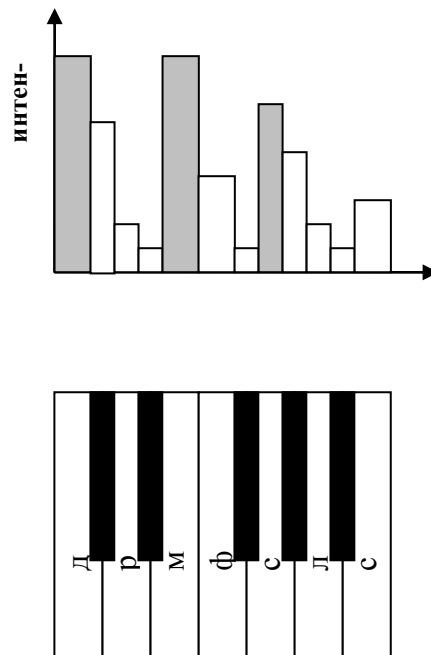


Рис. 3. Профиль тональной классификации

ности каждого полутона на всех октавных диапазонах (рис. 3). На втором этапе выполняется сравнение каждого  $p_i$  с образцами музыкальных аккордов  $C[n]$  по методу ближайшего соседа или по методу взвешенных сумм (на выбор пользователя).

На выходе формируется массив аккордов.

Этот подход обладает следующими недостатками:

1. Имеются высокие показатели качества и производительности для идеальных аудио сигналов, в случае реальных звуковых сигналов

требуются более качественные алгоритмы распознавания.

2. Наличие шума значительно ухудшает процесс распознавания аккордов.

3. На границах октав происходит небольшое наложение предыдущих нот на те, которые звучат в данный момент.

4. Система не позволяет отделять партии и распознавать тембральные характеристики инструментов.

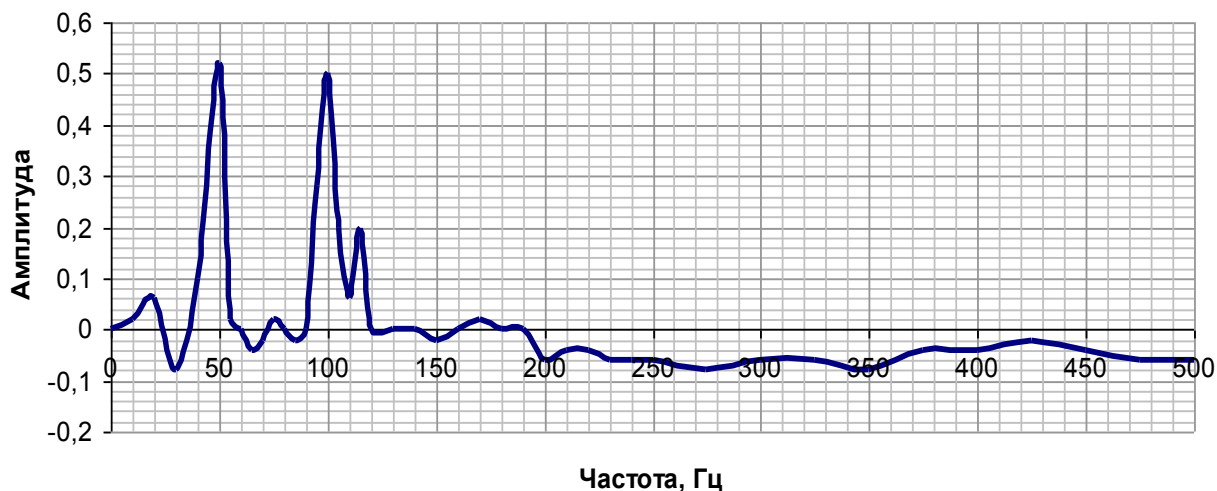


Рис. 4. ДПФ-спектр с четко выраженными локальными экстремумами

В разработке *Энсси Клапури* [10] сначала производится шумоподавление, сглаживая спектры  $X_i$  и убирая мелкие осцилляции. Затем, в каждом  $X_i$  определяются наиболее доминирующие частоты  $F_0$  и оцениваются их характеристики (рис. 4). На третьем этапе эти частоты  $F_0$  вычитаются из спектра. Однако, применение этого подхода имеет свои недостатки (рис. 5а):

1. Определение тактов и долей размера оказывается недостоверным при анализе музыкальных произведений без четко выраженных амплитудных акцентов.

2. Тональная оценка имеет низкие показатели при нотных переходах и вибрато.

3. Данный подход не позволяет отделять партии и распознавать тембральные характеристики инструментов.

Разработка *Валентина Эмия* [11] отличается тем, что в ней каждый  $X_i$  разбивается на два подспектра – фонового цветного шума  $Y_i$  и спектральной кривой  $Z_i$ .

С помощью статистического метода оценивания максимальных взвешенных правдоподобий авторегрессионной модели для анализа каждой  $Z_i$  и модели скользящего среднего для анализа каждой  $Y_i$  определяются частоты  $F_{0i}$  для каждого  $X_i$ , которые одновременно и в максимальной степени выравнивают  $Y_i$  и  $Z_i$ . Эти частоты соответствуют звуковым объектам, звучащим в данный момент. (рис. 5б).

Этот подход имеет свои существенные недостатки:

1. С увеличением степени музыкальной полифонии, резко увеличивается коэффициент ошибок распознавания (в связи с наложениями частичных тонов в частотной области), что ограничивает применение системы к распознаванию реальных звуковых сигналов;

2. В анализе полифонических мелодий, возникают ошибки в диапазоне самых низких и самых высоких тонов.

3. Аккорды, между которыми несколько октав распознаются с малой точностью.

4. Данный метод не позволяет отделять партии и распознавать тембральные характеристики инструментов.

*Эггинка и Браун* [12] ограничились тем, чтобы определять исключительно доминирующий инструмент и соответствующую ему нотную партию (рис. 6).

Блок *тональной оценки* состоит из трех этапов. На первом этапе каждый спектр  $X_i$  перемножается с функцией Гаусса, в результате получаются сглаженные версии первоначальных спектров  $X'_i$ , с четко выраженными пиками. На втором этапе оцениваются спектральные пики  $X'_i$  для выделения доминирующей музыкальной партии инструмента. Для этого используется эвристический алгоритм, эффективность которого приемлемая для большинства музыкальных инструментов. На третьем этапе

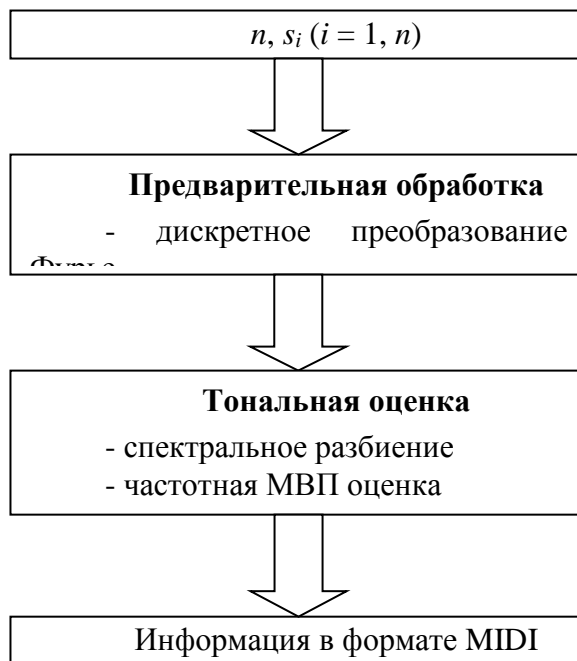
происходит извлечение доминант на фоне одновременно-звучащих тонов методом оценки соседних гармоник, который заключается в оценке относительной энергии доминирующих нот. Затем, формируется вектор признаков.

В блоке *инструментальной оценки* выполняется идентификация, путем передачи вектора

признаков в Гауссовскую смешанную модель, где строится плотность распределения вероятностей и определяется доминирующий инструмент вместе с принадлежащей ему нотной партией.



а)



б)

Рис. 5. Фрагменты алгоритмов работы систем:

а) Энци Клапури,  
б) Валентина Эмия

К недостаткам следует отнести следующее:

1. Вследствие того, что окно фиксированного размера, разрешение на высокочастотных компонентах сигнала достаточно слабое, что приводит к качественно-низким показателям работы;

2. При проектировании системы было сделано ограничение на пять музыкальных инструментов: флейта, кларнет, гобой, скрипка и контрабас.

Еще одним масштабным подходом к определению звукового содержания музыкальных произведений является применение *систем экспертных знаний*, которые содержат знания из различных предметных областей (рис. 7).

Центральный блок системы представляет собой вычислительную область, в которой вход расположен на самом нижнем уровне (ДПФ-спектр), а экспертные заключения определяются на самом высшем уровне [13]. При этом, процесс принятия решений идет снизу-вверх. Данный блок представляет собой *иерархию гипотез знаний*.

Вместе с тем, имеется множество *источников знаний*, определенные по принципу «если-то», с помощью которых выполняется комплексная экспертная оценка. Все операции планируются и координируются *диспетчером*.



Рис. 6. Фрагмент алгоритма работы системы Эггинки-Брауна

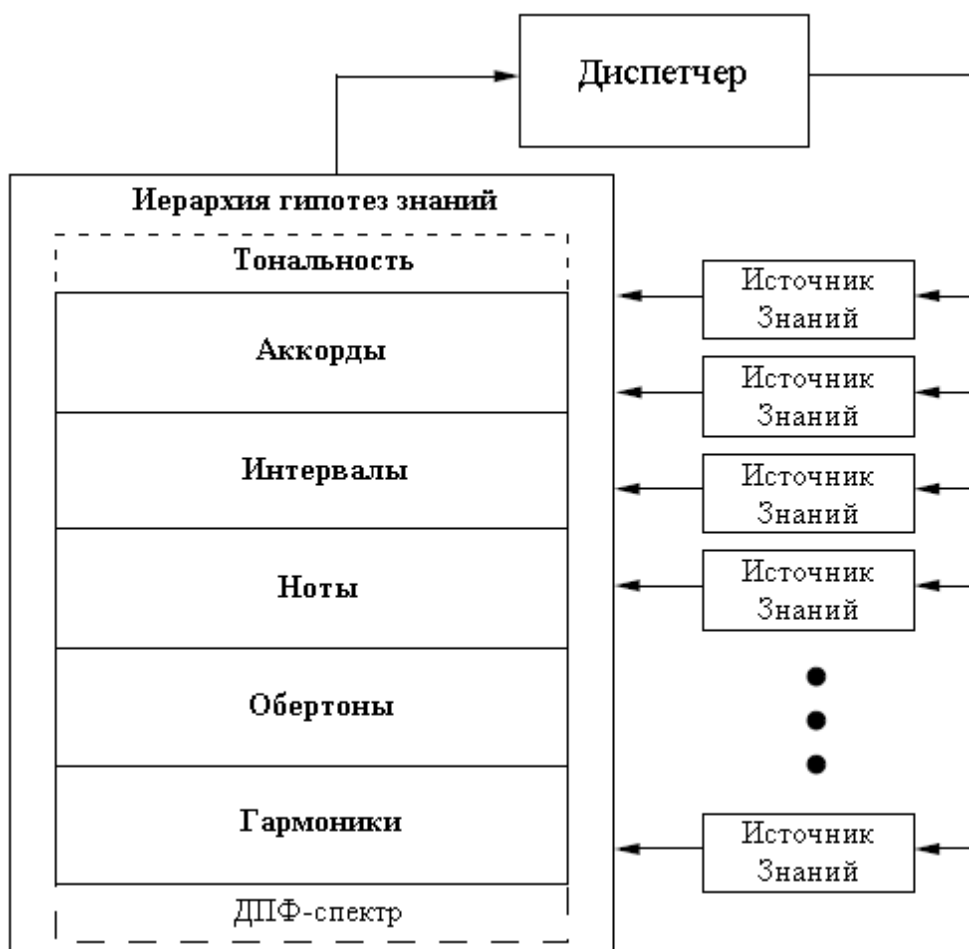


Рис. 7. Структура управления системы экспертных знаний



Реализацию подобных систем можно найти в разработках *Кейта Мартина* [3] и *Белла Сандлера* [14]. Следует отметить, что в них используется лишь небольшая часть от общей структуры системы экспертных знаний, включая малый размер экспертных знаний (рис. 7). Это, в первую очередь, связано с тем, что чем больше количество источников знаний, тем детальней требуется их уровневая структуризация. А если учесть еще тот факт, что знания из различных областей требуют, своего рода, смыслового соответствия, то задача реализации подобной системы сильно усложняется.

Недостатки подходов с применением систем экспертных знаний:

1. В случае аккордов, частичные тоны нот накладываются, возникает ошибка определения основных тонов и резко падает качественная оценка работы системы.

2. Наличие шума значительно ухудшает процесс определения основных тонов.

3. В случае, когда частичные тоны имеют ярко-выраженную окраску, появляются дополнительные ноты, которые отсутствуют в оригинальной аудиозаписи (для их удаления необходимо разработать алгоритмы, которые могут отразиться на производительности системы).

4. Данный метод не позволяет отделять партии и распознавать тембральные характеристики инструментов.

Некоторые разработчики [15-16] применяют модель линейной регрессии и различные обучающие методы для разделения гармонических наложений.

Другим довольно эффективным подходом является применение вейвлет-преобразований с последующей обработкой искусственных нейронных сетей. Этот метод идентификации имеет много общего с алгоритмами распознавания образов. Вначале, выполняется построение вейвлет-спектра входного сигнала (рис. 8б), а затем, эти данные обрабатываются с помощью искусственных нейронных сетей различных структур.

Базисные функции вейвлет-преобразования 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 (рис. 8а).

Благодаря масштабированию и смещению, вейвлеты позволяют проводить анализ с высоким разрешением по времени, но с низким раз-

решением по частоте и, наоборот, с низким разрешением по времени, но зато с высоким разрешением по частоте. А этот принцип удовлетворяет главному требованию музыкального анализа.

Таким подходом обладают системы *Ливингстона-Шепарда* [17] и *Александра Фадеева* [4]. Их качественно-количественные показатели на порядок выше, чем у предыдущих АСУ, но, несмотря на это, присутствуют следующие недостатки:

1. Отсутствие адаптации к конкретному звуковому сигналу состоит в определении пользователем во время выполнения числа анализируемых масштабов.

2. В качестве базисных функций используются вейвлеты Морле, которые не в максимальной степени показывают «сходство» с анализируемым сигналом (благодаря тому, что их форма не совпадает с большинством инструментальных тембральных образов), вследствие этого присутствуют неточности при распознавании музыкальных инструментов.

3. Модули предварительной обработки не содержат алгоритмов шумоподавления, которые могут качественно снизить процесс идентификации.

### Выводы

Было выполнено знакомство с таким направлением звуковой обработки сигналов, как автоматическая идентификация музыкальных произведений, изучением которой сегодня занимаются множество исследователей из области информационных технологий. Это объясняет столь широкое разнообразие АСУ, направленное на достижение максимальных результатов. Следует отметить, что более поздние разработки значительно эффективнее справляются с этой задачей, чем ранее разработанные. Это доказывает потребность в применении новых методологий и подходов в данной проблемной области.

Обзор современных АСУ показал, что необходимо выполнить исследования в области звуковой обработки сигналов для разработки нового подхода, который включил бы в себя все достоинства существующих алгоритмов. Также, следует использовать более эффективную математическую модель для повышения качества и производительности преобразования звуковой информации в объектно-ориентированный формат.

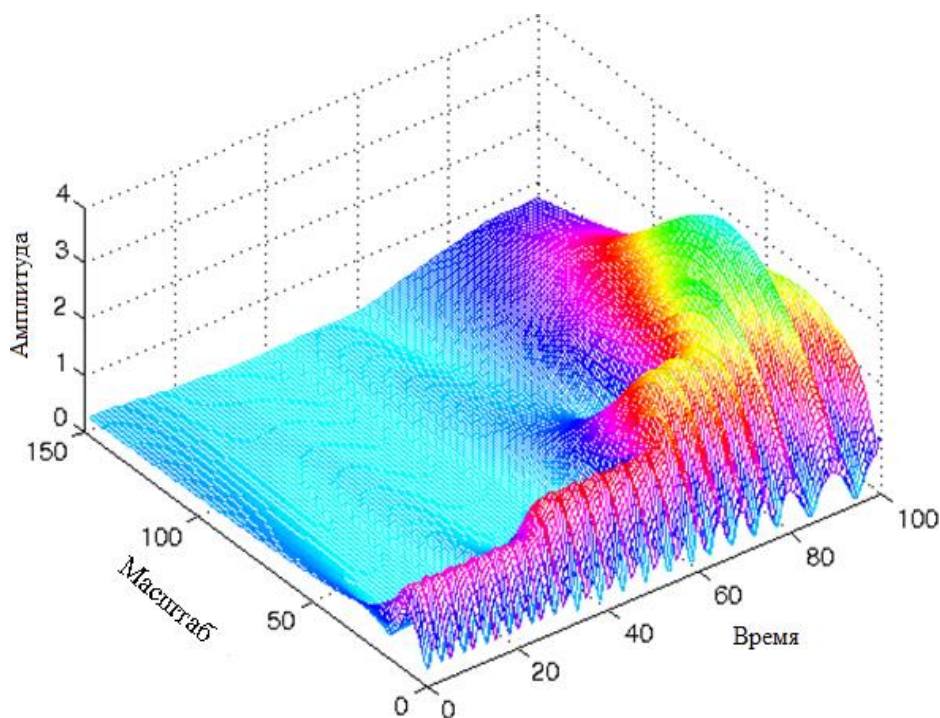
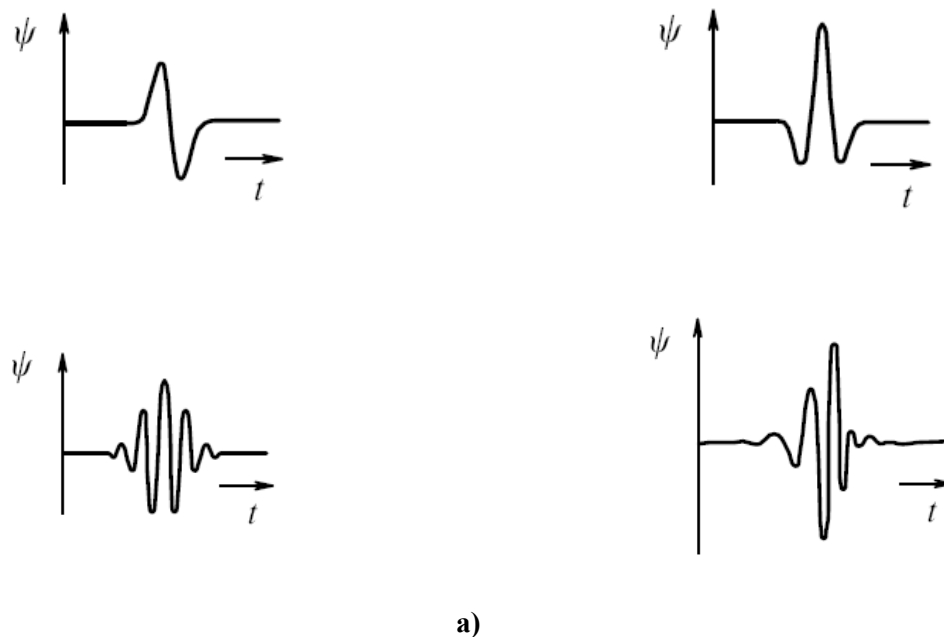


Рис. 8. Вейвлет-преобразование:  
а) базисные функции;  
б) вейвлет-спектр

#### Список использованной литературы

1. Klapuri A. Signal Processing Methods for Music Transcription / A. Klapuri, M. Davy. — Springer, New York, 2006.

2. Martin, K. D. Automatic Transcription of Simple Polyphonic Music / K. D. Martin // Computer Music Journal. — 2002. — No 1(7).

3. Ellis, D. Extracting Information from Music Audio / D. Ellis // LabROSA, Dept. of Electrical Engineering Columbia University, NY, March 15, 2006.



4. Фадеев, А. С. Идентификация музыкальных объектов на основе непрерывного вейвлет-преобразования / А. С. Фадеев // Диссертация. — Томский политехнический университет. — 2008.
5. Every, M. Separation of musical sources and structure from single-channel polyphonic recordings / M. Every // PhD thesis. — Department of Electronics. — University of York. — 2006.
6. Maher, R. C. Development and evaluation of a method for the separation of musical duet signals / R. C. Maher // Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, NY, Mohonk, October, 1989. — P. 1 — 3.
7. Emiya, V. Automatic transcription of piano music based on HMM tracking of jointly-estimated pitches / V. Emiya, R. Badeau, B. David // Proc. Int. Conf. Audio, Speech and Signal Processing (ICASSP). — 2009.
8. Kashino, K. Music recognition using note transition context / K. Kashino, H. Murase // Proc. of the 1998 IEEE ICASSP. Seattle. 1998.
9. Fujishima, T. Realtime Chord Recognition of Musical Sound: a System Using Common Lisp Music / T. Fujishima // Proc. of the International Computer Music Conference, Beijing: International Computer Music Association, China, 1999. — P. 464 — 467.
10. Klapuri, A. Musical meter estimation and music transcription / A. Klapuri // Proc. of the Cambridge Music Processing Colloquium, Cambridge University, UK, 2003.
11. Emiya, V. Multipitch Estimation of Quasi-Harmonic Sounds in Colored Noise / V. Emiya, R. Badeau, B. David // Proc. of the 10th Int. Conference on Digital Audio Effects (DAFx-07), Bordeaux, France, 2007. — P. 93 — 98.
12. Eggink, J. Instrument recognition in accompanied sonatas and concertos / J. Eggink, G. J. Brown // Proc. of IEEE Int. Conf. on Audio, Speech and Signal Processing (ICASSP). — V-4. — 2004. P.: IV-217 — IV-220.
13. Martin, K. D. A blackboard system for automatic transcription of simple polyphonic music / K. D. Martin // Technical Report #385, MIT Media Lab, Perceptual Computing Section, July, 1996.
14. Bello, J. Blackboard system and top-down processing for the transcription of simple polyphonic music / J. Bello, M. Sandler // Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00), Verona, Italy, December 7-9, 2000.
15. Chafe, C. Source separation and note identification in polyphonic music / C. Chafe, D. Jaffe // Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP. Apos 86. — V. 11, 1986. — P. 1289 — 1292.
16. Chien, Y. An Automatic Transcription System with Octave Detection / Y. Chien, S. Jeng // National Taiwan University, Graduate Institute of Communication Engineering and Department of Electrical Engineering, Taipei, 2002.
17. Livingston, J. Musical Instrument Identification using wavelets and Neural Networks / J. Livingston, N. Shepard // The University of Texas at Austin. — Austin, Texas, USA, 2005. — 20 p.

## 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.