

На правах рукописи

Фадеев Александр Сергеевич

**ИДЕНТИФИКАЦИЯ МУЗЫКАЛЬНЫХ ОБЪЕКТОВ
НА ОСНОВЕ НЕПРЕРЫВНОГО
ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ**

Специальность 05.13.01
Системный анализ, управление и обработка информации
(отрасль: промышленность)

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук

Томск — 2008

Работа выполнена на кафедре Автоматики и компьютерных систем Государственного образовательного учреждения высшего профессионального образования «Томский политехнический университет».

Научный руководитель: доктор технических наук, профессор
Цапко Геннадий Павлович

Официальные оппоненты: доктор технических наук, профессор
Гончаров Валерий Иванович

кандидат технических наук, доцент
Конев Антон Александрович

Ведущая организация: Государственное образовательное
учреждение высшего профессионального
образования «Новосибирский
государственный технический университет»,
г. Новосибирск

Защита диссертации состоится 10 декабря 2008 г. в 14:30 на заседании совета по защите докторских и кандидатских диссертаций Д 212.269.06 при Томском политехническом университете по адресу: г. Томск, ул. Советская, 84, институт «Кибернетический центр» ТПУ.

С диссертацией можно ознакомиться в Научно-технической библиотеке Томского политехнического университета по адресу: г. Томск, ул. Белинского, 55.

Автореферат разослан «07» ноября 2008 г.

Ученый секретарь Совета
кандидат технических наук, доцент

Сонькин М.А.

Общая характеристика работы

В современных мультимедийных системах для хранения и обработки музыкальных произведений используются различные форматы. Большинство форматов музыкальных произведений можно классифицировать на амплитудно-временные и объектные. К амплитудно-временным форматам относятся сигналы в аналоговом виде и цифровые форматы импульсно-кодовой модуляции сигнала. К объектным форматам относятся форматы нотных записей и такие, как форматы, основанные на цифровом интерфейсе музыкальных синтезаторов *MIDI (Musical Instrument Digital Interface)*. Форматы *MIDI*, так же как и нотная запись, позволяют описать ноты как объекты музыкального произведения с определенным набором свойств и отношений между ними.

В соответствие двум классам форматов введено и два типа преобразований музыкальных сигналов между ними. В настоящее время задача получения амплитудно-временных форматов музыкальных произведений по существующим объектным решена, и это решение активно используется, например, при звукозаписи произведений, исполняемых оркестром по нотам.

Актуальность работы обусловлена потребностью отдельных отраслей промышленности в автоматизированном получении объектного формата многоголосной мелодии по амплитудно-временному. Далее в работе эта задача названа задачей *идентификации музыкальных объектов*. Идентификация востребована в современной индустрии разработки программно-аппаратных средств, требующих графического и аудио-оформления событий пользовательского интерфейса. Малый объем файлов *MIDI*-формата позволяет включать элементы музыкального оформления без существенного увеличения физического объема разрабатываемых программных продуктов и требований к аппаратной части ЭВМ. Также высокую потребность в автоматизированных системах идентификации музыкальных объектов испытывают производители систем караоке, и отрасли промышленности связанные с реставрацией звукозаписей, нотным представлением игры виртуозов-исполнителей и композиторов, уменьшением объемов звуковых данных, процессом обучения и др.

К настоящему моменту было предпринято большое количество попыток автоматизировать процесс идентификаций мелодий музыкальных произведений и формирования по ним объектных форматов. Среди таких систем стоит отметить системы Мартинса Л. (*Martins L.*), Серафина С. (*Serafin S.*), Фуджишима Т. (*Fujishima T.*), Левью П. (*Leveau P.*), Эмия В. (*Emiya V.*) и др. Несмотря на высокие показатели работы систем в отдельных режимах, все они обладают существенными ограничениями в применении к реальным сигналам.

Некоторые авторы (Райен С. (*Rein S.*), Рейслин М. (*Reisslein M.*), Ливингстон Дж. (*Livingston J.*) и др.) предлагают использовать аппарат непрерывного вейвлет-преобразования (НВП) в качестве математического аппарата, позволяющего сформировать амплитудно-частотно-временное представление сигнала. Затем на основании наборов гармоник в каждый момент времени предлагается установление соответствия одному из тембральных базисов музыкальных инструментов и идентификация частотно-временных параметров объекта-ноты. Однако в качестве базиса НВП они используют стандартные вейвлет-функции. Применение же искусственно созданных базисов НВП на основе звучания музыкальных инструментов должно позволить настраивать само НВП на идентификацию нот музыкальных инструментов с

теми же частотными свойствами, что и свойства базиса. Определение наличия таких возможностей потребовало исследований в области НВП в аспекте эффективности и целесообразности его применения для идентификации музыкальных объектов.

Целью диссертационной работы является идентификация объектов-нот одно- и многоголосных мелодий отдельных музыкальных инструментов, а также идентификация партий определенных музыкальных инструментов на фоне звучания остальных в звукозаписи музыкальных произведений с использованием модификаций аппарата непрерывного вейвлет-преобразования.

Для реализации цели работы были поставлены следующие **задачи исследования:**

1. Провести анализ характеристик музыкальных сигналов, способов их описания, форм представления, форматов записи и возможных преобразований.
2. Провести обзор и анализ существующих систем распознавания музыкальной информации.
3. Оценить целесообразность применения НВП к идентификации музыкальных объектов, выявить технические границы возможностей его применения.
4. На основе проведенного анализа внести изменения и дополнения к аппарату НВП, позволяющие формировать базис НВП с заданными избирательными способностями. Разработать алгоритм повышения информативности результатов НВП, позволяющий автоматизировать дальнейшую обработку сигнала в среде искусственной нейронной сети (ИНС).
5. Модифицировать ИНС с прямой передачей типа «*MaxNet*» для решения задачи классификации музыкальных сигналов. Обеспечить возможность задания степени полифонической идентификации сети за счет изменения числа скрытых слоев.
6. Реализовать разработанные подсистемы в виде программно-технического комплекса. Разработать механизмы сопряжения форматов входных и выходных сигналов с применяющимися в настоящее время в мультимедийных системах.
7. Провести качественную и количественную оценку работы отдельных компонентов и системы в целом.

Методы исследования: методы статистической обработки результатов экспериментов; непараметрические методы идентификации; метод графо-аналитического исследования интерпретаций сигналов; метод слухового анализа музыкальных сигналов; методологии разработки и тестирования программного обеспечения.

Научная новизна работы:

1. Предложен метод формирования частных вейвлет-функций заданной избирательности, использующий фрагмент сигнала одной ноты различных музыкальных инструментов.
2. Обоснован способ формирования базисного вейвлета для задачи идентификации объектов в музыкальном произведении, заключающийся в формировании вейвлета на основе 16 периодов основного тона ноты «Ля субконтроктавы» музыкального инструмента.
3. Предложен способ изменения масштабного коэффициента НВП, заключающийся в формировании значений частоты вейвлетов, соответствующих частотам 96 тонов европейского равномерно темперированного строя музыки, что

позволило создать систему автоматизированной идентификации музыкальных объектов с диапазоном распознаваемых нот, составляющим 8 октав.

4. Разработан алгоритм повышения информативности результатов НВП, использующий наложение секущих на их частотные срезы с длиной, обратно пропорциональной значению частоты вейвлета текущего масштаба.
5. Для ИНС прямой передачи «*MaxNet*» с прямыми связями предложено в качестве настроечного параметра использовать число слоев сети. Изменение числа слоев ИНС позволило изменять степень полифонической идентификации системы в целом.

Практическая значимость работы:

1. Разработан метод идентификации одноголосных и многоголосных мелодий с количеством одновременно звучащих двух и более нот.
2. Разработан пакет прикладных программ, реализующих алгоритм непрерывного вейвлет-преобразования для сигналов и вейвлетов, представленных решетчатыми функциями. Пакет позволяет в качестве варьируемых параметров использовать частотно-временное окно исследования сигнала, базисный вейвлет, методы изменения коэффициентов НВП.
3. Программно реализован алгоритм повышения информативности результатов НВП в системах автоматизированной обработки.
4. Разработана программа, имитирующая работу ИНС поиска максимумов с варьируемым числом слоев.
5. На базе персонального компьютера реализован программно-технический комплекс, включающий программные реализации НВП, ИНС, других алгоритмов и методов, описанных в работе, и ряда коммерческих программных продуктов.
6. Разработаны формальные показатели оценки качества систем идентификации музыкальных объектов.

Основные положения диссертации, выносимые на защиту:

1. Метод формирования материнских вейвлетов заданной избирательности.
2. Метод формализации и дискретизации НВП в задаче идентификации объектов музыкальных произведений с целью уменьшения вычислительной избыточности.
3. Способ изменения масштабного коэффициента при формировании семейства вейвлет-функций, позволяющий классифицировать музыкальные объекты в соответствии с 96 полутонами европейского равномерно темперированного строя музыки.
4. Метод повышения информативности результатов НВП путем использования секущей на интервалах, кратных периоду гармонических составляющих сигнала.
5. Метод изменения степени полифонической идентификации ИНС путем варьирования числа скрытых слоев сети.
6. Структура и элементы программно-технического комплекса идентификации музыкальных объектов.

Личный вклад соискателя заключается в проведении теоретических исследований, их практической реализации, получении основных результатов, изложен-

ных в работе, их объяснении и интерпретации, авторском сопровождении при внедрении.

Сведения об апробации результатов диссертации:

Основные результаты диссертации были доложены на XII международной научно-практической конференции студентов и молодых ученых «Современные техника и технологии» (Томск, 2006 г.); международной открытой научной конференции «Современные проблемы информатизации в моделировании и программировании» (Воронеж, 2006 г.); IV, V и VI Всероссийских научно-практических конференциях студентов, аспирантов и молодых ученых «Молодежь и современные информационные технологии» (Томск, 2006 — 2008 гг.).

Внедрения:

Результаты работы были использованы при выполнении проекта внедрения системы идентификации музыкальных объектов в ООО Группа компаний «Интант» (г. Томск) и ООО «Тропикана» (г. Томск).

Результаты используются при выполнении научных исследований и выпускных квалификационных работ студентов кафедры автоматики и компьютерных систем Томского политехнического университета.

Сведения о публикациях по теме диссертации:

По теме диссертации опубликовано 9 работ. Из них 4 статьи — в журналах и 5 — в сборниках трудов конференций. 3 статьи опубликованы в ведущих рецензируемых журналах, рекомендованных Высшей аттестационной комиссией Министерства образования и науки Российской Федерации.

Структура диссертации, ее объем:

Диссертация состоит из введения, четырех глав, заключения, приложения и списка цитируемой литературы. В работе содержится 180 страниц печатного текста, 93 рисунка, библиография из 118 наименований на 12 страницах. Общий объем диссертации составляет 191 страницу.

Содержание работы

Во **введении** приводится обоснование актуальности работы, определяются цели и задачи исследования, научная новизна полученных результатов и их теоретическая и практическая значимость, формулируются положения, выносимые на защиту, приводятся основные характеристики работы.

Первая глава посвящена вопросам изучения предметной области: музыкальному сигналу, его форматам, описаниям и преобразованиям, формализации задачи идентификации музыкальных объектов. В главе вводятся основные понятия и термины, используемые в работе.

Музыкальным сигналом называется сигнал, полученный из звуков музыкальных инструментов путем аудиозаписи, воспроизведения ранее записанного сигнала, генерацией или иным способом. Наименьшей составляющей музыкального сигнала является звуковой объект — нота (O_i) одного музыкального инструмента. Каждая нота O_i обладает набором свойств: тон (m), тембр, определяемый музыкальным инструментом (k), время начала звучания (θ), продолжительность (Θ), громкость (A): $O_i = O(A_i, m_i, \theta_i, \Theta, k_i)$.

Совокупность нот O одного инструмента k_i образует партию. Партию, исполненную на одном музыкальном инструменте, называют мелодией. Совокупность партий разных инструментов образует партитуру. Партитура музыкального произ-

ведения O содержит описание всех нот всех музыкальных инструментов $O = \{O_1, O_2, \dots, O_i, \dots, O_N\}$. O представляет собой музыкальное произведение в объектном виде. Для музыкального произведения также характерны темп (скорость следования метрических счетных единиц) и степень многоголосности (полифонии) — максимальное количество нот, присутствующих в звучании одновременно.

Тембр ноты определяется музыкальным инструментом и отражается на ее амплитудно-частотном спектре (рисунок 1). В спектре ноты присутствует основная гармоника наибольшей амплитуды, определяющая тон ноты и ряд обертонов на кратных частотах выше основного тона.

Исследования амплитудно-частотных характеристик показали, что для музыкальных инструментов, звучание которых производится свободными колебаниями (гитара, фортепиано, колокольчик, скрипка пиццикато и др.), амплитуда обертонов значительно меньше амплитуды основного тона, и их общее количество относительно невелико. Однако для музыкальных инструментов, звучание которых формируется автоколебаниями под действием сторонней силы (орган, труба, скрипка и др.), амплитуда и общее количество обертонов значительно превышают амплитуду и количество обертонов музыкальных инструментов со свободными колебаниями.

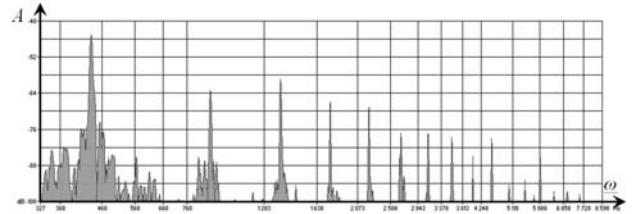


Рисунок 1 — Амплитудно-частотный спектр звука ноты фортепиано

Сигналы большинства музыкальных инструментов, представленные в амплитудно-временном формате, обладают свойством автомодельности. Это свойство позволяет из временной функции одной (базовой) ноты $n_0(t)$ определенного музыкального инструмента получить временную функцию любой другой ноты $n_i(t)$ (с требуемой частотой основного тона) этого же музыкального инструмента масштабированием функции $n_0(t)$: $n_i(t) = n_0(t/m_i)$, где m_i — коэффициент масштабирования, i — положение ноты $n_i(t)$ по высоте относительно ноты $n_0(t)$. Для равномерно темперированного строя европейской музыки $m_i = 2^{\frac{-i}{12}}$, $i = 0, \pm 1, \pm 2, \dots$. Это свойство применено во всех современных музыкальных синтезаторах, использующих метод волнового табличного синтеза.

В общем виде для формирования амплитудно-временного формата $f(t)$ музыкального произведения музыкальный синтезатор использует музыкальное произведение в объектном виде O и базис — набор базовых нот $N_0 = \{n_0^k(t)\}$ для всех k -инструментов: $f(t) = T(N_0, O)$. Здесь T — аналог математического оператора, выполняющий *прямое преобразование* музыкального сигнала из объектного O в амплитудно-временной формат $f(t)$. Процесс звукозаписи оркестра, исполняющего музыкальное произведение по нотным партитурам, представляет собой прямое преобразование: в качестве объектного формата O используется нотная запись, базис N_0 образуют музыкальные инструменты. Формируемая звукозапись представляет собой амплитудно-временной формат произведения.

Аналогично прямому T , введем и *обратное преобразование* (T^{-1}) амплитудно-временного формата музыкального сигнала $f(t)$ в объектный O : $O = T^{-1}(f(t), N_0)$. В результате обратного преобразования из музыкального сигнала $f(t)$ на основе базовых нот N_0 формируются объекты $O_i = O(A_i, m_i, \theta_i, \Theta, k_i)$. Решение задачи обратного

преобразования представляет собой решение задачи идентификации музыкальных объектов.

В мировой практике в основном используются качественные (субъективные) показатели оценки идентификации музыкальных композиций. Для комплексной оценки обратного преобразования T^{-1} был предложен ряд показателей: относительное количество распознанных нот ΔK_P , относительное количество нераспознанных нот ΔK_H , относительное количество нот, совпавших в распознаваемом и оригинальном произведениях ΔK_C , среднее относительное отклонение времени начала звучания каждой ноты $\Delta \theta$, среднее относительное отклонение времени продолжительности звучания каждой ноты $\Delta \Theta$.

Однако музыкальные произведения как произведения искусства принято оценивать качественно. В некоторых случаях качественные и количественные оценки сильно расходятся. Например, разница в продолжительности звучания нот слушателем может быть не замечена, однако в количественном эквиваленте отражена.

Обзор систем идентификации объектов музыкального сигнала показал, что существуют методы, которые обеспечивают высокие показатели при идентификации отдельных звуков. Однако при применении этих методов к более сложным задачам сегментации многоголосных фрагментов мелодий, в которых звучит больше одной ноты (созвучия, аккорды), или идентификации партий отдельных инструментов на фоне звучания других, показатели значительно ухудшаются. Кроме того, высокие показатели в искусственной среде значительно снижаются при обработке естественных сигналов, используемых в мировой практике обработки звука.

Тенденции последних лет в разработке систем идентификации музыкальных сигналов ведут к популяризации методов, способных идентифицировать составляющие сигнала в два этапа, без разделения сигнала на временные компоненты:

- 1) классификация присутствующих в музыкальном сигнале объектов по классам музыкальных инструментов;
- 2) идентификация параметров каждого отдельного объекта-ноты.

В решении задачи классификации музыкальных инструментов предлагается использовать базу данных об их тембральных образах. Большинство современных разработок предлагает использовать амплитудно-частотно-временные образы звучания нот музыкальных сигналов. Решения о соответствии конкретного тембрального образа сигналу производятся на основании их сравнения в амплитудно-частотно-временном представлении (АЧВП).

Переход от амплитудно-временного представления к АЧВП осуществляется методами оконного и быстрого преобразований Фурье, частотной фильтрации, методами вейвлет-преобразований и др.

Само сравнение АЧВП сигнала с АЧВП образа музыкального инструмента, принятие на его основе решения, а также решение задачи идентификации параметров каждого отдельного объекта (ноты) производятся, как правило, эвристическими алгоритмами и искусственными нейронными сетями.

В диссертационной работе для перехода к АЧВП музыкального сигнала принято решение применить достаточно гибкий и перспективный аппарат НВП. Особенностью предлагаемого подхода является использование не стандартных, а искусственно сконструированных из сигналов отдельных музыкальных инструментов вейвлетов.

Вторая глава посвящена описанию специфики непрерывного вейвлет-преобразования, возможностей его применения в практических задачах и задачах анализа музыкальных сигналов, представлению и интерпретации результатов.

Вейвлет-преобразования на сегодняшний день являются достаточно мощным инструментом исследования структуры нестационарных данных. Вейвлеты получили применение при решении широкого круга задач, связанных с подавлением шумов, сжатием больших объемов информации, анализом изображений, синтезом сигналов и т.д. С ростом возможностей вычислительной техники вейвлет-преобразования приобрели большое значение в различных областях науки. К возможностям вейвлет-преобразования можно отнести способность определять мгновенную амплитуду, фазу и частоту периодических компонент нестационарных широкополосных сигналов.

НВП сигнала $f(t)$ состоит в его разложении по некоторому базису, сконструированному из функции базисного вейвлета $w(t)$, посредством ее масштабирования и сдвигов вдоль оси времени:

$$[W_w f](\tau, s) = |s|^{-1/2} \int_{-\infty}^{\infty} f(t) w^* \left(\frac{t - \tau}{s} \right) dt .$$

Здесь $[W_w f](\tau, s)$ — коэффициенты преобразования, s — масштаб вейвлета, τ — параметр смещения вейвлета вдоль оси времени, символом «*» обозначена операция комплексного сопряжения. Выбор базисной функции $w(t)$ и формируемого на ее основе семейства вейвлетов $w_s(t) = |s|^{-1/2} w\left(\frac{t - \tau}{s}\right)$ определяется целями исследования. С помощью выбора базисных вейвлет-функций можно добиться выявления требуемых свойств рассматриваемого сигнала. Выбор осуществляется эмпирически.

Высокую популярность получил вейвлет Хаара. Наиболее часто этот вейвлет применяется в системах дискретного вейвлет-преобразования, фильтрации, сжатия аудио- и видеосигналов (рисунок 2, а). В системах непрерывных преобразований, как правило, применяются вейвлеты, имеющие аналитические выражения, легко интегрируемые и дифференцируемые численными методами. К таким вейвлетам относятся вейвлеты Гауса, Морле, Добеши и др. (рисунок 2, б, в). Аналитическое выражение для вейвлета Морле имеет вид: $w_{Morle}(t) = e^{-t^2/2} \cos(5t)$.

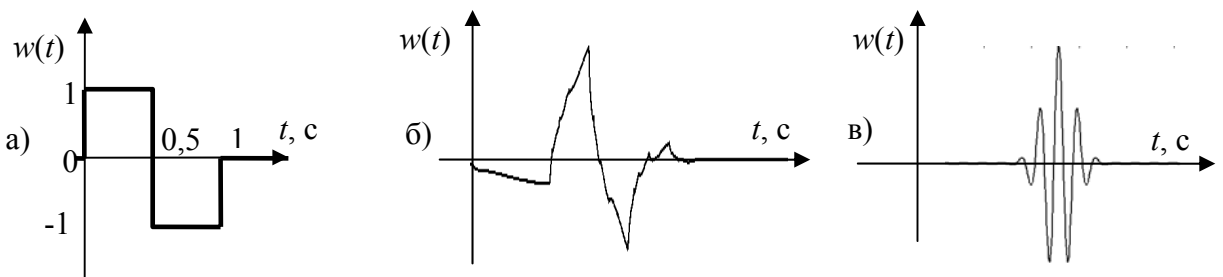


Рисунок 2 — Вейвлеты: а) Хаара; б) Добеши первого порядка; в) Морле
Базисные вейвлеты должны удовлетворять ряду требований:

- частотной локализации: $\hat{w}(\omega) \rightarrow 0$, при $\omega \rightarrow \infty$;
- временной локализации: $w(t) \rightarrow 0$, при $t \rightarrow \infty$;
- нулевого среднего: $\int_{-\infty}^{\infty} w(t) dt = 0$;

- ограниченности энергии: $\int_{-\infty}^{\infty} |w(t)|^2 dt \leq \infty$;
- кусочной непрерывности;
- автономности.

Функции, удовлетворяющие этим требованиям, могут быть использованы в качестве базиса вейвлет-преобразований.

Результатом НВП является поверхность $W(\tau, s)$, отражающая степень корреляции конкретного вейвлета $w_{\tau, s}(t)$ из одного семейства с функцией сигнала $f(t)$. Однако $W(\tau, s)$ может интерпретироваться как АЧВП сигнала. Многие базисные вейвлеты обладают узким (локализованным) частотным спектром $\varphi(\omega)$ вокруг основной частоты ω_0 . При масштабировании таких вейвлетов параметром s изменяется и значение основной частоты ω_0 : $\omega_s = s \cdot \omega_0$, где ω_0 — основная частота вейвлета $w_{\tau, s}(t)$ при $s=1$. Параметр τ определяет смещение вейвлета вдоль оси времени t в абсолютных единицах времени. В массиве результатов НВП $W(s, \tau)$ параметр τ можно заменить на переменную времени t , а параметр s — на отношение $s = \omega_s / \omega_0$. Если величина ω_0 для конкретного базисного вейвлета известна, то результаты НВП $W(\omega_s, t)$ могут быть представлены как в пространстве $W-s-\tau$, так и в пространстве частота-время: $W-\omega_s-t$. При графической интерпретации результатов на плоскости ω_s-t ось t представляет собой ось времени, а ось ω_s — ось относительной частоты ($\omega_s = s \cdot \omega_0$). При интерпретации результатов НВП в пространстве $W-\omega_s-t$ ось W представляет собой ось значений НВП в каждой точке $\omega_s = s \cdot \omega_0$, и $t = \tau$ (рисунок 3).

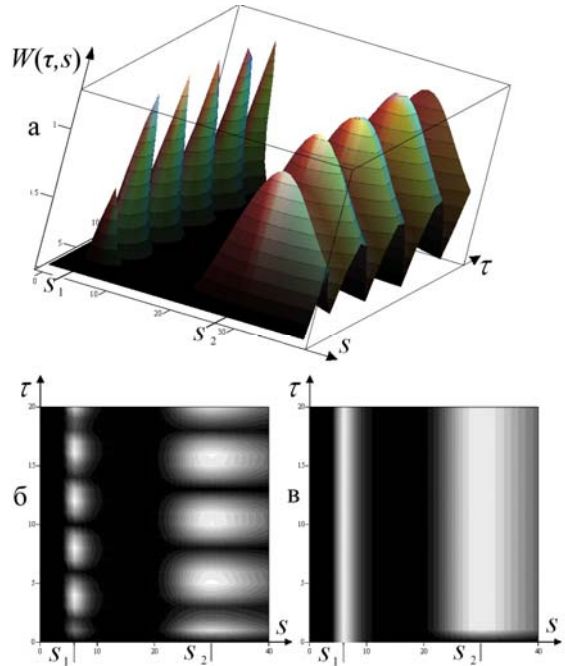


Рисунок 3 — Представления результатов НВП для сигнала $f(t) = \sin(250t) + \sin(300t)$: а) объемная фигура (скалограмма); б) карта проекций изолиний; в) результат сглаживания

Переход к частотно-временной плоскости при построении графических изображений по значениям $W(\tau, s)$ позволяет проследить за поведением частотных составляющих сигнала в каждый момент времени на интервале наблюдения. На рисунке (рисунок 4) приведена амплитудно-частотно-временная характеристика сигнала модельной системы $f(t) = p(t) + h(t)$. В состав сигнала $f(t)$ входит «полезный» сигнал $p(t)$, частота которого $\omega_p(t)$ изменяется по закону расходящегося колебательного процесса (сигнал частотно модулирован расходящимися колебаниями). В начальный момент времени частота сигнала $p(t)$ составляет 400 Гц.

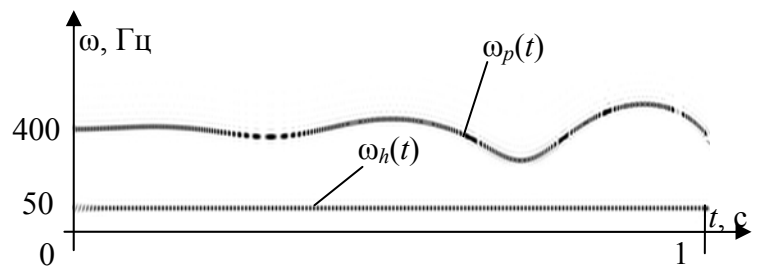


Рисунок 4 — Графическая интерпретация результатов НВП для сигнала $f(t) = p(t) + h(t)$

В исследуемом сигнале $f(t)$ также присутствует сигнал помехи $h(t)$. Частота сигнала помехи на всем интервале наблюдения постоянна $\omega_h(t)=50\text{с}^{-1}$.

НВП сигнала $f(t)=p(t)+h(t)$ позволяет построить графическое представление результатов НВП $W(\tau,s)$ (рисунок 4).

В работе приводятся примеры графической интерпретации НВП ряда сложных модельных сигналов, а также НВП реальных сигналов солнечной активности и электрических сигналов радиоволн. Широкие возможности графической интерпретации результатов вейвлет-преобразований позволяют выполнять анализ частотно-временной структуры сигнала и принимать решения относительно частотной динамики сигнала. Однако само НВП не решает задач идентификации и принятия решений и требует дальнейшей обработки результатов другими системами.

Третья глава представляет собой описание основных теоретических и практических исследований в области адаптации математического аппарата НВП к задачам идентификации музыкальных объектов.

Для исследования влияния формы материнского вейвлета на результат НВП был проведен ряд экспериментов с использованием модельного сигнала $F(t) = \sin(2 \cdot \pi \cdot 25 \cdot t) + \sin(2 \cdot \pi \cdot 120 \cdot t)$ и стандартных вейвлет-функций Хаара, Добеши, Гауса, Мейера, Морле и др. Установлено, что схожесть частотных спектров исследуемого сигнала и вейвлет-функции дает наилучшую частотную локализацию гармонических составляющих сигнала в массиве выходных значений $W(\tau,s)$. И наоборот, для получения наилучших результатов при использовании НВП следует выбирать материнский вейвлет таким образом, чтобы его амплитудно-частотная характеристика была как можно ближе к амплитудно-частотной характеристике элементарных составляющих сигнала.

Было сделано предположение, что дальнейшее повышение информативности результатов возможно при абсолютном совпадении вейвлет-функции и исследуемого сигнала в каждый момент времени. Так как в основе НВП лежит корреляция сигнала $f(t)$ и вейвлета

$w_s(t) = w\left(\frac{t}{s}\right)$: $Wf_s(\tau) = \int_{-\infty}^{\infty} f(t)w_s(t-\tau)dt$, то значения

$Wf_s(\tau)$ будут максимальным для таких τ , при которых функции $f(t)$ и $w_s(t-\tau)$ равны: $f(t)=w_s(t-\tau)$. Такое равенство может быть обеспечено при применении вейвлет-функции, сформированной из локального участка исследуемого музыкального сигнала: $w_s(t)=f(t+\tau)$. Для задачи идентификации нот музыкального инструмента материнский вейвлет формируется из локального участка функции сигнала базовой ноты этого инструмента $n_0(t)$: $w(t)=n_0(t+t_0)$, $0 \leq t \leq T$ (рисунок 5).

Для соответствия требованиям, предъяв-

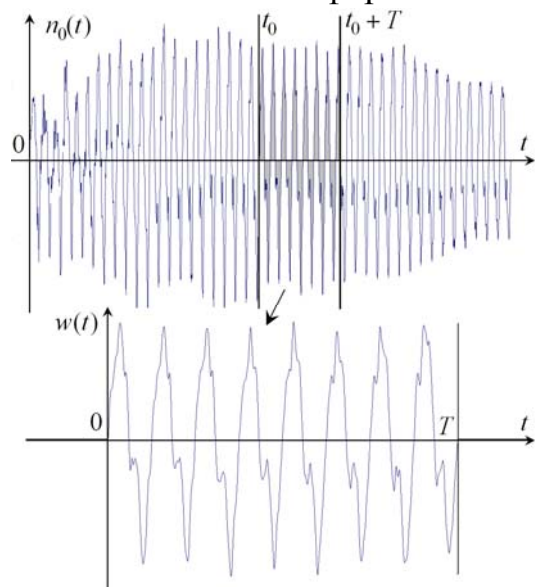


Рисунок 5 — Функция $n_0(t)$ и сформированный на основе ее фрагмента вейвлет $w(t)$

ляемым к базисным вейвлетам при выборе значений t_0 и T , необходимо использовать периодический участок сигнала $n_0(t)$ с нулевыми начальным и конечным моментами $n_0(t_0)=0$ и $n_0(T+t_0)=0$, такой, чтобы $\int_{t_0}^{T+t_0} n_0(t)dt = 0$. Кроме того $w(t)=0$, при $t \notin [0, T]$.

Для оценки избирательной способности создаваемых вейвлетов были созданы базисные вейвлеты из сигналов некоторых музыкальных инструментов (скрипка, скрипка пиццикато, орган, колокольчик, фортепиано, труба и др.). Каждый из сформированных вейвлетов был использован в НВП с каждым из сигналов тех же музыкальных инструментов.

Анализ графических интерпретаций полученных значений $W(\tau, s)$ позволил разделить исследуемые сигналы на две группы: группу музыкальных инструментов со свободными колебаниями и группу музыкальных инструментов с автоколебаниями. Результаты НВП сигналов музыкальных инструментов со свободными колебаниями содержали четко выраженную основную гармонику исследуемого сигнала на всем временном интервале, что соответствует наличию одной ноты в исследуемом интервале времени (рисунок 6).

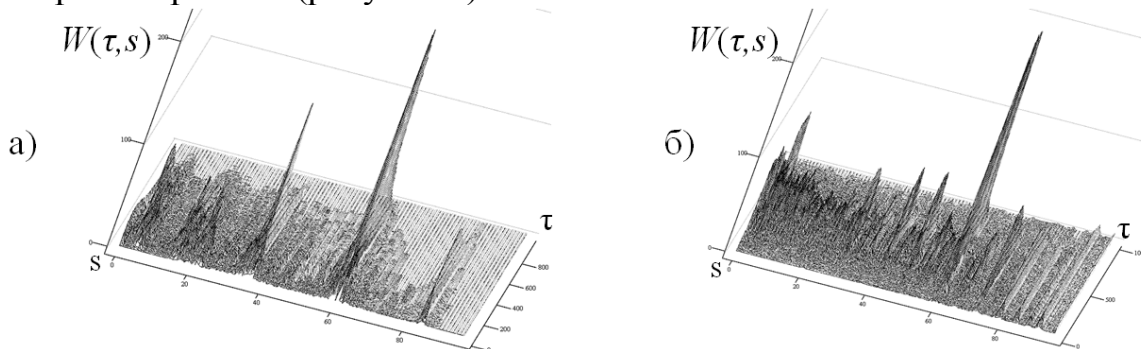


Рисунок 6 — Результаты НВП: а) сигнал — звук колокольчика, материнский вейвлет — вейвлет-колокольчик; б) сигнал — звук гитары, материнский вейвлет — вейвлет-гитара

При НВП сигналов музыкальных инструментов с автоколебаниями с применением различных вейвлетов (из обеих групп) идентифицировать частоту основной гармоники на фоне остальных удавалось только в случае применения к сигналу одноименного вейвлета. В остальных случаях, как показывают графические интерпретации НВП (рисунок 7), амплитуда основной гармоники равна или значительно меньше амплитуд других частотных составляющих.

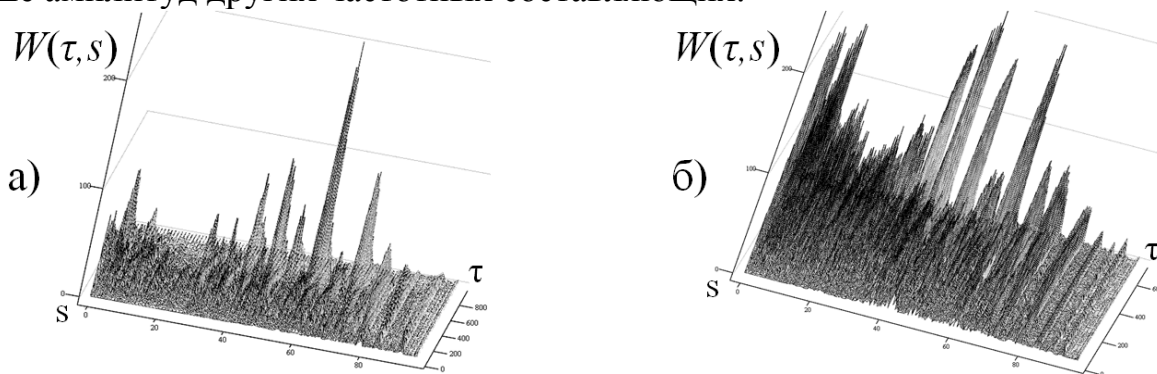


Рисунок 7 — Результаты НВП: а) сигнал — звук органа, материнский вейвлет — вейвлет-орган; б) сигнал — звук трубы, материнский вейвлет — вейвлет-скрипка

Эксперимент показывает способность выявления объекта (ноты определенного инструмента) в музыкальном фрагменте при помощи одноименного вейвлета. Причем НВП выявляет не частотные составляющие (набор и поведение гармоник) в сигнале, а частоту основной гармоники самого сигнала (следовательно, и тон ноты музыкального инструмента), его временное положение и длительность. Тогда как стандартные вейвлеты позволяют получить лишь представление об изменениях частотных составляющих сигнала во времени, а не об объектах — нотах определенного инструмента в целом.

Одним из свойств базисного вейвлета, определяющего его избирательную способность, является его локализация в частотной и временной областях. При НВП каждый вейвлет одного семейства (полученный из одного материнского вейвлета) образует частотно-временное окно конечного размера. Площадь окна для вейвлетов одного семейства на плоскости s – τ всегда остается постоянной.

Однако в ходе проведения ряда экспериментов выяснилось, что при изменении самого вейвлета геометрические характеристики окна изменяются. Так, с увеличением количества периодов в материнском вейвлете (а следовательно, и во всех вейвлетах семейства) окно вытягивается вдоль оси времени, сужаясь относительно оси масштабов, и наоборот, с уменьшением числа периодов в материнском вейвлете окно вытягивается вдоль оси масштабов s , сужаясь вдоль оси времени (рисунок 8).

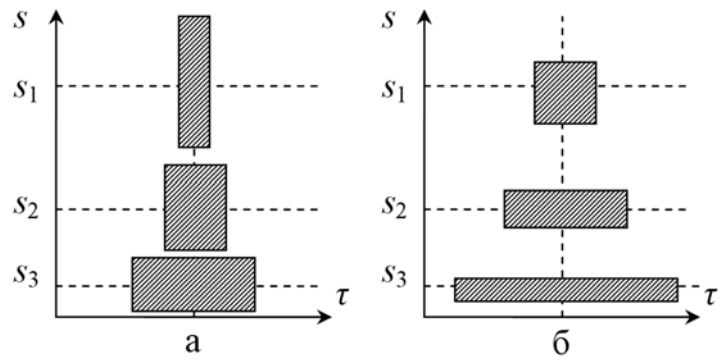


Рисунок 8 — Окно частотно-временной локализации вейвлетов с числом периодов а) $N=2$ и б) $N=8$

С целью оценки разрешающей способности искусственно создаваемых материнских вейвлетов был проведен эксперимент. Целью эксперимента являлось определение количества периодов в вейвлете, позволяющего идентифицировать частотные масштабы m_i всех нот, присутствующих в сигнале одновременно. В эксперименте использовано идеализированное трезвучие «до-мажор» первой октавы: $f(t) = \sin(2\pi 261,6t) + \sin(2\pi 329,6t) + \sin(2\pi 392t)$.

Для исследования тестового сигнала были созданы материнские вейвлеты $fw_i(t)$. Количество периодов гармонического сигнала в вейвлетах составляло 1, 2, 4, 8 и 16 периодов соответственно (рисунок 9).

Результаты НВП сигнала $f(t)$ с каждым из созданных вейвлетов приведены на рисунке (рисунок 10). Визуальный анализ графических интерпретаций $W(s, \tau)$ позволяет сделать вывод, что вейвлет с одним периодом ($N=1$) позволил точно определить локализацию сигнала во времени (от τ_0 до τ_1), но в частотной области идентифицировать три гармонических составляющих не представля-

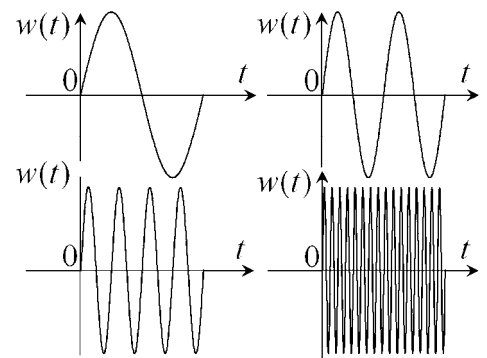


Рисунок 9 — Материнские вейвлеты с числом периодов гармонического сигнала 1, 2, 4 и 16

ется возможным (высокое разрешение по времени, но низкое по частоте).

Вейвлет с 16 периодами ($N=16$) определил локализацию всех трех гармонических составляющих, однако ошибка временной локализации ($\Delta\tau$) стала сравнима с длительностью сигнала (высокое разрешение по частоте, но низкое по времени).

Кроме того, установлено, что уже при 16 периодах вейвлета НВП из-за низких временных показателей не позволяет идентифицировать длительность коротких нот (шестнадцатых) при достаточно больших темпах исполнения музыкальных произведений, характерных для танцевальных ритмов (120 уд./мин и выше). Дальнейшее увеличение числа периодов в вейвлете привело бы к еще более низким показателям.

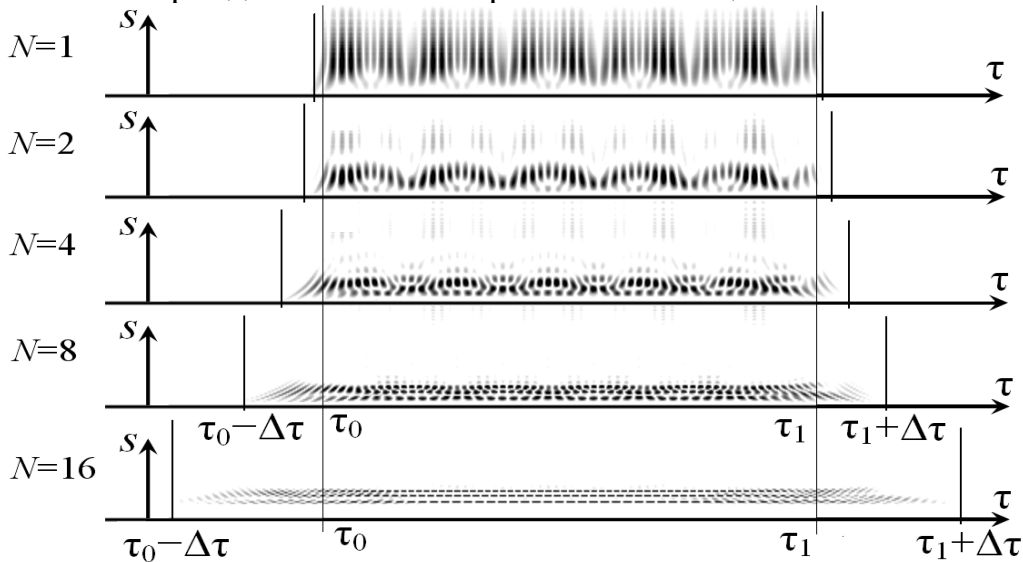


Рисунок 10 — Графические интерпретации результатов НВП для вейвлетов с различным числом периодов N гармонического сигнала

Для вейвлетов с числом периодов $N=16$ была произведена оценка способности идентифицировать временные составляющие ноты наименьшей длительности t_d . Было вычислено условие, задающее продолжительность T вейвлета во времени: $t_d \geq T \geq \frac{16}{v}$, где v — частота основного тона идентифицируемой ноты. Зависимость $v(t_d)$ (рисунок 11) определяет минимальную (граничную) частоту v основного тона ноты продолжительностью t_d , которая будет однозначно идентифицирована вейвлетом с 16 периодами. Так, для музыкальных композиций с темпом воспроизведения 120 уд./мин точная идентификация времени начала и окончания звучания ноты возможна только для нот выше «до» большой октавы с частотой основного тона 130,8 Гц и длительностью не меньшей 0,125 с.

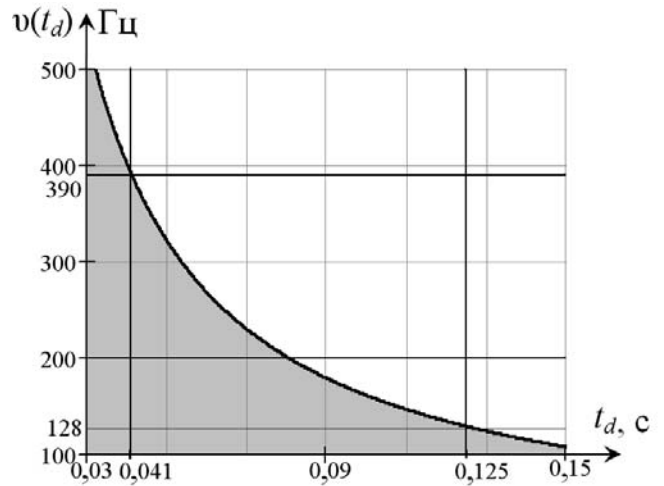


Рисунок 11 — Зависимость граничной частоты идентификации основного тона ноты от ее продолжительности

НВП основано на корреляции сигнала и вейвлета, что порождает в результирующем массиве $W(\tau, s)$ положительные, отрицательные и нулевые значения даже в тех частотно-временных участках, где сигнал присутствует явно. На рисунке это отображается чередованием светлых и темных пятен (рисунок 3, а и б). Для визуального анализа такое чередование не создает больших проблем, однако автоматизированная обработка результатов существенно осложняется.

В работе был предложен эвристический алгоритм, использующий для каждого масштабного коэффициента s_i наибольшие значения секущих $F(t)$ с длиной, равной периоду сигнала и вейвлета T :

$$F(t) = \frac{A(t_j + T/2) - A(t_j)}{T/2} t + A(t_j), \text{ где } t \in [t_j; t_j + T/2], T = \frac{1}{\omega_0 s_i}, \omega_0 \text{ — основная частота вейвлета } w(s, \tau, t) \text{ при } s=1; A'(t) = \max[A(t), F(t)].$$

Применение этого алгоритма позволило сформировать амплитудные огибающие $A'(t)$ (рисунок 12), сгладившие последовательности пиков в результатах НВП (рисунок 3, в).

Для уменьшения избыточности НВП было принято использовать в вычислениях дискретные значения параметра масштабирования s и параметра временного сдвига τ .

Большинство современных мультимедийных форматов импульсно-кодовой модуляции (например, *CDDA* и *Microsoft Waveform Data*) сохраняют сигнал с частотой дискретизации $\nu=44100$ Гц. Поэтому целесообразно в исследованиях использовать частоту дискретизации $\nu=44100 \text{ с}^{-1}$ для сигналов и вейвлетов. Приращение параметра временного сдвига τ принято равным минимально возможному значению — периоду квантования сигналов: $T_k = \Delta\tau = \frac{1}{\nu} = \frac{1}{44100} \text{ с}$.

В европейском хорошо темперированном строе частоты основных гармоник всех нот дискретизованы. Значения параметра масштабирования s должны строго соответствовать частотам основных тонов и могут быть описаны аналитически: $s_i = 27.5 \cdot 2^{\frac{i}{12}}$. Здесь 27,5 — частота ноты «ля» субконтроктавы (Гц) и частота основного тона базисного вейвлета, а $i = [1, 96]$ — порядковый номер ноты относительно ноты «ля» субконтроктавы (например, номер клавиши фортепиано). Соответственно количеству частотных уровней s_i , в выходном массиве $Wf(\tau, s)$ количество частотных уровней $S = 96$. Это значение обусловлено наибольшим значением нот, которое способен воспроизвести музыкальный инструмент.

НВП, осуществляя преобразование сигнала, не решает задач идентификации. Визуально выявляемые в сигнале образы нот необходимо идентифицировать при помощи автоматизированного компонента.

НВП, осуществляя преобразование сигнала, не решает задач идентификации. Визуально выявляемые в сигнале образы нот необходимо идентифицировать при помощи автоматизированного компонента.

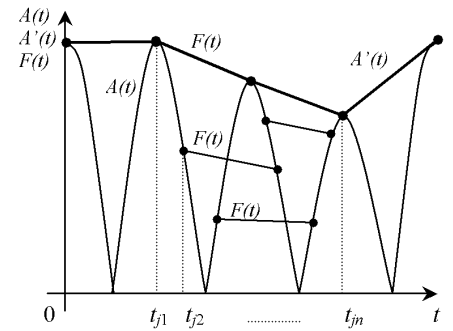


Рисунок 12 — Наложение секущих $F(t)$ на функцию $A(t)$ и формирование огибающей $A'(t)$

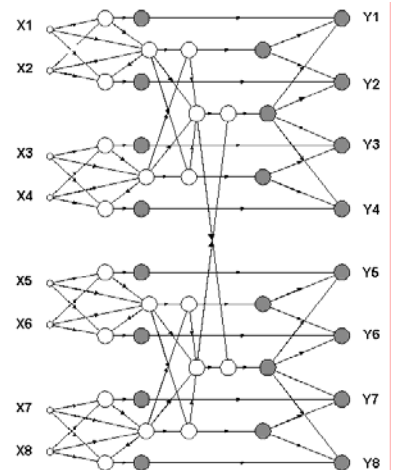


Рисунок 13 — Структура ИНС MaxNet

В качестве такого компонента была выбрана стандартная искусственная нейронная сеть (ИНС) прямой передачи *MaxNet* (рисунок 13). Сеть функционирует таким образом, что в установившемся режиме в выходном слое ненулевое значение присутствует на выходе только одного нейрона. Значение на выходе сети соответствует максимальному входному значению. Сеть состоит из множества нейросетевых компараторов и имеет число слоев $n = 2 \log_2 I + 1$, где I — количество входных сигналов. Для $I=128$ $n=15$.

Однако сеть не способна выявить больше одного локального максимума и не может использоваться при идентификации нескольких нот, звучащих одновременно.

В связи с этим было принято решение модифицировать ИНС. Модифицированная ИНС состоит из множества нейросетевых блоков сравнения (рисунок 14), включенных на каждом слое параллельно и осуществляющих выбор сигнала максимального значения из поданных на вход каждого блока. На выходе Y_i каждого нейросетевого блока сравнения формируется сигнал равный X_i , если $X_i \geq X_{3-i}$. На втором выходе Y_{3-i} формируется нулевое значение сигнала. Так, на выходе Y_1 установится значение сигнала $Y_1 = X_1$ тогда и только тогда, когда $X_1 \geq X_2$.

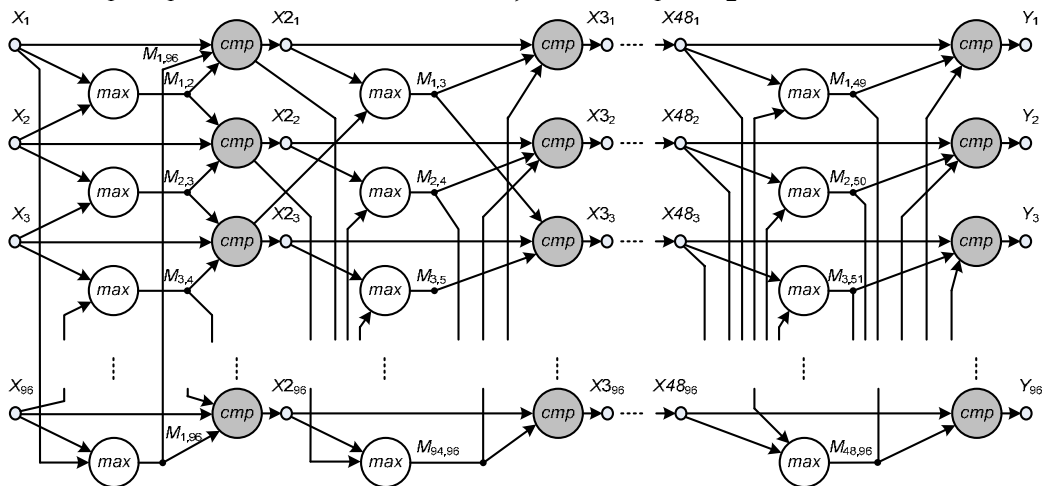


Рисунок 14 — Структура модифицированной ИНС *MaxNet*

Передаточная функция нейронов *cmp* имеет вид в общем случае:

$$\text{cmp}(X_i, M_{ij}, M_{ik})_i = \begin{cases} X_i, & \text{если } M_{ij} \leq X_i \text{ и } M_{ik} \leq X_i \\ 0, & \text{если } M_{ij} > X_i \text{ или } M_{ik} > X_i \end{cases}$$

где X_i — вход в блок сравнения, Y_i — соответствующий X_i выход блока сравнения, $M = \{M_{ij}, M_{ik}\}$ — набор выходных сигналов соседних нейронов выбора сигналов максимального значения ($M_{ij} = \max(X_i, X_j)$).

Число слоев для модифицированной сети поиска максимумов определяется по формуле: $n = \frac{I}{2}$, где I — число входов, n — число слоев ИНС.

Модифицированная сеть поиска максимумов обладает рядом особенностей.

- Сеть избыточна. Количество слоев n модифицированной сети больше количества слоев сети *Maxnet*.
- Количество нейронов N во всех слоях одинаково, $N = 2 \cdot I$, структура всех слоев идентична, что делает возможным реализовать сеть с помощью процедур на алгоритмических языках для ЭВМ с архитектурой Фон-Неймана, а не для нейронных компьютеров.

- Количество ненулевых сигналов Y' на выходе неполной сети зависит от количества слоев сети и определяется по формуле: $Y' \geq E\left(\frac{I}{n+1}\right)$, где $E(x)$ — функция, определяющая целую часть числа x .

Для обработки результатов НВП музыкальных сигналов многоголосных мелодий в задачах идентификации амплитудно-частотных составляющих объектов-нот ИНС использует число входов и выходов $I=96$. На вход сети подаются значения частотного среза τ , соответствующие состоянию музыкального сигнала в момент времени t :

$$X(s) = Wf'_\tau(s), \quad s=1, 2, \dots, 96.$$

В результате работы ИНС на выходе сети формируется массив значений $Y(s)$ с числом ненулевых значений сигналов Y' . Число слоев n является настроечным параметром сети, позволяющим изменять степень полифонической идентификации ИНС.

Изменение количества слоев n ИНС приводит к изменению минимально возможного интервала p между двумя ненулевыми выходными значениями сети: $p = n + 1$. Этот интервал характеризует количество полутонов между двумя нотами, звучащими одновременно. Так, например, для сети с $n=6$ число ненулевых значений сети $Y'=13$, а минимальный музыкальный гармонический интервал, определяемый сетью, — квинта с числом полутонов $p=7$.

Четвертая глава работы содержит описание программно-технического комплекса и его практического применения к задачам идентификации музыкальных объектов. Основу комплекса представляет НВП и ряд алгоритмических решений, предложенных в третьей главе работы.

ПТК состоит из трех модулей и позволяет выполнять прямое преобразование сигнала, обратное преобразование, а также осуществлять контроль над сигналами на каждой стадии обработки (рисунок 15).

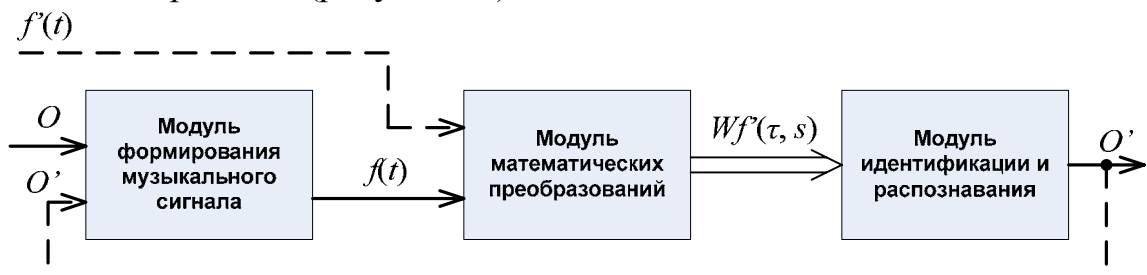


Рисунок 15 — Структура программно-технического комплекса

ПТК использует амплитудно-временное представление музыкального сигнала $f(t)$ и $f'(t)$ в виде файлов *Microsoft Waveform Data* и объектное представление сигнала O и O' в виде файлов стандарта *Standart Midi File*.

Контур обратной связи (по O') используется для воспроизведения, отображения фрагментов или корректировки полученных *MIDI*-объектов, а также для оценки качественных и количественных характеристик работы системы в целом.

Аппаратными компонентами комплекса являются персональный компьютер и *MIDI*-клавиатура *Yamaha PSR-620*. В качестве программных компонентов используются коммерческие программные продукты: Секвенсор *Cakewalk Sonar 4.0*, музыкальный синтезатор *Yamaha S-YXG50*, цифровой музыкальный процессор *Sony Sound*

Forge 7.0. и ряд разработанных программ, реализованных в средах *Delphi 2007 for Win32* и *MS Visual Studio*.

Реализованные в работе программные компоненты выполняют функции НВП сигнала $f(t)$ с произвольным вейвлетом $w(t)$, сглаживания выходных результатов НВП $Wf(\tau, s)$ методом секущих, формирования графических изображений для $Wf(\tau, s)$, идентификации частотных и временных параметров объектов и формирования файла O в формате *SMF*.

Апробация программно-технического комплекса проходила как на модельных сигналах, сформированных в объектном виде, так и на реальных фрагментах музыкальных произведений, представленных в формате решетчатых функций.

В качестве модельных были использованы одно- и многоголосные мелодии популярных композиций в *MIDI*-формате. При помощи музыкального синтезатора с каждым *MIDI*-сигналом O было произведено прямое преобразование $f(t) = T(N_0, O)$, сигналы в амплитудно-временном формате $f(t)$ были представлены в виде файлов *Microsoft Waveform Data*. На основе базисной ноты музыкального инструмента $n_0^k(t)$, используемого в мелодиях, формировался одноименный вейвлет $w^k(t)$.

Для идентификации нот в сигнале $f(t)$ осуществлялось обратное преобразование $O' = T^{-1}(f(t), N_0)$: сигнал $f(t)$ подвергался НВП с использованием вейвлета $w^k(t)$, обработке алгоритмом наложения секущих и модифицированной ИНС. Полученный сигнал O' по цепи обратной связи (рисунок 15) передавался в секвенсор *Cakewalk Sonar 4.0*, где производились сравнение с оригинальным сигналом O (рисунок 16) и оценка качества работы комплекса в целом.

Рисунок 16 — Нотная запись мелодии «*Money-money-money*» группы *ABBA*: а) сигнала O ; б) сигнала O'

Результаты экспериментов позволили выявить ряд особенностей системы.

1. Частотный диапазон распознаваемых объектов составляет 96 полутонов (8 октав) от ноты «ля» субконтроктавы до ноты «ля» пятой октавы.
2. При идентификации нот одноголосных мелодий для музыкальных инструментов, звук которых формируется автоколебаниями, относительное количество нот, совпавших в распознаваемом и оригинальном произведениях $\Delta K_c \in [80, 90]\%$, среднее относительное отклонение времени начала звучания каждой ноты $\Delta \theta \in [10, 20]\%$; для музыкальных инструментов, звук которых формируется свободными колебаниями, $\Delta K_c \in [95, 100]\%$, $\Delta \theta \in [1, 2]\%$.
3. Степень полифонической идентификации системы составляет 3 — 4 голоса. При полифонии в четыре голоса, $\Delta K_c \in [80, 90]\%$, $\Delta \theta \in [1, 5]\%$ для музыкальных инструментов со свободными колебаниями и $\Delta K_c \in [50, 70]\%$, $\Delta \theta \in [20, 40]\%$ для инструментов с автоколебаниями. При идентификации мелодий с количеством одно-

временно звучащих нот больше четырех, $\Delta K_c < 80\%$ для инструментов со свободными колебаниями.

4. Количество идентифицируемых музыкальных инструментов составляет 16. Это количество соответствует количеству сформированных и опробованных в работе системы базисных вейвлетов (фортепиано, колокольчик, электропианино, скрипка, и др.). Количество вейвлетов музыкальных инструментов может дополняться новыми.
5. При идентификации звучания отдельного инструмента на фоне других для музыкальных инструментов, звук которых формируется автоколебаниями, $\Delta K_c=35\%$, для музыкальных инструментов, звук которых формируется свободными колебаниями, $\Delta K_c=95\%$.

Заключение

В работе предложен и обоснован метод формирования частных вейвлет-функций (базисных вейвлетов и семейств на их основе). Показана возможность формирования базиса с заранее заданными частотно-временными свойствами и его применения в задачах анализа нестационарных сигналов. Предложена система изменения масштабного коэффициента семейства вейвлетов в НВП, позволяющая классифицировать музыкальные объекты в соответствии с 96 полутонами европейского равномерно темперированного строя музыки. Разработан алгоритм увеличения информативности результатов НВП, позволяющий применять их для дальнейшей автоматизированной обработки.

На основе искусственной нейронной сети *MaxNet* предложена модифицированная ИНС поиска максимумов с прямыми связями. Количество слоев сети предложено варьировать в качестве настроечного параметра степени полифонической идентификации системы в целом.

Разработан программно-технический комплекс, интегрировавший исследования в области НВП и ИНС. Комплекс позволяет решать задачи идентификации партии определенного музыкального инструмента с идентификацией в ней объектов: их высотного положения и временных характеристик. Экспериментально доказана возможность идентификации партии полифонического исполнения одним инструментом (когда в каждый момент времени может звучать более одной ноты), а также партии одного инструмента на фоне звучания других.

Сформированы материнские вейвлеты шестнадцати музыкальных инструментов и сигналов несинусоидальной формы. Отработана методика формирования материнских вейвлетов, позволяющая получать определенный вейвлет для требуемого музыкального инструмента и сигнала с заданными характеристиками.

Разработаны пять программных приложений, выполняющих функции основных этапов идентификации музыкальных объектов. В процессе отладки программного обеспечения была расширена функциональность разработанных программных компонентов, получившая применение в задачах спектрального анализа в отраслях промышленности, не связанных с обработкой музыкальных сигналов.

Перечень публикаций по теме диссертации

1. **Фадеев, А. С.** Метод преобразования форматов музыкальной информации [Текст] / А. С. Фадеев, Е. А. Кочегурова // Цифровая обработка сигналов. — 2007. — № 3. — С. 46—51.

2. **Фадеев, А. С.** Формирование вейвлет-функций в задаче идентификации музыкальных сигналов [Текст] / А. С. Фадеев // Известия Томского политехнического университета. — 2007. — Т. 311. — № 5. — С. 81—86.
3. **Фадеев, А. С.** Подготовка результатов непрерывного вейвлет-преобразования к автоматизированной обработке [Текст] / А. С. Фадеев, Е. А. Кочегурова // Известия Томского политехнического университета. — 2006. — Т. 309. — № 7. — С. 32—35.
4. **Фадеев, А. С.** К вопросу о применении искусственных нейронных сетей и непрерывных вейвлет-преобразований в задачах распознавания музыкальных образов [Текст] / А. С. Фадеев // Информационные технологии моделирования и управления. — 2007. — Вып. 7 (41). — С. 790—794.
5. **Фадеев, А. С.** Применение вейвлет-анализа в задаче исследования естественных радишумов звукового диапазона [Текст] / А. С. Фадеев, И. А. Заикин // Молодежь и современные информационные технологии: Сборник трудов VI Всеросс. научно-практи. конф. студентов, аспирантов и молодых ученых, 26—28 фев. 2008 г. — Томск: СПб Графикс, 2008. — С. 96—97.
6. **Фадеев, А. С.** Формирование базисного вейвлета в задачах анализа музыкальной информации [Текст] / А. С. Фадеев, Е. А. Кочегурова, И. А. Заикин // Молодежь и современные информационные технологии: Сборник трудов V Всеросс. научно-практи. конф. студентов, аспирантов и молодых ученых, 27 фев. — 1 мар. 2007 г. — Томск: Изд-во ТПУ, 2007. — С. 179—181.
7. **Фадеев, А. С.** Вейвлет анализ в задаче идентификации музыкальной информации [Текст] / А. С. Фадеев, Е. А. Кочегурова // Молодежь и современные информационные технологии: Сб. трудов IV Всеросс. научно-практи. конф. студентов, аспирантов и молодых ученых 28 фев. — 2 мар. 2006 г. — Томск: Изд-во ТПУ, 2006. — С. 149—151.
8. **Фадеев, А. С.** Выбор вейвлет-функций для анализа музыкальной информации [Текст] / А. С. Фадеев, Е. А. Кочегурова // Современные техника и технологии: Труды XII Междунар. научно-практи. конф. студентов и молодых ученых — г. Томск, 26—30 мар. 2007 г. — Томск: Изд-во ТПУ, 2006 — Т. 2. — С. 194—196.
9. **Фадеев, А. С.** К вопросу о преобразовании музыкальных форматов [Текст] / А. С. Фадеев, Е. А. Кочегурова // Современные проблемы информатизации в моделировании и программировании: Труды XI Междунар. открытой научн. конф. — г. Воронеж, нояб. 2005 г. — янв. 2006 г. — Вып. 11. — Воронеж: Научная книга, 2006. — С. 255—257.