

Министерство Российской Федерации по связи и информатизации
Московский технический университет связи и информатики

На правах рукописи

Кудинов Александр Александрович

ИСПОЛЬЗОВАНИЕ РАСПОЗНАВАНИЯ ОБРАЗОВ ДЛЯ
ОБРАБОТКИ И ВОССТАНОВЛЕНИЯ МУЗЫКАЛЬНЫХ
СИГНАЛОВ

Специальность 05.12.04. : *Радиотехника, в том числе системы и
устройства радионавигации, радиолокации и телевидения*

Автореферат
диссертации на соискание учёной степени
кандидата технических наук

Москва 2003

Работа выполнена на кафедре «Радиовещание и электроакустика» Московского технического университета связи и информатики (МТУСИ)

Научный руководитель - д. т. н., профессор С. Л. Мишенков

Официальные оппоненты – д. т. н., профессор Б. П. Хромой

к. т. н., доцент Л. В. Шитов

Ведущая организация - Федеральное государственное унитарное предприятие
«Научно-исследовательский институт радио» (ФГУП
НИИР)

Защита состоится «____» _____ 2003 года в _____ часов на заседании диссертационного совета К.219.001.02 в Московском техническом университете связи и информатики по адресу: 111024, г. Москва, ул. Авиамоторная, д. 8а, ауд. А-455

С диссертацией можно ознакомиться в библиотеке МТУСИ.

Автореферат разослан «____» _____ 2003 года

Ученый секретарь диссертационного совета _____

кандидат технических наук, профессор _____

Общая характеристика работы

Актуальность тематики

Современное состояние проблемы реставрации музыкальных фонограмм характеризуется определённым противоречием между возможностями, предоставляемыми техникой цифровой обработки сигналов, и результатами, получаемыми на практике. Современные аппаратные и программные средства реставрации сигналов позволяют повысить эффективность подавления импульсных помех и широкополосных шумов. В то же время, нередко субъективные оценки квалифицированных экспертов свидетельствуют о потерях в эстетических свойствах фонограмм. В связи с этим возникает проблема принципиального характера о целесообразности реставрации архивных звукозаписей при современном уровне развития аудиотехнологий. В частности, некоторые специалисты предлагают ограничиться простой «консервацией» оригинальных фонограмм. В рамках подобной технологии звукозапись преобразуется в цифровую форму с использованием высококачественных аналого-цифрового преобразователей и переносится на долговечный носитель типа компакт-диска. Существующие способы восстановления музыкальных сигналов основываются на субъективных методах обнаружения дефектов и *уменьшения их влияния*. Актуальность тематики определяется необходимостью сохранения музыкальных звукозаписей, представляющих историческую и культурную ценность. Для этого необходима разработка объективных методов восстановления, нацеленных на *устранение последствий* нежелательных искажений.

Цель и задачи работы

Целью работы является повышение качества (по критерию субъективной оценки) восстановления музыкальных сигналов.

Для достижения этой цели в работе ставятся и решаются следующие **задачи**:

1. Разработка управляемого системой распознавания алгоритма цифровой обработки сигнала на уровне структуры отдельного звукового объекта.
2. Разработка системы распознавания образов для управления автоматической и автоматизированной обработкой сигнала в процессе реставрации.

В задачи восстановления музыкальных сигналов входят:

- а) устранение компонентов спектра, не составляющих полезный сигнал (продукты нелинейных искажений, компоненты шумов и помех);
- б) восстановление компонентов сигнала, утраченных из-за изменений, вносимых трактом формирования-запись и/или носителем записи (линейные искажения), из-за длительного и/или неправильного хранения носителя записи.

Для решения этих задач необходимо в процессе *наблюдения реализации* сигнала *классифицировать* звуковые объекты, что по определению является задачей распознавания образов.

Структура музыкального звукового объекта — соотношение амплитуд, начальных фаз и частот основного тона и обертонов как функций времени.

Распознавание музыкального сигнала — принятие решения о частоте основного тона, длительности, громкости и тембре звуковых объектов.

Сведения о частоте основного тона и длительности звукового объекта позволяют:

- а) классифицировать компоненты спектра не только как полезные, но и как относящиеся к конкретному звуковому объекту;
- б) на основе априорных знаний или эвристических моделей синтезировать утраченные компоненты (обертоны) звуковых объектов и, при необходимости, управлять соотношением параметров (амплитуд, фаз) сохранившихся компонентов.

Методы исследования

При исследовании свойств музыкальных сигналов использованы методы математической статистики и методы спектрального анализа на основе дискретного преобразования Фурье (ДПФ).

При разработке алгоритма распознавания музыкального сигнала использованы методы спектрального и корреляционного анализа на основе ДПФ, численные методы анализа дискретных последовательностей, положения общей теории распознавания образов, сведения о психофизиологии восприятия звука, сведения о физике звукоизвлечения в музыкальных инструментах, положения элементарной теории музыки.

Научная новизна работы

1. Для управления **обработкой** сигнала в процессе реставрации используется информация, полученная при предварительном **распознавании** сигнала.

2. Алгоритмы сегментации и распознавания музыкального сигнала построены с учётом **априорной информации** о распознаваемых реализациях образов. Априорная информация о музыкальном сигнале состоит в заранее известном соотношении ОТ звуковых объектов, подлежащих распознаванию. В европейской музыке частоты основных тонов звуков составляют множество f_i^{i0} , зная один из членов которого (например, эталонный тон $f_{\check{y}}$, по которому настраиваются инструменты), можно с незначительными отклонениями восстановить все остальные. Например, для равномерного строя: $f_i^{OT} = f_{\check{y}} \cdot \sqrt[12]{2^i}$, где $f_{\check{y}}$ — эталонная частота, для натурального строя $f_i^{OT} = f_{\check{y}} \cdot \frac{m}{n}$, где $m, n \in \mathbb{Z}, m, n \geq 1$.

Практическая значимость

1. Автоматическое распознавание музыкальных сигналов с расширением множества распознаваемых сигналов найдёт своё применение в современной звукозаписи. Автоматическое распознавание музыкального сигнала позволяет звукозапись музыкального произведения поставить в соответствие с его нотной записью. Такая возможность используется:

- при создании электронных архивов партитур музыкальных произведений;
- аранжировщиками и композиторами для сокращения рутинной работы по написанию партитур.

Соответственно, практическую значимость представляет разработанный алгоритм распознавания музыкальных сигналов.

2. Развитие систем распознавания слуховых и зрительных образов и их объединение в единые комплексы приведёт к тому, что наиболее эффективная обработка и видео-, и звуковых сигналов будет проводиться именно на основе распознавания. Результаты работы разработанной системы восстановления несложных звучаний — подтверждение целесообразности и реализуемости распознающих систем обработки, один из первых шагов в данном направлении.

3. Использование для автоматической сегментации априорной информации в виде распределения значений одного из признаков потенциально эффективнее обучения и самообучения без предоставления этой информации.

4. Предложенная методика определения параметров основного тона и обертонов звуковых объектов позволяет повысить качество аддитивного синтеза музыки и речи по критерию натуральности звучания.

5. Разработанное для анализа квазипериодических сигналов программное обеспечение (ПО) может быть использовано для объективной оценки качества звучания музыкальных инструментов.

Апробация результатов работы

Основные результаты и предложения, полученные в ходе работы над диссертацией, докладывались и опубликованы в виде тезисов докладов 55-ой студенческой НТК МТУСИ, профессорско-преподавательских конференций МТУСИ 2001-2003 гг., 10-ой межрегиональной конференции МНТОРЭС им А. С. Попова, 4-ой международной НТК «Перспективные технологии в средствах передачи информации» (Владимир). Также результаты опубликованы в статьях, депонированных ЦНТИ «Информсвязь», докладывались на заседаниях кафедры РВ и ЭА, НТС МТУСИ.

Всего по теме работы опубликовано 8 тезисов докладов на конференциях, депонировано 5 статей.

Положения, выносимые на защиту

1. Целью реставрации фонограмм должно быть повышение субъективных оценок качества сигнала путём устранения нежелательных последствий прохождения сигнала через тракты формирования, записи и воспроизведения, а также последствий длительного и/или неправильного хранения фонограмм.

2. Для повышения эффективности реставрации фонограмм необходимо применение распознавания образов.

3. Использование для автоматической сегментации априорной информации в виде распределения значений одного из признаков потенциально эффективнее методов, применяемых в обучаемых и самообучающихся системах распознавания.

4. В задачах автоматического распознавания музыкальных сигналов целесообразно использовать предлагаемую в работе классификацию музыкальных инструментов, основанную на типе возбуждаемых колебаний (автоколебания или свободные затухающие колебания).

5. Натуральность звучания сигналов, получаемых при воссоздании звуков музыкальных инструментов и вокализованных звуков речи аддитивным синтезом, достигается при наличии информации о каждом полном цикле колебаний.

Краткое содержание работы

Введение

Описано и проанализировано современное состояние техники реставрации музыкальных фонограмм. Обсуждены универсальность понятия «реставрация» и необходимость общего для разных отраслей науки и техники понимания её целей и задач. Сформулированы цели и задачи реставрации и на их основе выявлены проблемы развития устройств и систем, используемых для реставрации и восстановления сигналов звукового вещания.

Обсуждена *целесообразность* выполнения реставрации фонограмм. Построена общая схема процесса реставрации на основе распознавания (рис.1).

Цели и задачи

Обсуждается актуальность тематики, некоторые аспекты субъективной и объективной оценок качества сигналов звукового вещания (ЗВ). Формулируются цели и задачи диссертационной работы (см. выше).

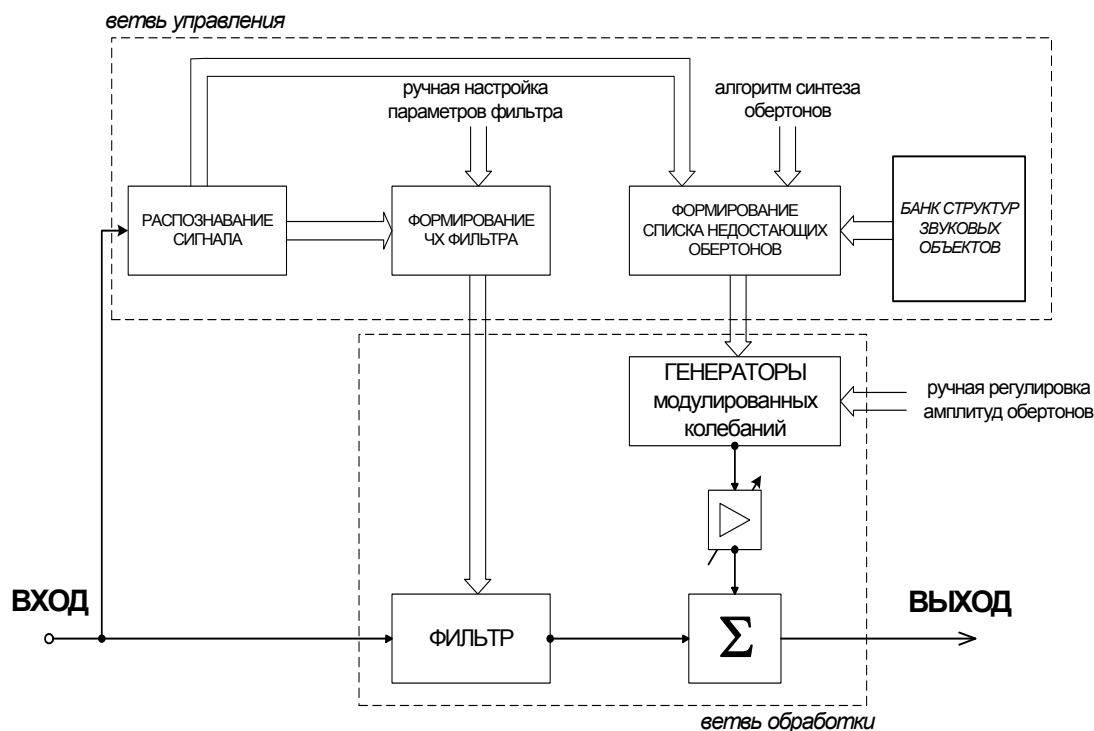


Рис. 1 Схема управления и последовательность процессов обработки при восстановлении сигнала на основе распознавания

Принципы построения систем распознавания образов

■ В параграфе «Общая теория распознавания образов» систематизированы применительно к сигналам звукового вещания основные понятия теории распознавания образов, приведена классификация систем распознавания, обсуждаются задачи распознавания образов, а также назначение и цель создания систем распознавания образов. Кроме того, обсуждается в общем виде последовательность распознавания образов.

Образ — некоторое структурированное приближённое описание (эскиз) изучаемого объекта, явления или процесса.

Образ представляется некоторым набором *признаков*. Выбор признаков — *эвристическая операция*. Признаки могут подразделяться на:

- детерминированные;
- вероятностные;
- логические;
- структурные.

Классы — объединения объектов (явлений), отличающиеся общими свойствами, интересующими человека.

Для классификации систем распознавания (СР) используются следующие принципы:

1. Однородность информации для описания распознаваемых объектов или явлений (простые и сложные СР).
2. Способ получения апостериорной информации (одноуровневые и многоуровневые СР).
3. Количество первоначальной априорной информации (системы без обучения, обучающиеся (ОСР) и самообучающиеся системы (ССР)).
4. Характер информации о признаках распознавания (детерминированные, вероятностные, логические, комбинированные СР).

Назначение СР — получение информации, необходимой для принятия решения о принадлежности неизвестных объектов (явлений) к тому или иному классу.

Цель создания СР — обеспечение высокой эффективности принимаемых решений в управлении.

■ В параграфе «Распознавание образов с информационной точки зрения» обсуждаются два основных вопроса: информативность музыкальных произведений (т.е. содержание и характер информации, переносимой музыкальным сигналом) и преобразования информации в процессе распознавания образов.

Различают два типа информации:

а) семантическая информация, имеющая структуру, допускающая точное представление, записываемая символами;

б) эстетическая информация, «непереводимая», относящаяся к набору знаний, общих для приёмника и передатчика.

Семантическая информация в музыке частично описывается с помощью нотной записи. При распознавании происходит выделение семантической информации, переносимой звуковым сигналом.

■ Параграф «Статистический подход к распознаванию образов» посвящён постановке задачи и общим принципам выполнения распознавания образов как статистической задачи. В большинстве случаев имеют дело с вероятностными признаками, при этом области значения признаков перекрываются. В такой ситуации принимается решение о вероятности принадлежности к тому или иному классу. В случае, если области значений признаков не перекрываются вероятность ошибки распознавания становится сколь угодно малой, и сама процедура распознавания является тривиальной: по значениям признаков принимается решение о принадлежности к классу. Частота основного тона — величина случайная. Однако области значений основных тонов нот для равномерного и натурального строев не перекрываются.

■ В параграфах «Распознавание звуковых образов» и «Системы распознавания музыкальных сигналов» обсуждаются проблемы распознавания звуковых сигналов, определяется набор важнейших признаков, характеризующих звуковые объекты, кратко описывается история развития систем распознавания музыкальных сигналов, обсуждаются принципы их построения.

Теория автоматического распознавания звуковых сигналов разработана недостаточно, взамен ее иногда применяются методы распознавания зрительных образов, но такой перенос не всегда обоснован. Проблема автоматического распознавания звуковых образов оказалась замкнутой на распознавание звуков речи, другие приложения развивались мало. Современные успехи теории распознавания музыкальных сигналов не убедительны — на сегодняшний день не создано ни одной коммерчески используемой системы распознавания многоголосных мелодий.

Очень важен вопрос о том, содержит ли звуковое поле все особенности, необходимые (с учетом современного уровня техники) для автоматического распознавания речи или музыки? Возможно, звуковое поле несет некоторые коды, действующие в процессе слухового восприятия по неизвестным науке процедурам в нейронных сетях, ограничивая множество автоматически распознаваемых сигналов.

Формулируется в виде предположения основание физической теории распознавания образов: *основные признаки, необходимые для автоматического распознавания образа, содержатся в самом сигнале.*

Основные признаки звуковых образов:

1. Энергетические параметры:

— размах сигнала;

— мощность сигнала;

— относительная средняя мощность;

2. Спектральные характеристики:

- характер кратковременного спектра: непрерывный или дискретный (соответственно, шумовой или тональный сигнал);
- положение на частотной оси формант кратковременного амплитудного спектра, соотношение амплитуд соответствующих спектральных составляющих;
- соотношение амплитуд и фаз гармоник;
- частота основного тона;
- кратковременная автокорреляционная функция сигнала.

Исторически первыми появились системы распознавания одноголосных мелодий. Задача их распознавания сведена к слежению за мелодией основного тона (fundamental frequency tracking) и моментами её смены. Первая система распознавания двухголосных мелодий создана в 1975 году Джеймсом Андерсоном Мурером, в настоящее время возглавляющим подразделение компании Sonic Solutions, занимающееся системой реставрации «NoNoise». В разделе приведён аналитический обзор современных систем распознавания музыкальных сигналов, выявлены достоинства и недостатки этих систем, обоснован выбор предлагаемого подхода.

Развитие систем распознавания музыкальных сигналов (CPMC) происходит в рамках более общего направления — компьютеризированного анализа звуковых картин (КАЗК, CASA – computational auditory scene analysis). Бурное развитие КАЗК связано с развитием многоканальных систем звукозаписи. Основными тенденциями в развитии CPMC является:

1. *моделирование восприятия человеком громкости* для сегментации сигнала и выделения ритма;

2. *использование экспертных систем* («блэкборд»-архитектура, методология «классной доски» (blackboard — классная доска)) при анализе амплитудных спектров созвучий на предмет определения основных тонов отдельных звуков — учёт законов гармонии, особенностей музыкального строя.

Сложность распознавания многоголосных музыкальных сигналов определяется перекрытием звуковых объектов во временной и частотной областях. Традиционные методы определения высоты (например, по кратковременному амплитудному спектру, кратковременной АКФ) неприменимы.

■ Теоретическим подходам к разделению созвучий тональных звуков посвящён следующий параграф — «Сложности распознавания многоголосных мелодий». Формулируются основные свойства созвучий тональных звуков — соотношение частот основных тонов и гармоник. На основе этих свойств формулируются требования к алгоритмам разделения созвучий тональных звуков. Приводятся правила, на которых строятся экспертные системы для разделения созвучий. Все сведения — результат анализа публикаций зарубежных авторов.

■ Выводы:

1. Система распознавания всегда является элементом системы управления, цель создания системы распознавания — повышение эффективности решений, принимаемых при управлении.

2. Частота основного тона музыкального звуков — вообще говоря, случайная величина. Области значений частот ОТ звуков, отличающихся на полутон, не перекрываются, поэтому вероятность принятия ошибочного решения о высоте тона ноты по частоте основного тона звукового объекта равна 0. Ошибки возникают при определении значения данного признака, а не при классификации.

3. Основные направления развития систем распознавания музыкальных сигналов — моделирование восприятия звука человеком, объединение при принятии решения источников разнородных сведений о распознаваемом сигнале с помощью методологии «классной доски»

Современная техника реставрации и восстановления музыкальных сигналов

Рассматриваются возможности современных программных и аппаратных средств, используемых для реставрации музыкальных сигналов. Выявлены достоинства и недостатки этих систем.

■ Рассматриваются возможности и алгоритм работы двух компьютеризированных систем реставрации: «NoNoise» компании Sonic Solutions и «Канонъ» российской компании Digiton. Операции, выполняемые с помощью этих систем:

- подавление импульсных помех (щелчков);
- подавление широкополосного шума с предварительным обучением по фрагменту шума;
- устранение последствий ограничения мгновенных значений сигнала.
- подавление импульсных помех;
- маскировка шума с помощью многополосной АРУ («Канонъ»).

Эффективность подавления импульсных помех достаточна для реставрации грамзаписей. Эффективность же подавления широкополосных шумов в значительной степени зависит от статистических свойств самого шума, а также опытности реставратора. Нередко причиной снижения качества реставрации по субъективным критериям является именно «грубое» шумоподавление.

■ В параграфе «Психоакустические процессоры» рассматриваются структурные схемы, и обсуждается принцип работы отдельного класса комбинированных устройств обработки.

Психоакустические процессоры широко используются при реставрации фонограмм и восстановлении сигналов с пониженным техническим качеством. Используются они, в частности, для устранения субъективного ощущения глухости звучания или ограниченности полосы частот сигнала.

■ Заключительный параграф раздела — «Общая характеристика современной техники реставрации». Для подведения итогов раздела построена обобщённая схема процесса реставрации, отражающая современное состояние данной области техники. Схема приведена на рисунке 2, показательно её сравнение со схемой на рисунке 1.

Принципиальные отличия этих двух схем:

1. В схеме на рис. 2. формирование ЧХ фильтра производится на основе сравнения амплитудных спектров, полученных с помощью БПФ, обрабатываемой и обучающей выборок, а не на основе распознавания звуковых объектов.

2. Синтез гармоник целочисленным умножением частоты сигнала не идентичен синтезу модулированных колебаний с управляемыми параметрами.

3. В схеме на рис. 2 чётко разделяются процессы шумоподавления и синтеза гармоник, в схеме на рис.1 эти процессы объединены управляющим воздействием процесса распознавания.

■ Выводы:

1. Работу современных алгоритмов шумоподавления можно назвать *шумоподавлением на основе обнаружения сигнала* (в отличие от предлагаемого *шумоподавления на основе распознавания* звуковых объектов): спектральные компоненты классифицируются как «шумовые» и «нешумовые».

2. В современных подходах к реставрации целью синтеза новых спектральных составляющих является создание у слушателя ощущения наличия высокочастотных составляющих, широкой полосы сигнала, а не восстановления структуры звукового объекта, т.е. спектральной структуры сигнала.

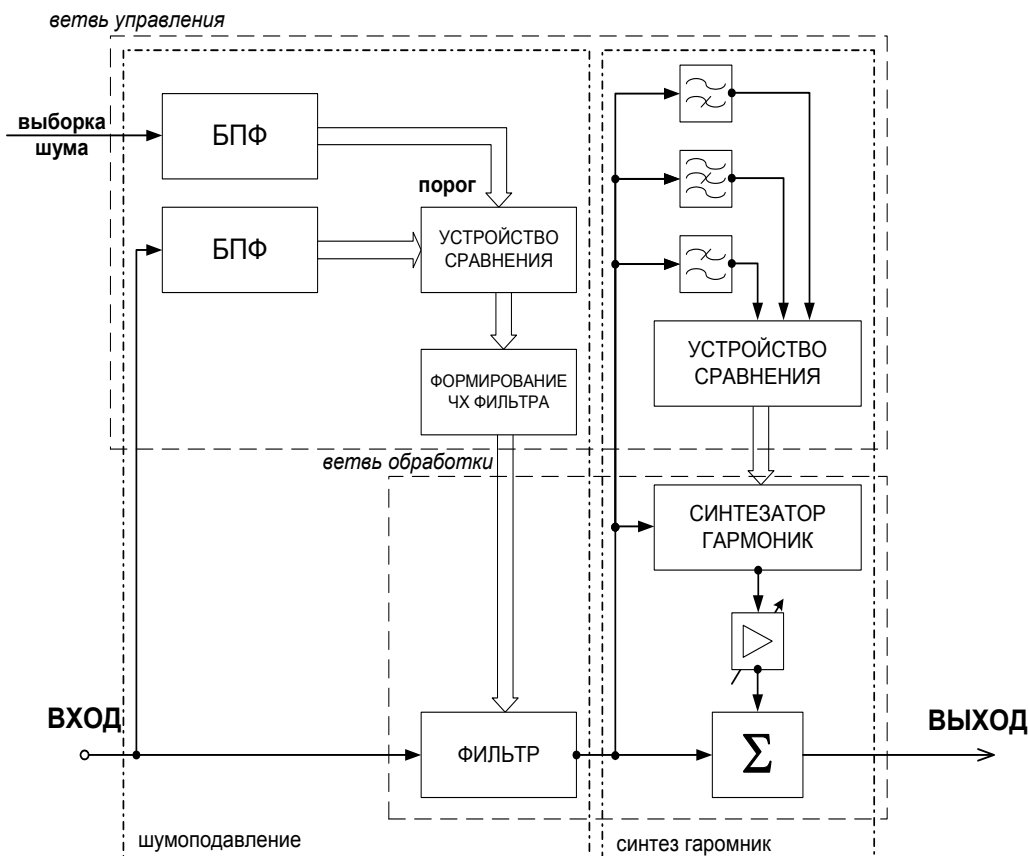


Рис. 2 Обобщённая схема процесса реставрации современными средствами

Свойства звуковых объектов

Этот раздел посвящён изучению свойств сигналов, задачу распознавания которых предстоит решать. Исследуемым материалом стали записи реализаций отдельных нот, исполненных на механико-акустических музыкальных инструментах.

■ Параграф «Исследование структуры звуковых объектов» посвящён постановке задачи и целей исследования, описанию инструментов исследования. Основным инструментом исследования стало разработанное программное обеспечение (в среде Delphi 5 для IBM PC с ОС Windows), выполняющее следующие операции: чтение файлов формата RIFF WAVE, определение периода фрагмента сигнала, разложение в комплексный ряд Фурье, формирование траекторий амплитуд и фаз гармоник, их графическое отображение и сохранение в файл. Кроме того, сохраняется текстовый отчёт о параметрах исследуемого сигнала и результатах анализа.

Траектория амплитуды (начальной фазы, частоты) гармонической составляющей — зависимость амплитуды (начальной фазы, частоты) от времени

Последовательность анализа звукового объекта:

1. определение ориентировочной продолжительности полного цикла колебаний сигнала;
2. разбиение всего звукового объекта на непересекающиеся фрагменты, соответствующие полным циклам колебаний;
3. для каждого полного цикла колебаний выполняется разложение в гармонический ряд Фурье;
4. средние значения амплитуд и фаз колебаний основной частоты и обертонов на данном цикле колебаний принимаются равными значениям амплитуд и фаз коэффициентов комплексного ряда Фурье.

Всего проанализировано около 150 фрагментов звучания различных музыкальных инструментов различной длительности и с различным периодом

колебаний. Исследовались, в основном, звуки инструментов симфонического оркестра (струнных и духовых).

■ Обобщению результатов анализа посвящён параграф «Результаты исследования». По итогам обобщения предложена *альтернативная классификация музыкальных инструментов*. Для классификации предлагается использовать следующие факторы (в порядке убывания уровня классификации):

- характер колебаний (свободные или автоколебания);
- способ возбуждения (щипок, удар молоточка, трение, вдувание воздуха и т.д.);
- колеблющееся тело (струна, пластинка, тарелка, цилиндр, столб воздуха и т.д.);
- конструкция колеблющегося тела (закрытая или открытая труба, струна с обмоткой и без обмотки и т.д.);
- материал колеблющегося тела (кожа, металл, дерево, стекло, жила и т.д.).

Каждый из используемых для классификации факторов имеет определяющее значение именно для *свойств* возбуждаемых *колебаний*, но при этом характеризует колебательную систему (музыкальный инструмент).

Также в параграфе в виде таблиц и графиков приводятся некоторые статистические данные о параметрах колебаний составляющих звуковых объектов:

- соотношения амплитуд обертонов;
- отклонения частот колебаний обертонов от значений, кратных частоте основного тона (средние значения и дисперсии);
- матрицы корреляции траекторий амплитуд обертонов;
- отклонения длительности полного цикла колебаний от среднего значения (максимальные значения, дисперсии).

В заключении параграфа предложена *модель музыкального звукового объекта*. Во временной области музыкальный звуковой объект с достаточной точностью описывается суммой модулированных гармонических колебаний:

$$s(t) \approx \sum_{n=0}^N A_n(t) \cdot \sin[2\pi \cdot f_n(t) + \varphi_n(t)], \text{ где} \quad (1)$$

$A_n(t)$ — мгновенное значение амплитуды n -ого обертона (основного тона);

$f_n(t) = F_0(t) \cdot n + \Delta f_n$ — мгновенное значение частоты n -ого обертона,

$F_0(t)$ — мгновенное значение частоты основного тона,

Δf_n — постоянная величина, отражающая отклонение частоты обертона от значения, кратного частоте ОТ;

$\varphi_n(t)$ — начальная фаза колебаний n -ого обертона (основного тона).

Функции $A_n(t)$, $f_n(t)$, $\varphi_n(t)$ описывают *структуру звукового объекта*.

В частотной области музыкальный звуковой объект может быть представлен следующим образом:

$$S(j\omega) = S_{\text{возб}}(j\omega) \cdot H_{\text{КС}}(j\omega) \cdot H_{\text{рез}}(j\omega), \text{ где} \quad (2)$$

$S_{\text{возб}}(j\omega)$ — комплексный спектр функции возбуждения;

$H_{\text{КС}}(j\omega)$ — частотная характеристика колебательной системы музыкального инструмента;

$H_{\text{рез}}(j\omega)$ — частотная характеристика резонатора музыкального инструмента.

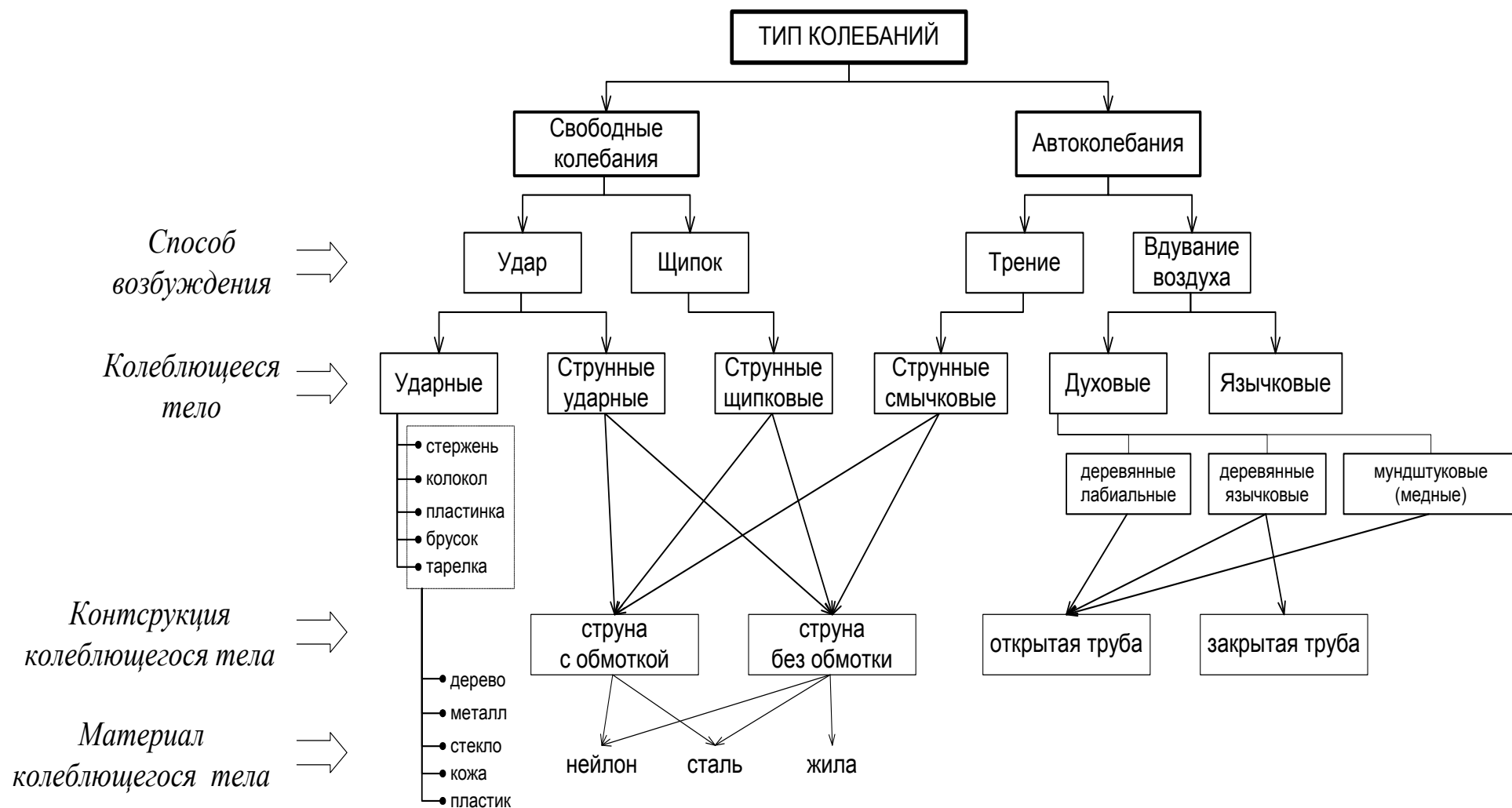


Рис. 3 Классификация музыкальных инструментов по типу колебаний

■ Выводы

1. В задачах распознавания музыкальных сигналов и идентификации музыкальных инструментов целесообразно использовать классификацию музыкальных инструментов, основанную на типе возбуждаемых колебаний (*автоколебания или свободные затухающие колебания*).

2. Стабильность автоколебаний в музыкальных инструментах определяется в основном стабильностью возбуждающего воздействия, стабильность свободных колебаний — стабильностью колебательной системы.

Разработка алгоритма распознавания многоголосных музыкальных сигналов

■ В первом параграфе раздела «Предлагаемая общая концепция» формулируются требования к алгоритму распознавания, являющемуся основой системы управления обработкой сигнала. Предлагается способ достижения этих требований.

Основное требование: *система распознавания должна сформировать поток данных для формирования АЧХ фильтра подавления шума и сигналы, управляющие синтезом обертонов.*

При решении задачи *сегментации* предлагается использовать априорные сведения о распознаваемом сигнале, при решении задачи определения высот тонов в созвучиях звуковых объектов предлагается использовать *гипотезы*, сформированные на стадии сегментации.

■ Современное понимание задачи сегментации музыкального сигнала и существующие способы её решения описаны в параграфе «Современные подходы к сегментации звукового сигнала».

Наиболее развитый подход к сегментации — *сегментация по громкости*, состоящая в моделировании восприятия человеком громкости звука и локализации её максимумов. Используемая модель слухового восприятия основана на наборе полосовых фильтров, таких, что полоса пропускания каждого фильтра близка к критической полосе слуховой системы. В огибающей каждого узкополосного сигнала ищутся локальные максимумы, при дальнейшем сопоставлении интенсивности и положении во времени которых, принимается решение о появлении в сигнале атаки, т.е. интервала нестационарности с быстрым нарастанием интенсивности колебаний.

При сегментации многокомпонентных сигналов использование *параллельного спектрального анализатора* неизбежно. Чем уже полоса анализа, тем меньше звуковых объектов модулируют энергию колебаний в ней и тем проще «найти след» каждого звукового объекта. Системы сегментации по громкости используются для выделения ритма музыкального сигнала и ориентированы на локализацию моментов начала звучания нот (*sound onset detection* в зарубежных публикациях.)

Другой подход основан на *визуализации* звукового сигнала. Для визуализации используется графическое представление спектрограммы (сонограммы) сигнала — временной зависимости модулей коэффициентов комплексной формы дискретного преобразования Фурье. Анализ и обработка сигнала осуществляются на основе *анализа изображения*. Отображение интервалов нестационарности сигнала на спектрограмме имеет характерные признаки.

■ «Разработка алгоритма сегментации многоголосных музыкальных сигналов». Очевидно, для разделения звуковых объектов, перекрывающихся во времени необходимо их разделение в частотной области. Разделяемые сигналы квазипериодичны, при этом, множество примерных значений периодов звуковых объектов известно заранее. Поэтому желательно использовать *фильтрацию по периоду*.

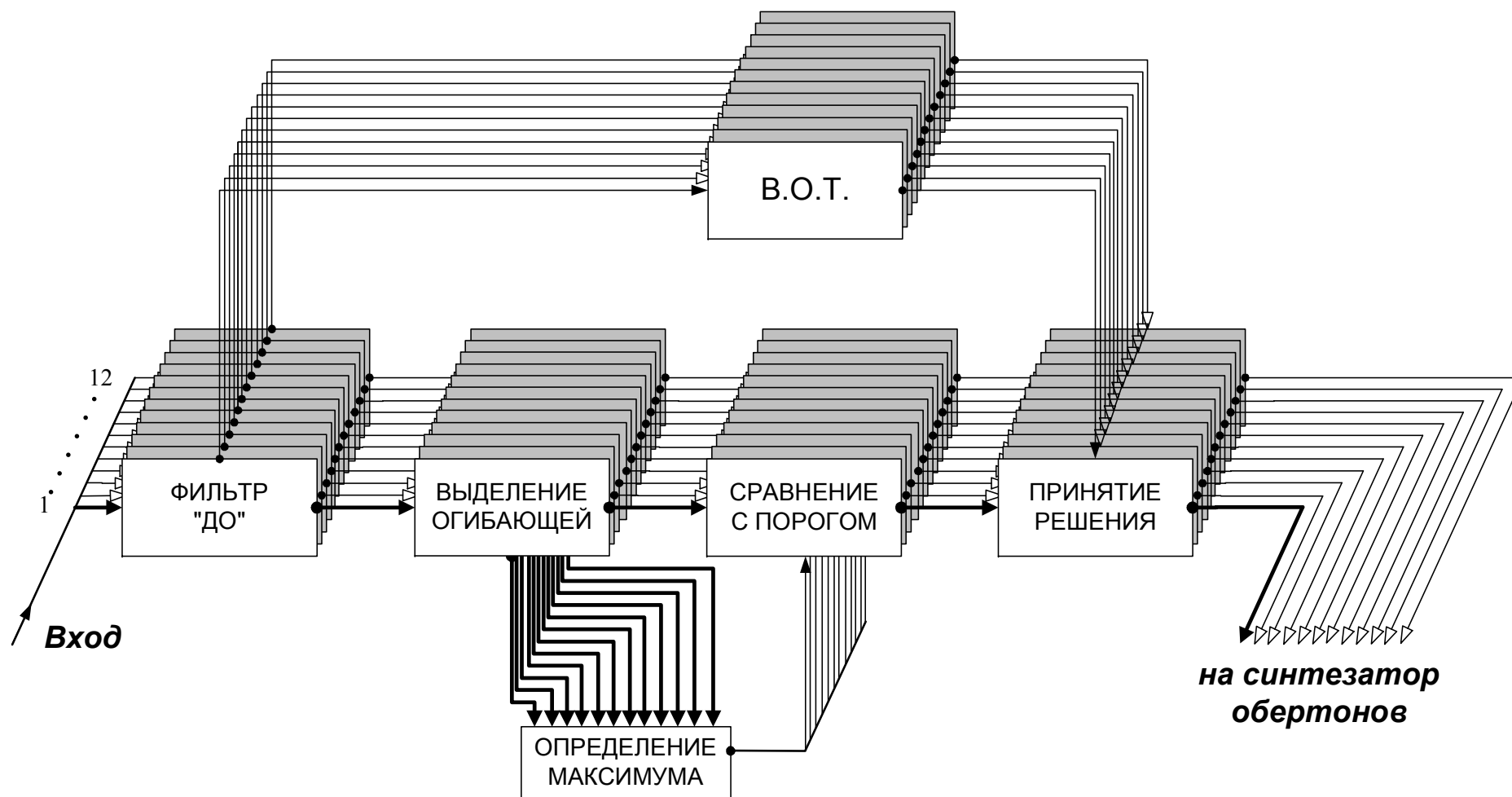


Рис. 4 Схема процесса сегментации многоголосных мелодий

Фильтром, *настроенным на определённый период* сигнала будем называть фильтр, такой, что интенсивность его отклика тем больше, чем ближе период входного сигнала к периоду, на который настроен фильтр. Группа параллельно включённых полосовых фильтров, чьи центральные частоты образуют ряд $F_{ц}, 2F_{ц}, 4F_{ц}, 8F_{ц}, 16F_{ц}, 32F_{ц}, \dots$, будет в наименьшей степени подавлять сигналы с периодами, образующими ряд $\dots, \frac{1}{32F_{ц}}, \frac{1}{16F_{ц}}, \frac{1}{8F_{ц}}, \frac{1}{4F_{ц}}, \frac{1}{2F_{ц}}, \frac{1}{F_{ц}}$, т.е. сигналы с основными тонами, отличающимися на октаву. Это предположение подтверждено теоретическими рассуждениями и опытными данными.

Для сегментации используется набор из 12 гребенчатых фильтров — по числу полутонов в октаве. Каждый гребенчатый фильтр настроен на периоды ОТ, отличающихся на целое число октав: например, на все «ля», на все «до».

Если частоты основных тонов соотносятся как $\frac{m}{n}$, $m, n \in \mathbb{Z}$, $m, n \geq 1$, то некоторые их гармоники имеют совпадающие частоты. Такие соотношения имеют место для натурального строя, с некоторым приближением — для равномерного. Поэтому идеального разделения звуковых объектов с разными основными тонами в описанной системе фильтров не происходит, необходимо использование дополнительных признаков.

Таковыми признаками выбраны *основной тон* и *огibaющая* выходного сигнала составного фильтра.

Для каждой временной выборки определяется максимальное значение E_{MAX} *огibaющей* откликов фильтров и происходит полное подавление сигнала тех откликов, интенсивность которых меньше порога $k \cdot E_{MAX}$ ($k < 1$).

Далее на каждой временной выборке определяется основной тон отклика фильтра и если он не относится к набору основных тонов, на которые настроен фильтр, выходной сигнал фильтра также полностью подавляется.

Схема процесса представлена на рисунке 4. «В.О.Т» — выделитель основного тона.

Результатом работы системы по предложенному алгоритму являются 12 сигналов, представляющие собой *огibaющие* откликов составных фильтров, полностью подавленные на интервалах несоответствия указанным условиям. Использование именно *огibaющей* откликов необходимо по двум причинам:

1. отличие *огibaющей* отклика от нуля позволяет определить временные границы звукового объекта и сформировать гипотезы для алгоритма определения высоты тона;
2. *огibaющую* удобнее использовать для управления амплитудами синтезируемых гармонических составляющих.

Таким образом, требования к алгоритму сегментации выполнены.

■ «Разработка алгоритма определения основных тонов компонентов созвучий».

Определение высоты тона происходит комбинированным методом. Первоначально амплитудный спектр текущей выборки сигнала анализируется на наличие локальных максимумов. Локальные максимумы, меньшие определённого порога ($0,125 \cdot A_{MAX}$, где A_{MAX} — максимальное значение амплитудного спектра на данной выборке), игнорируются (см. рис. 5).

Для проверки гипотезы $F0'_1$ в исходном амплитудном спектре приравняются нулю значения на частотах основных тонов $F0'_2$, $F0'_3$, на частотах их гармоник, а также в некоторой окрестности основных тонов и гармоник. Если в полученном амплитудном спектре не обнаруживается ни одного максимума из списка, полученного при анализе полного спектра, гипотеза *признаётся ошибочной*.

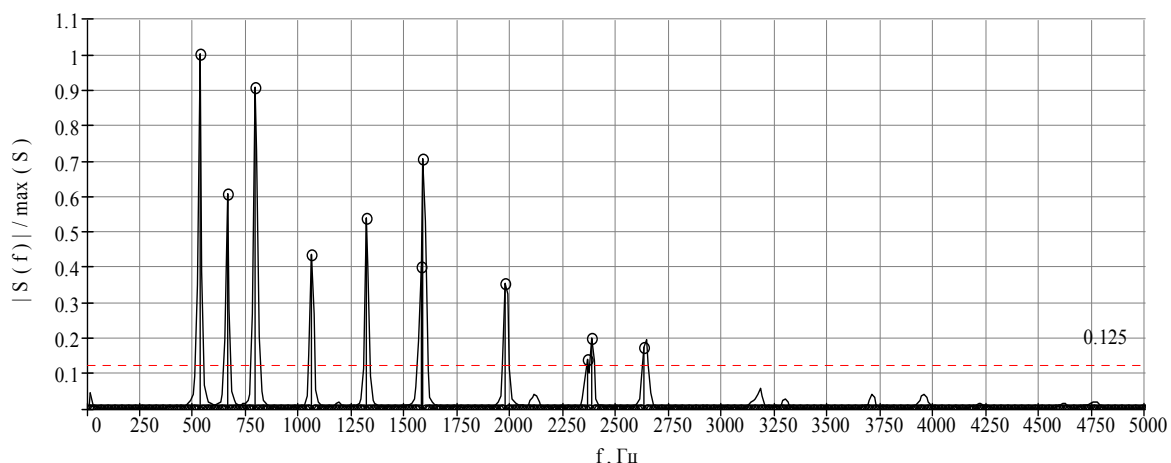


Рис. 5 Исходный амплитудный спектр

Соответствующие поправки вносятся и в выходные сигналы алгоритма сегментации: на интервалах времени, соответствующих ошибочным гипотезам приравнивается нулю огибающая отклика, сформировавшего ошибочную гипотезу.

Обратным БПФ квадрата полученного амплитудного спектра вычисляется соответствующая автокорреляционная функция (АКФ). За истинное значение $F0_1$ принимается значение $\frac{1}{\tau_M}$, где τ_M — положение на оси времени наибольшего

локального максимума АКФ. Работа алгоритма иллюстрируется рисунком 6. Рисунки а), б) и в) соответствуют проверкам гипотез $F0'_1$, $F0'_2$, и $F0'_3$ соответственно. На рисунке 6 жирными сплошными линиями изображены амплитудные спектры после подавления всех составляющих, не относящихся к проверяемой гипотезе, тонкими пунктирными — исходный амплитудный спектр.

■ Выводы

1. Подход к сегментации, основанный на модели слухового восприятия (сегментация «по громкости») может быть использован для любого типа звучания, но ориентирован на определение моментов начала звучания, поэтому, в основном, используется для выделения ритма.

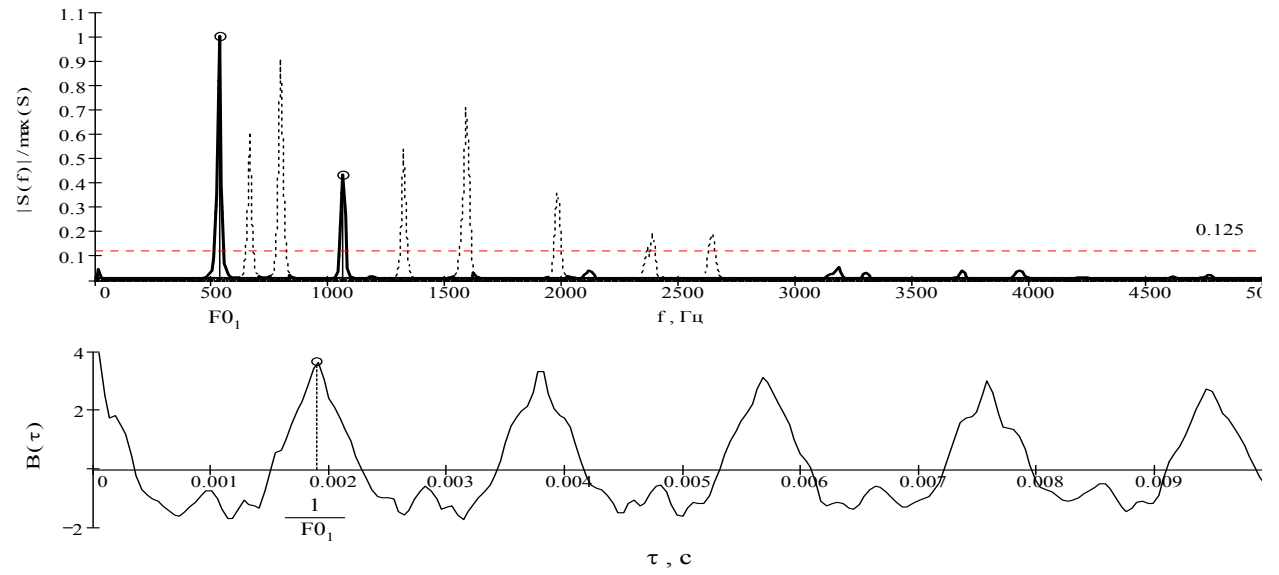
2. Использование «фильтрации по периоду» может быть использована только для сегментации сигналов, состоящих из звуковых объектов с основными тонами из заранее известного набора, но такой подход позволяет определять длительность нот, формировать гипотезы для алгоритма определения основного тона.

3. Взаимодействие алгоритмов сегментации и распознавания, при котором алгоритм сегментации формирует гипотезы частот основных тонов для алгоритма распознавания, а алгоритм распознавания исправляет ошибки алгоритма сегментации, повышает достоверность распознавая.

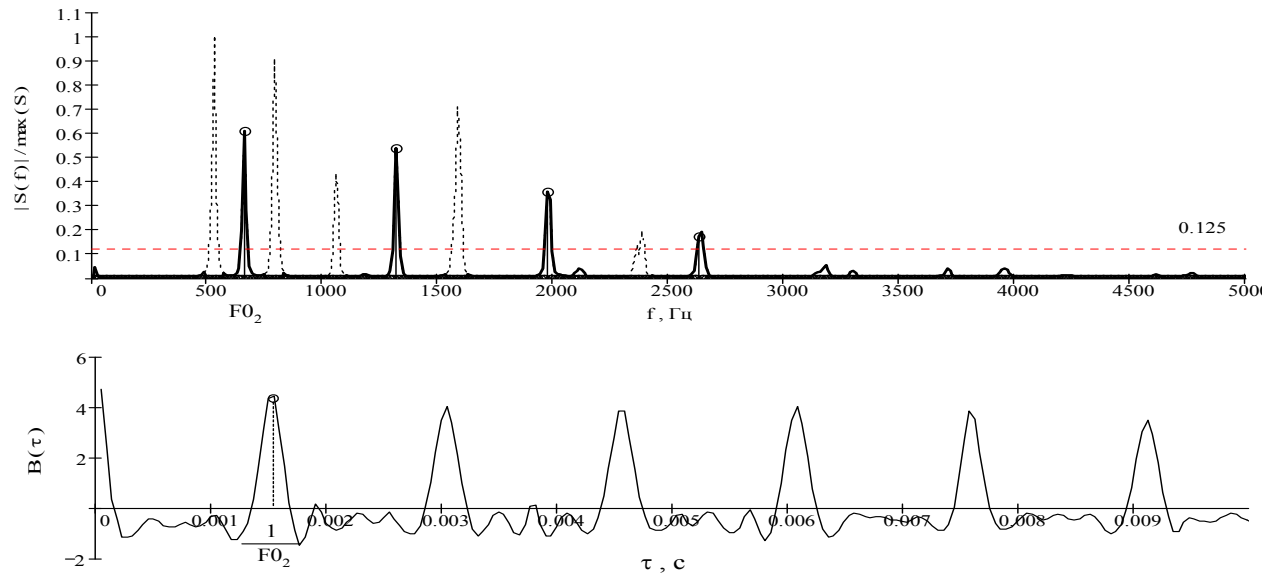
4. Использование для сегментации музыкального сигнала описанной системы фильтров позволяет сформировать сигналы, управляющие амплитудами обертонов при синтезе.

Разработка алгоритма обработки сигнала на основе распознавания

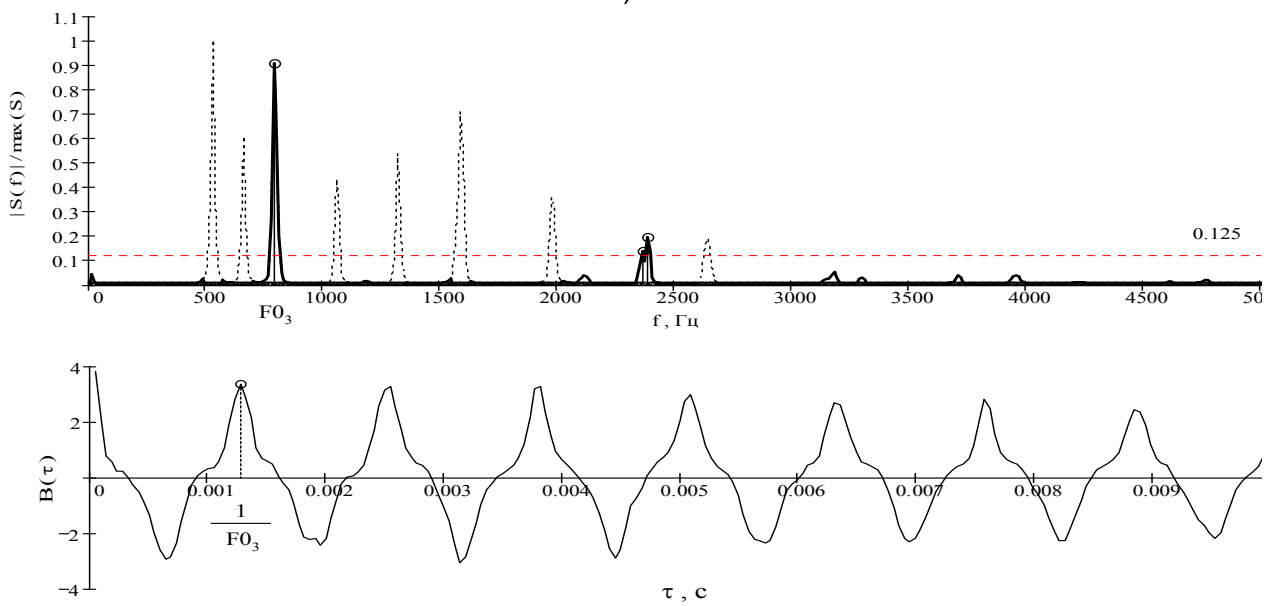
Для шумоподавления используется БПФ-фильтр — простейший способ сформировать АЧХ сложного вида при линейной ФЧХ. Формирование АЧХ фильтра происходит следующим образом. Из частот основных тонов распознанных звуковых объектов формируется массив частот, соответствующих основным тонам и их гармоникам. АЧХ фильтра приравнивается 1 в окрестности частот этого массива.



а)



б)



в)

Рис. 6 Нормированные амплитудные спектры и соответствующие АКФ при проверке гипотез частот ОТ

Ширина такой окрестности задаётся пользователем величиной эквивалентной добротности. В остальной области частот значение АЧХ фильтра приравняется значению $H_{\text{ПОД}} < 1$, также задаваемому пользователем (см. рис. 7)

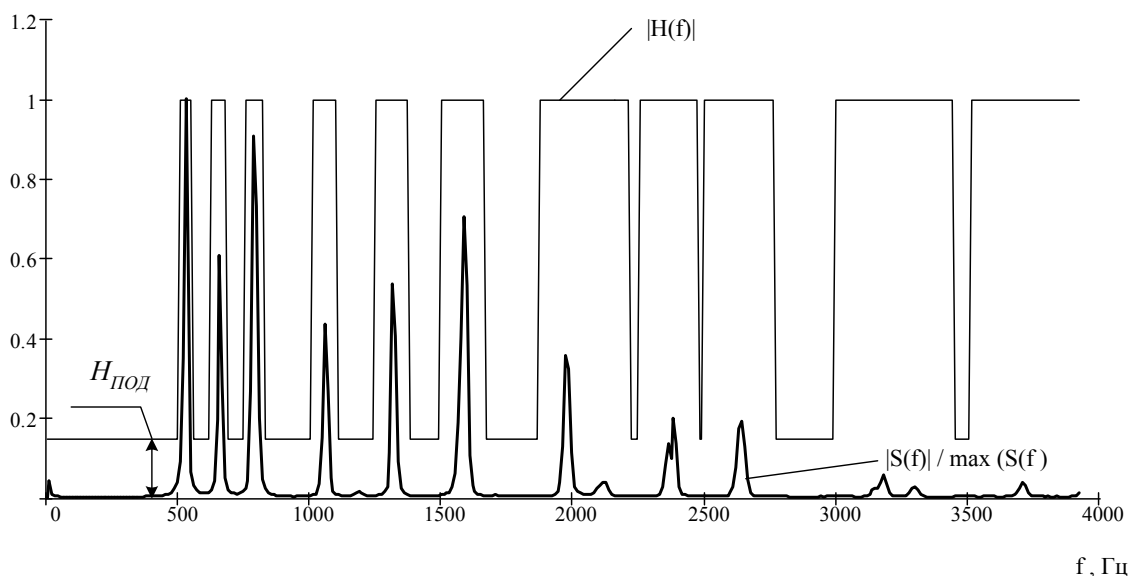


Рис. 7 Формирование АЧХ фильтра шумоподавления

Проведённый (глава «Свойства звуковых объектов») анализ результатов исследования структуры звуковых объектов показывает, что отклонения частот обертонов от среднего значения увеличиваются с ростом порядка обертона. Поэтому и «степень свободы» обертона, т.е. ширина полосы пропускания фильтра в окрестности его средней частоты должна быть не одинаковой во всём диапазоне частот.

БПФ фильтрация осуществляется с перекрытием временных окон для избежания скачкообразных изменений спектра при переходе от выборки к выборке.

Наконец, синтез гармоник осуществляется под управлением пользователя на основе полученных при распознавании сведений о сигнале. Синтезируемые обертоны — гармонические колебания с управляемой амплитудой, начальной фазой и частотой.

Частота синтезируемого обертона определяется частотой основного тона звукового объекта, определённой при распознавании, и целочисленным множителем — порядком обертона:

$$f_{n,k} = F0_n \cdot k, \text{ где}$$

$F0_n$ — частота основного тона n -ого распознанного звукового объекта ($n=1\dots4$).

k — порядковый номер обертона ($k=2\dots20$).

Амплитуда обертона определяется значением огибающей выходного сигнала и множителем, установленным пользователем при ручном регулировании. При этом необходимо коммутирование нужного сигнала огибающей на генератор нужного обертона. Т.е. для синтеза обертонов «до-диез» третьей октавы нужно выбрать сигнал с фильтра «до-диез».

Начальная фаза обертона меняется при переходе от одной выборки к другой. Для каждой выборки запоминается мгновенное значение фазы обертона на последнем отсчёте выборки. Это значение запоминается и используется при начале синтеза на новой выборке. Таким образом, обеспечивается непрерывность мгновенной фазы обертонов.

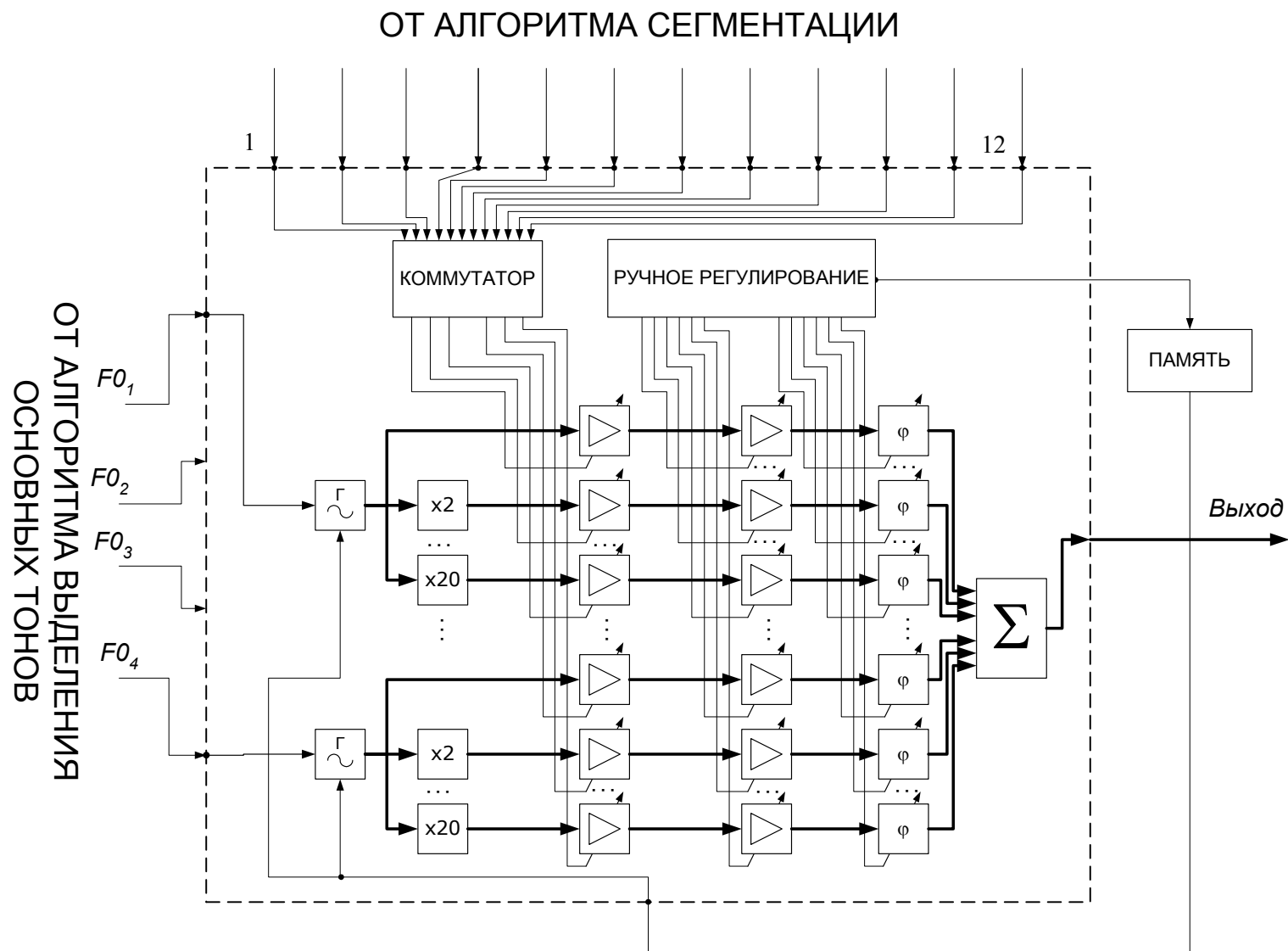


Рис.8 Схема синтеза обертонов на основе распознавания сигнала.

Схема синтеза обертонов представлена на рисунке 8. Выходной сигнал синтезатора суммируется с выходным сигналом шумоподавителя, согласно схеме на рис. 1.

Синтезируются только обертоны, не обнаруженные в исходном сигнале, т.е. мощность которых меньше определённого порога. При ручном регулировании амплитуды обнаруженных обертонов происходят изменения АЧХ фильтра шумоподавления: в окрестности частот регулируемых обертонов модуль коэффициента передачи устанавливается в соответствии с выбором пользователя и, соответственно, отличен от 1. Значение коэффициента передачи в области подавления при этом не изменяется.

■ Выводы

1. Разработанные алгоритмы сегментации и распознавания музыкального сигнала формируют весь набор данных, необходимых для обработки сигнала на уровне структуры отдельного звукового объекта, поэтому фильтрация сигнала и синтез обертонов становятся чисто технической операцией.

2. Процесс обработки сигнала с использованием распознавания образов становится полностью контролируемым и управляемым пользователем, в отличие от работы многих психоакустических процессоров.

Заключение

В заключительном разделе подводятся итоги работы, обсуждается полнота выполнения поставленных задач. Обсуждается применение достигнутых результатов. Выявляются недостатки разработанных систем распознавания и обработки сигнала, намечаются перспективы их развития.

Основные результаты работы

1. Повышено качество восстановления музыкальных сигналов. Проведённые субъективно-статистические экспертизы показали, что при сравнении результатов работы двух систем: разработанной и одной из существующих, 65% экспертов отдали предпочтение звучанию, восстановленному с помощью разработанной системы

2. Оператору системы реставрации предоставлена возможность осуществлять ранее невыполнимые операции: с минимальными искажениями тембра изменять баланс громкостей звуковых объектов, по-отдельности обрабатывать звуковые объекты, перекрывающиеся во времени.

3. Процесс реставрации фонограмм — процесс творческий. Реставратору предоставлен инструмент творческой работы, инструмент, для овладения которым не требуются специальной технической подготовки.

4. Предоставлена возможность сократить время обучения неподготовленных или неопытных пользователей систем восстановления музыкальных сигналов за счёт наделения устройств обработки интуитивно более понятными функциями, например, регулировка амплитуд отдельных обертонов и мощности отдельных звуковых объектов вместо регуляторов АЧХ эквалайзера.

Список публикаций соискателя

1. Восстановление музыкальных сигналов с использованием распознавания образов, «Научноёмкие технологии» №3, 2003, том 4.
2. Реставрация без иллюзий Broadcasting, №2, 2003.
3. Подавление взаимного влияния сигналов, соответствующих разным группам инструментов, при многомикрофонной звукозаписи. Депонировано в ЦНТИ «Информсвязь». М.: — 2001.
4. Распознавание музыкальных сигналов. Депонировано в ЦНТИ «Информсвязь», М.: — 2000.
5. Исследование структуры звуковых объектов. Депонировано в ЦНТИ «Информсвязь». М.:— 2002.
6. Распознавание музыкальных образов. Тезисы докладов 55 студенческой научно-технической конференции МТУСИ. «Информсвязьиздат» М.: — 2000.
7. Автоматическое определение типа звучания. Депонировано в ЦНТИ «Информсвязь». М.:— 2003.
8. Автоматическая сегментация многоголосных мелодий. Депонировано в ЦНТИ «Информсвязь». М.:— 2003.
9. Автоматическая сегментация речевых и музыкальных сигналов Тезисы докладов 55 студенческой научно-технической конференции МТУСИ. «Информсвязьиздат» М.: — 2000.
10. Основной тон неречевых сигналов. Тезисы докладов 10-й Межрегиональной конференция МНТОРЭС им А. С. Попова «обработка сигналов в системах связи».М.: : 2000
11. От вокодера к музкодеру. Тезисы докладов IV международной НТК «Перспективные технологии в средствах передачи информации». — Владимир, 2001.
12. Обработка звукового сигнала на основе распознавания. Тезисы докладов IV международной НТК «Перспективные технологии в средствах передачи информации». — Владимир, 2001.
13. Узнавание тембра: что важно и что – нет? Тезисы докладов конференции профессорско-преподавательского состава МТУСИ. М: — 2002.
14. Классификация музыкальных инструментов. Тезисы докладов конференции профессорско-преподавательского состава МТУСИ. — М. : 2002.
15. Аддитивный синтез вокализованных звуков речи. Тезисы докладов конференции профессорско-преподавательского состава МТУСИ. — М.: — 2003.