

61:03-5/3066-6

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*

Научный руководитель: *д.т.н., профессор*

С. Л. Мишенков

Д и с с е р т а ц и я
на соискание учёной степени
кандидата технических наук

Москва 2003

ОГЛАВЛЕНИЕ

1. ВВЕДЕНИЕ.....	5
2. ЦЕЛИ И ЗАДАЧИ.....	11
2.1. АКТУАЛЬНОСТЬ ТЕМАТИКИ	11
2.2. ЦЕЛИ И ЗАДАЧИ.....	13
3. ПРИНЦИПЫ ПОСТРОЕНИЯ СИСТЕМ РАСПОЗНАВАНИЯ ОБРАЗОВ	15
3.1. ОБЩАЯ ТЕОРИЯ РАСПОЗНАВАНИЯ ОБРАЗОВ.....	15
3.1.1. Основные определения.....	15
3.1.2. Классификация систем распознавания	17
3.1.3. Цели и задачи систем распознавания образов.....	22
3.1.4. Последовательность распознавания образов в общем виде	24
3.1.5. Определение полного перечня признаков, характеризующих объекты.....	25
3.2. РАСПОЗНАВАНИЕ ОБРАЗОВ С ИНФОРМАЦИОННОЙ ТОЧКИ ЗРЕНИЯ	27
3.2.1. Информативность музыкальных произведений	27
3.2.2. Преобразование информации при распознавании музыкального сигнала.....	31
3.3. РАСПОЗНАВАНИЕ ЗВУКОВЫХ ОБРАЗОВ	34
3.4. СТАТИСТИЧЕСКИЙ ПОДХОД К РАСПОЗНАВАНИЮ ОБРАЗОВ.....	38
3.4.1. Реализации и признаки.....	38
3.4.2. Процедуры принятия решения	42
3.5. СИСТЕМЫ РАСПОЗНАВАНИЯ МУЗЫКАЛЬНЫХ СИГНАЛОВ	45
3.5.1. Развитие систем распознавания музыкальных сигналов.....	45
3.5.2. Задачи распознавания музыкальных сигналов	50
3.6. СЛОЖНОСТИ РАСПОЗНАВАНИЯ МНОГОГОЛОСНЫХ МЕЛОДИЙ.....	56
3.6.1. Общие соображения.....	56
3.6.2. Созвучия тональных звуков.....	58

3.7. СОВРЕМЕННЫЕ ПОДХОДЫ К РАСПОЗНАВАНИЮ ПОЛИФОНИЧЕСКИХ МЕЛОДИЙ.....	61
3.7.1. Представление данных в системах распознавания музыкальных сигналов.....	61
3.7.2. Методология «классной доски»	68
3.7.3. Недостатки современных систем распознавания музыкальных сигналов.....	72
3.8. ВЫВОДЫ	74
4. СОВРЕМЕННАЯ ТЕХНИКА ВОССТАНОВЛЕНИЯ МУЗЫКАЛЬНЫХ СИГНАЛОВ	75
4.1. КОМПЬЮТЕРИЗИРОВАННЫЕ СИСТЕМЫ ВОССТАНОВЛЕНИЯ	75
4.1.1. Система “NoNoise”	75
4.1.2. Система «Канонь».....	76
4.2. ПСИХОАКУСТИЧЕСКИЕ ПРОЦЕССОРЫ.....	77
4.2.1. Эксайтер (Exciter).....	78
4.2.2. Энхэнсер (Enhanser).....	80
4.2.3. Максимайзер (Maximizer).....	81
4.2.4. Виталайзер (Vitalizer)	83
4.3. ОБЩАЯ ХАРАКТЕРИСТИКА СОВРЕМЕННОЙ ТЕХНИКИ ВОССТАНОВЛЕНИЯ И ОБОСНОВАНИЕ ПРЕДЛАГАЕМОГО ПОДХОДА	85
4.4. ВЫВОДЫ	90
5. ИССЛЕДОВАНИЕ СТРУКТУРЫ ЗВУКОВЫХ ОБЪЕКТОВ.....	91
5.1. СТРУКТУРА ЗВУКОВОГО ОБЪЕКТА И ТЕМБР	91
5.2. ОПИСАНИЕ ИССЛЕДОВАНИЯ.....	92
5.2.1. Цели и задачи исследования.....	92
5.2.2. Объект исследования	93
5.2.3. Методика и последовательность исследования	93
5.2.4. Инструменты исследования	95

5.3. РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ	99
5.3.1. Классификация музыкальных инструментов	99
5.3.2. Гармоничность музыкальных звуков. Модель синтеза звукового объекта	108
5.3.3. Узнавание тембра: что важно и что нет? Модель анализа звукового объекта	114
5.4. ВЫВОДЫ	118
6. РАЗРАБОТКА АЛГОРИТМА РАСПОЗНАВАНИЯ МНОГОГОЛОСНЫХ МУЗЫКАЛЬНЫХ СИГНАЛОВ	118
6.1. РАЗРАБОТКА АЛГОРИТМА СЕГМЕНТАЦИИ	118
6.1.1. Современные подходы к сегментации музыкальных сигналов	120
6.1.2. Сегментация музыкального сигнала с использованием априорной информации	125
6.2. РАЗРАБОТКА АЛГОРИТМА ОПРЕДЕЛЕНИЯ ОСНОВНЫХ ТОНОВ ЗВУКОВЫХ ОБЪЕКТОВ ПЕРЕКРЫВАЮЩИХСЯ ВО ВРЕМЕНИ	137
6.3. ВЫВОДЫ	144
7. РАЗРАБОТКА АЛГОРИТМА ОБРАБОТКИ СИГНАЛА НА ОСНОВЕ РАСПОЗНАВАНИЯ	145
8. ЗАКЛЮЧЕНИЕ	155
9. СПИСОК ЛИТЕРАТУРЫ	159
10. СПИСОК ПУБЛИКАЦИЙ СОИСКАТЕЛЯ	167
ПРИЛОЖЕНИЕ 1. ПРОТОКОЛ ИСПЫТАНИЙ ПО ОЦЕНКЕ ЭФФЕКТИВНОСТИ РАЗРАБОТАННОГО АЛГОРИТМА ПОДАВЛЕНИЯ ШУМОВ.....	168

1. ВВЕДЕНИЕ

Процесс реставрации фонограмм длителен и трудоёмок. Начинается он почти всегда с обработки носителя записи — механического или магнитного, затем проводится обработка сигнала. Обработка носителя записи — процесс не менее важный, чем обработка сигнала, однако именно работа с сигналом, точнее проблемы, возникающие в её ходе, требуют разработки новых методов. Субъективные оценки качества восстановленных сигналов заставили всерьёз задуматься о целесообразности реставрации фонограмм на современном этапе развития техники. Нередко организации, занимающиеся хранением и реставрацией фонограмм, представляющих художественную ценность, для продления срока хранения фондов «консервируют» копии «до лучших времён». Для этого сигналы преобразуются в цифровую форму с использованием высококачественных АЦП (например, 24-битное квантование и дискретизация с частотой 192 кГц) и переносятся на долговечный носитель (например, компакт-диск)

Как правило, восстановления требуют сигналы, некоторые технические параметры которых деградировали: ограничена полоса частот, утрачены фрагменты фонограммы, уменьшено отношение сигнал/шум за счёт шумов, например, магнитного носителя и старения магнитного слоя, присутствуют нелинейные искажения. Задачи реставрации фонограмм можно сформулировать следующим образом:

1. устранение «лишних» спектральных составляющих, т.е. компонентов спектра, не содержащихся в исходной фонограмме по окончании процесса её формирования, — продуктов нелинейных искажений, всевозможных помех и шумов и т.п.
2. восстановление компонент спектра, утраченных из-за линейных искажений — ограничение полосы частот сигнала, «провалы» АЧХ тракта запись-воспроизведение, и т.п.

Очевидно, исходная форма сигнала неизвестна и потому достижима с некоторой вероятностью. Поэтому процесс реставрации фонограммы — такой же творческий, как и процесс её создания.

Реставрация — понятие универсальное, используемое различными областями человеческих знаний. Не разумно ли сформулировать и общие подходы, принципы и задачи реставрации? Восстановлению подлежат и памятники письменности, и живописные полотна, и уникальные звукозаписи, материалы кинохроники. В любом случае, имеют дело с носителями звуковой или визуальной информации. Восстановление становится возможным при условии достаточности информации, содержащейся в носителе, требующем реставрации, а также информации, полученной об объекте реставрации из разных источников.

Утрата частей физического носителя информации или части информации, переносимой носителем не всегда исключает достоверную передачу этой информации. Очевидно, существует такое количество информации (*предел*), при утрате которого делается невозможным восприятие остальной части. Обратимся к рисунку 1. Допустим, нам предлагают восстановить исходное изображение по одному из фрагментов — большему или меньшему. Априорная информация в данном случае состоит в том, что восстановлению подлежат фрагменты симметричного изображения (человеческого лица). Чтобы решить поставленную задачу, мы должны, прежде всего, понять, что изображено на предложенной нам части листа, т.е. узнать (*распознать*) часть объекта и, согласно своим представлениям домыслить недостающие части. Распознать — значит принять решение о принадлежности изучаемого объекта к одному из известных классов. Таким образом, мы приходим к мысли о том, что достоверность восстановления зависит от достоверности распознавания, достоверность распознавания (т.е. вероятность принятия верного решения о принадлежности распознаваемого объекта к тому или иному классу) зависит, в частности:

1. от количества информации, содержащейся в исследуемом фрагменте носителя информации,
2. от количества **априорной** информации, которой располагает распознающий субъект о распознаваемом объекте, т.е. от набора разнородных знаний, которыми обладает человек.

При восстановлении музыкального сигнала мы всегда обладаем меньшим количеством информации, чем содержалось в исходном объекте, иначе реставрация бессмысленна. Одним из путей восстановления утраченной информации является её «интерполяция» и «экстраполяция» на основе данных, полученных при распознавании.

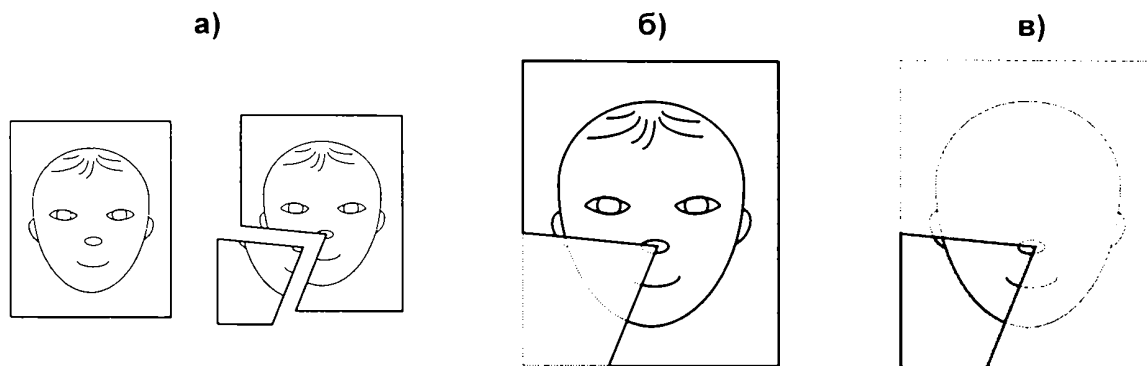


Рис. 1. Пример восстановления зрительного образа при наличии априорной информации: а) исходный образ и фрагменты, подлежащие восстановлению; б) восстановление с большой достоверностью; в) восстановление с малой достоверностью.

Реставратор, восстанавливающий живописное полотно, прежде всего, решает задачу **распознавания образов**, которые он восстанавливает — он принимает решение о том, что изображено на полотне и, только исходя из этих сведений, устраняет ненужные фрагменты изображения и добавляет недостающие. При этом «недостачу» и «ненужность» фрагментов определяет сам реставратор в зависимости от того, как он распознал восстанавливаемые образы и насколько он себе представляет то, что было изображено на полотне.

Первым этапом восстановления сигнала является фильтрация, поскольку проще сначала подавить шум, а затем провести синтез спектральных составляющих, чем наоборот. Какие сведения необходимо получить при распознавании сигнала и как сформировать АЧХ фильтра на основе этих данных? Очевидно, необходимо локализовать в частотной области полезные составляющие — основные тоны и обертоны. Соответственно, при таком подходе считается, что область частот, не занятая полезным сигналом, занята шумом.

После фильтрации можно переходить непосредственно к восстановлению утраченных компонентов спектра — синтезу обертонов. Известны прибли-

зительные частоты недостающих обертонов, а их амплитуды должны либо назначаться пользователем, либо определяться в результате анализа спектра (предсказание, интерполяция, экстраполяция), либо определяться в соответствии с априорными сведениями об обрабатываемом сигнале.

Важно отметить, что применение распознавания образов предоставляет возможность обрабатывать отдельный звуковой объект (реализацию отдельной ноты). По сути, отдельный звуковой объект является минимальным элементом – «кирпичиком» восприятия звукового сигнала, и возможность работать со звуковым сигналом на «элементарном» уровне пока не предоставляет ни один подход к обработке сигнала.

Теперь можно изобразить схему действий и управления процессами при восстановлении музыкального сигнала с использованием распознавания образов (рис. 2).

В схеме выделим ветвь управления и ветвь обработки. Процессы, входящие в ветвь управления формируют данные, управляющие процессами ветви обработки. Процесс распознавания сигнала – наиболее важный, эффективность восстановления почти полностью зависит от его результатов. Остальные процессы управления, действуя по намного более простым алгоритмам, механически исполняют указания процесса распознавания. Процесс распознавания формирует информацию, необходимую для работы остальных процессов.

Распознавание – задача статистическая, и достоверность распознавания современных систем не превышает 80%. Ошибки распознавания (неверное определение высоты тона, ошибочная сегментация), вообще говоря, приведут к непредсказуемым последствиям, как на стадии фильтрации, так и на стадии синтеза обертонов. Поэтому внедрение любых устройств обработки на основе распознавания станет возможным лишь при условии создания устойчивых систем распознавания. Современные системы распознавания звуковых сигналов не универсальны, создание и использование же универсальных систем распознавания, отличающих речь от музыки и, тем более, распознающих смесь речевого и музыкального сигналов значительно удорожат устройства обработки.

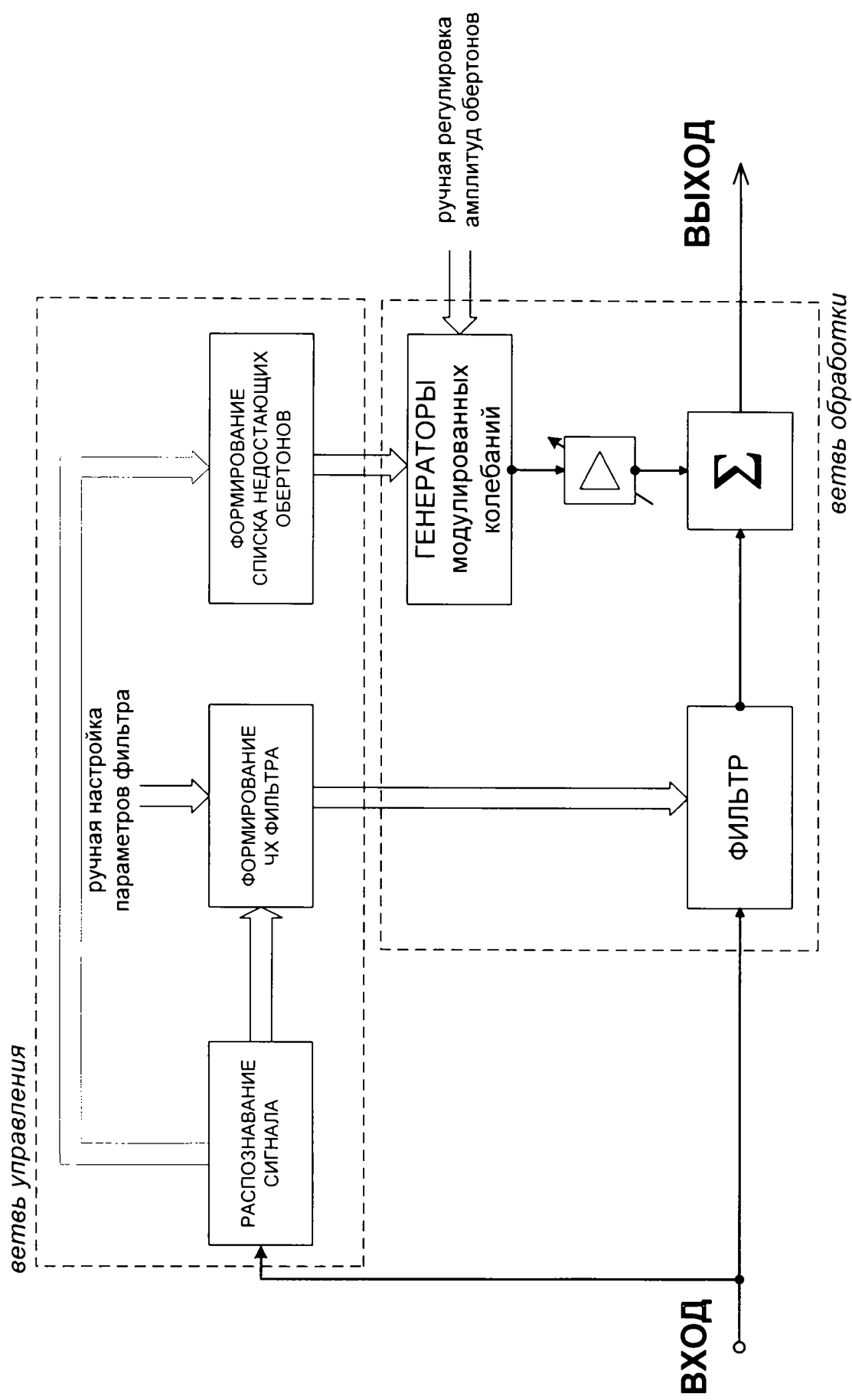


Рис 2. Схема управления и последовательность процессов обработки при восстановлении распознавания

Обратим внимание на то, что применение распознавания к восстановлению сигналов — это попытка провести действительно *восстановление* исходной спектральной структуры сигнала, а не *создать иллюзию* восстановления, обмануть слуховую систему.

- Научная новизна работы

1. Для управления **обработкой** сигнала в процессе реставрации используется информация, полученная при предварительном **распознавании** сигнала. Т.е. к решению задачи восстановления музыкальных сигналов по неполной или искажённой информации применены методы распознавания образов.

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^{OT} , зная один из членов которого (например, эталонный тон f_3 , по которому настраиваются инструменты), можно с незначительными отклонениями восстановить все остальные. Например, для равномерного строя: $f_i^{OT} = f_3 \cdot \sqrt[12]{2^i}$, где f_3 — эталонная частота, для натурального строя $f_i^{OT} = f_3 \cdot \frac{m}{n}$, где $m, n \in Z, m, n \geq 1$.

- 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 статье

ях, депонированных ЦНТИ «Информсвязь», докладывались на заседаниях кафедры РВ и ЭА, НТС МТУСИ.

Всего по теме работы опубликовано 2 статьи в открытой печати, 8 тезисов докладов на конференциях, депонировано 5 статей.

2. ЦЕЛИ И ЗАДАЧИ

2.1. АКТУАЛЬНОСТЬ ТЕМАТИКИ

Документальная или художественная ценность произведений искусства в значительной степени зависит от событий, происходящих в культурной и политической сферах жизни общества. Нередко период создания произведения искусства и период, когда оно является востребованным, не совпадают во времени. С одной стороны, происходит естественное необратимое старение и разрушение носителя информации. Соответственно часть информации теряется, и, очевидно, существует *предел* количества потерянной информации, делающий невозможным восприятие оставшейся части. С другой стороны, происходит развитие техники, стимулирующее человеческие потребности, стимулирующие, в свою очередь, развитие техники. Если информации (смысловой или эмоциональная), сохранившейся в носителе, становится недостаточно для удовлетворения соответствующих потребностей человека, возникает необходимость *реставрации*. Если возможностей технических средств достаточно для восполнения такого дефицита информации, реставрация становится *возможной*.

Под реставрацией (*restauratio* (лат.) — восстановление) понимается восстановление чего-либо в первоначальном (или близком к первоначальному) виде. Первые носители звуковой информации были механическими, и восстановления требовал, в первую очередь, сам носитель. Поэтому часто под восстановлением звуковой информации понимают *реставрацию фонограмм* (т.е. носителей записи). Однако, с появлением магнитной записи, по крайней мере, не меньшее внимание стало уделяться обработке сигнала, воспроизведённого с носителя. Кроме того, иногда существенные потери полезной информации происходят не только в процессе записи или хранения носителя, но и в процес-

се формирования сигнала записи. Причиной таких потерь является некорректная работа с устройствами тракта формирования сигнала, а искажения могут выражаться и в повышении уровня шума, и в ограничении полосы частот, и в обогащении спектра продуктами нелинейных искажений. В подобной ситуации восстановление звуковой информации необходимо, однако реставрации фонограммы не требуется.

Целесообразно наряду с термином «реставрация» (понимая под ним реставрацию фонограмм) использовать термин «восстановление», понимая восстановление утраченной информации.

Действительно, конечной целью реставрации фонограммы является не восстановление исходного состояния носителя записи, а восстановление информации (смысловой и эмоциональной), переносимой сигналом, записанным на этот носитель. Для восстановления информации, переносимой сигналом не обязательно восстановление формы сигнала.

При реставрации стремятся повысить *качество* сигнала по критериям и объективных, и субъективных оценок. Безусловно, субъективные и объективные оценки качества коррелированы, но не для любых оценок найдены функциональные зависимости. По существующим правилам технической эксплуатации (ПТЭ) для радиовещания и телевидения оценке не подлежат видео- и звукозаписи, имеющие срок хранения более 15 лет, выпуск в эфир таких материалов допускается и без обработки. Таким образом, формальная необходимость повышать при реставрации качество по критерию объективной оценки может возникнуть лишь в случае, если фонограмма «моложе» 15 лет, а при современных параметрах носителей записи, это в большинстве случаев означает, что искажения возникли при формировании сигнала, а не при его хранении. Повышение качества по субъективным критериям бывает необходимо для достижения режиссёрского замысла или повышения потребительских свойств продукта вещания. Почти всегда для повышения качества сигнала по субъективным критериям необходимо повышение качества и по объективным критериям. В то же время, нередко имеет место и обратный эффект — обработка

сигнала, направленная на подавление шума и синтез высокочастотных составляющих спектра, приводит к снижению субъективных оценок качества.

Таким образом, актуальность тематики в основном определяют два аспекта:

1. Потери информации (смысловой и эмоциональной), переносимой звуковым сигналом, неизбежно происходят не только при хранении носителей записи, но и при формировании и передаче сигнала, что делает проблему восстановления утраченной информации актуальной.

2. Применение современных средств восстановления информации искажённых сигналов нередко приводит к снижению субъективной оценки их качества, что делает актуальным поиск новых подходов к задаче реставрации.

2.2. ЦЕЛИ И ЗАДАЧИ

Разрабатывая систему восстановления информации, переносимой музыкальным сигналом, будем преследовать ту же цель, что и реставрация фонограмм (см. выше), но при этом учтём следующий факт. Архивные и фондовые записи нередко используются при формировании программ теле- и радиовещания наравне с современными, и при восстановлении утраченной информации необходимо подготовить её к этому использованию, сделать возможными дальнейшую обработку, передачу и хранение сигнала. Широко распространённым способом достижения целей реставрации музыкальных записей является уменьшение заметности искажений — снижение уровня шума до порога маскировки полезным сигналом, синтез высокочастотных составляющих для устранения ощущения узости полосы сигнала, глухого звучания. Именно при таком подходе часто снижаются субъективные оценки качества сигнала, в первую очередь по критерию натуральности звучания. Одной из причин снижения субъективных оценок, вероятно, является обогащение спектра продуктами нелинейных искажений, что влечёт снижение оценки качества и по объективным критериям.

- Целью данной работы является повышение качества восстановления музыкальных сигналов по критериям объективной и субъективной оценок.

Во «Введении» была описана общая концепция восстановления информации, переносимой музыкальным сигналом, с использованием распознавания образов. Именно распознавание образов предлагается использовать для достижения поставленной цели работы. Для подавления компонентов спектра, не относящихся к полезному сигналу, очевидно, необходимо *классифицировать* компоненты спектра как полезные и как посторонние. Чтобы синтезировать недостающие компоненты спектра, необходимо, во-первых, установить недостачу этих компонентов, и, во-вторых, определить параметры генерируемых компонентов. Взаимосвязаны ли задачи шумоподавления и синтеза обертонов? Взаимосвязаны – с точки зрения *структуры звукового объекта*. В контексте данной работы под звуковым объектом будем понимать реализацию отдельной ноты. Под структурой звукового объекта будем понимать соотношение амплитуд, фаз и частот колебаний основной частоты и обертонов (далее в разделе «Свойства звуковых объектов» будут даны определения введённых понятий). Рассматривая музыкальный сигнал как набор звуковых объектов, будем считать, что компоненты звуковых объектов – это полезные компоненты спектра, а все остальные обнаруженные компоненты – посторонние. Далее, утрата полезных компонентов спектра означает изменение исходной структуры звуковых объектов, т.е., восстановив, структуру звукового объекта, мы восстановим утраченные компоненты. Но чтобы работать с сигналом на уровне структуры звукового объекта, необходимо разделить входной сигнал на звуковые объекты и классифицировать каждый звуковой объект по длительности и частоте основного тона, **что по определению является задачей распознавания образов.** Т.е. для достижения поставленной цели необходима разработка двух взаимосвязанных систем: системы распознавания музыкального сигнала и системы его обработки. При этом особенностью системы распознавания должна быть её направленность на обработку сигнала: система распознавания должна сформировать воздействия, управляющие системой обработки. В данной ситуации система распознавания – основа *системы управления* обработкой сигнала. В свою очередь, система обработки должна выполнять функции собственно *восстановления* информации, переносимой музыкальным сигналом, т.е. выполнять:

1) подавление компонентов спектра, на классифицированных как полезные, т.е. относящиеся к звуковым объектам, составляющим полезный сигнал,

2) восстановление структур звуковых объектов, т.е. приведение их к виду, заданному априорными сведениями или эвристическими правилами (*автоматическое* восстановление), или к виду, заданному пользователем (*автоматизированное* восстановление).

Важное замечание: по соображениям облегчения моделирования и реализуемости разрабатываемых алгоритмов целесообразно для анализа и обработки представить сигнал в цифровом виде. Итак, сформулируем ставящиеся в работе 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.

3. ПРИНЦИПЫ ПОСТРОЕНИЯ СИСТЕМ РАСПОЗНАВАНИЯ ОБРАЗОВ

3.1. ОБЩАЯ ТЕОРИЯ РАСПОЗНАВАНИЯ ОБРАЗОВ

3.1.1. Основные определения

В современном распознавании и особенно искусственном интеллекте понятие образа употребляют в самом широком смысле, имея в виду, что

«образ» – это некоторое структурированное приближённое описание (эскиз) изучаемого объекта, явления или процесса. То есть, частичная определенность описания является принципиальным свойством образа.

Основное назначение описаний (образов) – это их использование в процессе установления соответствия объектов, то есть при доказательстве их

идентичности, аналогичности, подобия, сходства и т.п., которое осуществляется путем сравнения (сопоставления). Два образа считаются подобными, если удастся установить их соответствие. Можно, в частности, считать, что имеет место соответствие, если достигнута их идентичность.

Сопоставление образов представляет собой основную задачу распознавания и играет существенную роль в информатике в целом. Эта задача возникает, в частности, в различных разделах искусственного интеллекта, например в понимании естественного языка компьютером, символьной обработке алгебраических выражений, экспертных системах, преобразовании и синтезе программ ЭВМ.

В различных задачах образу придается различный смысл. Это определяется тем, какие характеристики объекта входят в описание образа, какой аппарат используется для представления этих характеристик. Чем большее число свойств и качеств объекта отражено на принятом языке в образе рассматриваемого объекта, тем полнее это описание, тем полнее этот образ характеризует описываемый объект. Однако, в любом случае мы имеем дело с описанием, а не с самим объектом, который всегда богаче описания. Итак, любой образ представляется некоторым набором *признаков*.

Достаточно наглядно и теоретически и практически понимается различие между объектом и образом, если рассмотреть различия между картиной (художественное полотно), являющейся плоским объектом, и таким ее изображением как фотографическое или полученное телекамерой или сканером.

Теперь сконцентрируем внимание на следующем важном термине распознавания образов - «класс». Здесь, прежде всего, обратим внимание на то, что как человек, так и автомат принимают решение на основе отождествления совокупности конкретных значений характеристик объектов или явлений не просто друг с другом, а обычно с некоторым классом, в который объединяются объекты или явления, имеющие общие свойства. Таким образом,

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

Всегда, имея в виду цель распознавания, в конечном итоге принятое решение об отнесении объекта к тому или иному классу определяет реакцию соответствующей системы на данную входную ситуацию однозначно.

Таким образом, в самых общих чертах распознавание можно определить как соотнесение объектов или явлений на основе анализа их характеристик, представляющих образы этих объектов, с одним из нескольких, заранее определенных классов. И следует обратить внимание на то, что термин «распознавание» в равной мере относится как к процессам восприятия и познания, свойственным человеку и живым организмам, так и к техническим попыткам человека реализовать «электронные» или «вычислительные» аналоги этих процессов, то есть к решению задач в рамках предмета распознавания как раздела информатики.

3.1.2. Классификация систем распознавания

Для классификации СР будем использовать следующие принципы:

1. Однородность информации для описания распознаваемых объектов или явлений.
2. Способ получения апостериорной информации.
3. Количество первоначальной априорной информации.
4. Характер информации о признаках распознавания.

Рассмотрим *первый принцип (однородность информации)*. Здесь под однородностью следует понимать различную или единую физическую природу информации (признаков).

По этому принципу СР делятся на:

- простые;
- сложные.

Простые СР характеризуются единой физической природой признаков. Например:

1) только масса — для систем распознавания жетонов и монет в автоматах;

2) геометрические размеры – для таких СР, как всякого рода замки.

Ясно, что для простых систем распознавания не обязательно иметь компьютер. Достаточно их реализовать в виде механических или электромеханических устройств. Сложные СР характеризуются физической неоднородностью признаков. Конечно, простота всегда предпочтительнее.

Второй принцип классификации СР (способ получения апостериорной информации). По этому принципу сложные системы распознавания делятся на:

- одноуровневые;
- многоуровневые.

Многоуровневые сложные системы распознавания отличаются от одноуровневых тем, что не все признаки от разнородных физических измерителей используются непосредственно для решения задачи распознавания. Здесь на основе объединения признаков нескольких измерителей и соответствующей обработки могут быть получены вторичные признаки, которые могут как использоваться в алгоритме классификации, так и сами в свою очередь служить основой для объединения. То есть, получаем 2-й, 3-й и др. уровни признаков, определяющие многоуровневость СР. Причем подсистемы, которые осуществляют объединение признаков, в свою очередь могут представлять собой также устройства распознавания (локальные СР).

Таким образом, в одноуровневых СР информация о признаках распознаваемого объекта (апостериорная информация) формируется непосредственно на основе обработки прямых измерений; в многоуровневых СР информация о признаках формируется на основе косвенных измерений как результат функционирования вспомогательных распознающих устройств (пример: измерение дальности радиолокатором по времени задержки излученного импульса).

Третий принцип классификации (количество первоначальной априорной информации). Здесь вопрос касается того, достаточно или недостаточно априорной информации для определения априорного алфавита классов, построения

априорного словаря признаков и описания каждого класса на языке этих признаков в результате непосредственной обработки исходных данных.

Соответственно этому СР делятся на:

- системы без обучения;
- обучающиеся (ОСР) и самообучающиеся системы (ССР).

Сразу заметим, что многоуровневые сложные СР однозначно нельзя разделить на указанные классы, так как каждая из локальных СР, входящих в их состав, сама может представлять как систему без обучения, так и систему обучающуюся или самообучающуюся.

Системы без обучения. Для построения таких систем необходимо располагать полной первоначальной априорной информацией.

Обучающиеся системы (ОСР). Обучающиеся системы необходимы в ситуации, когда априорной информации не хватает для описания распознаваемых классов на языке признаков (возможны случаи, когда информации хватает, однако делать упомянутое описание нецелесообразно или трудно). На стадии формирования ОСР работают с «учителем», осуществляющим указание о принадлежности предъявленного для обучения объекта. И прежде, чем система будет применяться, должен пройти этап обучения.

Самообучающиеся системы (ССР). В отличие от систем без обучения и систем, обучающихся с учителем, для самообучающихся систем характерна недостаточность информации для формирования не только описаний классов, но даже алфавита классов. То есть, определен только словарь признаков распознавания. Однако для организации процесса обучения задается все-таки некоторый набор правил, в соответствии с которым система сама вырабатывает классификацию.

Для ССР также как для ОСР существует период обучения, характерно наличие периода самообучения, когда ей предъявляются объекты обучающей последовательности. Только при этом не указывается принадлежность их к какому-либо классам.

Завершая рассмотрение классификации СР по количеству первоначальной априорной информации, заметим, что СР, в которых недостаточно информации для назначения словаря признаков, не существует. Без этого не создается никакая система.

Четвертый принцип классификации (характер информации о признаках распознавания). С характеристикой информации о признаках распознавания мы уже имели дело при изучении задач создания СР. В соответствии с ее отличительными особенностями СР подразделяются на

- детерминированные;
- вероятностные;
- логические;
- структурные (лингвистические);
- комбинированные.

Отметим характерные особенности этих систем, а именно: метод решения задачи распознавания и метод априорного описания классов.

Детерминированные системы.

а) Метод решения задачи распознавания: использование геометрических мер близости;

б) Метод априорного описания классов: координаты векторов-эталонов по каждому из классов или координаты всех объектов, принадлежащих классам (наборы эталонов по каждому классу).

Вероятностные системы.

а) Метод решения задачи распознавания: вероятностный, основанный на вероятностной мере близости (средний риск);

б) Метод априорного описания классов: вероятностные зависимости между признаками и классами.

Логические системы.

а) Метод решения задачи распознавания: логический, основанный на дискретном анализе и исчислении высказываний;

б) Метод априорного описания классов: логические связи, выражаемые через систему булевых уравнений, где признаки — переменные, классы — неизвестные величины.

Структурные (лингвистические) системы.

а) Метод решения задачи распознавания: грамматический разбор предложения, описывающего объект на языке неприводимых структурных элементов с целью определения его правильности.

б) Метод априорного описания классов: подмножества предложений, описывающих объекты каждого класса.

Комбинированные системы.

а) Метод решения задачи распознавания: специальные методы вычисления оценок;

б) Метод априорного описания классов: табличный, предполагающий использование таблиц, содержащих классифицированные объекты и их признаки (детерминированные, вероятностные, логические).

После проведенной классификации обратим внимание на термин «достаточное» или «недостаточное» количество информации. С этой меркой мы подошли к разделению СР на три большие класса: СР без обучения, обучающиеся и самообучающиеся СР. То есть, для СР без обучения имели дело с полной информацией, а для ОСР — с неполной (нет описания классов на языке признаков), а для ССР — еще с большей неполнотой (отсутствует даже алфавит классов). Однако заметим, что само понятие «неполнота информации» — качественное, относительное. Для СР без обучения при прочих равных условиях этой информации просто больше. Это означает, что результативность СР при имеющемся объеме априорной информации значительно выше, чем в той ситуации, которая требует создания ОСР. О результативности СР, для которой невозможно априорно назначить алфавит классов, говорить вообще нельзя.

При построении как ОСР, ССР, так и просто СР необходимо всегда использовать принцип обратной связи для расширения объема информации. То есть, результаты решения задачи распознавания неизвестных объектов после апостериорного подтверждения правильности их классификации необходимо использовать для уточнения описания классов в простых СР без обучения и для дополнительного обучения в ОСР и ССР.

3.1.3. Цели и задачи систем распознавания образов

«Центральную задачу распознавания образов представляет построение на основе систематических теоретических и экспериментальных исследований эффективных вычислительных средств (объединяемых в понятия «системы распознавания») для отнесения описаний объектов, явлений, процессов к соответствующим классам» ([14], стр. 22).

Широкий круг задач, возлагаемых на такие системы, определяется самим понятием «распознавание» и включает выяснение по разнородной, часто неполной, нечеткой, искаженной и косвенной информации факта, обладают ли изучаемые объекты, явления, процессы, ситуации фиксированным конечным набором свойств, позволяющим отнести их к определенному классу.

К задачам распознавания относятся задачи технической и медицинской диагностики, геологического прогнозирования, прогнозирования свойств химических соединений, распознавания свойств динамических и статических объектов в сложной фоновой обстановке и при наличии активных и пассивных помех, прогнозирования урожая, обнаружения лесных пожаров, управления производственными процессами. Разработки систем распознавания, начатые с 50-х годов, исчисляются тысячами. Сегодня уже трудно назвать такую отрасль науки и сферы производства, где системы распознавания (СР) не используются.

Основные этапы процесса распознавания (подробнее поговорим об этом ниже):

- прием на входе СР образа распознаваемого объекта;

- сопоставление апостериорной информации поступившего объекта с имеющимся в СР априорным описанием классов всех объектов, подлежащих распознаванию (объектов, на которые рассчитана система);
- принятие решения об отнесении объекта, образ которого был принят, к одному из классов.

Правило, согласно которому объекту, образ которого принят, ставится в соответствие наименование класса, называется *решающим правилом*. «В литературе широко распространено мнение, что суть проблемы распознавания и состоит в определении такого решающего правила. То есть, центральной задачей часто считается нахождение в признаковом пространстве таких границ, которые некоторым оптимальным образом (например, по критерию минимума ошибок распознавания) разделяют это пространство на области, соответствующие классам» ([14], стр. 37).

При этом нами четко установлено, что в зависимости от объема априорной информации возможно два подхода к определению решающих правил (границ между классами в признаковом пространстве):

1. Непосредственное предварительное определение при достаточном количестве априорной информации (СР без обучения).
2. Постепенное уточнение в ходе работы СР по назначению при наборе достаточного количества информации (обучающиеся СР).

Каждый из подходов основан на том, что априорный словарь признаков и алфавит классов известны. При отсутствии априорного алфавита классов применяется подход, реализуемый в самообучающейся СР. Однако при этом заранее должны быть известны словарь признаков и, кроме того, набор некоторых правил назначения классов в процессе самообучения. Решающие правила здесь определяются как итог нахождения алфавита классов.

«Исторически сложилось, что первые теоретические и прикладные работы в области распознавания основывались на полной определенности алфавита классов и словаре признаков. При этом проблема распознавания сводилась обычно к проблеме оптимального в некотором смысле определения ре-

шающих правил, решающих границ между классами. Широкая практика создания СР в последующие годы (особенно в военных приложениях) и дальнейшее развитие теории распознавания показали, что приведенное отношение устарело. При построении реальных СР даже при известных признаках и классах приходится решать сложную и дорогостоящую задачу разработки, ввода и использования специальных измерительных средств и комплексов таких средств с ЭВМ. Эти средства и комплексы оказываются главным элементом в получении признаков распознавания» ([14], стр. 37).

В результате представляется возможным сформулировать назначение любой СР.

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

Но и принятие решений о принадлежности – не самоцель. Поэтому второй момент, на который нам необходимо обратить внимание – это необходимость понимания того, что любая СР является частью какой-либо системы управления (автоматической или автоматизированной). Отсюда легко формируется цель создания СР.

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

Из рассмотрения содержания задач построения СР следует, что словарь признаков распознавания и алфавит классов формируются человеком на основе его знаний, опыта, интуиции. В то же время ответственность за решения может быть очень велика. Поэтому очень важно найти объективные методы оценки такого рода человеческой деятельности.

3.1.4. Последовательность распознавания образов в общем виде

Если у нас имеется некоторая совокупность объектов или явлений, которые необходимо распознавать (классифицировать), последовательность решения соответствующих задач следующая:

- 1) в соответствии с выбранным принципом совокупность объектов или явлений подразделяется на ряд классов (говорят: назначается алфавит классов);
- 2) разрабатывается совокупность признаков (говорят: словарь);
- 3) на языке словаря признаков описывается каждый класс;
- 4) выбираются и (или) создаются средства определения признаков;
- 5) на вычислительных средствах реализуется алгоритм сопоставления апостериорных и априорных данных и принимается решение о результатах распознавания.

3.1.5. Определение полного перечня признаков, характеризующих объекты

В решении этой задачи — главное найти все признаки, характеризующие существо распознаваемых объектов (явлений). Любые ограничения, любая неполнота приводят к ошибкам или полной невозможности правильной классификации объектов (явлений).

Реально даже целая группа признаков может оказаться неэффективной. Поэтому для решения задачи создания СР необходимо найти все возможные признаки, описывающие объекты распознавания, с тем, чтобы при оценке эффективности решений системы не возвращаться к этой задаче, обнаружив ограниченность выбранных признаков на последующих этапах разработки.

Но чтобы назначать признаки распознавания, необходимо, во-первых, понять, что не существует способов их автоматической генерации. На сегодня это под силу только человеку. Поэтому говорят, что **выбор признаков — эвристическая операция**. Во-вторых, выбор признаков можно осуществлять, имея представление об их общих свойствах. Признаки могут подразделяться на:

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

Детерминированные признаки – это такие характеристики объектов или явлений, которые имеют конкретные и постоянные числовые значения.

Необходимо иметь в виду, что в задачах распознавания с детерминированными признаками ошибки измерения этих признаков не играют никакой роли, если, например, точность измерений признака, значительно выше, чем различие этого признака у разных классов. Распознавание осуществляется путем сравнения полученных значений признаков с имеющимися в базе данных характеристиками.

Вероятностные признаки – это характеристики объекта (явления), носящие случайный характер. С такими признаками в основном и имеют дело в природе и технике. Отличаются эти признаки тем, что в силу случайности соответствующей величины признак одного класса может принимать значения из области значений других классов, каждый из которых подлежит распознаванию в системе. Для того, чтобы можно было в условиях случайности говорить о возможности распознавания, следует потребовать, чтобы вероятности наблюдения значений признака в своем классе были как можно больше, чем в чужих. В противном случае данный признак не позволит построить СР, использующую описание классов на его основе. Эффективность его недостаточна для достоверного решения, и необходимо искать другие признаки, имеющие большую разделительную способность. Вспомним из теории вероятностей, чем характеризуется случайная величина – законом распределения вероятностей. То есть, точно так же законом распределения должен характеризоваться каждый вероятностный признак. В качестве законов распределения вероятностей выступают интегральная функция $F(x)$ – интегральный закон или плотность распределения вероятностей (ПРВ) – дифференциальный закон $f(x)$.

Логические признаки распознавания – это характеристики объекта или явления, представленные в виде элементарных высказываний об истинности («да», «нет» или «истина», «ложь»). Эти признаки, как мы понимаем, не имеют количественного выражения, то есть являются качественными суждениями о наличии, либо об отсутствии некоторых свойств или составляющих у объектов или явлений. К логическим можно отнести также такие признаки, у которых ис

важна величина, а лишь факт попадания или непадания ее в заданный интервал.

Структурные признаки — производные (то есть, элементарные, не производимые из других элементарных признаков) элементы символов. Появление структурных признаков обязано возникновению проблемы распознавания изображений с ее специфическими особенностями и трудностями.

3.2. РАСПОЗНАВАНИЕ ОБРАЗОВ С ИНФОРМАЦИОННОЙ ТОЧКИ ЗРЕНИЯ

Очевидно, что процесс распознавания связан со сбором, обработкой и, самое главное, отображением информации. Для более полного понимания процесса распознавания образов, представляется целесообразным обсудить два вопроса: 1) содержание информации в музыкальном сигнале (как наиболее интересующем нас) и 2) преобразования информации, происходящие при распознавании.

3.2.1. Информативность музыкальных произведений

При обсуждении информации, которую несёт звукозапись музыкального произведения, необходимо уточнить некоторые понятия и дать необходимые определения. При этом используем определения, приведённые французским учёным Абрахамом Модем [13].

«...сообщение — последовательность элементов набора, несущих информацию, пропорциональную оригинальности сообщения, т.е. его непредвиденности ... по сравнению с максимально возможным значением оригинальности, достигаемым в том случае, когда все элементы набора равновероятны» ([13], стр.195).

«... всякий раз, когда мы говорим об информации, содержащейся в рукописном или печатном тексте, мы подразумеваем информацию, которую извлекает средний «читатель», читающий слова и связывающий их друг с другом в зависимости от общего запаса знаний, который мы предполагаем у людей, принадлежащих к определённой социальной группе со статистически примерно одинаковым культурным уровнем.» ([13], стр.197). Очевидно, что один и

тот же текст несёт разное количество информации для людей, говорящих на разных языках, умеющих или не умеющих читать.

«... в случае художественных сообщений, обращённых к целой общественной группе, у каждого принимающего индивидуума имеется своя собственная личная «таблица» знаний (социально-культурная таблица), определяющая, какую информацию он извлекает из сообщений, полученных из внешнего мира или от других людей. В предельном случае, когда приёмник обладает полным знанием всех свойств сообщения, которое ему будет передано, ... полученная информация равна нулю, а её избыточность равна 100%, т.е. сообщение не представляет интереса, оно банально» ([13], стр.197).

Необходимость разделения двух видов информации возникает в связи с потребностью объяснить сохранение оригинальности сообщения, смысловая информативность которого уже исчерпана. Исполнитель, знающий наизусть музыкальное произведение, всё равно слушает его в чужом исполнении. Понравившийся нам спектакль мы можем ходить смотреть даже тогда, когда наизусть знаем все реплики актёров. Что же заставляет нас вновь и вновь возвращаться к уже неоднократно прочитанным литературным произведениям, просмотренным кинофильмам? Иными словами, что сохраняется оригинального в сообщении, смысловая информация которого известна приёмнику априори? Очевидно, что человек ищет в сообщении не только смысловую, но и эстетическую оригинальность.

«... для заданной совокупности сообщений существуют две точки зрения на сообщение, соответствующие двум типам информации:

семантическая точка зрения, в соответствии с которой в сообщении выделяется логическая информация, связанная со структурой, поддающаяся точной формулировке, переводимая, вызывающая определённые действия, и

эстетическая точка зрения, согласно которой выделяется неперевоаемая информация, вызывающая определённые состояния.»([13], стр.200).

«Итак, ... множеству сообщений соответствуют два типа информации. Эти два типа связаны с теми способами, которыми наблюдатель ... группирует

последовательные элементы из последовательности, составляющей сообщение, относя их к тому или иному набору:

а) семантическая информация, подчиняющаяся универсальной логике, имеющая структуру, допускающая точное представление, переводимая на другие языки; согласно концепции бихевиористов, она предполагает действия;

б) эстетическая информация, «непереводимая», относящаяся не к универсальному набору символов, а только к набору знаний, общих для приёмника и передатчика;» ([13], стр.202-203).

Теперь важно выяснить, каково содержание эстетической и семантической информации в музыкальных сообщениях. С одной стороны, музыкальное произведение можно записать с помощью набора определённых символов, и набор этот не единственен: нотная запись возникла очень давно и не всегда имела современную общепринятую форму. С другой стороны, мы понимаем, какое ничтожное представление о музыкальном произведении можно получить, изучив его ноты, даже профессиональному музыканту, представляющему, как оно будет звучать. Музыкальное произведение нужно услышать, но, кроме того, нужно понимать язык образов создателя пьесы, его эмоциональное состояние и, может быть, его менталитет. Ведь среднестатистический европеец вряд ли будет наслаждаться индийской или африканской музыкой, также как индеец вряд ли станет слушать Баха.

Как было сказано выше, объектом нашего внимания является семантическая информация, заключённая в музыкальном произведении, её выделение. Поэтому, в дальнейшем разговор будет идти о символьном описании музыкальных произведений. Итак, содержание семантической информации в музыке мало, и она частично описывается с помощью нотной записи. Любому исполнителю очевидно несовершенство нотной записи. Ноты ни в коем случае не отражают всех нюансов исполнения произведения. Но в этом, по мнению автора, и их достоинство — для исполнителя всегда остаётся свобода творческих действий. Своим исполнением он может, как украсить, так и обезличить музыкальное произведение. Если бы композитор в точности описывал исполнителю, что тот должен сыграть, нотная запись была бы невероятно загромождена, ибо

язык эмоций, как мы говорили — язык непереводаемый, не отражаемый символами. Исполнитель или понимает, что хотел выразить композитор, или не понимает. Или понимает это по-своему.

Из сказанного следует очень важный для нас вывод:

- нотная запись представляет лишь схему музыкального произведения, но не точное руководство к действиям для исполнителя;
- ни в коем случае нельзя ставить знак равенства между тем, что исполнитель сыграл и нотами, по которым он это сыграл; «Музыкальное произведение, как правило, является коллективным произведением, оно имеет по меньшей мере двух авторов, почти в равной степени «участствующих» в его создании: композитора, создающего схему действий, и исполнителя, «реализующего» ... эту схему.» ([13], стр.213).

На самом деле, исполнителю всегда предоставлено право интерпретации музыкального произведения, начиная с транспонирования мелодии, т.е. повышения или понижения высоты тона (изменения тональности), до изменения ритмического рисунка и всяческих украшений мелодии (мелизмов). Более того, если бы исполнитель в точности следовал нотной записи, произведение могло бы частично потерять свою эстетическую ценность: чёткий выверенный ритм, точное поддержание громкости нот вряд ли усилят впечатление слушателя.

«Музыкальный сигнал, музыкальная материя богаче схемы, представленной партитурой, благодаря многочисленным «степеням свободы»:

Во-первых, из-за несовершенства самой системы записи, используемой в партитуре;...

Во-вторых, варианты оркестровки могут целиком изменить звуковую материю, поскольку партии можно перераспределить между инструментами, не изменив партитуры. ...

В-третьих, разнообразие конструкции инструментов: не существует двух одинаковых фортепьяно, двух одинаковых скрипок. ...

В-четвёртых, отступления или вольности интерпретации (исполнения), допускаемые как инструментом (оркестром), так и исполнителем (дирижёром).» ([13], стр. 212-213).

Существует и ещё одно важное для нас обстоятельство: нотная запись никак не описывает тембра звучания. Это означает, что одной и той же ноте, записанной, например, композитором в партитуре, соответствует множество «звуковых реализаций», сыгранных на разных инструментах, разными исполнителями и разными приёмами. То есть мы обязаны изучить свойства звуковых колебаний, создаваемых различными инструментами, обобщить эти сведения и учесть их при построении системы распознавания музыкальных образов.

3.2.2. Преобразование информации при распознавании музыкального сигнала

Теперь поговорим о том, что мы будем понимать под образами. В данном случае имеются в виду не те образы, которые возникают в сознании человека при прослушивании им музыкального произведения. Речь идёт о звуковых объектах, составляющих музыкальное произведение, по аналогии со зрительными образами, составляющими изображение текста или графики. При оптическом распознавании текста мы преобразуем изображение, несущее смысловую информацию, заключённую в этом тексте, в форму, позволяющую работать с текстом именно как с текстом, а не как с изображением.

При распознавании речи мы выделяем из звукового сигнала смысловую часть информации, которую он несёт, эстетическая же часть при этом теряется — задача определения интонации, с которой произнесена фраза, не ставится. Напрашивается аналогия процесса передачи смысловой и эстетической информации с процессом передачи информационных сигналов с помощью сигналов-переносчиков в системах телекоммуникаций. Только в нашем случае передаётся сама смысловая или эстетическая информация, источником которой является человеческий мозг, а переносчиком являются или звуковые колебания, то есть речь, или оптическое изображение. При этом язык, на котором говорит человек, определяет способ модуляции. Распознавание же составляет обратный процесс — демодуляцию, детектирование. В музыке по аналогии с речью мож-

но выделить смысловую и эстетическую информации. Смысловая часть записывается в виде нот — общепризнанных символов. Эстетическая же информация, закладываемая в музыку, частично отображается в нотной записи с помощью специальных символов (обозначений громкости, знаков крещендо и т.п.). Таким образом, распознавание музыкальных образов (музыкальных сигналов) представляет собой выделение из сигнала-переносчика, то есть звукового сигнала музыкального произведения, смысловой информации, заключённой в нём и представление её в виде нотной записи или виде, удобном для преобразования в нотную запись.

Итак, при распознавании образов происходит выделение (своего рода демодуляция) смысловой информации из переносчика — физического объекта (например, поверхностное распределение яркости — в случае распознавания изображения) или процесса (например, звуковых колебаний, создаваемых музыкальным инструментом). Смысловая информация описывается символами. Под символом понимается нечто, существующее само по себе и воспринимаемое отдельно от других символов, некая самостоятельная единица. И мы можем смело утверждать, что *символы*, существующие отдельно друг от друга, и *переносятся отдельно*, т.е. в процессе-переносчике информации можно выделить части, соответствующие отдельным символам. Например, печатный текст мы воспринимаем как совокупность расположенных отдельно букв, и мы можем выделить части поверхности листа, заключающие каждый отдельный символ. Каждая буква передаётся отдельным символом. Отдельная нота переносится отдельным же звуковым объектом. Естественно, возможны варианты перекрытия символов: в рукописном тексте буквы могут «налезать» друг на друга, музыкальные произведения в большинстве своём многоголосны. Такое перекрытие усложняет процесс распознавания, ибо сложнее отнести элементы перекрывающихся символов к какому-либо из них.

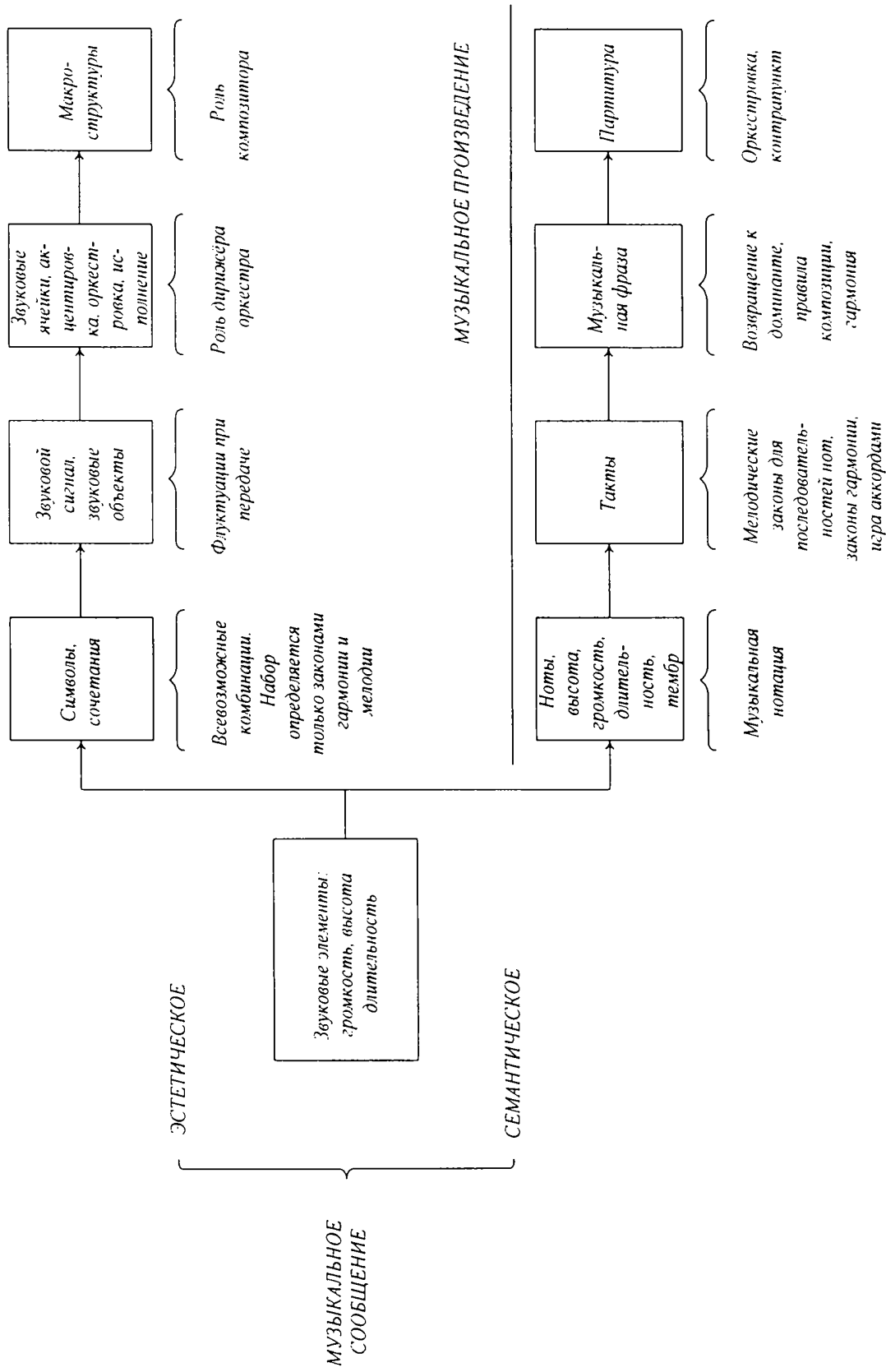


Рис. 3 Информативность музыкальных сообщений ([13] стр. 220)

Таким образом, распознавание становится возможным в случае, если возможно разделение отдельных символов в процессе-переносчике информации. А разделение символов возможно благодаря различию их свойств, которые мы наблюдаем в процессе распознавания. Т.е. распознавание происходит на основе совокупности поступающих наблюдений свойств, признаков объектов.

3.3. РАСПОЗНАВАНИЕ ЗВУКОВЫХ ОБРАЗОВ

Задача автоматического распознавания звуковых образов сопряжена с одной из самых увлекательных и важных проблем новой техники – с осуществлением устной речевой связи человека с машиной. Однако теория автоматического распознавания звуковых сигналов разработана недостаточно. Взамен ее обычно применяются методы распознавания зрительных образов (например, [8]), но такой перенос не всегда обоснован. Кроме того, проблема автоматического распознавания звуковых образов оказалась замкнутой на распознавание звуков речи: другие приложения развивались мало.

«Нельзя ждать, пока проблема автоматического распознавания речи получит полное решение: необходимо уже достигнутое применять в других задачах» ([14], стр. 8).

Прежде, чем приступать к распознаванию, как мы уже говорили, необходимо составить алфавит распознаваемых символов и определить для каждого символа его специфические свойства, выражаемые через признаки, общие для всего алфавита символов, но проявляющихся по-разному у каждого из символов. В теории распознавания образов этот процесс называется *формированием признакового пространства*.

При формировании признаков мы должны определить количество признаков, с помощью которых можно описать каждый символ и определить значимость каждого из признаков. Увеличение количества признаков увеличивает избыточность описания отдельного символа и повышает вероятность правильного опознания. Но в то же время, увеличивается время сбора и обработки информации об отдельном фрагменте процесса-переносчика. Таким образом,

очевидна важность правильного формирования признакового пространства: количество признаков необходимо минимизировать, но оно должно быть достаточным для распознавания с требуемой достоверностью, т.е. признаки должны быть выбраны таким образом, чтобы они несли как можно больше информации и свойства символа. Естественно, признаки должны характеризовать различные, не перекрывающиеся взаимно свойства символа.

Какие же основные признаки могут быть взяты для целей *автоматического* распознавания акустических сигналов? Этот вопрос требует некоторого разграничения задач распознавания и *автоматического* распознавания. При распознавании, в конечном счете, участвует сам человек; так, сонограммы речи могут быть узнаны при просматривании их человеком: он может судить, какому звуку речи соответствует представленная сонограмма. В процессе же автоматического (объективного) распознавания акустического сигнала все поручается автоматам, и решение должно выдаваться независимо от наблюдателя — человека. Системы признаков, по которым производится распознавание, могут быть в этих двух случаях различными.

В какой мере вычислительные машины могут заменить для объективного распознавания функции коры головного мозга? Известны ли те признаки, которые наиболее существенны для субъективного распознавания, известен ли механизм такого распознавания, и можно ли его моделировать? К сожалению, в настоящее время на оба эти вопроса еще нельзя дать удовлетворительного ответа. Процесс опознания слуховых образов до сих пор недостаточно изучен физиологами. Можно сделать вывод, что вопросы моделирования слухового восприятия и опознания образов человеком (т. е. субъективно) до настоящего времени еще не получили достаточного развития.

«В задачах автоматического распознавания сложных акустических сигналов следует ... ограничиться изучением чисто физических признаков акустического образа и связи признаков с системой, излучающей звук» ([14], стр. 58).

Укажем 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.

Необходимо отметить, что именно перечисленные признаки или признаки, являющиеся производными от перечисленных, составляют признаковое пространство современных систем распознавания. Разработано большое количество алгоритмов для выделения этих признаков, системы решающих правил на основе перечисленных признаков. Подробнее о признаках используемых при распознавании музыкальных сигналов – ниже. Опыт построения систем распознавания звуковых образов позволяет выделить особенности распознавания музыкальных и речевых сигналов. В распознавании речевых сигналов наибольшие трудности вызывает *определение набора* наиболее информативных *признаков* (наборы эти различны, например, в случаях узнавания диктора и распознавания слитной речи). В задачах распознавания музыкальных сигналов набор наиболее информативных параметров известен, существуют большие сложности с *выделением* параметров реальных сигналов (например, определение количества одновременно звучащих нот и высот их тонов).

Важно иметь в виду, что увеличение количества признаков ведёт к усложнению функции решения, увеличению количества правил вынесения решения. Поэтому процесс обучения должен, кроме всего прочего, выявить несо-

вершенство признакового пространства, т.е. показать, какие признаки являются избыточными, не востребуемыми, а каких признаков не хватает.

Нет никакого сомнения в том, что акустическое поле должно в необходимой мере отражать особенности процессов в системе. Так, например, при артикуляции голосовым органом звука *a* создается поле, которое должно обладать всеми особенностями, необходимыми для распознавания *a* человеком. В противном случае было бы немыслимо распознавание фонем по звуку.

Однако вопрос состоит в том, содержит ли это звуковое поле все особенности, необходимые (с учетом современного уровня техники) для автоматического распознавания фонемы? Ведь может оказаться, что это звуковое поле несет некоторые коды, действующие в процессе слухового восприятия по неизвестным еще науке процедурам в нейронных сетях, и поэтому не имеется еще возможности осуществлять автоматическое распознавание. Иными словами, не «зашифрованы» ли слуховые образы так, что они недоступны для опознания машиной?

Вопрос является достаточно серьезным, и только практика может его осветить. Однако та же практика показывает, что набор слов (пусть небольшой) может быть распознан автоматами, равно как и отдельные фонемы (в особенности гласные), одногласные и несложные многогласные мелодии.

Выскажем следующее предположение: **основные признаки, необходимые для автоматического распознавания образа, содержатся в самом сигнале.**

Этот принцип при всей его схематичности является некоторым основанием физической теории распознавания образов. Но нельзя, конечно, преуменьшать значение исследования слухового восприятия акустических образов, которые могут внести соответствующие коррективы в физическую теорию.

При решении задачи распознавания приходится решать и ещё одну задачу, не имеющую отношения собственно к распознаванию образов. На практике поток данных наблюдений поступает неразделённым на фрагменты, соответствующие отдельным символам, объектам. Поток этот, прежде всего, не-

обходимо разделить. Печатный текст необходимо разделить на фрагменты, соответствующие отдельному знаку (букве или знаку препинания), речевой сигнал необходимо разделить на сегменты, соответствующие отдельным фонемам или слогам. Задача эта отнюдь не проста. Она легко решается человеком, но машина не обладает способностями и опытом человека. Поэтому необходима также разработка алгоритма *сегментации* (фрагментации) входного потока данных (наблюдений). Сегментация должна проводиться на основе составленных нами эталонных описаний распознаваемых символов. В процессе сегментации мы должны разделить входной поток на две части: ту, которую мы сможем распознать, т.е. которая имеет свойства, общие с распознаваемыми символами, и часть, которую мы не сможем опознать, — своеобразный шум. При распознавании зрительных образов желательно отчистить распознаваемое изображение от «грязи» — ненужных чёрточек, точек и т.п. Из речевого сигнала желательно устранить фрагменты, соответствующие паузам или придыханию.

3.4. СТАТИСТИЧЕСКИЙ ПОДХОД К РАСПОЗНАВАНИЮ ОБРАЗОВ

3.4.1. Реализации и признаки

Сформулируем задачу распознавания акустического образа с позиций статистического подхода. Как было установлено, образ (иногда его называют классом объектов) выражается как собственная функция, описывающая этот объект.

Можно построить некоторый алфавит акустических образов, содержащий множество собственных функций:

$$\Psi = \{\psi_1, \psi_2, \dots, \psi_m\} \quad (1)$$

Здесь ψ_i — один из элементов множества Ψ , т.е. некоторый определенный образ; m — число образов во взятом алфавите.

Образ имеет не одно изображение, а целый ряд изображений. Можно считать, что каждый образ имеет некоторое число реализаций, которые для каждого образа образуют некоторое множество. Обозначим его буквой

$$A = \{a_1, a_2, \dots, a_r\} \quad (2)$$

a_j представляет отдельную реализацию, T – число различных реализаций. Обычно $T \gg m$, т. е. число реализаций гораздо больше числа образов или классов. Приведем для пояснения несколько примеров.

Пусть множество классов составляют русские гласные фонемы. Число гласных b : $a, o, y, u, ы, э$. Каждая из гласных, например гласная o , включает множество реализаций: ведь эта фонема может быть произнесена на множество ладов (и записана множеством почерков). Множество произносимых звуков o и есть множество реализаций этого образа или этого класса. Для другого образа, например образа u , будет другое множество реализаций. Справедливость приведенного неравенства очевидна: число реализаций заведомо больше числа классов.

Если образ выражается собственной функцией, то каждая реализация может быть изображена некоторой пси-функцией. Отсюда благодаря статистическому смыслу волновой функции следует, что всякая задача распознавания является статистической.

Она ставится так: дана некоторая пси-функция $\Psi(\lambda)$; спрашивается, реализацией какого образа она является? Этот вопрос должен решаться по схеме процедуры распознавания. Чаще всего решение дает лишь вероятность того, что сигнал относится к какому-то классу.

Множество реализаций данного образа образуется по эталону. За эталон принимается реализация, в каком-то смысле наиболее близкая образу. Исходя из эталона, строится совокупность реализаций. Иногда имеется набор масок, соответствующих разным образам. Эти маски принимаются за эталонные. Практически исключен случай, чтобы сигнал в точности соответствовал эталону, т. е. данная реализация совпала с эталонной. Однако представляется возможным путем наложения масок и оценки близости отнести реализацию к определенному классу, образу если не с достоверностью, то с достаточной вероятностью. Математический смысл этого утверждения основан на статистическом характере пси-функции, квадрат модуля которой выражает плотность вероятности иметь те или иные значения признаков. Признаки составляют опре-

деленный набор, описывающий динамику объекта. Совокупность признаков составляет некоторый алфавит:

$$\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_N\} \quad (3)$$

Здесь α_k – некоторый конкретный признак, входящий в этот алфавит; N – число признаков.

Каждый признак, т. е. каждое собственное значение α_k , может иметь целый ряд градаций. Например, для признака α_k можно построить множество градаций: $\alpha_1^k, \alpha_2^k, \dots, \alpha_p^k$ (p – число градаций признака α_k).

Теперь можно привести в соответствие признаки и реализации. Каждой реализации a_j соответствует совокупность признаков:

$$\alpha_j = \{\alpha_{j1}, \alpha_{j2}, \dots, \alpha_{jN}\} \quad (4)$$

Обратимся к геометрической интерпретации. Реализация образа математически выражается через пси-функцию, которая есть вектор в пространстве Гильберта. Множество реализаций – это также множество векторов в пространстве признаков. Указанные векторы получаются, исходя из некоторого базиса, т. е. некоторого набора коммутирующих операторов, и системы собственных векторов этих операторов, которые и образуют базисную систему. Собственные значения операторов являются аргументами пси-функции. Образ соответствует тому вектору или той собственной функции, которые принадлежат взятому набору операторов.

Это геометрическое представление позволяет сформулировать основную задачу распознавания. Задан какой-то вектор в пространстве Гильберта; по нему нужно определить, какому собственному вектору он соответствует или к какому собственному вектору он наиболее близок.

В большинстве случаев признаки, которые содержатся в распознаваемой реализации, могут быть извлечены из сигнала. Так, например, частотные и амплитудные признаки могут быть без труда извлечены из заданной сонограммы. В других случаях определение более сложных признаков, например формантных признаков гласных или шумовых признаков согласных, требуют

выполнения сложной программы анализа. Во всяком случае, после первого этапа процедуры распознавания получается некоторая последовательность признаков, некоторое слово, состоящее из цепочки признаков. Такое слово можно представить как вектор в многомерном пространстве признаков. Мы встречаемся здесь с наиболее распространенной геометрической интерпретацией реализации, которая использует многомерное пространство, измерения которого суть признаки. Для случая N-мерного пространства признаков существует координатная система, на N осях которой откладываются значения этих признаков. Признаки, извлеченные из данного сигнала, определяют координаты

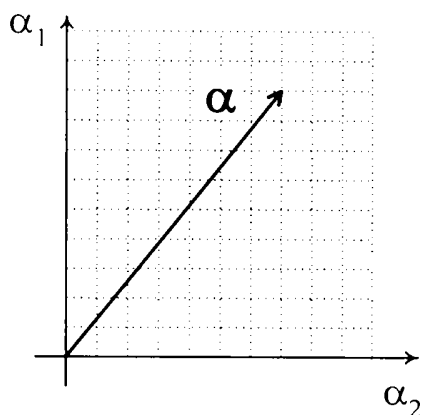


Рис. 4 Область двумерного пространства признаков некоторой точки в этом N-мерном пространстве. На рис. 4 для наглядности изображено двумерное пространство признаков и построена точка, координатами которой служат признаки, извлеченные из сигнала. Из начала координат проведем вектор, конец которого совпадает с этой точкой. Положение любой точки может быть фиксировано таким вектором.

Обозначим общее число признаков через N, а число градаций признаков через P; считая, что каждый признак имеет одинаковое число градаций, получим общее число реализаций признаков.

$$T = P^N \quad (5)$$

На рис. 4 приведено двумерное пространство признаков. Один признак обозначается α_1 и нанесён по оси абсцисс; пускай число его градаций будет десять. Другой признак, α_2 , отложен по оси ординат и имеет такое же число градаций. Налицо дискретная совокупность признаков, которая дает некоторую сетку на плоскости; узлы соответствуют реализациям. Для каждой реализации

признаков может быть построен вектор реализации. Мы видим, что каждой реализации, поскольку $N = 2$, соответствуют два признака.

Число точек в пространстве признаков может быть достаточно велико, и их плотность (число точек в единице объема) достаточно велика, чтобы можно было перейти к пределу и считать распределение признаков непрерывным.

3.4.2. Процедуры принятия решения

Если получен набор апостериорных вероятностей того, что распознаваемая реализация соответствует тому или иному акустическому образу, то требуется принять решение, т. е. сделать заключение, к какому образу относится данная реализация. Необходимо иметь в виду и тот случай, когда реализация не относится ни к одному из известных классов. Этот случай называется нулевым и такой класс называется нулевым.

Решающие процедуры выражаются определенными схемами алгоритмов. Эти алгоритмы или правила позволяют определять, к какому классу или образу относится данная реализация. В ряде случаев весь алфавит реализаций, относящихся ко всему множеству классов, можно разбить на подмножества, каждое из которых относится к определенному классу.

Перейдем к вероятностям значений признаков. Построим (рис. 5) на плоскости координатную систему двух каких-либо признаков α_1 и α_2 . Разумеется, двумерное изображение системы признаков берется только для простоты. На плоскости изображены контуры (сплошной линией), соответствующие различным классам или образам. Внутри каждого из изображенных контуров точки дают значения признаков, относящиеся к данному классу. Эти области предполагаются непересекающимися. Для каждой области показаны также пунктиром расширенные границы. Внутри пунктирного контура содержатся точки, координа-

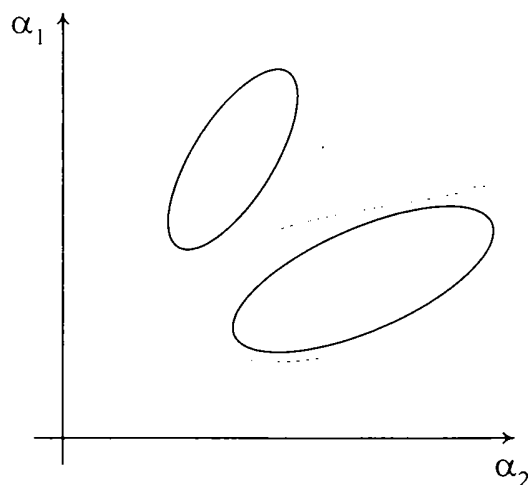


Рис. 5 Плоскость признаков

тами которых служат признаки, соответствующие реализации. Области реализаций превосходят области классов (образов), которые в них включены. Мы уже говорили о том, что число элементов реализации больше числа элементов образов. Реализации перекрывают всю плоскость.

На границе между областями вероятности реализации должны быть равны нулю. Точки на границе не должны относиться ни к той ни к другой области, а на самом деле относятся к обеим смежным, поэтому вероятность $P(a_j)$ равна нулю. В идеальном случае разделения признаков по областям вероятность ошибки распознавания становится сколь угодно малой, и сама процедура распознавания является тривиальной: по координатам признаков прямо находят область. Иными словами, строя вектор, составляющими которого служат заданные признаки сигнала, находят образ в соответствии с тем, куда придется конец вектора. Однако в действительности практически во всех задачах распознавания области пересекаются, и потому процедура принятия решения усложняется. Возникают и ошибки распознавания. Тогда решающими могут быть различные процедуры, между которыми иногда трудно сделать обоснованный выбор. Наиболее часто применяемыми критериями являются следующие:

- 1) минимизация средней вероятности ошибки распознавания;
- 2) минимизация вероятности ошибки распознавания каждого образа;
- 3) минимизация среднего риска распознавания;
- 4) минимизация риска распознавания каждого образа; и некоторые другие, в том числе комбинации из указанных.

Для простоты возьмем такой случай, когда имеется только один признак, так как задача получается одномерной, а заданы всего два класса. Обозначим признак через α , а классы — через ψ_1 и ψ_2 . Для каждого класса может быть указано распределение вероятности признака. Обозначим функцию распределения вероятностей признаков первого и второго образов через $f_1(\alpha)$ и $f_2(\alpha)$.

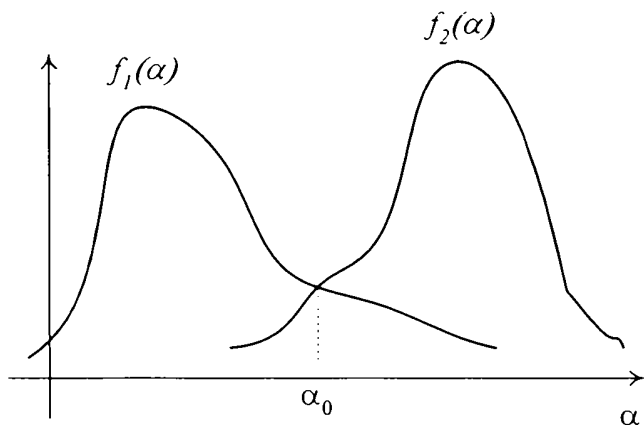


Рис. 6 Разделение классов

Знание распределений предполагается априорным, т. е. заданным перед выполнением процедуры принятия решения. На рисунке 6 изображены эти распределения вероятности. Пунктирная вертикаль должна быть построена так, чтобы разделить классы один и два, что означает решение задачи опознания образов ψ_1 и ψ_2 если это действительно граница, которая приходится на ординату α_0 , то значение признака $\alpha < \alpha_0$ дает класс ψ_1 , а значение признака $\alpha > \alpha_0$ даёт класс ψ_2 . Задача состоит в том, чтобы найти эту границу при условии пересечения кривых распределения вероятностей. Вот здесь могут быть применены указанные выше критерии.

Если принять разграничение классов таким, как показано на рисунке 6 с границей α_0 , то вероятности ошибок могут быть описаны следующим образом. Решение в пользу образа ψ_1 для $\alpha < \alpha_0$ связано с ошибкой, вероятность которой

$$p_1(e) = \int_{\alpha_0}^{\infty} f_1(\alpha) d\alpha \quad (1)$$

Решение в пользу образа ψ_2 для значений признака $\alpha > \alpha_0$ связано с ошибкой, вероятность которой

$$p_2(e) = \int_{-\infty}^{\alpha_0} f_2(\alpha) d\alpha \quad (2)$$

Следуя критерию В. А. Котельникова, можно потребовать, чтобы для оптимальных условий соблюдался минимум вероятности суммарной ошибки

$$p(e) = P(\psi_1) \cdot p_1(e) + P(\psi_2) \cdot p_2(e) \quad (3)$$

3.5. СИСТЕМЫ РАСПОЗНАВАНИЯ МУЗЫКАЛЬНЫХ СИГНАЛОВ

3.5.1. Развитие систем распознавания музыкальных сигналов

Сложно сказать, когда именно началась история распознавания музыкальных сигналов. Идея, очевидно, зародилась вместе с идеей «разумной» машины, выполняющей команды человека — работа. Если кто-то хотел, чтобы робот подносил еду и убирал помещение, то почему бы композитору или исполнителю не мечтать об автоматическом аккомпаниаторе? Действительно человек давно мечтал научить машину говорить, понимать речь, сочинять и исполнять музыку. Однако осуществляться эти мечты стали лишь с развитием теории гармонического анализа, появлением акустикоэлектрических преобразователей (микрофонов) и звукозаписи. Создание первых вокодеров и первые попытки распознавания человеческой речи относятся к 40-вым годам XX столетия. На распознавание (в отличие от синтеза) музыкальных сигналов внимание инженеров и исследователей обратилось существенно позже — в середине 70-х годов. Даже сегодня можно утверждать, что распознаванию речевых сигналов уделяется гораздо больше внимания и средств. Следует признать, что и успехов распознавание речи добились более значительных. Основные причины такого положения дел, по мнению автора, две:

1) распознавание речи было и остаётся востребованным промышленностью, в том числе оборонной (голосовое управление автоматами), в результате чего распознавание музыкальных сигналов, целесообразность которого не очевидна, до последнего времени не финансировалось на государственном уровне;

2) современные успехи теории распознавания музыкальных сигналов не убедительны — на сегодняшний день не создано ни одной коммерчески используемой системы распознавания многоголосных мелодий.

Кроме того, принципы и математический аппарат, использованные при построении систем речевого распознавания оказались неприменимы при анализе музыкальных сигналов, прежде всего, полифонических (особенности музыкальных сигналов обсудим ниже). Поэтому исследователям пришлось разра-

батывать новые методы анализа для музыкальных сигналов и глубже изучать психофизиологию восприятия звука человеком.

Первая известная система распознавания многоголосных мелодий была разработана сотрудником Стэнфордского университета Джеймсом Андерсеном Мурером (James Anderson Moorer) в 1977 году. Система была описана в статье под названием «О распознавании музыкальных звуков с помощью компьютера» (On the transcription of musical sound by computer) и стала темой для докторской диссертации Мурера.

Необходимо особо отметить две особенности. Во-первых, система, созданная Мурером распознавала *многоголосные* мелодии. Задача распознавания одноголосных мелодий (в зарубежной литературе, также называемая «fundamental frequency tracking» - выделение мелодии основного тона) к концу 1970-х годов не представляла особого интереса. Во-вторых, распознавание производилось с помощью компьютера, т.е. звуковой сигнал был представлен в **цифровой форме**. Действительно, попыток распознавания аналогового музыкального сигнала не известно. Дело в том, что современные представления о решении задачи требуют выполнения операций, которые с многих точек зрения (экономической, реализуемости, массы и габаритов) целесообразней производить именно в цифровом виде. Возможно, начало развития теории распознавания музыкальных сигналов было связано именно с бурным развитием теории цифрового анализа и обработки сигналов в конце 1960-х годов.

Система Мурера имела весьма ограниченные возможности:

- количество голосов мелодии было ограничено двумя,
- набор инструментов, звуки которых подлежали распознаванию, состоял из скрипки и гитары,
- диапазон выделяемых высот тонов составлял 2 октавы.

Кроме того, имелись жёсткие ограничения на высотный интервал между двумя одновременно звучащими нотами. Построение алгоритма не имело какой-либо теоретической основы.

Группа исследователей Стэнфордского университета, в которой работал Мурер, под руководством Криса Чейфа (Chris Chafe) продолжила начатые исследования и в 1982 опубликовала результаты своей работы. Результаты были менее убедительны, чем результаты Мурера: распознавались только звуки фортепиано, диапазон высот тонов сокращён до 19 полутонов (1,5 октавы).

Следующая попытка создания системы распознавания многоголосных мелодий, результаты которой были опубликованы, была предпринята в 1989 году Робертом Кроуфордом Махером (Robert Crawford Maher), сотрудником Иллинойского университета. Полифония по-прежнему ограничивалась двумя голосами. Сохранились жёсткие ограничения на диапазон распознаваемых высот тонов. Перечень инструментов, чьи звуки подлежали распознаванию, был расширен: кларнет, фагот, труба, туба.

Первой системой, чей алгоритм имел теоретическую основу, стала система, разработанная в 1994 году сотрудником университета Дюрхэма (Великобритания) Дугласа Нанна (Douglas Nunn). При построении алгоритма распознавания были учтены особенности слухового восприятия человека. Такой подход позволил расширить диапазон опознаваемых высот тонов до 48 полутонов (4 октавы), а количество голосов мелодии — до 8. Однако распознаванию подлежали звуки электронного органа. Достоверность распознавания низкая.

Настоящим прорывом в теории распознавания музыкальных сигналов стала система исследователей Токийского университета под руководством Куньо Кашино (Kunio Kashino) (1993 год). В системе были применены совершенно новые подходы. Во-первых, были чётко сформулированы, формализованы и реализованы алгоритмы разделения звуковых источников человеком, а точнее процессы разделения и слияния спектральных компонент одновременно звучащих источников. Кроме того, они использовали «тембральные модели» (tone models) звуков, т.е. при обработке звукового сигнала использовалась заранее заложенная информация о спектральной структуре звуков распознаваемого музыкального инструмента. Основой системы распознавания стала так называемая «блэкборд» (blackboard — дословно: классная доска) архитектура (о блэкборд архитектуре см. ниже), позволяющая при вынесении решения использо-

вать разнородную информацию. Вероятности гипотез определялись по схеме Байеса. Полифония: 3 голоса, диапазон высот тонов: 1,5 октавы. Распознавались звуки флейты, фортепиано и трубы. Имелась возможность автоматической адаптации к звукам новых музыкальных инструментов, т.е. автоматического выделения тембральной модели инструмента.

Другая недавняя система распознавания, также использующая блэк-борд архитектуру, разработана сотрудницей Массачусетского технологического института Кейт Даной Мартин (Keith Dana Martin) в 1996 году. В отличие от системы Кашино, алгоритм Кейт Мартин не использует статистический подход. При сегментации сигнала учитывались особенности слухового восприятия человека, а также новый математический аппарат анализа сигналов — так называемая *коррелограмма* (см. ниже). Такой подход существенно расширил возможности системы распознавания: число голосов мелодии до 4, диапазон высот тонов 33 полутона (2,5 октавы). Распознаваемый инструмент — фортепиано.

Сведения о наиболее известных системах распознавания, отчёты о создании которых были опубликованы и используются в настоящее время специалистами в области распознавания музыкальных сигналов, сведены в таблицу 1.

Направление развития теории распознавания музыкальных сигналов очевидно: от простых алгоритмов, основанных на данных об амплитудном спектре, происходит развитие по пути внедрения статистических методов, нового математического аппарата спектрального и корреляционного анализа сигналов. Современные системы распознавания музыкальных сигналов широко используют сведения о психофизиологии слухового восприятия. В последние годы, видимо, в связи с развитием систем многоканальных систем звукозаписи (например, Dolby 5.1) много внимания уделяется алгоритмам искусственной реверберации, воссоздания реальных звуковых обстановок и сцен. Соответственно, потребовались интенсивные исследования свойств человеческого слуха, отвечающих за пространственную локализацию звука, разделение сигналов, соответствующих разным источникам звука. Известны попытки создания систем распознавания стереофонических музыкальных сигналов, разделение голо-

сов, соответствующих разным инструментам, в которых производится именно на основе анализа пространственной информации — локализации источников звука.

Иными словами, разработчики систем распознавания музыкальных сигналов стремятся симитировать работу органов слуха человека (от наружного уха до коры головного мозга). Оправдан ли такой подход? Покажет практика. Очевидно лишь то, что такой подход не является единственно правильным и возможным. Процессы слухового восприятия до конца не изучены и зачастую новые исследования приводят к неожиданным результатам, противоречащим устоявшимся понятиям. Кроме того, задача распознавания музыкальных сигналов, как один из видов автоматического распознавания сигналов вообще, не ставит перед собой цели моделировать работу органа слуха, задача формулируется иначе: **дана реализация звукового образа, спрашивается, к какому классу она относится.** Задача распознавания музыкальных сигналов — это, прежде всего, задача анализа сигнала, задача выделения признаков, необходимых и достаточных для принятия обоснованного решения, т.е. для достоверного распознавания образа. В конечном итоге, стоит задача распознать то, что было сыграно исполнителем, а не то, что услышит «среднестатистический слушатель». Именно в такой формулировке решение задачи распознавания музыкального сигнала может быть востребовано при формировании сигналов вещания и звукозаписи.

Табл. 1 Системы распознавания музыкальных сигналов

Автор	Научное учреждение	Возможности системы	Использованный подход
Дж. А. Мурер	Стэнфордский университет	<i>Полифония: 2 голоса. Распознаваемые инструменты: скрипка, гитара. Диапазон высот тонов: 2 октавы.</i>	Эвристический
К. Чейф	Стэнфордский университет	<i>Полифония: 2 голоса. Распознаваемые инструменты: фортепиано. Диапазон высот тонов: 1,5 октавы (19 полутонов).</i>	Эвристический

Автор	Научное учреждение	Возможности системы	Использованный подход
Р. К. Махер	Иллинойский университет	<i>Полифония: 2 голоса. Распознаваемые инструменты: кларнет, фагот, труба, туба. Диапазон высот тонов: жёстко ограничен. Диапазоны голосов не должны перекрываться.</i>	Эвристический
Катайозе	Университет Осаки	<i>Полифония: 5 голосов. Распознаваемые инструменты: фортепиано, гитара. Диапазон высот тонов: 2,25 октавы (32 полутона).</i>	Эвристический
Д. Нанн	Дюрхэмский университет	<i>Полифония: 8 голосов. Распознаваемые инструменты: орган. Диапазон высот тонов: 4 октавы.</i>	Учтены особенности восприятия.
К. Кашино	Токийский университет	<i>Полифония: 3 голоса. Распознаваемые инструменты: флейта, фортепиано, труба. Диапазон высот тонов: 1,5 октавы (18 полутонов).</i>	Учтены особенности восприятия, использованы «тембральные модели». Блэкборд-архитектура, схема Байеса для определения вероятности гипотезы
К. Д. Мартин	Массачусетский технологический институт	<i>Полифония: 4 голоса. Распознаваемые инструменты: фортепиано. Диапазон высот тонов: 2,25 октавы (33 полутона).</i>	Учтены особенности восприятия. Блэкборд-архитектура.

3.5.2. Задачи распознавания музыкальных сигналов

Итак, как мы уже неоднократно говорили, в контексте теории распознавания образов задачу распознавания музыкального сигнала (как и любого звукового сигнала) составляет вынесение решения о принадлежности реализации звукового образа к тому или иному классу. Первый возникающий вопрос — что понимать под звуковым образом?

- Звуковой объект

Теория распознавания определяет «образ» как некое структурированное, упрощённое описание объекта (параграф 3.1). В данном случае речь идёт о звуковом объекте. Соответственно, определив понятие звукового объекта, мы определимся с тем, что именно мы будем понимать под «образом». Очевидно, что для удобства построения системы распознавания желательно выделить

элементарный звуковой объект (ему будет соответствовать элементарный образ), обладающий минимальным набором свойством и одновременно являющийся составным элементом более сложных объектов (и, соответственно, более сложных образов). В случае речевого сигнала таким элементарным объектом является фонема: «фонема – это минимальная фонетическая единица, которая служит для различения и отождествления значимых единиц языка (морфем и слов)» ([19], стр. 135). Следует также отметить, что звуковые сигналы, соответствующие фонемам и является теми элементарными звуковыми объектами, из которых строятся более сложные объекты – слоги, слова и т.д. Образам, составляющим распознаваемые классы при пофонемном распознавании речевого сигнала, соответствуют фонемы.

Полная аналогия наблюдается в музыкальном сигнале. Элементарным, фундаментальным понятием в музыкальной теории является нота. Ноты, образуя последовательные или одновременные интервалы, формируют соответственно мелодию или созвучия (мелодические и гармонические интервалы). Т.е. из нот формируются аккорды (звуковые объекты, выполняющие гармоническую функцию) и музыкальные фразы (звуковые объекты, выполняющие мелодическую функцию). При этом под нотой понимают и символ нотной записи и соответствующий этому символу звук, точнее – звуковой объект.

Итак, обобщив сказанное выше относительно элементарных звуковых объектов речи и музыки, мы можем сформулировать определение звукового объекта:

звуковой объект – минимальный по длительности звуковой сигнал, воспринимаемый как самостоятельное единое целое.

Восприятие звукового объекта как самостоятельного единого целого означает, что объект имеет длительность, достаточную для определения всех необходимых для распознавания (человеком или автоматом) свойств: высоты тона, тембра, соответствующей фонемы.

Как мы уже говорили, нотная запись описывает высоту тона, длительность ноты, но не описывает тембра. Важно отметить, что на сегодняшний день не дано чёткого определения тембра и не выделены параметры сигнала, опре-

деляющие восприятие тембра. К понятию тембра мы вернёмся и поговорим более подробно в следующей главе.

Совокупность всех музыкальных звуковых объектов соответствует совокупности звуков, извлечённых различными музыкальными инструментами. Среди музыкальных инструментов выделяют инструменты с определённой высотой тона и инструменты с неопределённой высотой тона (т.н. ударные и перкуссия). Поэтому сделаем важное замечание, касающееся всего, сказанного далее: будем иметь дело с тональными звуковыми объектами, т.е. звуковыми объектами, для которых определена высота тона, шумовые звуковые объекты распознаванию в данной системе не подлежат. Звуковыми объектами будем для краткости называть тональные звуковые объекты.

Теперь выделим свойства музыкальных звуковых объектов, подлежащих распознаванию:

1) **Высота тона** (ей соответствуют частота основного тона или период основного тона колебания).

2) **Длительность** (условное понятие, определяющее интервал времени от момента извлечения ноты, т.е. возбуждения звуковых колебаний, до момента затухания). Условность понятия «длительность» в том, как определяются моменты возбуждения и затухания: за эти моменты принимаются моменты времени, в которые интенсивность звукового сигнала (акустического или электрического) достигает определённого, условно выбранного уровня. Длительность ноты в нотной записи может описывать как количественно (указывается количество нот определённой длительности за единицу времени – общепринято указание количества четвёртых нот в минуту), так и качественно (темп исполнения описывается, как правило, итальянскими словами, например: Largo (широко), Moderato (умеренно), Prestissimo (очень быстро)).

3) **Тембр** (понятие, позволяющее отличать звуковые объекты, соответствующие одной и той же ноте, но исполненные разными музыкальными инструментами, в разных акустических обстановках).

4) **Громкость** (также условное понятие, описывающее субъективно воспринимаемую интенсивность звукового сигнала). Громкость зависит и от длительности звукового объекта, и от высоты тона, и от состава гармоник. Кстати нотное письмо описывает громкость качественно: *ff* («форте фортиссимо», т.е. очень громко), *mp* («мечшо пьяно», т.е. не слишком тихо). При автоматическом же распознавании громкость может быть оценена только количественно. Существуют примерные шкалы соответствия между количественными и качественными оценками громкости.

Перечисленными свойствами обладают звуковые объекты, составляющие как акустические сигналы (т.е. акустические колебания в воздухе), так и соответствующие электрические сигналы (в аналоговой форме), полученные с помощью электромеханических преобразователей. Совпадают ли одноимённые свойства акустических и электрических звуковых объектов? Второй вопрос, возникающий в связи с общей постановкой задачи распознавания образов, — вопрос о соотношении звукового образа и его реализации, т.е. звукового объекта.

- Реализация звукового образа

Обсуждая распознавание образов с информационной точки зрения (параграф 3.2), мы установили, что нотная запись не описывает однозначно реализацию звукового образа. Т.е. одному образу всегда соответствует бесконечное множество реализаций, прежде всего, из-за вольности интерпретации партитуры. Второй важнейший фактор — многообразие типов музыкальных инструментов и их свойств. Музыкальный инструмент порождает звуковой объект, он является источником информации в системе исполнитель-слушатель. Между приёмником информации и её источником всегда есть канал передачи, и его влияние нельзя не учесть. Обычно музыка прослушивается в помещении, поэтому ещё одним инструментом, участвующим в исполнении музыки является акустическая обстановка. Специфическим влиянием помещения на звуковой образ является реверберация (затухание энергии звукового поля), отражение от поверхностей. Такое влияние формирует у человека ощущение удалённости от источника звука, ощущение комфортности прослушивания. Одновременное воздействие прямого и отражённого сигналов приводит

к искажению и тембра (помещение имеет резонансные частоты), и длительности («затягивается» процесс нарастания колебаний, медленнее происходит затухание), и громкости. Единственный неизменный параметр — высота тона.

Если в зале используется система звукоусиления, то весь электрический тракт от микрофонов до громкоговорителей вносит свои изменения в музыкальный сигнал. Т.е. акустическая реализация звукового образа (ноты) неэквивалентна его электрической реализации. К специфическим влияниям электрического тракта, будь то система звукоусиления или канал междугороднего вещания, можно отнести искажения тембра, динамики (соотношение громкостей) музыкального сигнала, нелинейные искажения, сопровождаемые появлением новых спектральных составляющих.

На пути от внешнего уха человека до слуховой коры мозга музыкальный сигнал также претерпевает изменения. К свойствам слуха относится его нелинейность, неравномерная чувствительность по диапазону частот, явление предмаскировки и послемаскировки. Важным свойством слухового восприятия, как и вообще восприятия является ограниченность количества одновременно воспринимаемых стимулов. Человек может одновременно распознавать и обрабатывать 7-8 внешних стимулов. Проведённые исследования показывают, что даже музыканты не могут определить количество нот в аккорде, если их больше семи. Это важно с точки зрения свойства полифоничности музыкального сигнала, его «многослойности» воздействия. Как известно, состав больших ансамблей, например, полного состава симфонического оркестра с большим хором, может включать несколько сотен исполнителей — до 500-600. Трудно себе представить, что творилось бы со слушателями, если бы каждый исполнитель имел свою отдельную партию, не похожую на все остальные.

Итак, неоднозначность соответствия звукового образа и звукового объекта (как реализации образа) очевидна и перечислены факторы, определяющие эту неоднозначность. Становится понятно, что одноимённые свойства акустической и электрической реализаций звукового образа не совпадают.

- Задачи распознавания музыкального сигнала

Распознать реализацию звукового объекта значит отнести её к какому-либо классу. Классифицируют музыкальные звуковые объекты по перечисленным выше четырём их свойствам. В общем виде в задачи распознавания музыкального сигнала входит:

1. Определение моментов появления и длительностей нот.
2. Определение высот тонов нот.
3. Определение громкостей нот.
4. В случае ансамблевого исполнения — разделение партий инструментов по различиям в тембре или иным признакам.

Задача определения громкости не нашла пока даже приближённого решения. Попытки разделения партий инструментов, как мы говорили, имеются, и опираются они в основном на результаты исследований способности человека локализовать источники звука в пространстве, времени и частоте или на априорные сведения о тембрах музыкальных инструментов. Действительно, исследования в области распознавания музыкальных сигналов и исследования в области автоматического опознавания звуковых источников (в зарубежной литературе принято сокращение CASA – Computational Auditory Scene Analysis) пока ведутся отдельно. Результаты и тех и других исследователей пока не столь значительны, чтобы ими делиться. Большую часть сил и средств исследователи, берущиеся за распознавание музыкальных сигналов, тратят на разработку алгоритмов определения высот тонов одновременно звучащих нот. Таким образом, решение всех четырёх названных задач непосильно, обычно ограничиваются первыми двумя. Мы также ограничимся определением длительностей и высот тонов нот. При распознавании музыкальных сигналов частота основного тона и длительность ноты квантуются: естественно, в реальном сигнале эти два параметра представляют собой непрерывную величину, но общепринято в современных системах распознавания результат выводить в виде, удобном для преобразования в нотную запись (в формате MIDI). При определенном (т.е. заранее оговорённом) строе (натуральном, равномерно темперированном и т.п.) и заданной частоте эталонного тона (как правило, частота

От «ля» первой октавы) существует соответствие между высотой ноты и частотой основного тона соответствующего звукового объекта. Стандарт MIDI для кодирования высоты тона использует целые числа от 0 до 127, для кодирования длительности ноты задаётся длительность минимальной единицы времени (квантиля), а длительность нота выражается целым числом длительностей квантиля. Таким образом, очевидно, что при распознавании целесообразнее квантовать длительность и высоту нот.

Стоящую перед нами задачу распознавания сформулируем следующим образом:

дана реализация многоголосной мелодии, требуется классифицировать длительности и высоты тонов составляющих её звуковых объектов.

Количество голосов мелодии определяется максимальным количеством одновременно звучащих нот, многоголосной считается мелодия, число голосов которой больше одного.

3.6. СЛОЖНОСТИ РАСПОЗНАВАНИЯ МНОГОГОЛОСНЫХ МЕЛОДИЙ

3.6.1. Общие соображения

При решении поставленной выше задачи априорной информации о входном сигнале, как правило, очень мало. Возможен вариант, когда известен состав ансамбля, т.е. известно количество звучащих инструментов и каким-либо образом описаны свойства сигналов, производимых данными инструментами. Однако такая ситуация искусственна — в большинстве задач сведения о составе оркестра отсутствуют, если распознавание происходит с помощью данных о модели звуков музыкального инструмента, то эти сведения приходится получать из самого сигнала уже в процессе распознавания.

Источником основных сложностей распознавания музыкальных сигналов является то, что в многоголосной мелодии звуковые объекты перекрываются. Перекрываются звуковые объекты во времени, т.е. перекрываются интервалы существования звуковых объектов, моменты их возникновения могут совпадать, за время существования одного звукового объекта могут возникнуть и за-

тухнуть ещё несколько. Перекрываются звуковые объекты и в частотной области, т.е. звуковые объекты могут иметь спектральные компоненты с примерно совпадающими частотами. Сложнейший вариант — т.н. исполнение в унисон, когда несколько инструментов одновременно исполняют одну и ту же мелодию. Перекрытие почти стопроцентное, тем не менее, человеческое ухо способно определить их количество. Такое разделение возможно даже в случае минимальных различий звуковых объектов по скорости переходных процессов, частоте основного тона, отклонениям частот обертонов от значений, кратных частоте ОТ, по параметрам модуляции обертонов.

Случай распознавания одноголосной мелодии гораздо проще — звуковые объекты не перекрываются, время затухания одного звукового объекта совпадает со временем возникновения следующего. Следует отметить, что пауза также является звуковым объектом, не имеющим, правда, высоты тона, но имеющим длительность, и в нотной записи паузы обозначаются специальными символами. Иными словами, одноголосная мелодия представляет собой линейную последовательность звуковых объектов. Распознавание сводится к определению моментов смены высоты тона, сведения об интенсивности колебаний используются лишь в случае разделения двух последовательных нот с одинаковой высотой тона и для обнаружения пауз. Поэтому распознавание одноголосных мелодий называют «слежением за мелодией основного тона» (*fundamental frequency tracking*).

Для определения длительности и высоты тона звукового объекта, очевидно, прежде всего, необходимо определить его временные границы. Такой процесс называется сегментацией. После выполнения задачи сегментации образуется ритмическая структура распознаваемой мелодии. Ритмическая структура мелодии не менее важна, чем её высотная структура. В этом состоит одно из отличий задачи распознавания музыкального сигнала от распознавания речевого сигнала. При распознавании речи ошибки сегментации могут привести лишь к снижению достоверности распознавания фонемы, возникшие ошибки можно исправить автоматически, имея словарь языка. Ошибки сегментации музыкального сигнала повлекут и неверное определение высоты тона (если на вход алгоритма определения частоты основного тона попадут фрагменты по-

следовательных звуковых объектов с разной высотой тона), и искажение ритмической структуры, в результате чего мелодия может стать неузнаваемой. Сегментация многоголосной мелодии сложнее сегментации одноголосной мелодии и речевого сигнала.

3.6.2. Созвучия тональных звуков

Перекрытие звуковых объектов в частотной области делает невозможным применение алгоритмов определения высоты основного тона, используемых при распознавании речи или одноголосных мелодий. Чтобы понять причину этого, обсудим кратко структуру амплитудного спектра звуковых объектов (более подробно свойства звуковых объектов обсуждаются в главе 6). Рассмотрим свойства созвучия двух тональных звуков (по [73], стр. 31)

Утверждение 1 Если гармоника звука Z_1 перекрывает гармонику звука Z_2 , то частоты основных тонов этих звуков соотносятся как $\frac{m}{n}$, $m, n \in Z$, $m, n \geq 1$.

Доказательство: Действительно, пусть звуки Z_1 и Z_2 имеют частоты основных тонов соответственно $F0_1$ и $F0_2$, а перекрывающиеся гармоники имеют номера i и j , тогда:

$$i \cdot F0_1 = j \cdot F0_2$$

$$\frac{F0_1}{F0_2} = \frac{j}{i}, \quad i, j \in Z, \quad i, j \geq 1$$

Утверждение 2 Если частоты основных тонов звуков Z_1 и Z_2 соотносятся как $\frac{m}{n}$, $m, n \in Z$, $m, n \geq 1$, то гармоники звука Z_1 с номерами $P(k) = n \cdot k$, где $k = 1, 2, 3, \dots$ перекрывают гармоники звука Z_2 с номерами $Q(k) = m \cdot k$, где $k = 1, 2, 3, \dots$

Доказательство: Запишем условие совпадения i -той гармоники звука Z_1 с j -той гармоникой звука Z_2 :

$$i \cdot F0_1 = j \cdot F0_2, \text{ при этом } \frac{F0_1}{F0_2} = \frac{m}{n}, \text{ т.е. можем записать:}$$

$$i \cdot \frac{m}{n} \cdot F0_2 = j \cdot F0_2 \Rightarrow i \cdot m = j \cdot n, \text{ что справедливо для любых}$$

$$i = n \cdot k \text{ и } j = m \cdot k, \text{ где } k = 1, 2, 3, \dots$$

Теперь выясним, в каком соотношении могут находиться основные тональных звуков. В современной европейской и русской музыкальных практиках применяются два строя – натуральный (пифагорова гамма) и равномерно темперированный.

Натуральным звукорядом называют звукоряд, образованный звуками, частоты основных тонов образуют ряд $F, 2F, 3F, 4F$, и т.д., где F – некоторая основная частота. Понятно, что частоты, соответствующие высотам тонов звуков *натурального строя* будут соотноситься как целые числа, большие единицы ($\frac{F0_m}{F0_n} = \frac{m}{n}$, где $m, n \in Z, m, n \geq 1$). Соотношение частот ОТ звуков, отличающихся на полутон, не одинаково для всего звукоряда.

Темперированный строй характеризуется тем, что октава делится на 12 одинаковых интервалов – полутонов. Любые два звука темперированного строя, отличающиеся на полутон, имеют частоты ОТ, соотносящиеся как $\sqrt[12]{2} \approx 1,0595$.

Для сравнения натурального и темперированного строев обратимся к табл. 2. В ней приводятся соотношения частот основных тонов нот, отличающихся на целое число полутонов.

Табл. 2 Различия натурального и равномерно темперированного строев

Интервал, полутонов	$\frac{f0_1}{f0_2}$				Отклонение натурального строя от темперированного, %	Процент совпадающих гармоник
	темперированный строй	натуральный строй				
		m/n	m	n		
1	1,0595	1,0667	16	15	-0,68	6,45
2	1,1225	1,125 (1,143)	9 (8)	8 (7)	-0,23(1,8)	11,8
3	1,1892	1,2	6	5	-0,91	18,33
4	1,2599	1,25	5	4	0,79	22,5
5	1,3348	1,3333	4	3	0,11	29,15

Интервал, полутонов	$\frac{f_{0_1}}{f_{0_2}}$				Отклонение натурального строя от темперированного, %	Процент совпадающих гармоник
	темперированный строй	натуральный строй				
		m/n	m	n		
6	1,4142	1,4	7	5	1,0	17,13
7	1,4983	1,5	3	2	-0,11	41,65
8	1,5874	1,6	8	5	-0,79	16,25
9	1,6818	1,6667	5	3	0,91	26,65
10	1,7818	1,778 (1,75)	16 (7)	9 (4)	0,23 (-1,8)	8,68
11	1,8877	1,875	15	8	0,68	9,58
12	2	2	2	1	0	75

Как видно, соотношения частот ОТ нот звуков темперированного строя можно с незначительной погрешностью можно выразить отношением двух целых чисел. Для нас это важно с учётом утверждения 2: чем меньше числа m и n , соотношением которых приближается соотношение частот ОТ двух музыкальных звуков, тем больший процент обертонов этих звуков будет иметь приблизительно совпадающие частоты. Можно назвать и интервалы, имеющие наибольший процент «общих» обертонов: октава, кварта (5 полутонов) и квинта (7 полутонов).

Итак, соотношения частот ОТ музыкальных звуков точно или приблизительно могут быть выражены соотношением целых чисел. Это означает, что если бы количество гармоник звуков было бесконечным, то не было бы двух звуков, таких, чтобы у них не совпали частоты некоторых гармоник. Именно это усложняет процесс распознавания созвучий — при распознавании необходимо иметь в виду, что каждая спектральная составляющая может быть элементом нескольких звуковых объектов.

Очевидно, возможен вариант совпадения частот гармоник у трёх и более звуков, это создаёт ещё одну проблему. Допустим, анализируется спектр созвучия двух звуков с ОТ, имеющими совпадающие по частоте гармоники. Тогда гармоники третьего звука, имеющего с этими двумя общие гармоники, могут быть обнаружены в спектре даже если этот третий звук не входит в созвучие

(не звучит). Конкретный пример — созвучие нот «до» и «ми» первой октавы — поясняется рисунком 7. Вертикальные штриховые линии соответствуют частотам основного тона и гармоник третьей ноты — «соль» первой октавы. Кружками обозначены обертоны нот «до» и «ми», имеющие частоты, примерно равные частотам гармоник «соль».

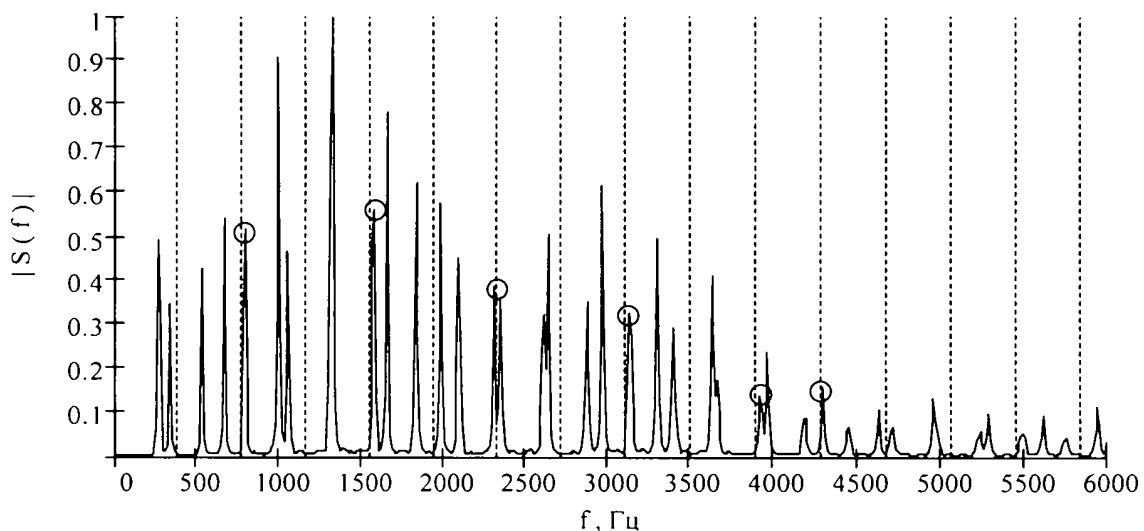


Рис. 7 Нормированный амплитудный спектр созвучия нот «до» и «ми»

В результате возможно принятие ошибочного решения при сегментации — о наличии несуществующей ноты. Казалось бы, важным признаком при обнаружении звукового объекта с данным основным тоном может служить наличие в спектре с частотой этого основного тона. Однако, во-первых, как известно, не все звуки, извлекаемые акустическими музыкальными инструментами, в силу физических особенностей содержат основной тон, а, во-вторых, отсутствие основного тона в спектре звука может быть следствием линейных искажений.

3.7. **СОВРЕМЕННЫЕ ПОДХОДЫ К РАСПОЗНАВАНИЮ ПОЛИФОНИЧЕСКИХ МЕЛОДИЙ**

3.7.1. **Представление данных в системах распознавания музыкальных сигналов**

Обсуждая классификацию систем распознавания, мы разделили СР по способу получения апостериорной информации на одноуровневые и много-

уровневые. Также при рассмотрении набора признаков, необходимых для распознавания музыкальных сигналов, мы говорили о том, что определение этого набора не представляет особой сложности. Гораздо более сложной задачей является выделение этих признаков, т.е. получение апостериорной информации о реализациях звуковых объектов. Систематизируем наши рассуждения о получении и обработке апостериорной информации и выявим сложности, возникающие при её сборе и обобщении.

На «входе» системы распознавания — звуковой сигнал, на её выходе структурированная информация о длительностях и ОТ обнаруженных звуковых объектов. Примером такой структуризации может быть нотная запись. Первым наблюдаемым признаком звукового сигнала является его амплитудный спектр, вычисляемый из ДПФ. Очевидно, что амплитудный спектр сигнала и отображающая сигнал нотная запись — понятия *разного уровня абстракции*, при распознавании происходит обобщение информации, абстрагирование. Как показывает опыт создания СР музыкальных сигналов, не существует алгоритмов, позволяющих от амплитудного спектра непосредственно перейти к нотной записи. Иными словами, при распознавании музыкальных сигналов неизбежно использование нескольких уровней представления апостериорной информации, т.е. СР музыкальных сигналов — многоуровневые.

Рассмотрим уровни представления данных при распознавании музыкальных сигналов (см. рис. 8):

- 1) наиболее конкретное понятие — амплитудный спектр сигнала;
- 2) по наличию превышающих некий порог локальных максимумов амплитудного спектра судят о наличии гармонических (в отличие от шумовых) *спектральных составляющих*;
- 3) отдельные спектральные составляющие могут быть как компонентами звуковых объектов — *основными тонами* или *оберттонами*, так и продуктами нелинейных искажений, помехами;
- 4) звуковые объекты являются реализациями нот, при этом «нота» — понятие теории музыки, понятие более абстрактное, чем «звуковой» объект;

5) ноты объединяются в последовательные *интервалы* (мелодические интервалы) и созвучия (гармонические интервалы), определяемые соотношением высот их тонов, выраженным количеством полутонов или количеством ступеней гаммы;

6) последовательные интервалы составляют мелодию *голосов*, созвучия образуют *аккорды*, выполняющие гармоническую функцию.

7) наконец, характер мелодии и аккомпанемента определяется выбранной тональностью (мажорной или минорной).

В схеме на рис. 8 стрелки, соединяющие уровни абстракции данных двунаправленные. Очевидно, что процесс обработки информации может быть как *восходящим* (bottom-up в англоязычной литературе), так и *нисходящим* (top-down): переход возможен как от менее абстрактных данных к более абстрактным, так и наоборот. При чём восходящий и нисходящий анализ информации характерен и для человеческого мышления, и для вычислительных алгоритмов. Например, при распознавании музыкантом пьесы с быстрыми пассажами или сложной гармонией, нераспознанные фрагменты мелодии или аккорды аккомпанемента воссоздаются исходя из сведений о тональности музыкального произведения — если проще определиться на слух с тональностью. Т.е. при принятии решения имеет место объединение сведений *разного уровня абстракции*.

Для всех современных систем распознавания музыкальных сигналов высшим уровнем абстракции данных являются последовательные и одновременные интервалы. Методы, позволяющие объединить интервалы в мелодии отдельных голосов и аккорды, пока не разработаны. Основное внимание разработчиков сосредоточено на промежуточных формах представления сведений о распознаваемом сигнале. В современных системах данные промежуточного уровня — это признаки, полученные на основе анализа признаков первого уровня (спектра сигнала, огибающей, мощности и т.д.), а также признаков, полученных на основе признаков первого уровня. На основе признаков уровней выше первого — промежуточных признаков — принимаются решения о длительности и частоте ОТ звукового объекта.

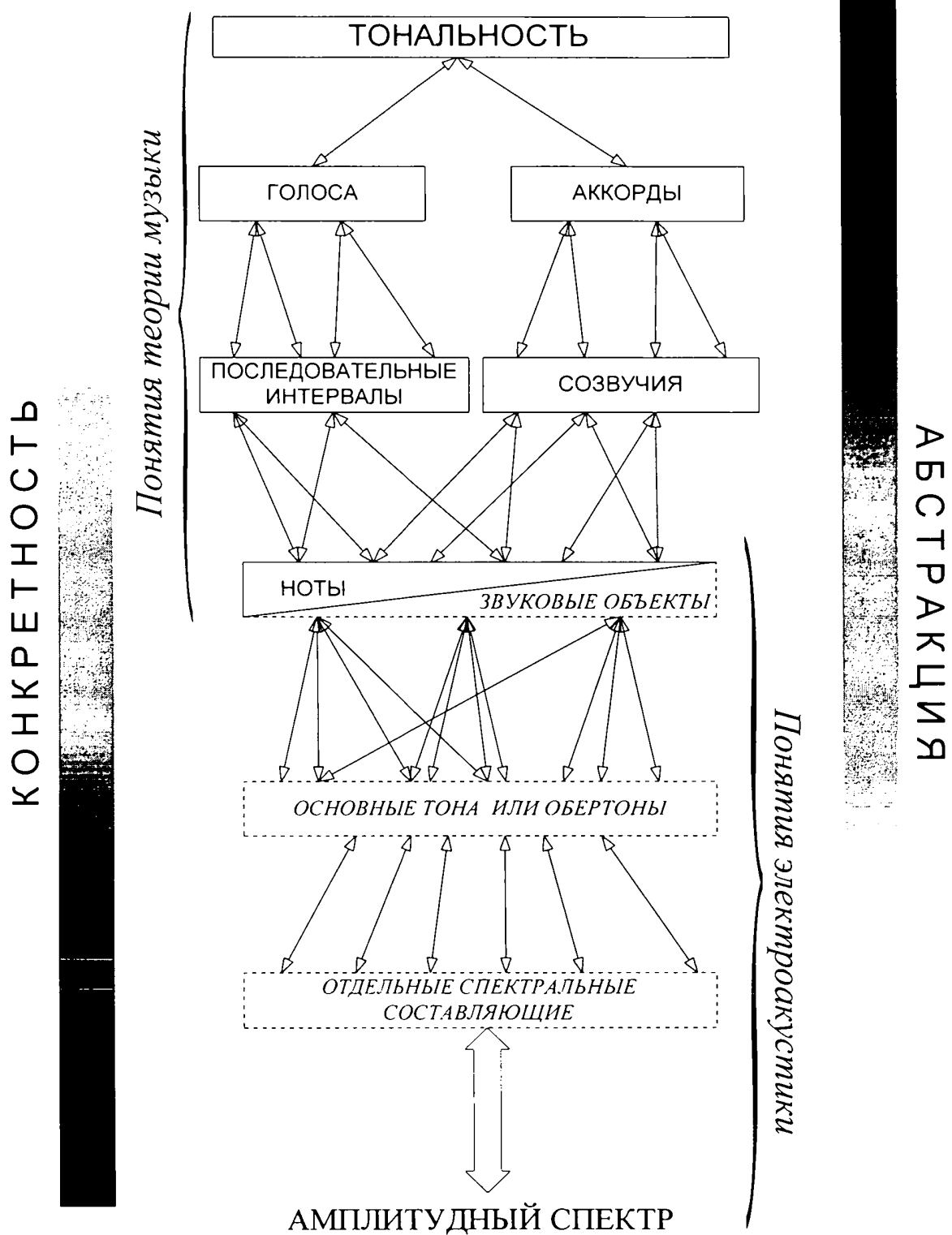


Рис. 8 Возможные уровни представления данных при распознавании музыкального сигнала

Мы не раз отмечали, что при распознавании многоголосных музыкальных сигналов наиболее сложной задачей является определение частот ОТ звуковых объектов, принятие решения о высоте тона (в терминах теории музыки) по известной частоте ОТ — процедура тривиальная. Именно поэтому идёт постоянный поиск новых промежуточных признаков, новых форм промежуточного представления данных, которые могли бы повысить достоверность распознавания.

Одной из промежуточных форм представления сведений о звуковом сигнале является т.н. *коррелограмма*. Строится коррелограмма следующим образом: исходный сигнал с помощью набора фильтров делится на узкополосные составляющие $b_i(t)$, выделяются огибающие этих составляющих $e_i(t)$, затем вычисляется кратковременная автокорреляционная функция (АКФ) $R_i(\tau)$ *огибающей* каждого узкополосного компонента. Коррелограмма определяется так:

$$C(f_i, \tau) = R_i(\tau), \text{ где} \quad (4)$$

f_i — центральная частота i -ого полосового фильтра.

С помощью коррелограммы моделируется восприятие человеческим ухом периодических звуков. Полосы пропускания фильтров выбирают так, чтобы они совпадали с критическими полосками слуха. АКФ отражает периодичность сигнала: АКФ квазипериодического сигнала имеет ярко выраженные максимумы, шумовой сигнал, напротив имеет спадающую АКФ.

Важно отметить, что для построения коррелограммы используется АКФ не самих узкополосных компонентов, а их огибающих. Огибающую получают однополупериодным выпрямлением сигнала. Обратимся к рис. 9. На нём изображены начальные циклы колебаний сигнала (пунктирная линия), соответствующего букве «а», произнесённой мужским голосом, а также огибающие верхних и нижних полувольт. Однополупериодное выпрямление необходимо, чтобы период огибающей сигнала был равен периоду сигнала, при двухполупериодном выпрямлении, период огибающей в 2 раза больше.

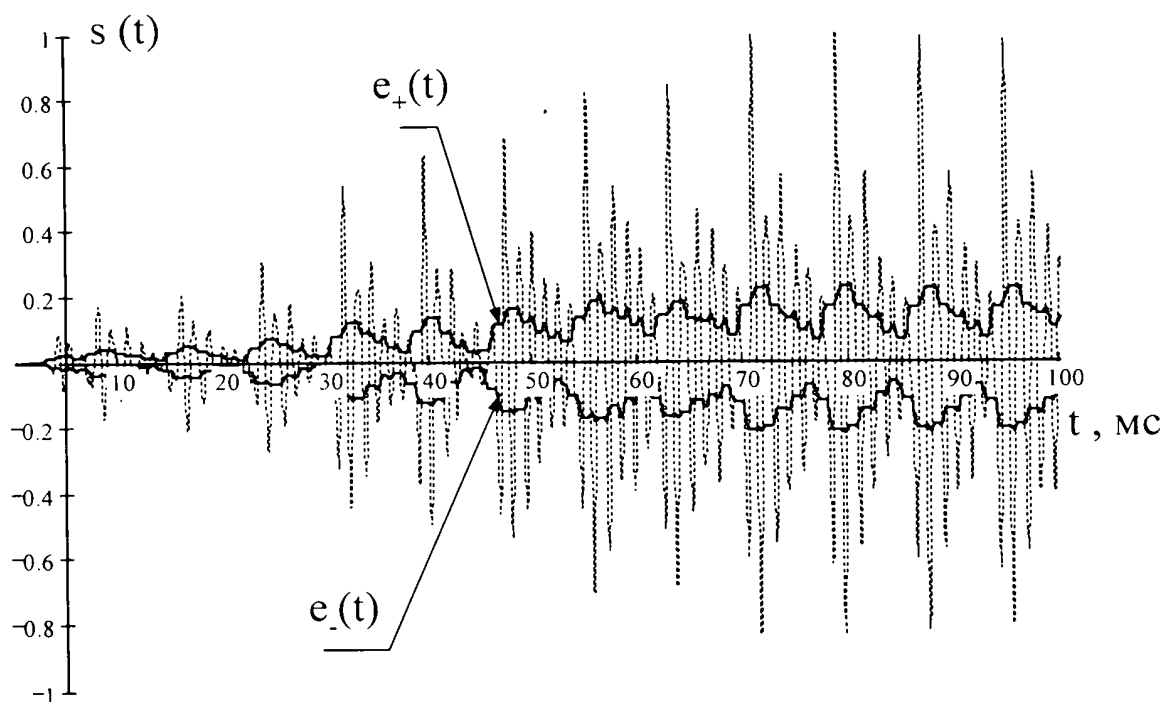


Рис. 9 Фрагмент речевого сигнала и огибающие верхних и нижних полувольт

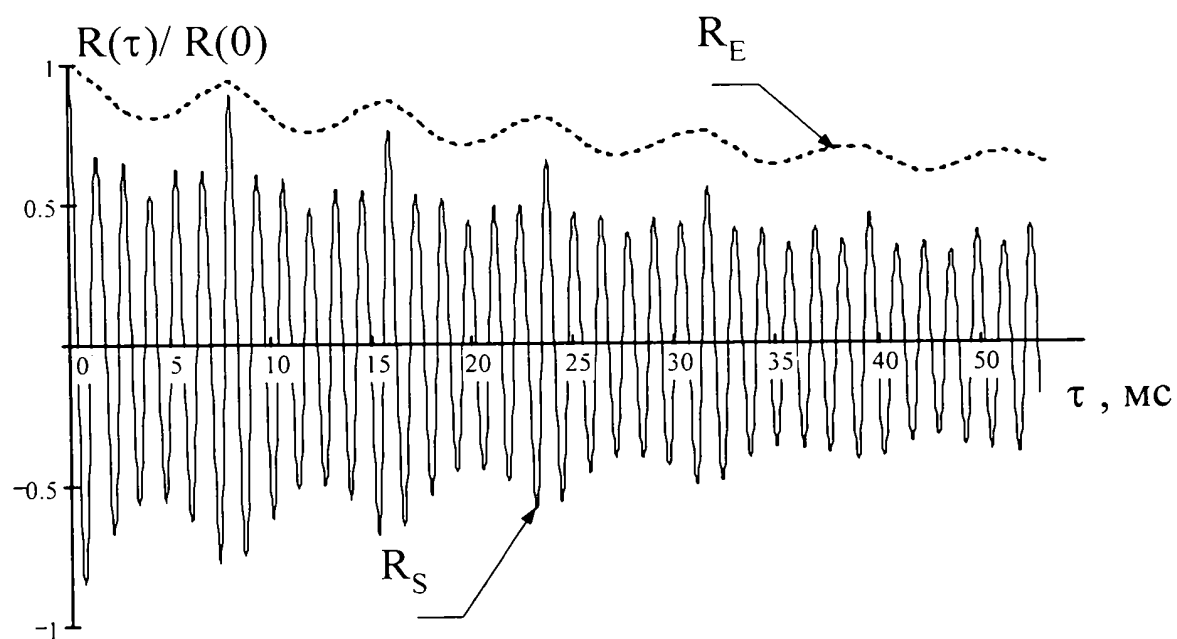


Рис. 10 Нормированные кратковременные АКФ речевого сигнала и огибающей его верхних полувольт

Преимущество использования АКФ огибающей состоит в том, что огибающая — функция с более узким спектром, её АКФ R_E имеет меньше локальных максимумов (см. рис. 10), соответственно, принять решение о периоде огибающей проще.

Таким образом, коррелограмма отражает периодичность амплитудной модуляции колебаний в узких частотных полосах. Также используется т.н. суммарная автокорреляционная функция $R_\Sigma(\tau)$:

$$R_\Sigma(\tau) = \sum_i \frac{R_i(\tau)}{R_i(0)} \quad (5)$$

АКФ узкополосных сигналов нормируются по своим максимальным значениям $R_i(0)$, затем нормированные значения суммируются по всем частотным полосам. В результате максимумы суммарной АКФ соответствуют периодам амплитудной модуляции, наиболее часто встречающимся по всему диапазону частот (см. рис. 11).

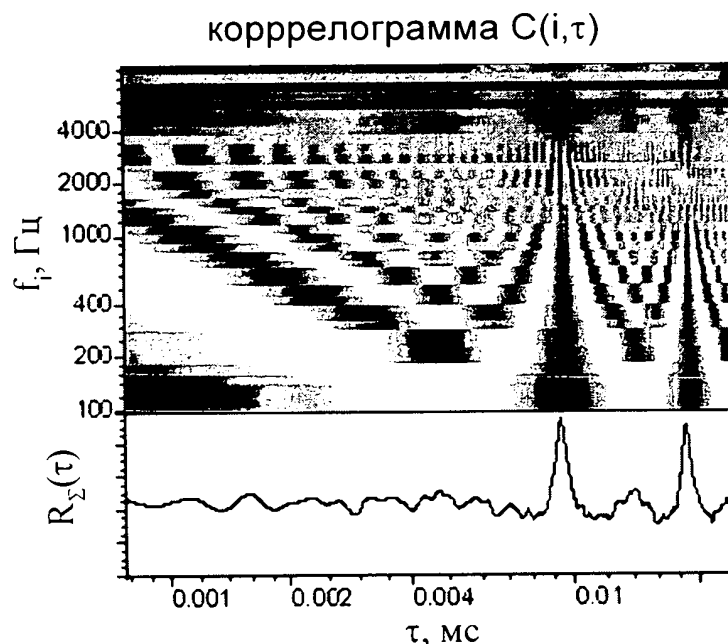


Рис. 11 Коррелограмма и соответствующая суммарная АКФ (по [84])

Зависимость суммарной АКФ от начального момента временной выборки, на которой рассчитывается коррелограмма, называется *периодограммой*. Периодограмма позволяет проследить изменение во времени периода сигнала.

Заметим, что периодограмма получена путём вычислений без принятия решений и логических операций.

Суммарная АКФ – признак промежуточного уровня, используемый при принятии решения о содержании в сигнале периодических (квазипериодических) составляющих – например, соответствующих вокализованным звукам речи или тональным звукам музыкальных инструментов. При обнаружении периодических компонентов сигнала суммарная АКФ используется при принятии решения о частоте основного тона этих компонентов. Периодограмма может использоваться для сегментации сигнала: появление нового звукового объекта с основным тоном, отличным от ОТ предыдущего звукового объекта сопровождается смещением локальных максимумов суммарной АКФ, что наглядно отображается на графическом изображении периодограммы.

Описанные промежуточные признаки, получают из более конкретных признаков (например, кратковременного амплитудного спектра сигнала) *обобщением*, выделением общих свойств узкополосных составляющих спектра и отбрасыванием их индивидуальных свойств. Тем самым сокращается набор свойств, описывающих распознаваемый сигнал, выбираются наиболее информативные свойства, и, следовательно, упрощается процедура принятия решения, например, о высоте тона и длительности ноты, достоверность распознавания повышается.

3.7.2. Методология «классной доски»

В предыдущем параграфе мы рассмотрели наиболее современные подходы к формированию пространства признаков для распознавания музыкальных сигналов. Теперь рассмотрим один из современных подходов к построению процедуры принятия решения.

В англоязычной литературе этот подход называется *blackboard architecture*, а система принятия решения, использующая данный подход – *blackboard system*. *Blackboard* – дословно: классная доска. Название происходит от аналогии между работой такой системы и работой группы специалистов (экспертов), решающих совместно некоторую задачу, поочерёдно подходящих к доске, на которой отображается решение, и вносящих свои коррективы. Решение нахо-

дится после определённого количества «итераций» — этапов решения с поправками, внесёнными экспертами. По сути, система, использующая методологию «классной доски» является специфической экспертной системой, используемой для принятия решения.

Одно из базовых понятий методологии «классной доски» — *гипотеза*. Гипотезы поступают на вход экспертной системы, проверяются «экспертами», отбрасываются как ложные, формируются новые гипотезы, и снова проверяются. Различают *уровни* гипотез, соответствующие степени абстракции данных, описываемых гипотезой. Каждая гипотеза связана со сформировавшими её гипотезами более низкого уровня, а также с гипотезами более высокого уровня, в формировании которых она участвовала. На выходе системы — подтверждённые «экспертами» абстрактные гипотезы, принимаемые за верное решение.

«Эксперты» в вычислительно-логической системе, работающей по методологии «классной доски» — т.н. *источники знаний* (сведений). Источники знаний — также базовое понятие методологии. Важно то, что источники знаний могут быть разнородными по характеру предоставляемых сведений. На каждом уровне обобщения гипотез источники знаний свои. Каждый «эксперт» принимает решение о достоверности проверяемой гипотезы. Принятие решение источником знаний сводится к проверке выполнения какого-либо *условия*. При выполнении условия источник знаний выполняет определённое *действие*. По решениям, вынесенным источниками знаний, формируется *рейтинг* гипотезы. Рейтинг формирует т.н. *рейтинговая функция*. Если на данном уровне абстракции гипотез существует несколько источников знаний, рейтинговая функция суммирует взвешенные рейтинги, сформированные разными «экспертами». Возможные действия, выполняемые источниками знаний на каждом уровне абстракции гипотез:

1. гипотезы данного уровня признаются ложными или достоверными;
2. даются указания на перепроверку гипотез более низкого уровня, сформировавших ложную гипотезу более высокого уровня;
3. на основе анализа ложных и достоверных гипотез могут быть сформированы гипотезы более низкого уровня в случае, если сведения «экспертов»

данного уровня требуют наличия гипотезы низкого уровня, но она не была сформирована.

Работой системы в целом руководит *планировщик*. Система работает пошагово, итерациями. На каждой итерации планировщик проверяет выполнимость условий «экспертов», начиная с самого низкого уровня. По итогам формируется список действий, соответствующих выполненным условиям. Эти действия выполняются и происходит переход к следующему шагу. Обобщённая схема системы для распознавания звуковых сигналов, работающей по методологии «классной доски» приведена на рис. 12.

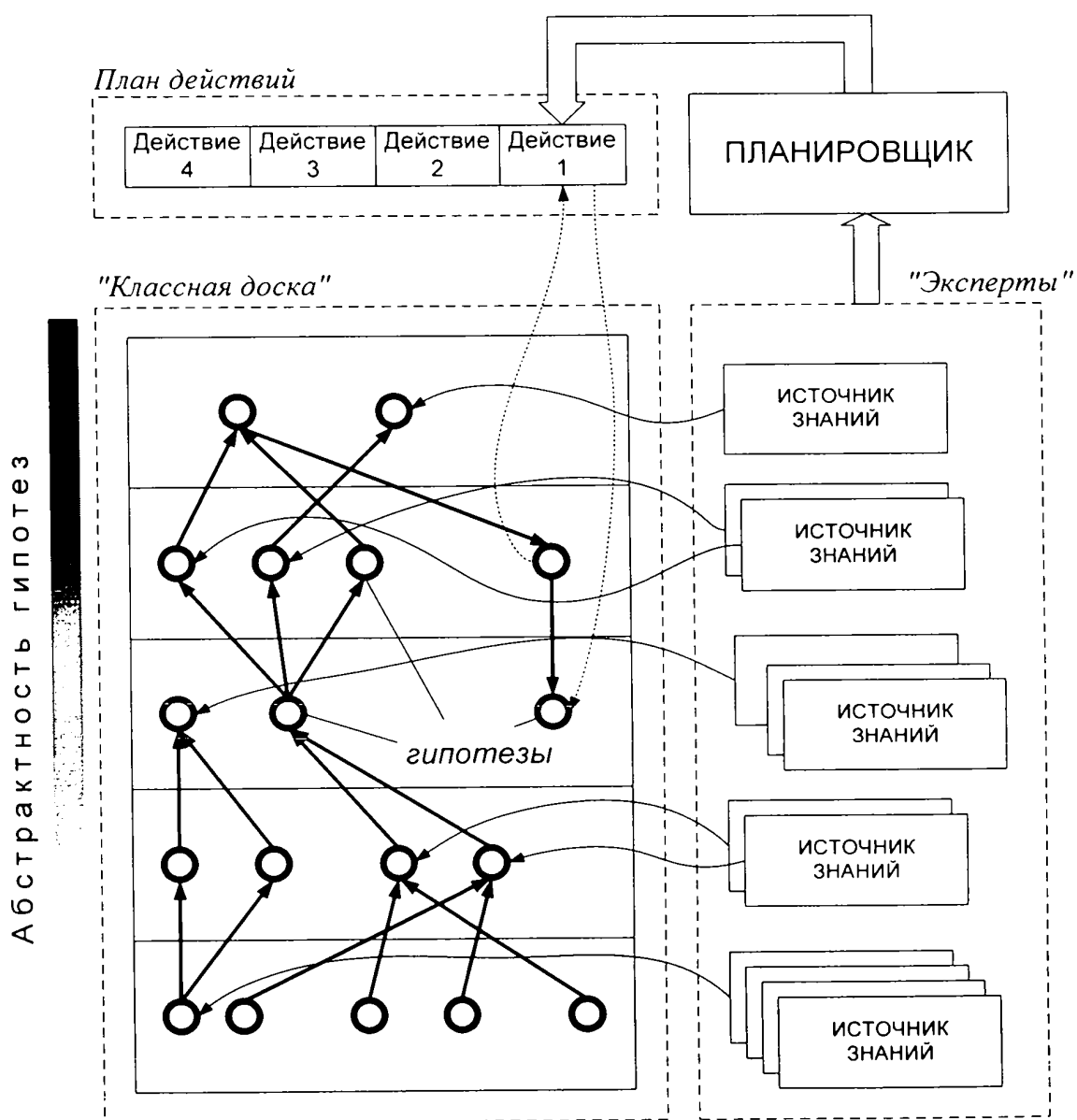


Рис. 12 Общая схема реализации методологии «классной доски»

По рассмотренной нами классификации систем распознавания системы, использующие методологию «классной доски» относятся к сложным многоуровневым комбинированным системам распознавания. Рассмотрим конкретную систему [84], использующую данный подход, – систему распознавания музыкальных сигналов, разработанную Кейт Даной Мартин (Keith Dana Martin), сотрудницей Массачусетского технологического института.

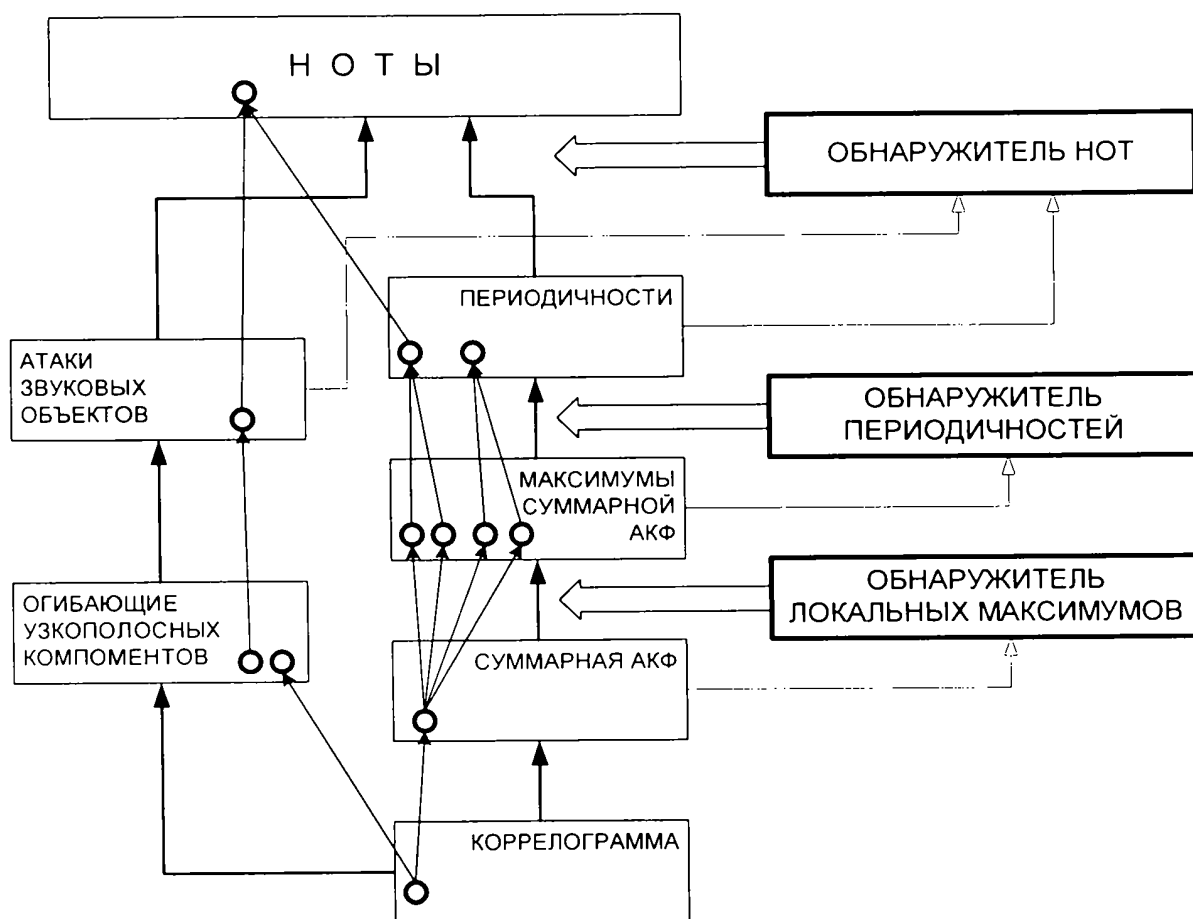


Рис. 13 Схема реализации методологии классной доски в системе распознавания музыкальных сигналов (К. Д. Мартин)

В системе используются следующие уровни представления данных (см. рис. 13):

1. Коррелограмма.
2. Суммарная АКФ.
3. Локальные максимумы суммарной АКФ.

4. *Периодичности* (periodicities) сигнала — интервалы времени, соответствующие предполагаемым периодам колебаний в распознаваемом сигнале.

5. *Огибающие узкополосных компонентов* — элементы коррелограммы.

6. Предполагаемые моменты начала звучания звуковых объектов — *атаки узкополосных компонентов*.

7. *Ноты* — результат распознавания.

Гипотезы оцениваются следующими «экспертами»:

1. *Обнаружитель локальных максимумов*, формирующий список локальных максимумов суммарной АКФ, превышающих некий порог. Каждый интервал времени, соответствующий такому максимуму, выдвигается как гипотеза о периодичности.

2. *Обнаружитель периодичностей*, анализирующий список максимумов АКФ и отбрасывающий гипотезы, соответствующие кратным периодам. В суммарной АКФ звукового объекта с периодом основного тона T всегда имеются локальные максимумы вблизи интервалов времени $2T$, $3T$ и т.д. Если этот факт не учесть, будут сформированы ложные гипотезы основных тонов.

3. *Обнаружитель нот*, объединяющий сведения об атаках узкополосных компонентов и их периодичностях и принимающий решение о высотах тонов и длительностях нот, отображающих распознаваемый музыкальный сигнал.

3.7.3. Недостатки современных систем распознавания музыкальных сигналов

Практика внедрения модели слухового восприятия (описанной, например в [99]) для получения промежуточных признаков при распознавании музыкальных сигналов показала, что такой подход не всегда оправдан. Действительно, моделирование слухового восприятия стало первым *систематизированным* подходом к распознаванию музыкальных сигналов, сменившим различные эвристические методы. Однако, большинство сложностей, описанных в разделе 3.6, преодолеть так и не удалось. Применение модели слухового восприятия, строящейся на наборе полосовых фильтров, не позволило ни повысить разре-

шающую способность алгоритмов выделения основного тона, ни повысить достоверность разделения перекрывающихся во времени звуковых объектов.

Психофизиологические модели были позаимствованы из другой области анализа звуковых сигналов — идентификации источников звука (см., например, [42]). Действительно, использование периодограммы в качестве признака позволило в смеси тональных и шумовых звуков (например, шум улицы) определять временные границы сигналов, соответствующих различным источникам (человеческая речь, лай собаки, сигнал автомобиля и т.п.). В то же время, механизм восприятия человеком высоты тона до конца не изучен, модель, описывающая восприятие тональных звуков не создана. Поэтому, основным применением существующей модели восприятия к распознаванию музыкальных сигналов, можно считать их *сегментацию* (см. главу 6).

Разделение распознаваемого сигнала на узкополосные компоненты рассматривается как альтернатива разложению в гармонический ряд Фурье. Оправдан ли такой подход? Природа тональных звуков, производимых акустическими музыкальными инструментами такова, что они лучше всего аппроксимируются именно гармоническим рядом Фурье (в главе 6 это будет подтверждено результатами эксперимента). Поэтому можно говорить, что параметры коэффициентов комплексной формы ряда Фурье (модуль и аргумент) имеют *физический смысл* — амплитуды и начальные фазы гармонических составляющих звуков. Узкополосные компоненты спектра, получаемые разделением исходного сигнала набором фильтров с полосами пропускания, совпадающими с критическими полосками слуха, по сути, физического смысла не имеют. Кроме того, узкополосные компоненты несут информацию более общего характера, чем члены гармонического ряда Фурье, что затрудняет разделение звуковых объектов, имеющих близкие частоты основных тонов.

Таким образом, основным недостатком современных систем распознавания музыкальных сигналов является неоправданное внедрение модели слухового восприятия человека. Модель эта не отражает механизма восприятия многокомпонентных сигналов. В результате основные сложности распознава-

ния музыкальных сигналов, связанные с совпадением частот гармоник музыкальных звуков, не преодолены.

3.8. Выводы

1. Выбор, назначение признаков распознавания — эвристическая операция, зависящая от творчества, изобретательности разработчика.
2. Состав признаков должен быть как можно более разносторонним и полным, независимым от того, можно или нельзя эти признаки получить.
3. Система распознавания всегда является элементом системы управления, цель создания системы распознавания — повышение эффективности решений, принимаемых при управлении.
4. Частота основного тона музыкальных 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.
5. При распознавании музыкального сигнала происходит выделение переносимой семантической информации. Эта семантическая информация частично может быть отображена нотной записью.
6. Основная сложность распознавания многоголосных мелодий состоит в отсутствии достоверных методов разделения тональных звуковых объектов, некоторые обертоны которых имеют примерно совпадающие частоты.
7. 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»

8. Основным недостатком современных систем распознавания музыкальных сигналов состоит в неоправданном внедрении модели слухового восприятия, не описывающей разделение человеческим слухом звуковых объектов, перекрывающихся в частотной и временной областях.

4 . СОВРЕМЕННАЯ ТЕХНИКА ВОССТАНОВЛЕНИЯ МУЗЫКАЛЬНЫХ СИГНАЛОВ

4.1. КОМПЬЮТЕРИЗИРОВАННЫЕ СИСТЕМЫ ВОССТАНОВЛЕНИЯ

4.1.1. Система "NoNoise"

Система реставрации "NoNoise" – программный продукт компании Sonic Studio. "NoNoise" – наиболее известное и распространённое в профессиональной звукозаписи средство обработки сигналов в процессе реставрации звукозаписей. Система используется такими крупными звукозаписывающими компаниями, как ГП «Фирма Мелодия» и «EMI Records» Ltd. Подразделение компании, занимающееся разработкой новых версий и поддержкой пользователей «NoNoise», возглавляет Джеймс Андерсон Мурер – создатель первой системы распознавания полифонии. Основные возможности NoNoise:

- подавление импульсных помех (щелчков);
- подавление широкополосного шума с предварительным обучением по фрагменту шума;
- устранение последствий ограничения мгновенных значений сигнала.

Для подавления импульсных помех используется замена повреждённых фрагментов, интерполяцией формы сигнала. Различаются два типа импульсных помех, и, соответственно, используются два типа алгоритмов интерполяции:

- *щелчки (clicks)*, т.е. мощные одиночные импульсы;

– треск (crackle), т.е. последовательность коротких импульсов малой мощности.

Интерполяция формы сигнала, утраченной из-за щелчков, производится на основе анализа формы сигнала на интервалах времени, предшествующих и следующих за щелчком. Для интерполяции непериодических сигналов анализируются фрагменты длительностью 0,5 с до и после щелчка. Для интерполяции периодических сигналов анализируется форма соседних с щелчком 6-и «периодов» сигнала. За длительность периода сигнала принимается интервал времени между наибольшими локальными максимумами сигнала, т.е. **выделение основного тона не происходит**. Пользователь сам принимает решение, периодичен сигнал или нет, в соответствии с выбором пользователя система работает по первому или второму алгоритму.

Для интерполяции формы сигнала, утраченной из-за импульсных помех, используется медианная фильтрация сигнала и сглаживание методом скользящего среднего. Для устранения последствий ограничения мгновенных значений также используется интерполяция формы сигнала.

Для работы алгоритма шумоподавления необходимо обучение: пользователь выделяет фрагмент сигнала, содержащий только шум, и этот фрагмент анализируется системой реставрации. Выделенный фрагмент разбивается на выборки по 4096 отсчётов, над которыми выполняется ДПФ. Полученные 2048 коэффициентов усредняются по всем выборкам, и вычисляется усреднённый амплитудный спектр шума. В процессе шумоподавления сигнал также обрабатывается выборками по 4096 отсчёта. Каждое из 2048 значений амплитудного спектра сигнала $|S(j\omega_i)|$ сравнивается с соответствующим значением амплитудного спектра шума $|N(j\omega_i)|$. По результатам сравнения строится АЧХ БПФ-фильтра $|H(j\omega)|$: если $|S(j\omega_i)| \leq |N(j\omega_i)|$, то $|H(j\omega_i)| < 1$.

4.1.2. Система «Канонь»

Система реставрации «Канонь» — программный продукт российской компании Digiton. В настоящее время название «Канонь» является торговой

маркой российской звукозаписывающей компании «Русский компакт-диск».

Основные возможности системы:

- подавление импульсных помех;
- подавление широкополосного шума с предварительным обучением,
- маскировка шума с помощью многополосной АРУ;
- изменение длительности сигнала без изменения тональности и изменение тональности без изменения длительности сигнала;
- психоакустическая обработка сигнала («эксайтер») для устранения ощущения узости спектра сигнала, глухости звучания.

Маскировка шума основана на свойствах слухового восприятия человека: спектральные компоненты большей мощности маскируют близкие по частоте компоненты меньшей мощности. Настроив многополосную АРУ таким образом, чтобы мощные узкополосные сигналы ещё больше усиливались или оставались без изменения, а маломощные (шумовые составляющие) подавлялись, в некоторых случаях можно добиться эффекта маскировка.

Кроме перечисленных возможностей система «Канонъ» позволяет выполнять монтаж фонограмм. «Канонъ» является программно-аппаратным средством, т.е. некоторые операции выполняются аппаратно по алгоритмам, реализованным в микросхемах, установленных на звуковой карте компьютера, а не программно.

4.2. ПСИХОАКУСТИЧЕСКИЕ ПРОЦЕССОРЫ

Психоакустические процессоры – отдельный класс комбинированных устройств обработки, применяемых в радиовещании, на телевидении, в звукозаписи, в системах звукоусиления концертных залов и площадок. Принцип действия психоакустических процессоров основан на психофизиологических особенностях восприятия человеком звука (отсюда и название – «психоакустические»).

Структурными элементами психоакустических процессоров, как правило, являются устройства «традиционной» обработки сигналов: устройства

динамической обработки, фильтры, линии задержки, генераторы гармоник. Очевидно, что при одинаковом наборе структурных элементов, многообразие подобных устройств может определяться вариантами соединения и управления составляющими прибора. Взаимосвязи структурных элементов определяются фантазией разработчика и желаемым эффектом.

4.2.1. Эксайтер (Exciter)

Первое появившееся и выпускаемое серийно устройство психоакустической обработки. Первая модель была выпущена в 1979 году компанией Арhex. Эксайтер считается наиболее универсальным психоакустическим процессором, способным улучшить звучание любого инструмента. Варианты построения эксайтера разнообразны, разнообразны и дополнительные возможности различных моделей разных производителей, поэтому рассмотрим наиболее общую схему (рисунок 14) прибора и обсудим принцип его работы.

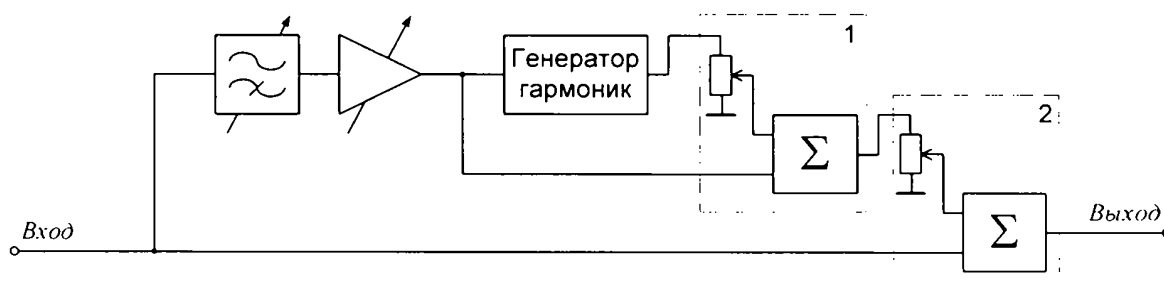


Рис.14 Структурная схема эксайтера (по [130])

Эксайтер – устройство, работающее по т.н. параллельному принципу. Это означает, что входной сигнал проходит по двум ветвям схемы: в одной он обрабатывается, в другой – остаётся без изменений. В выходном сумматоре обработанный и исходный сигнал смешиваются в устанавливаемом пользователем соотношении. Цепь обработки эксайтера состоит из регулируемого усилителя, перестраиваемого фильтра верхних частот, генератора гармоник и сумматора 1.

ФВЧ строится таким образом, чтобы намеренно внести в сигнал амплитудные и фазовые искажения. Цель таких искажений – увеличить субъективно воспринимаемую громкость коротких импульсов за счёт увеличения их длительности. «Удлинение» коротких импульсов достигается за счёт смешивания

необработанного и искажённого сигналов. Регулируемыми в фильтре являются граничная частота полосы пропускания и крутизна спада АЧХ в области подавления. Регулируя полосу пропускания фильтра, добиваются акцентирования, подчёркивания нужной области частот. Как правило, диапазон изменения граничной частоты ФВЧ ограничен несколькими сотнями герц снизу (300 Гц - 400 Гц) и несколькими килогерцами сверху (4 кГц - 7 кГц).

Усилитель с регулируемым коэффициентом усиления используется для изменения интенсивности возбуждения генератора гармоник. От интенсивности возбуждения зависит состав синтезируемых гармоник.

Генератор гармоник синтезирует гармоники *верхней* части спектра исходного сигнала, выделенной ФВЧ. Добавление гармоник приводит, кроме всего прочего, к расширению спектра сигнала. Самой мощной синтезируется вторая гармоника сигнала, как наиболее благозвучно сочетающаяся с исходным сигналом, мощность более высоких гармоник намного меньше. Присутствие нелинейного элемента (генератора гармоник) в цепи обработки с очевидностью означает, что эксайтер преднамеренно *искажает* сигнал.

Сумматор 1 формирует выходной сигнал цепи обработки, определяя долю содержания гармоник. Искажённый фильтром верхних частот сигнал в любом случае поступает для суммирования с исходным сигналом, т.е. описанный выше эффект увеличения громкости коротких импульсов присутствует всегда. Потенциометр, регулирующий долю гармоник, иногда называют Brightness, т.е. «яркость», «ясность»..

Сумматор 2 формирует выходной сигнал процессора. Потенциометр, управляющий долей обработанного сигнала, называют «регулятором глубины эффекта». Чем больше в выходной сумме обработанного сигнала, тем яснее ощущается эффект.

Итак, эксайтер – нелинейное устройство, искажающее сигнал таким образом, чтобы подчеркнуть «мелкие детали» звуковой картины, т.е. короткие импульсы, нередко маскируемые более мощными сигналами, верхние частоты спектра сигнала.

4.2.2. Энхэнсер (Enhanser)

Один из первых психоакустических процессоров. Цель применения — та же, что у эксайтера, однако принцип работы иной. Энхэнсер обладает самой простой структурной схемой из всех психоакустических процессоров. Обычно рекомендуется для обработки звучания ударных инструментов и перкуссии, поскольку подчёркивает короткие атаки звуков этих инструментов. Как и эксайтер, подчёркивает «мелкие детали звуковой картины».

Энхэнсер, как и эксайтер строится по параллельному принципу: в выходном сумматоре смешиваются обработанный и необработанный сигналы. Цепь обработки энхэнсера состоит из ФВЧ, амплитудного детектора (АД), управляющего элемента и регулируемого усилителя. Амплитудный детектор, управляющий элемент и регулируемый усилитель составляют устройство автоматической регулировки уровня. В разных моделях и у разных производителей эта обработка имеет свой уникальный алгоритм и делает различными по результату обработки всевозможные модели энхэнсеров.

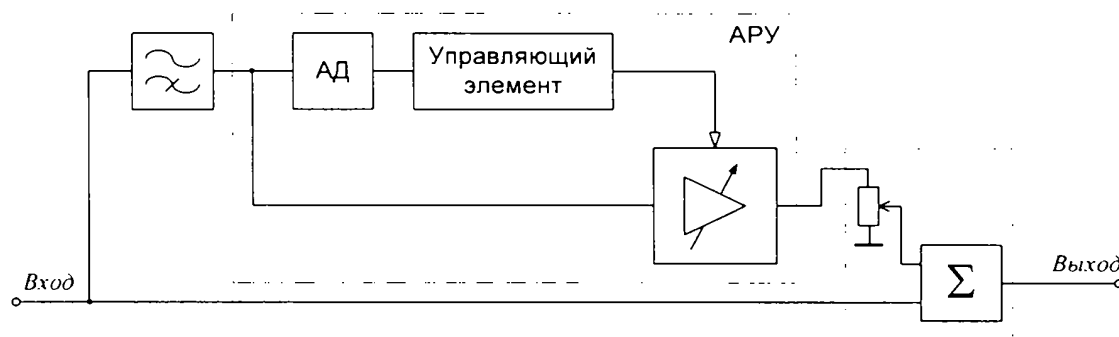


Рис 15 Структурная схема энхэнсера (по [130])

ФВЧ выделяет из спектра входного сигнала верхнюю часть для дальнейшей обработки. Характеристики фильтра (граничная частота и крутизна спада АЧХ в области подавления) различаются у разных производителей. В отличие от эксайтера, ФВЧ энхэнсера не перестраивается.

Амплитудный детектор формирует огибающую сигнала верхней части исходного спектра, выделенного ФВЧ.

Управляющий элемент анализирует огибающую, выделенную амплитудным детектором и на основе анализа управляет коэффициентом усиления

регулируемого усилителя. Алгоритмы управления могут быть различными. Наиболее простые цепи обработки энхэнсеров работают как пороговые подавители (gate): если энергия верхней части спектра меньше некоторого порога, то усилитель запирается и содержание верхних частот при суммировании в исходном сигнале остаётся неизменным, напротив, если содержание верхних частот в сигнале велико, они усиливаются. Есть цепи обработки, реагирующие на динамику изменения огибающей сигнала верхней части спектра: чем быстрее нарастает энергия в верхней части спектра, тем больше усиление и наоборот, тем самым, например, подчёркиваются «атаки» ударных металлических инструментов.

Выходной сумматор смешивает обработанный и необработанный сигналы, регулируя степень введения эффекта.

Энхэнсер является линейным устройством, регулирующим энергию высокочастотных составляющих сигнала. Особенность обработки состоит в том, что подавление верхних частот никогда не происходит: их содержание или остаётся неизменным, или они подчёркиваются. Кроме того, увеличивается крутизна атак звуков, обладающих быстро нарастающим и быстро затухающим характером.

4.2.3. Максимайзер (Maximizer)

Максимайзер — устройство, впервые выпущенное компанией ВВЕ и производшее, пожалуй, самое сильное впечатление на звукорежиссёров. Однако, несмотря на производимый эффект, максимайзер не является устройством универсальным, т.е. одинаково эффективно обрабатывающий сигналы всех типов звучания. Схема максимайзера (рисунок 16) построена по последовательному принципу: единственная цепь обработки сигнала состоит из последовательно соединённых элементов, управляемых элементами цепи управления.

Цепь обработки состоит из двух элементов — двух регуляторов тембра: верхних и нижних частот. Регулятор тембра — активный фильтр с переменным коэффициентом усиления; коэффициент усиления принимает значения

от 0 до нескольких десятков (на практике – до 10дБ-20дБ. Граничные частоты регуляторов тембра максимайзера не регулируются. Цепь управления состоит из двух фильтров: полосового фильтра средних звуковых частот и фильтра верхних частот.

Коэффициент усиления регулятора тембра НЧ устанавливается вручную пользователем. Поскольку основной и иногда очень глубокой обработке подлежат верхние частоты, регулировка энергии НЧ необходима для создания сбалансированного звучания верхней и нижней частей спектра.

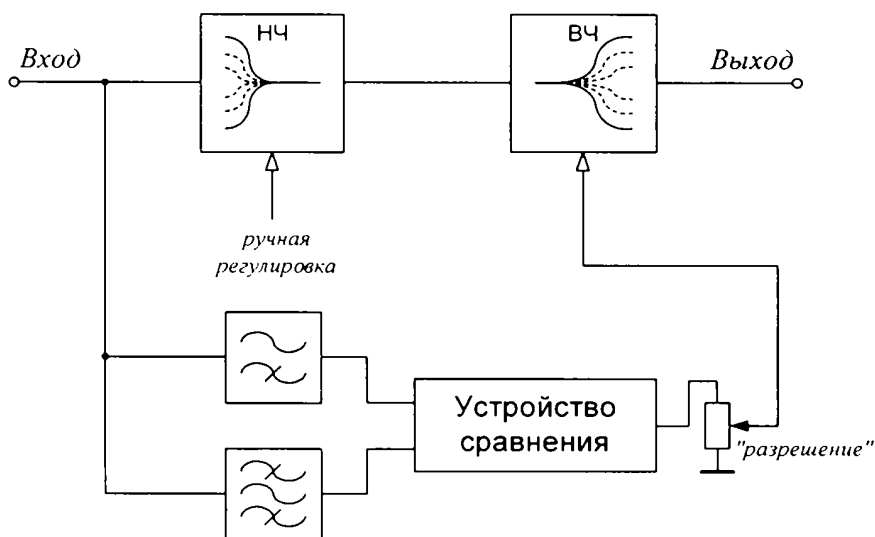


Рис. 16 Структурная схема максимайзера (по [130])

Полосовой фильтр и ФВЧ цепи управления выделяют из исходного сигнала области частот, имеющие общую границу. Сигналы с выхода этих фильтров необходимы для работы алгоритма обработки верхних частот. Отличие максимайзеров разных производителей, в частности, состоит в границе раздела полос пропускания этих фильтров. Фильтры делаются неперестраиваемыми, поскольку алгоритм обработки тесно связан с этой граничной частотой, её изменение могло бы привести к потере производимого эффекта.

Устройство сравнения выносит решение о регулировке энергии верхних частот спектра. Анализируя соотношение в спектре сигнала средних и верхних частот (сравнивая мощность выходных сигналов ФВЧ и фильтра средних частот) устройством сравнения принимается решение о коэффициенте усиления регулятора тембра ВЧ.

Коэффициент усиления регулятора тембра ВЧ управляется полностью автоматически. Вмешательство пользователя ограничивается диапазоном регулировки потенциометра «разрешение» (Definition), изменяющего уровень сигнала управления. Т.е. изменяя положение потенциометра, пользователь меняет пределы изменения коэффициента усиления регулятора тембра, т.е. степень введения эффекта, интенсивность обработки верхней части спектра.

Важно отметить, что в отличие от энхэнсера, решение об усилении или подавлении верхних частот происходит именно на основе сравнения энергий средних и верхних частот. Эта особенность работы алгоритма обработки делает максимайзер «узкоспециализированным» устройством. При прохождении сигналов, небогатых высокочастотными составляющими (например, звуки деревянных духовых инструментов), устройство сравнения «обманывается». Происходит усиление, по сути, мешающих высокочастотных звуков: шумы микрофона и электрического тракта, в случае духовых: звук вдуваемого воздуха, стук клапанов. Иначе работает максимайзер при обработке широкополосных сигналов: богатых обертонами музыкальных звуков, сведённых музыкальных композиций. Описанная особенность максимайзера ограничивает область его применения: как правило, он используется для окончательной обработки сведённых композиций, сбалансированных по тембру.

4.2.4. Виталайзер (Vitalizer)

Прибор разработан и выпускается немецкой фирмой SPL. Виталайзеры выпускаются этой фирмой в нескольких моделях и под разными названиями: от просто Vitalizer до Psycho Dynamic Processor – «психодинамический процессор». Виталайзер построен по параллельному принципу: верхняя и нижняя части спектра сигнала обрабатываются разными цепями независимо друг от друга.

Регулятор НЧ представляет собой активный фильтр нижних частот, коэффициент усиления которого больше или равен единице, т.е. возможно лишь усиление нижних частот. Особенность этого блока состоит ещё и в том, что от направления вращения регулятора зависит полярность выходного сигнала

блока. В результате сигнал НЧ составляющих подмешивается к исходному сигналу то синфазно, то противофазно.

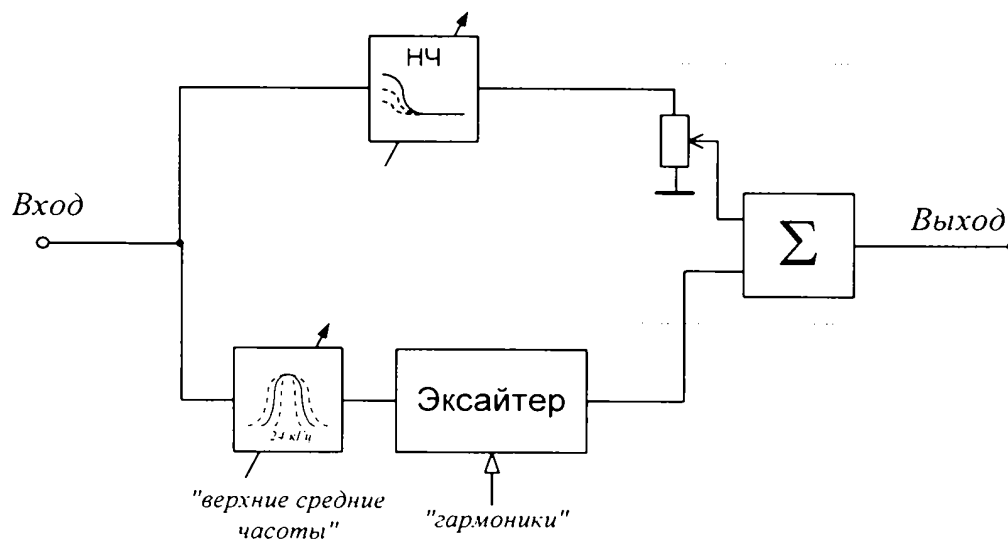


Рис 17 Структурная схема виталайзера (по [130])

Регулировка тембра на средних и высоких частотах в виталайзере осуществляется двумя регуляторами - «верхние средние частоты» (Mid-High Tune) и «гармоники» (Harmonics). Первый из этих регуляторов управляет добротностью полосового фильтра с центральной частотой примерно 24 кГц, которая при этом остаётся неизменной. При высокой добротности осуществляется подъем только высших частот звукового спектра. При снижении добротности (поворот регулятора Mid-High Tune в сторону более низких частот) - полоса пропускания фильтра расширяется, осуществляется подъем не только наивысших составляющих спектра, но и более низких. Очевидно, что при этом более высокие частоты усиливаются сильнее, чем более низкие. Второй регулятор виталайзера - ручка Harmonics. Это регулятор уровня выходного сигнала встроенного в виталайзер эксайтера, подмешиваемого в общий сигнал эффекта. Этот эксайтер - один из самых простейших, и кроме регулятора уровня имеет только регулятор частоты настройки, совмещенный конструктивно с ручкой регулятора Mid-High Tune.

4.3. ОБЩАЯ ХАРАКТЕРИСТИКА СОВРЕМЕННОЙ ТЕХНИКИ ВОССТАНОВЛЕНИЯ И ОБОСНОВАНИЕ ПРЕДЛАГАЕМОГО ПОДХОДА

Прежде всего, сформулируем подходы, на основе которых выполняется восстановление музыкальных сигналов современными средствами. Важной особенностью и программных, и аппаратных средств является их нацеленность на *уменьшение заметности* искажений. Иными словами, целью восстановления является создание у слушателя ощущения отсутствия искажений. Нередко эта цель достигается внесением в сигнал дополнительных нелинейных искажений (синтез гармоник сигнала с помощью нелинейного элемента). Обнаружение дефектов происходит посредством субъективной оценки звучания. По субъективными критериями руководствуются и при выборе устройств обработки сигнала, при настройке этих устройств. Действительно, квалифицированные звукорежиссёры, руководствуясь опытом, в большинстве случаев умеют обнаружить дефекты, установить их причину и правильно выбрать средства для уменьшения заметности искажений. Однако существующие средства, нередко подавляя компоненты сигнала, вызывающие у слушателей ощущение искажений, подавляют и компоненты, переносящие полезную информацию. Образно выражаясь, инструмент реставратора оказывается грубее реставрируемого объекта. Однозначной взаимосвязи параметров музыкального сигнала и его эстетической информативности на сегодняшний день не выявлено. Поэтому не оценён и отрицательный эффект применения современных средств обработки.

Современные комбинированные устройства обработки звуковых сигналов вызывают много споров и разногласий среди профессионалов: кому-то нравится, как «звучит» то или иное устройство, кому-то — нет. Кто-то считает, что современный процесс формирования сигналов вещания может и должен обходиться без психоакустических процессоров, кто-то — напротив, стремится максимально разнообразить и усложнять обработку сигнала. Тем не менее, описанные устройства пользуются неизменным спросом, постоянно идёт разработка выпуск всё новых и новых устройств. Интересная особенность разработки психоакустических процессоров, состоит в том, что производители, видимо опасаясь конкуренции, всячески скрывают информацию о принципе работы своих устройств, предоставляя потребителя «чёрный ящик» с одним -

боты своих устройств, предоставляя потребителя «чёрный ящик» с одним - двумя регуляторами, «улучшающий звучание». В результате, большинство звукорежиссёров и звукоинженеров даже не представляют суть процессов, происходящих внутри устройства. Поэтому нередко использование некоторых устройств приводит к обратному эффекту – искажению звучания и отказу от использования. При этом цена большинства описанных устройств составляет несколько сотен и даже тысяч долларов (в зависимости от набора функций, комплектации и исполнения).

Разнообразие комбинированных устройств обработки вызвано в основном двумя факторами: 1) развитием теории и практики цифровой обработки сигналов и 2) особенно возросшим в последнее время интересом к психофизиологическим особенностям слухового восприятия (в том числе к «зомбированию»). Технологии цифровой обработки позволяют производителям реализовывать свои самые смелые проекты, сложноосуществимые или неосуществимые в аналоговом виде. Кроме того, использование современных процессоров обработки сигнала позволяет создавать компактные устройства: большая часть полезного пространства занята входными и выходными усилителями и трансформаторами. Новейшие результаты исследований психоакустики позволяют, грубо говоря, ещё лучше обманывать слух потребителей звуковых программ, создавая впечатление «ещё более естественного и живого звучания». В действительности, как мы выяснили, почти все описанные устройства искажают сигнал, причём вводимые искажения ещё 15 лет назад считались недопустимыми (по объективным параметрам, таким, например, как коэффициент нелинейных искажений). Однако слух всё-таки обманывается, и искажённый сигнал звучит предпочтительнее, чем неискажённый. В связи с этим можно заметить, что подобное состояние дел и направление развития техники обработки звуковых сигналов *закономерно*, и вызвано необходимостью повышения именно эстетического качества звуковых программ в условиях ограниченных технических (а часто и профессиональных, квалификационных) возможностей.

Обсудим подход к подавлению широкополосных шумов, предлагаемый разработчиками системы «NoNoise» и используемый во многих других про-

граммных средствах (например, Cedar, DART). Речь идёт об обучении системы шумоподавления по короткой выборке. Известно, что большинство шумовых процессов, возникающих в трактах формирования программ звукового вещания, включающих тракты записывающих и воспроизводящих устройств, не являются стационарными. Т.е. подобное обучение, тем более с использованием в качестве признака амплитудного спектра, получаемого с помощью БПФ некорректно с точки зрения оптимизации фильтрации по критерию соотношения сигнал/шум. Для современных систем восстановления характерно, что фильтрация производится с использованием сведений о шуме, а не сведений о сигнале. Подобный подход можно назвать фильтрацией с использованием *обнаружения*. Т.е. по некоторому признаку (в данном случае амплитудному спектру) принимается решение о наличии или отсутствии обнаруживаемого процесса. Обнаруживаемым процессом является шум. Компоненты, не классифицированные (при простом пороговом сравнении) как компоненты шума признаются компонентами полезного сигнала. Компоненты шума подавляются. Недостатки такого подхода: компоненты полезного сигнала, перекрывающиеся в частотной области с компонентами шума и имеющие равную мощность, подавляются, в то же время, спектральные компоненты, не классифицированные как шумовые, но не являющиеся и компонентами полезного сигнала (например, продукты нелинейных искажений) подавлены не будут. Таким образом, очевидна необходимость классификации компонентов спектра не как «шумовых» и «нешумовых», но как «полезных» и «шумовых». Спектры музыкальных звуковых объектов крайне редко содержат один компонент, следовательно,

классификация компонентов спектра как «полезных» немыслима без их соотнесения со структурами звуковых объектов.

Т.е. необходима классификация не только компонентов спектра, но и классификация звуковых объектов по длительности и частоте основного тона. На основе этих сведений описывается структура объектов, составляющих восстанавливаемый сигнал. И именно с этой структурой должны соотноситься спектральные компоненты для классификации их как «полезных». Классифи-

кация звуковых объектов по длительности и частоте основного тона по определению представляет собой *распознавание образов*.

Итак, для классификации спектральных компонент как относящихся к полезному сигналу необходимо распознавание образов. Распознавание образов позволяет сформировать описание структуры звуковых объектов, переводимое на язык машины. Соотнесение спектральных компонент, составляющих музыкальный сигнал со структурами распознанных звуковых объектов, позволяет, кроме того, констатировать утрату компонент полезного сигнала. Т.е., зная основной тон звукового объекта, очевидно, можно установить, мощности каких обертонов ниже уровня, принимаемого за пороговый. Очевидно, становится возможным согласованное проведение фильтрации сигнала и синтеза утраченных компонент.

Для формирования общей картины состояния современной техники восстановления музыкальных сигналов, построим (рис. 18) обобщённую схему процесса реставрации современными средствами и сравним её с общей схемой предлагаемого подхода (рис. 2). В схеме на рис. 18 также как и в схеме на рис. 2 выделим ветвь управления и ветвь обработки сигнала. Формирование ЧХ фильтра шумоподавления производится под управлением устройства, сравнивающего амплитудные спектры обрабатываемого сигнала и обучающей выборки. При таком подходе в качестве порога при принятии решения о значении модуля коэффициента передачи в данном диапазоне частот используется мощность шума обучающей выборки в соответствующей области спектра. Возможен вариант, когда порог определяется динамически – на основе анализа спектра сигнала. При таком подходе «шумовыми» компонентами спектра, подлежащими подавлению признаются компоненты, мощность которых меньше этого динамического порога. В таком случае не требуется обучения, но и шумоподавление эффективно лишь в случае, когда отношение сигнал/шум много больше 1. В любом случае, формирование ЧХ фильтра шумоподавления на основе сравнения модулей коэффициентов комплексной формы ДПФ с неким порогом отражает существо современных подходов к подавлению широкополосных шумов и продуктов нелинейных искажений.

Синтез спектральных составляющих, как мы говорили, не имеет целью восстановление структуры объектов, составляющих сигнал. Цель такой обработки – создание ощущения наличия высокочастотных составляющих спектра. Характерно, что происходит синтез гармоник сигнала целочисленным умножением частоты.

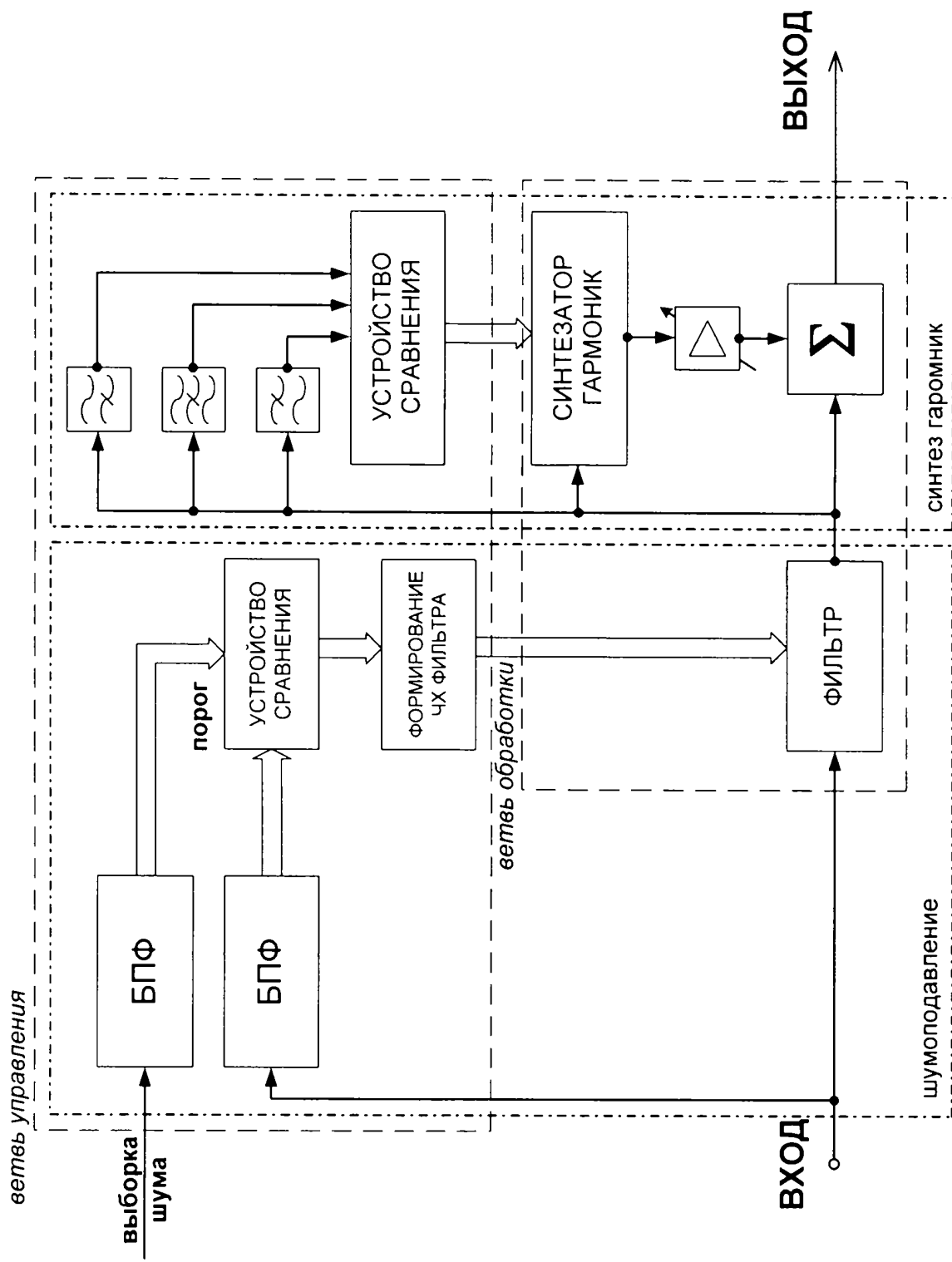


Рис. 18 Обобщённая схема процесса реставрации современными средствами

Для синтеза гармоник из сигнала может быть выделена полоса верхних или нижних частот. Мощность синтезируемых гармоник определяется при анализе распределения энергии сигнала по частоте. Сравнивается мощность выходных сигналов ФНЧ, полосового фильтра и ФВЧ, имеющих полосы пропускания, соответствующие низким, средним и высоким звуковым частотам. Недостатком предлагаемого способа синтеза является его «неизбирательность». На синтезатор гармоник поступает сигнал, содержащий кроме полезных спектральных составляющих и неподавленные на стадии шумоподавления продукты нелинейных искажений, компоненты шумов. Т.е. после суммирования исходного сигнала с выходным сигналом синтезатора гармоник исходный спектр обогащается новыми шумовыми составляющими.

4.4. Выводы

1. Работу современных алгоритмов шумоподавления можно назвать шумоподавлением на основе обнаружения шума (в отличие от предлагаемого шумоподавления на основе распознавания звуковых объектов): спектральные компоненты классифицируются как «шумовые» и «нешумовые», «шумовые» компоненты подавляются.
2. В современных подходах к реставрации целью синтеза новых спектральных составляющих является создание у слушателя ощущения наличия высокочастотных составляющих, широкой полосы сигнала, а не восстановление утраченных компонентов структур звуковых объектов.
3. Необходимость использования распознавания образов для восстановления музыкальных сигналов обусловлена потребностью в структурированном описании звуковых объектов на языке, понятном машине, для классификации компонентов спектра как «полезных» и «шумовых».

5. ИССЛЕДОВАНИЕ СТРУКТУРЫ ЗВУКОВЫХ ОБЪЕКТОВ

5.1. СТРУКТУРА ЗВУКОВОГО ОБЪЕКТА И ТЕМБР

Проведение данного исследования необходимо для изучения свойств звуковых объектов, подлежащих распознаванию. Необходимо оно и для разработки алгоритма распознавания музыкальных сигналов, и для формирования требований к разрабатываемой системе обработки сигнала. Однако существуют и ещё один важный вопрос, возникающий в связи с исследованием структуры звуковых объектов. Обсуждая перечень задач распознавания музыкальных сигналов, мы говорили, что к ним также относится классификация тембра, т.е. идентификация источника сигнала. Хотя мы и ограничились классификацией лишь длительности и высоты тона, целесообразно исследовать связь структуры звукового объекта с ощущениями тембра.

Современная акустика не даёт определения понятия «тембр». Связь тембра с объективными параметрами сигнала остаётся изученной не до конца. Результаты данного исследования позволяют отчасти ответить на вопрос, чем отличаются друг от друга звуковые колебания различных инструментов и чем похожи сигналы, синтезированные однотипными инструментами.

Уже Гельмгольцу, основоположнику классической теории тембра, было очевидно, что усреднённый амплитудный спектр ни в коем случае не описывает тембра, реальный звуковой сигнал не является строго стационарным даже на стадии установившихся колебаний. Соотношение амплитуд, частот и начальных фаз гармонических составляющих реальных звуков неустойчиво и постоянно изменяется. Более того, на сегодняшний день принято считать, что субъективное восприятие тембра звука в большей степени зависит от участков нестационарности сигнала: атак и спадов. Аддитивный синтез, не учитывающий этого обстоятельства, больших успехов не добился: попытки синтезировать реальные тембры с помощью немодулированных однотональных сигналов даже не приблизились к своей цели.

Сегодня изучение тембра направлено именно на исследование изменений внутренней структуры звуковых объектов, как во временной, так и в частотной областях. «Для оценки тембра звучания важен не только момент его распознавания (т.е. способность отличить один инструмент от другого), но и возможность оценить изменение тембра в процессе исполнения. Здесь важнейшую роль играет динамика изменения спектральной огибающей во времени на всех этапах звучания: атаки, стационарной части, спада.

Характер поведения каждого обертона во времени также несёт важнейшую информацию о тембре. ... Поведение различных обертонов во времени зависит от типа инструмента: в звучании рояля, органа, гитары и др. процесс изменения амплитуд обертонов имеет совершенно разный характер.

Опыт показывает, что аддитивный синтез, учитывающий специфику развёртывания отдельных обертонов во времени, позволяет получить значительно более «жизненное» звучание»([29,30]).

5.2. ОПИСАНИЕ ИССЛЕДОВАНИЯ

5.2.1. Цели и задачи исследования

Целью исследования ставится **построение адекватной модели звуковых объектов, пригодной для синтеза и идентификации тембра.**

Для достижения этой цели в процессе исследования решим следующие задачи:

1. наблюдение изменений параметров гармонических составляющих звуковых объектов, т.е. процесса изменения их амплитуд и фаз на всей длительности объекта;
2. выявление общих закономерностей в развёртывании гармонических составляющих у разных музыкальных инструментов;
3. изучение влияния интенсивности возбуждения и способа извлечения звука на процесс развития гармонических составляющих;
4. исследование взаимосвязи гармонических составляющих звуковых объектов.

Кроме того, попытаемся установить связь гармонической структуры звукового объекта с тембром.

5.2.2. Объект исследования

Под *звуковым объектом* будем понимать минимальный по длительности элементарный звуковой сигнал, воспринимаемый как самостоятельное единое целое (для каждого звукового объекта можно определить высоту тона и тембр). Музыкальным звуковым объектом является фрагмент сигнала, соответствующий отдельной ноте от момента извлечения до момента затухания. Речевым звуковым объектом может быть отдельная фонема или слог.

Объектом данного исследования являются звуковые сигналы, произведённые механико-акустическими инструментами, соответствующие отдельным нотам. Т.е. исследуемые звуковые объекты не являются фрагментами исполнения какой-либо мелодии и не являются результатом электронного или электромеханического синтеза.

Исследованию подлежат звуки различных музыкальных инструментов: инструменты симфонического оркестра (духовые, смычковые), фортепиано, гитара, ударные инструменты с определённой высотой тона. В свою очередь, звуки каждого инструмента отличаются по высоте тона, а для некоторых инструментов – ещё и по интенсивности возбуждения и приёму исполнения (стаккато, пиццикато).

5.2.3. Методика и последовательность исследования

При анализе сигналов будем опираться на следующие предположения:

1. Все исследуемые сигналы соответствуют одной ноте, а не созвучию.
2. Частоты натуральных призвуков основного тона (обертонов) примерно кратны частоте основного тона.
3. Сигнал является квазипериодическим: длительность полных циклов колебаний имеет незначительные отклонения от некоего среднего значения. На стадиях установления и затухания колебаний (атаки и спада) эти отклонения не нормируются, а на установившейся стадии отклонение периода соответствует изменению высоты тона не более четверти тона.

4. Средняя продолжительность полного цикла колебаний равна длительности одного такого цикла на установившейся стадии колебаний.
5. Сигнал можно неким способом разбить на участки, имеющие примерно одинаковую длительность и имеющие топологически одинаковую форму, т.е. определить временные границы каждого полного цикла колебаний.
6. Для каждого инструмента существуют закономерности развития обертонов звукового объекта, не зависящие от высоты тона и интенсивности возбуждения.

Структура любого объекта – это взаимосвязи составляющих его элементов. Мы будем исследовать поведение и взаимосвязь элементов звуковых объектов – его обертонов, поэтому исследование и названо «исследованием структуры звуковых объектов».

Исследование каждого звукового объекта разобьём на следующие этапы:

1. определение ориентировочной продолжительности полного цикла колебаний T ;
2. разбиение всего звукового объекта на непересекающиеся фрагменты, соответствующие отдельным полным циклам колебаний;
3. для каждого цикла колебаний выполним разложение в гармонический ряд Фурье на длительности этого цикла (количество членов ряда обсудим ниже);
4. средние значения амплитуд и начальных фаз (для краткости далее – просто фаз) обертонов в данном цикле колебаний примем равными соответственно модулям и аргументам коэффициентов комплексной формы полученного ряда Фурье (такое приближение столь же допустимо, сколь и аппроксимация колебаний интонированных музыкальных инструментов гармоническими функциями).

В итоге мы получим набор значений амплитуд и фаз обертонов на каждом полном цикле колебаний сигнала, зная при этом точное значение дли-

тельности каждого такого цикла. Введём следующее понятие, которым нам будет легче оперировать:

*назовём зависимость амплитуды (фазы) обертона от времени **траекторией** амплитуды (фазы).*

Мы определяем лишь среднее значение амплитуд и фаз обертонов для каждого цикла колебаний (функция, дискретная по времени), поэтому в нашем случае под траекторией амплитуды (фазы) обертона будем понимать зависимость среднего значения амплитуды (фазы) в данном цикле колебаний от номера цикла. По аналогии траекторией длительности полного цикла колебания будем называть зависимость длительности цикла сигнала от его номера.

Вычислив траектории амплитуд и фаз обертонов и выполнив их графическое изображение, приступим к их анализу:

1. выявим особенности траекторий амплитуд и фаз обертонов для каждого инструмента в зависимости от высоты тона и способа извлечения звука: соотношения амплитуд и фаз, особенности убывания, флуктуаций.
2. выявим схожие черты траекторий для каждого инструмента независимо от высоты тона и способа извлечения, а также различия в траекториях амплитуд и фаз обертонов разных инструментов;
3. выявим степень корреляции траекторий амплитуд обертонов как случайных функций времени;
4. попытаемся установить связь между способом извлечения звука на данном инструменте и особенностями изменения во времени параметров обертонов извлечённых звуков;
5. попытаемся установить связь между особенностями траекторий амплитуд и фаз обертонов и субъективным ощущением тембра.

На основе проведённого анализа опишем модель звукового объекта.

5.2.4. Инструменты исследования

Исследование проведено с помощью персонального компьютера. В качестве звукового материала взяты сэмплы — фрагменты звучания механико-акустических музыкальных инструментов, используемые для т.н. волнового

синтеза (Wavetable synthesis). Все исследуемые сэмплы удовлетворяют нашему определению звукового объекта. Для анализа они представлены в цифровом формате записи RIFF WAVE. Частота дискретизации для большинства имеющих в распоряжении автора фрагментов 44100 Гц, квантование – 16-битное.

Для решения поставленных задач автором было разработано программное обеспечение, снабжённое пользовательским интерфейсом и выполняющее следующие операции: чтение файлов формата RIFF WAVE, определение частоты основного тона, разложение в комплексный ряд Фурье, формирование траекторий амплитуд и фаз обертонов, их графическое отображение и сохранение в файле. Кроме того, имеется возможность сохранить текстовый отчёт о параметрах исследуемого сигнала и результатах анализа.

Определение частоты основного тона сигнала выполняется комбинированным методом:

1. вычисляется кратковременная автокорреляционная функция сигнала и определяется положение её максимума τ_0 ;
2. выполняется дискретное преобразование Фурье на длительности $2\tau_0$, вычисляется амплитудный спектр;
3. определяется положение максимумов полученного амплитудного спектра и находится наименьший общий делитель F' частот, соответствующих этим максимумам;
4. за период сигнала принимается значение $T=1/F'$.

Следующей стадией является разметка сигнала, т.е. определение временных границ каждого полного цикла колебаний. Разметка может производиться по выбору пользователя двумя способами: по нулям и по максимумам.

При разметке по максимумам за длительность данного цикла колебаний сигнала условно принимается расстояние между двумя локальными (местными) максимумами сигнала, отстоящими друг от друга на расстояние $T \pm \tau$ (T – длительность периода основного тона, найденное описанным выше способом). Для разметки находится наибольший локальный максимум сигнала – опорный (его положение по времени обозначим T_{on}), положение следующего ло-

кального максимума определяется поиском в интервале $[T_{оп} + T - \Delta t, T_{оп} + T + \Delta t]$ локального максимума. Величину Δt устанавливает пользователь. Положение остальных локальных максимумов находится аналогично: последовательно удаляясь от наибольшего максимума в обе стороны. Алгоритм разметки по максимумам поясняется рис. 19. За длительность данного цикла колебаний принимается величина T' .

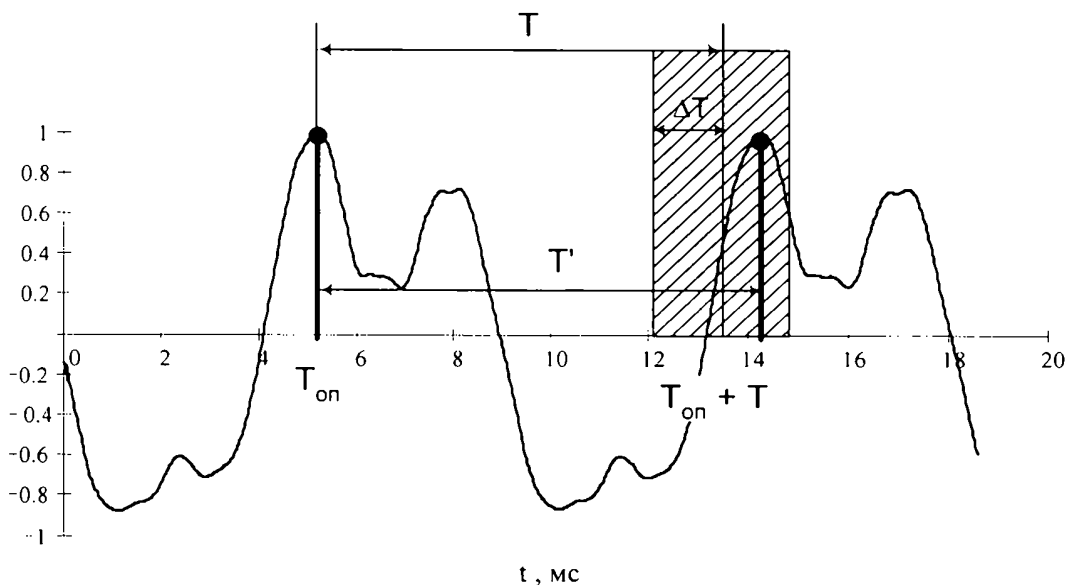


Рис. 19 Разметка по максимумам

При разметке по нулям за длительность данного цикла колебаний условно принимается расстояние между двумя нулевыми значениями сигнала, находящимися на расстоянии $T \pm t$. Для разметки находится ближайшее к наибольшему локальному максимуму нулевое значение (или переход через нуль) – опорное (его положение по времени обозначим $T_{оп0}$). Положение следующего нулевого значения (перехода через нуль) определяется поиском в интервале $[T_{оп0} + T - \Delta t, T_{оп0} + T + \Delta t]$ такого значения сигнала, для которого выполняются два условия:

1. значение сигнала точно или приблизительно равно опорному значению;
2. совпадают знаки производной сигнала в данной и в опорной точках.

Такой сложный алгоритм необходим потому, что сигнал не обязательно примет нулевое значение в интервале $[T_{оп} + T - \Delta t, T_{оп} + T + \Delta t]$, поскольку фор-

ма сигнала меняется от периода к периоду из-за изменения соотношению фаз и амплитуд обертонов. Кроме того, возможно два и более перехода через нуль в интервале $[T_{on} + T - \Delta t, T_{on} + T + \Delta t]$, тогда мы должны выбрать ту точку, в которой совпадает знак производной со знаком производной в опорной точке. Алгоритм иллюстрирует рис. 20

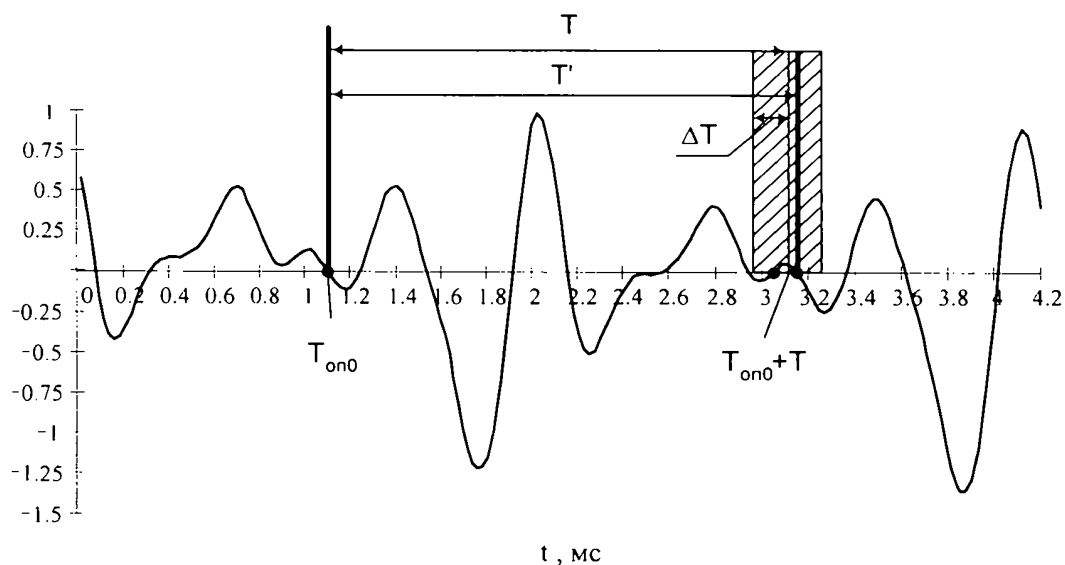


Рис. 20 Разметка по нулям

После разметки сигнала мы точно знаем временные границы каждого полного цикла колебаний, а значит, и длительность каждого цикла, т.е. уже можем составить траекторию длительности цикла колебаний сигнала.

Позиции (номера) граничных отсчётов каждого периода хранятся в памяти, пользуясь ими, мы считываем из входного файла массив отсчётов, соответствующих одному полному циклу и выполняем дискретное преобразование Фурье **на длительности данного цикла**. Взвешивание отсчётов (наложение временного окна) не производится. В этом нет необходимости, т.к. каждый период ограничен с обеих сторон примерно равными значениями. Количество членов ряда Фурье, т.е. принимаемых в расчёт обертонов может принимать по выбору пользователя значения от 1 до 40 (количество обертонов ограничено допустимым объёмом используемой оперативной памяти компьютера). По коэффициентам разложения вычисляются значения амплитуд и фаз обертонов в

данном цикле колебаний. Усреднённые по всем полным циклам значения амплитуд и фаз обертонов составят *усреднённый спектр* звукового объекта.

Для оценки статистической зависимости траекторий амплитуд обертонов как функций времени вычисляется матрица коэффициентов корреляции совокупности случайных величин, составленных траекториями амплитуд обертонов. Элемент матрицы R_{ij} равен коэффициенту корреляции траекторий амплитуд i -ого и j -ого обертонов. Коэффициент корреляции R_{ij} вычисляется следующим образом:

$$R_{ij} = \frac{\sum_{k=0}^{N-1} (A_k^{(i)} - a^{(i)}) \cdot (A_k^{(j)} - a^{(j)})}{\sigma_i \cdot \sigma_j}, \text{ где} \quad (6)$$

$A_k^{(i)}$ – значение траектории амплитуды i -го обертона на k -ом цикле колебаний;

$a^{(i)}$ – среднее значение траектории амплитуды i -ого обертона;

σ_i – дисперсия траектории амплитуды i -ого обертона;

N – количество полных циклов колебаний, уложившихся на длительности данного звукового объекта.

В текстовом отчёте выводятся сведения о входном звуковом файле (частота дискретизация, число разрядов квантования, длительность) и сведения о звуковом объекте (средняя длительность полного цикла колебаний, соответствующая ей высота основного тона, матрица коэффициентов корреляции, таблица амплитуд и фаз усреднённого спектра звукового объекта).

5.3. РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

5.3.1. Классификация музыкальных инструментов

Прежде всего, полученные результаты позволяют предложить *альтернативную классификацию* музыкальных инструментов. В основе предлагаемой классификации лежит тип колебаний звукопорождающего тела. На следую-

шем уровне классификация происходит по форме или материалу звукопорождающего тела. Будем различать два типа колебаний:

Свободные колебания — имеет место кратковременное поступление извне энергии в систему и затухание колебаний, вызванное всевозможными потерями.

Автоколебания — колебания, возникающие в замкнутой нелинейной системе. Параметры автоколебаний зависят от конструкции и параметров системы, но не от начальных условий. Автоколебательная система включает 4 элемента:

- источник энергии;
- клапан, регулирующий поступление энергии в колебательную систему;
- колебательная система;
- цепь обратной связи (между колебательной системы и клапаном).

Схема автоколебательной системы представлена на рис. 21.

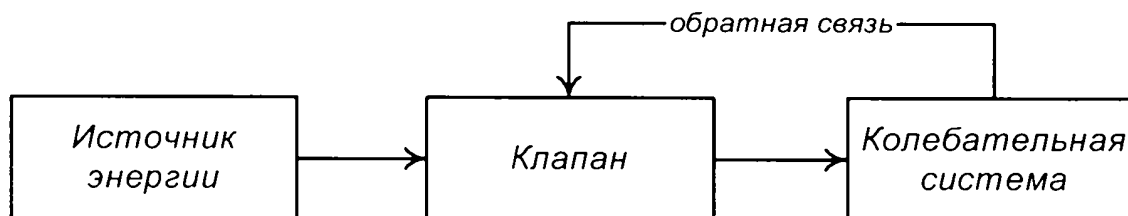


Рис. 21 Общая схема автоколебательной системы

Одна из особенностей автоколебаний в отличие состоит в том, что поступление энергии в колебательную систему не идёт непрерывно, оно регулируется клапаном, в свою очередь управляемым колебательной системой.

По мнению автора, именно такая классификация колебаний механико-акустических музыкальных инструментов наиболее удобна при решении задач узнавания тембра человеком, т.е. определения звучащего инструмента. Каждый человек, даже не имеющий музыкального образования и не знающий названий всех музыкальных инструментов, способен отделить колебания зату-

хающие от вынужденных. Каждый человек определённо скажет, колеблется струна или металлическая пластинка.

Предполагается, что при распознавании звучащего инструмента мы опираемся на возникающее представление о колеблющемся теле, происходящими с ним процессами и действиях исполнителя. У каждого из нас, когда мы слышим незнакомый или малознакомый звук, в первую очередь возникают ассоциации вроде «чем-то ударили по чему-то», «кто-то дует», «дёрнули струну», «что-то разбилось», «что-то упало». Возвращаясь к классификации типов колебаний, естественно, необходимо указать, колебания каких инструментов к какому классу относятся.

В основе принятой в инструментальной акустике классификации музыкальных инструментов (см. рис. 22) лежат особенности *конструкции* музыкальных инструментов. Все инструменты поделены условно на шумовые (ударные) и неударные (тональные). Ударные инструменты делятся на инструменты с определённой высотой тона и инструменты с неопределённой высотой тона. Тональные инструменты делятся по типу колеблющегося тела на струнные, духовые, и язычковые. Дальнейшая классификация идёт по конструкционным особенностям: духовые — условно делятся на медные и деревянные, струнные — на грифовые и безгрифовые. В случае духовых инструментов материал, из которого изготовлен инструмент неважен (саксофоны могут содержать лишь некоторые деревянные детали, а флейты могут быть полностью изготовлены из дерева, хотя и те и другие относятся к деревянным духовым). Конструкционное различие между медными и деревянными духовыми заключается в наличии мундштука у медных духовых инструментов. Медные духовые, в свою очередь различаются по конструкции клапанов и вентиляей, а также по конструкции самой трубы. Грифовые инструменты делятся на ладовые и безладовые. Далее, группы инструментов делятся по способу звукоизвлечения, т.е. возбуждения колеблющегося тела: струнные делятся на щипковые, смычковые и ударные, деревянные духовые — на лабиальные (губные) и язычковые. На самом высоком уровне классификации музыкальные инструменты делятся по диапазону высот воспроизводимых тонов.

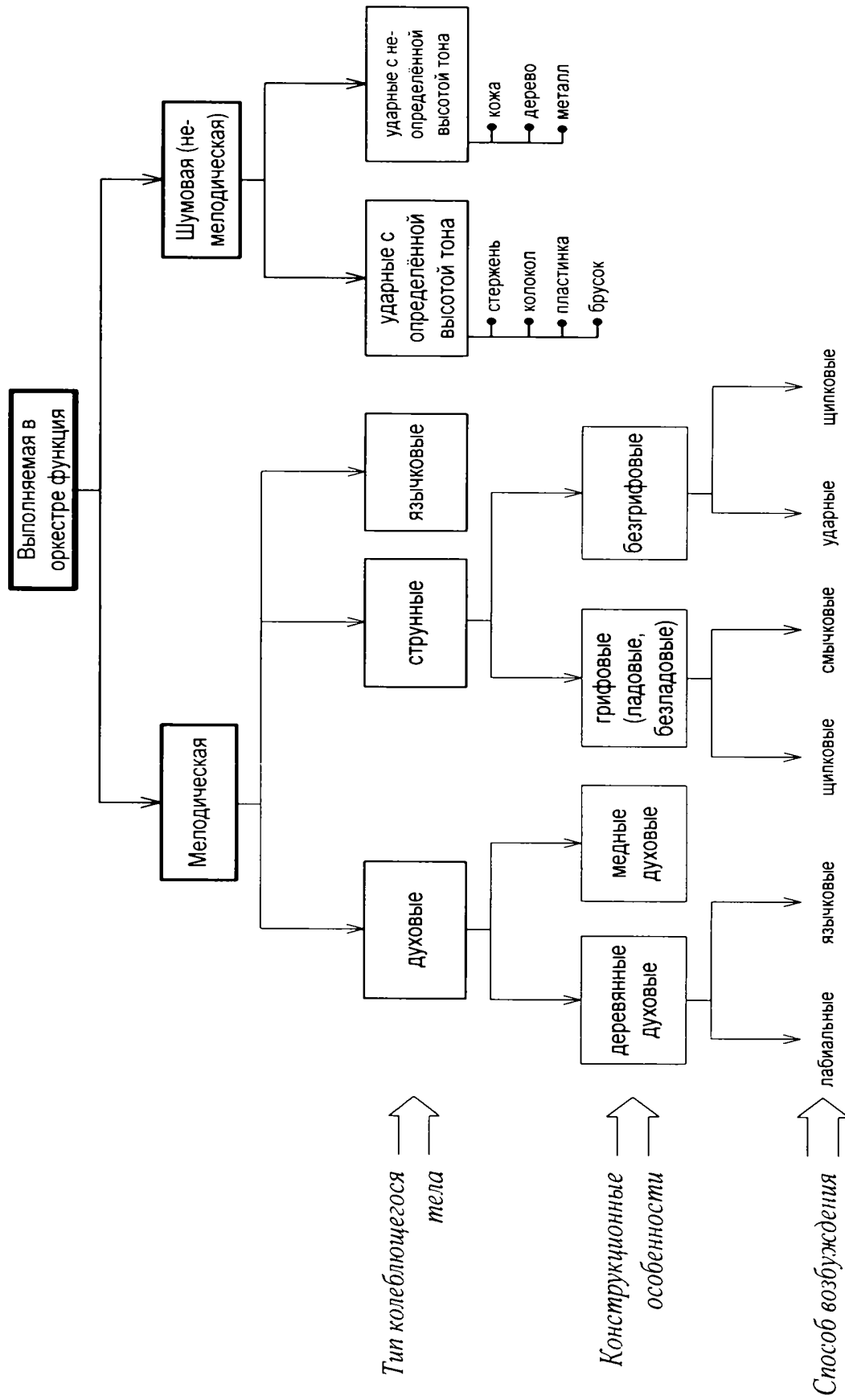


Рис. 22 Классификация музыкальных инструментов по конструкционным особенностям

Описанная классификация, во многом сложившаяся исторически, используется исполнителями, дирижёрами, аранжировщиками, конструкторами музыкальных инструментов. По сути, классифицируются именно музыкальные **инструменты**, и это обоснованно: эти люди работают с самими инструментами и им в первую очередь важно, как инструмент устроен.

В задачах объективного описания тембра, а также узнавания тембра (опознавание инструментов) приведённая выше классификация представляется *неприменимой*. Задачи объективного анализа тембра — это задачи анализа звуковых **сигналов**. Следовательно, и классифицировать целесообразно сигналы (колебания), точнее — музыкальные инструменты по типам колебаний.

Классификация предполагает выделение общих и индивидуальных свойств объектов или явлений, выявление их наиболее существенных свойств и группировку в *классы*, группы, подгруппы и т.д. На каждом уровне классификации объекты разделяются по одному из признаков или свойств. Классификация строится на основе изучения свойств объектов и явлений и предполагает наличие общей точки зрения на элементы из одного класса или группы.

На основе результатов проведённого анализа звуковых объектов предлагается альтернативная классификация музыкальных инструментов. В основе классификации лежит характер колебаний: свободные затухающие (релаксационные) колебания или автоколебания.

Для классификации предлагается использовать следующие факторы (в порядке возрастания уровня классификации):

- характер колебаний (свободные или вынужденные);
- 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.).

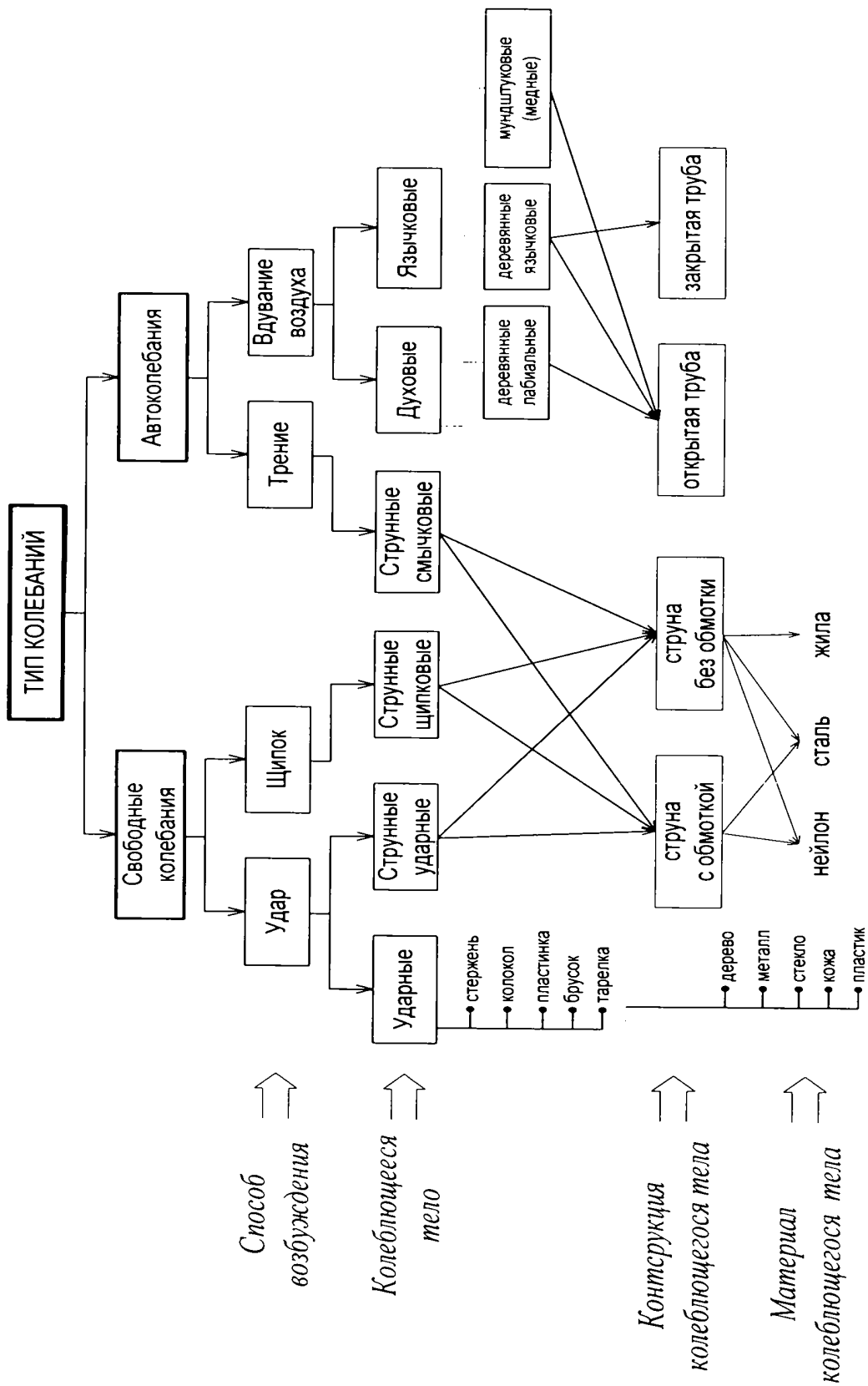


Рис. 23 Классификация музыкальных инструментов по типу колебаний

Именно такое разделение позволяет осмыслить свойства звуковых колебаний в музыкальных инструментах: выраженность высоты тона, периодичность (отклонения периода от среднего значения), характер возбуждения и затухания и т.д. Каждый из используемых для классификации факторов имеет определяющее значение именно для *свойств* возбуждаемых колебаний. Предлагаемая классификация позволяет с общих позиций подходить к колебаниям, например скрипки и флейты (в обоих случаях имеют место автоколебания), кларнета и тромбона (хотя, кларнет относится к деревянным, а тромбон — к медным духовым), фортепиано и ксилофона (возбуждение ударом молоточка), и различать колебания, порождённые скрипичной струной в случае исполнения *pizzicato* (т.е. щипком) и в случае обычного исполнения (с помощью смычка).

Очевидно, к инструментам с затухающими колебаниями относятся все струнные щипковые и все струнные ударные инструмента. Общая часть колебательной системы, узнаваемая любым человеком, — струна (или система струн, как, например, в фортепиано или рояле). Щипок или удар плектром, удар молоточка — кратковременный «впрыск» энергии в систему, после чего колебания затухают. Затухающими являются и колебания во всех ударных инструментах. Это и очевидно: звуки ударных носят импульсный шумовой характер. Различие в ударных инструментах заключается в материале и форме звукопорождающего тела: металлические или деревянные пластинки, кожаная мембрана или металлическая тарелка.

Более подробно следует остановиться на музыкальных инструментах, представляющих собой автоколебательные системы.

К музыкальным инструментам с вынужденными колебаниями относятся все духовые инструменты, струнные смычковые, а также язычковые инструменты (гармония, аккордеон, баян). Что же особенного в этих инструментах? Случай язычковых инструментов наиболее сложен, его описание заняло бы слишком много места, поэтому перейдём к смычковым и духовым инструментам.

Струнные смычковые. Прежде всего, определим элементы автоколебательной системы: источник энергии, очевидно, — лёгкие исполнителя, колебательная система — струна, роль клапана играет механизм сцепления струны и смычка, механизм обратной связи состоит в том, что упругость струны в определённом её положении расцепляет струну и смычок. При ведении смычком первоначально происходит сцепление. Совместное движение смычка и струны длится до тех пор, пока сила упругости струны, действующая в сторону, противоположную движению смычка, не заставит расцепиться смычок и струну. Некоторое время струна предоставлена сама себе, т.е. поступление энергии в колебательную систему не происходит. Далее та же упругость струны заставляет её двигаться в обратную сторону (при достижении максимального смещения — амплитуды колебаний струны). Струны также свободно движется до нового сцепления со смычком и начинается новый цикл колебаний. Определяющей в такой системе является струна, т.е. в установившемся режиме частота колебаний струны совпадает с её собственной частотой. Длительность переходного процесса определяется временем от касания смычка до того момента, когда интервал времени между сцеплениями смычка и струны станет равным периоду колебаний свободной струны.

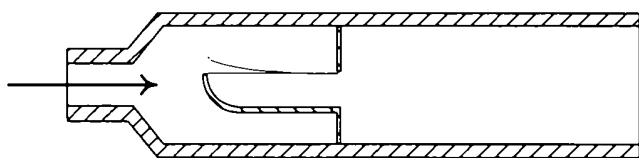


Рис. 24 Язычковый духовой инструмент

Духовые язычковые инструменты. Источник энергии — лёгкие исполнителя. Колебательная система — столб воздуха в «теле» инструмента. Роль клапана

играет язычок. Механизм обратной связи состоит в следующем: от разности давлений по обеим сторонам язычка (со стороны рта исполнителя и со стороны колеблющегося столба воздуха) зависит, пропускает ли язычок сжатый воздух изо рта или не пропускает. Таким образом, сжатый воздух поступает в «тело» инструмента порциями, а не постоянно. Кстати, собственная частота язычка много меньше частоты колебаний столба воздуха: задающей системой является столб воздуха, т.е. частота колебаний определяется его длиной, а язычок специально делается мягким, чтобы он мог выполнять функции клапана. Импульсы сжатого воздуха, распространяясь по трубе, отражаются от её конца.

Конец трубы открыт, поэтому импульс сжатия превращается в импульс разрежения, когда импульс разрежения дойдёт до клапана, то клапан не откроется, т.к. давление снаружи больше. Пройдя опять весь путь по трубе, отразившись от открытого конца и, импульс сжатия превращается уже в импульс разрежения, и когда он вновь достигнет клапана, язычок откроется, впуская новую партию сжатого воздуха.

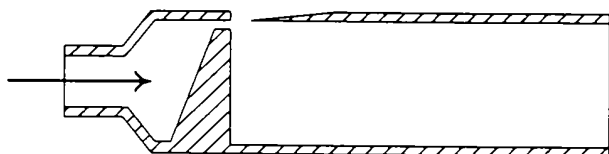


Рис. 25 *Лабиальный духовой инструмент*

Духовые лабиальные инструменты. Сжатый воздух поступает в камеру (см. рис. 25), выходное отверстие которой имеет форму узкой щели. Струю воздуха из этой щели попадает на клиновидно скошенную часть стенки трубы.

Случай открытой трубы: импульс сжатия при отражении от конца трубы превращается в импульс разрежения, когда он приходит к щели из-за избыточного внешнего давления струя воздуха из щели отклоняется внутрь и создаёт повышенное давление, в результате новый импульс сжатия побежит по трубе. Период колебаний равен удвоенному времени пробега по трубе. Случай закрытой трубы: импульс сжатия, отразившись от закрытого конца, возвращается опять же в виде импульса сжатия и отклоняет струю из щели наружу, т.е. начало трубы открыто и импульс сжатия, отразившись от него, превращается в импульс разрежения; пробежав по трубе туда и обратно без изменения, он приходит к началу трубы в виде импульса сжатия, струя из щели снова втягивается внутрь, и начинается новый цикл. Период колебаний равен учетверённому времени пробега импульса по трубе. Таким образом, основной тон трубы, закрытой с одного конца, вдвое ниже, чем у трубы той же длины, открытой с обоих концов.

Духовые мундштуковые инструменты. Роль клапана играют губы исполнителя. Процессы, происходящие в теле инструмента те же, что в духовых лабиальных инструментах. Разница состоит в форме импульсов возбуждения: в медных духовых инструментах импульсы возбуждения короче, а значит и спектр их шире. Поэтому тембр медных духовых богаче, чем у деревянных ду-

ховых и при этом обладает своеобразием: медный духовой инструмент на слух всегда можно отличить от деревянного духового инструмента при прочих равных условиях.

5.3.2. Гармоничность музыкальных звуков. Модель синтеза звукового объекта

Под гармоничностью будем понимать совпадение частот обертонов со значениями, кратными частоте основного тона. По сути, строгая гармоничность должна означать стабильность длительности полного цикла сложных колебаний (сумма периодических функций есть функция периодическая), негармоничность — напротив, отклонения и хаотические изменения длительности полного цикла колебаний. Мы обсудим гармоничность колебаний в музыкальных инструментах в свете всего вышесказанного относительно природы звукообразования, а также в свете полученных результатов исследования гармонической структуры колебаний.

По каким признакам можно судить о гармоничности сложного колебания? Мы имеем в распоряжении результаты исследования: траектории амплитуд и фаз обертонов, а также траекторию длительности полного цикла колебаний. Траектория длительности полного цикла колебаний характеризует колебание в целом, траектория же фаз — отдельные составляющие, т.е. обертоны. Воспользуемся известным соотношением, связывающим мгновенную фазу и частоту модулированного гармонического колебания:

$$\frac{d\varphi(t)}{dt} = \omega(t) \quad (7)$$

$$\text{В частном случае: } \varphi(t) = 2\pi \cdot F \cdot t + \Delta\varphi(t) \Rightarrow \omega(t) = 2\pi \cdot F + \frac{d\Delta\varphi(t)}{dt}$$

Траектория длительности полного цикла колебаний даёт нам зависимость $F(t)$, траектория фаз — $\Delta\varphi(t)$ для любого обертона. Изначально мы предполагаем, что частота k -ого обертона примерно равна kF_0 (F_0 — частота основного тона). Поправку к частоте мы можем определить, вычислив производную $\frac{d\Delta\varphi(t)}{dt}$. Иными словами, анализируя соотношение фаз обертонов и основной

частоты, мы можем определить и соотношение их частот. Так, если имеет место линейное по времени увеличение разности начальных фаз какого-либо обертона и основной частоты (на каждом периоде накапливается фиксированное опережение начальной фазы основной частоты), это означает, что частота данного обертона (обозначим её номер через n) равна nF_0 плюс фиксированное $\frac{d\Delta\varphi(t)}{dt} = \Delta F$. Естественно, линейный рост разности начальных фаз обертонов – идеальный случай, на деле зависимость сложнее, но пример важен для правильного понимания тенденций поведения начальных фаз отдельных обертонов.

Результаты проведённых автором исследований, показали, что колебания в духовых музыкальных инструментах более гармоничны, чем в струнных щипковых (гитара и мандолина) и струнных ударных (фортепиано). Наиболее гармоничными (в том смысле, в котором мы договорились понимать это слово) являются колебания в медных духовых (в проведённом исследовании – тромбон и труба). Есть основания предполагать, что вынужденные колебания в музыкальных инструментах более гармоничны, чем свободные затухающие колебания.

На рис. 26 приведены траектории фаз первых пяти обертонов гитары и тромбона для колебания, соответствующего ми первой октавы. По оси абсцисс отложен номер периода (фактически, время), по оси ординат – фаза в градусах. На траекториях тромбона легко выделить установившуюся стадию и нестационарные стадии установления и затухания колебаний (конечный «горб» траекторий).

Для колебаний гитары характерно следующее: траектория фазы второго обертона практически совпадает с траекторией основной частоты (лишь в данном конкретном случае), остальные обертоны (что характерно для всех высот тонов) опережают по начальной фазе основную частоту и это опережение со временем накапливается. Очевидно, это свидетельство того, что частоты обертонов отклоняются от значений, кратных частоте основного тона.

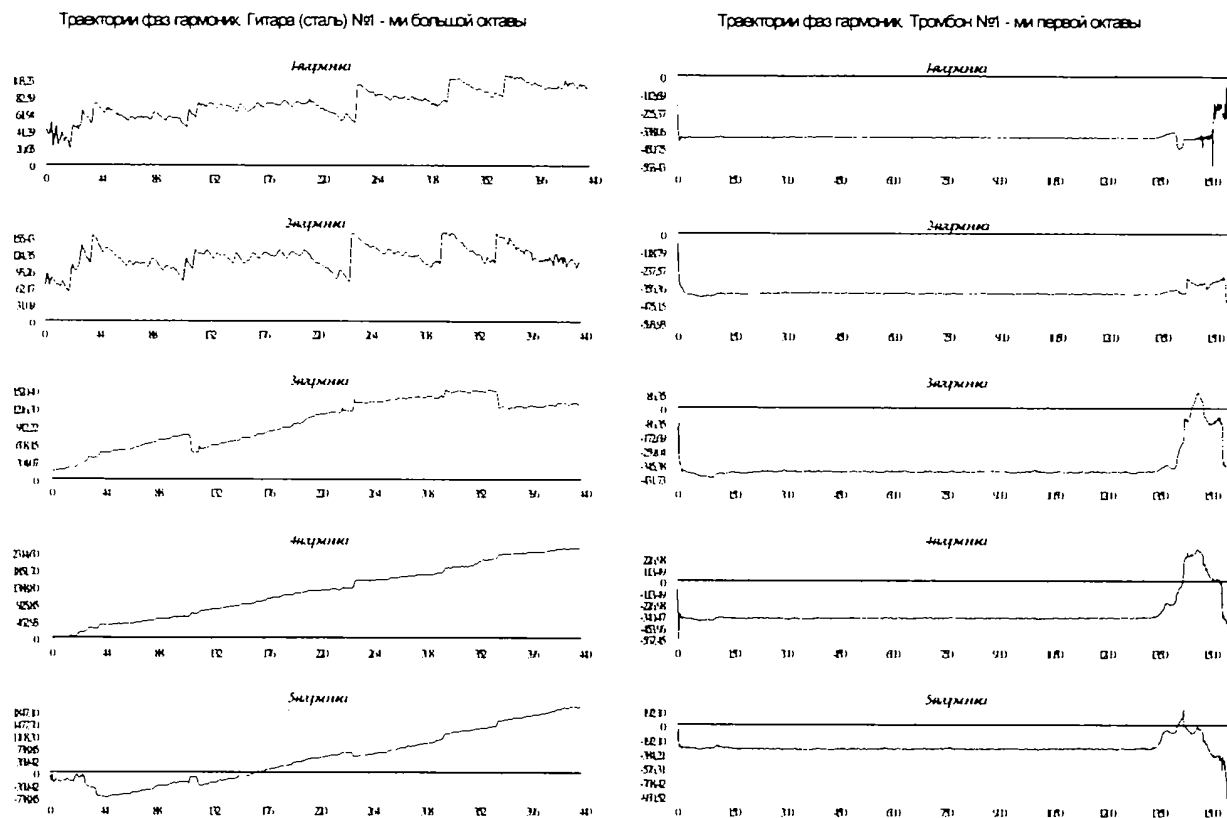


Рис.26 Траектории начальных фаз обертонов колебаний гитары(слева) и тромбона(справа)

В колебаниях тромбона (как и в колебаниях трубы и, видимо, во всех медных) наблюдается принципиально иная картина: после установления колебаний соотношение начальных фаз обертонов не изменится, начальные фазы обертонов практически постоянны на всей длительности стадии установившихся колебаний. Вывод из этого следующий: частоты обертонов с большой точностью кратны частоте основного тона, начальные фазы колебаний стабильны (об амплитудах поговорим ниже).

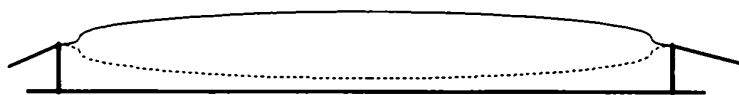


Рис.27 Колебания струны гитары (схематично)

Попробуем объяснить такую разницу происходящих процессов.

Одну из причин негармоничности колебаний струны автор склонен видеть в краевых эффектах. Обратимся к рисунку 27. На нём утрированно изображены колебания струны гитары, опирающейся на верхний и нижний порожки. Граничные условия в мес-

тах соприкосновения с порожками требуют, чтобы струна в этих точках была параллельна грифу. Выражаясь проще, при колебаниях струна не может иметь крутого излома на краю порожка, имеется плавный переход от части струны, расположенной параллельно грифу к изгибающейся части. Очевидно, что соотношение длин участков, участвующих и не участвующих в колебаниях определяется, в частности, амплитудой возбуждения. Т.е. «слабый» обертон имеет меньшую полезную длину для колебаний.

Есть и ещё одно соображение, связанное с негармоничностью колебаний струны. Частота её колебаний прямо пропорциональна натяжению и обратно пропорциональна длине. В процессе колебаний струна удлиняется, что приводит к увеличению натяжения: сила упругости стремится вернуть струну в состояние равновесия. Для колебаний основной частоты это не столь существенно — натяжение изменяется синхронно с колебаниями струны. Но колебания обертонов возникают на той же самой струне. Т.е. колебания обертонов происходят при натяжении, по крайней мере, отличном от натяжения для основной частоты (сложно точно говорить о распределении натяжения струны по её длине).

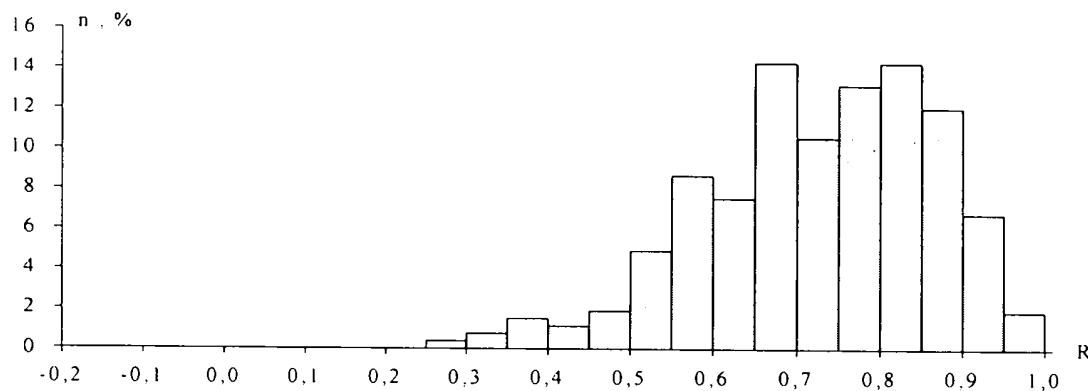
С другой стороны, в колебаниях струнных смычковых, эффект негармоничности, согласно полученным автором данным, проявляется меньше. Поэтому, есть основания считать, что

основным фактором, определяющим гармоничность колебаний механико-акустических музыкальных инструментов, является их тип — свободные колебания или автоколебания.

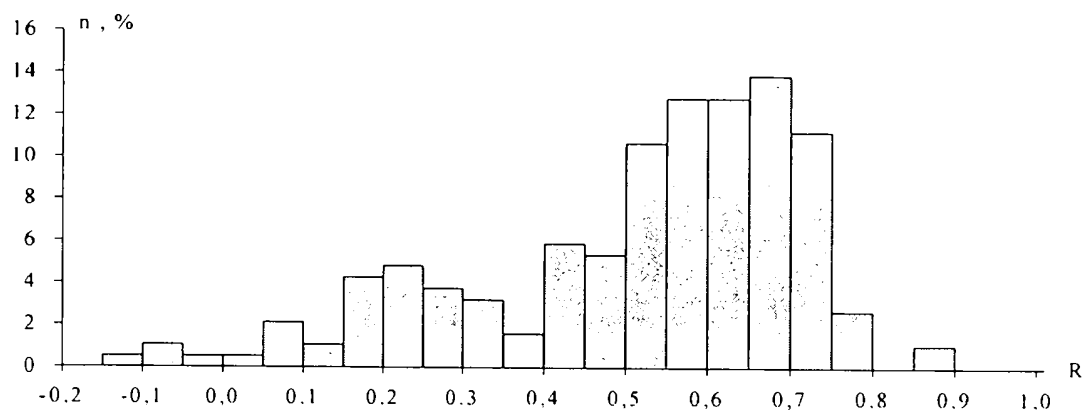
Очевидны факторы, определяющие стабильность автоколебаний: это, во-первых, стабильность параметров колебательной системы и, во-вторых, стабильность источника энергии (в случае духовых музыкальных инструментов — стабильность давления воздуха, создаваемого лёгкими исполнителя).

Стабильность автоколебаний отражается не только на траекториях фаз, но и на траекториях амплитуд. Одной из характеристик звукового объекта, использованной в исследовании была матрица коэффициентов корреляции траекторий амплитуд обертонов. Осмыслим реальные значения коэффициентов

корреляции траекторий амплитуд для различных инструментов, полученные в исследовании.



а)

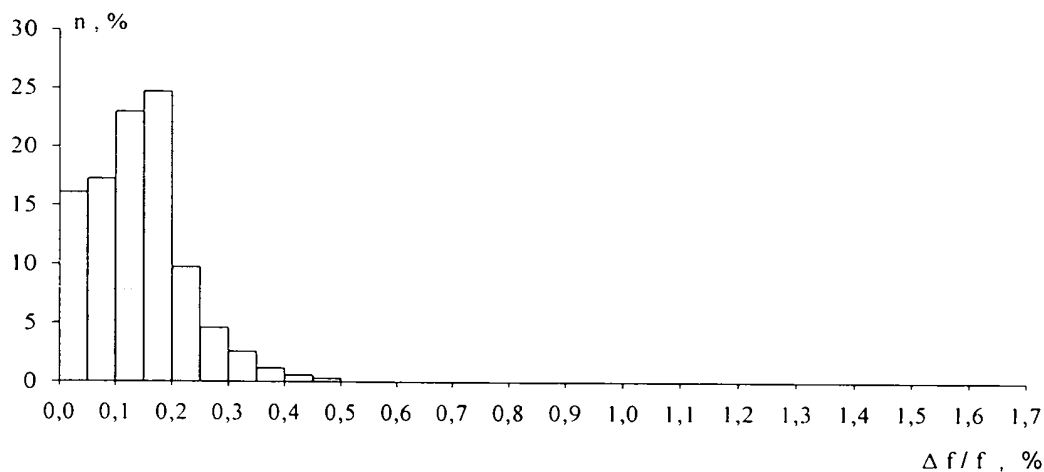


б)

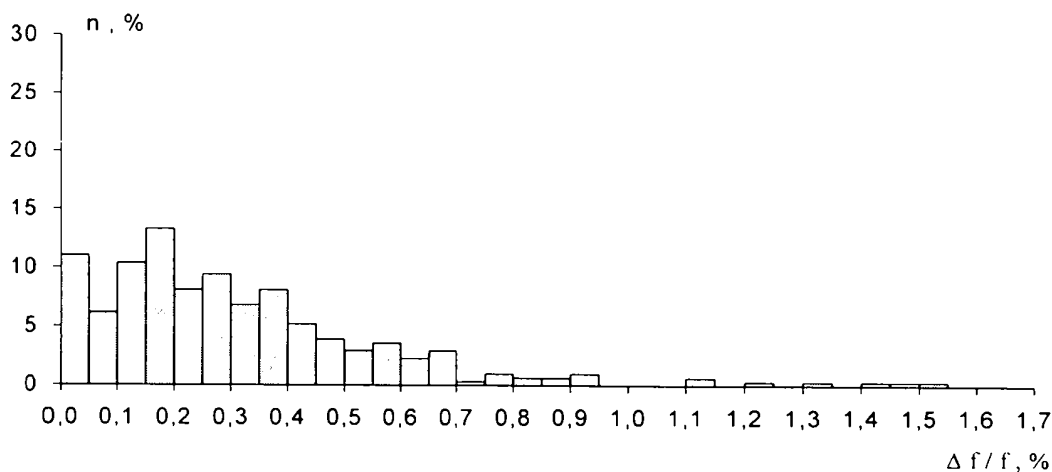
Рис. 28 Гистограммы коэффициента корреляции траекторий амплитуд обертонов
а) для инструментов с автоколебаниями, б) для инструментов со свободными колебаниями

Для гитары и фортепиано значения коэффициентов корреляции, полученные автором, распределены между 0 и 0,8, имеют место и отрицательные значения. Для деревянных духовых коэффициенты корреляции равны 0,5÷0,9. У медных духовых — 0,75÷0,99 (гистограммы значений коэффициентов корреляции приведены на рис. 28). Т.е. опытные данные подтверждают все наши предыдущие размышления: свободным колебаниям свойственна более слабая взаимосвязь отдельных гармонических составляющих, а для наиболее стабильных автоколебаний свойственна почти функциональная зависимость между амплитудами обертонов (коэффициент корреляции 0,99).

Важным свойством, свидетельствующим о стабильности колебаний является отклонение частот обертонов от значений, кратных частоте ОТ. Гистограммы значений относительных отклонений частот обертонов от значений, кратных частоте ОТ приведены на рис. 29.



а)



б)

Рис. 29 Гистограммы относительных отклонений частот обертонов от значений, кратных частоте ОТ а) для автоколебаний, б) для свободных колебаний

Выводы те же — стабильность автоколебаний при стабильном возбуждающем воздействии выше, чем стабильность свободных колебаний.

Целью исследования звуковых объектов является построение модели звукового объекта. Во временной области музыкальный звуковой объект с дос-

таточной точностью описывается суммой модулированных гармонических колебаний:

$$s(t) \approx \sum_{n=0}^N A_n(t) \cdot \sin[2\pi \cdot f_n(t) + \varphi_n(t)], \text{ где} \quad (8)$$

$A_n(t)$ – 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).

Действительно, синтез проанализированных звуковых объектов по имеющимся траекториям $A_n(t)$, $f_n(t)$ и $\varphi_n(t)$ показал, что звуки музыкальных инструментов могут быть синтезированы с помощью *аддитивного* синтеза, а проведённый анализ множества значений $A_n(t)$, $f_n(t)$, $\varphi_n(t)$, показал, что все параметры компонентов музыкальных звуковых объектов модулированы. Модель (8) целесообразно использовать при электронном синтезе звуков музыкальных инструментов. Поэтому можно назвать её *моделью синтеза* звукового объекта.

5.3.3. Узнавание тембра: что важно и что нет? Модель анализа звукового объекта

1. *Усреднённый амплитудный спектр колебания.* Прежде всего, поймём, как формируется амплитудный спектр звукового объекта, от чего зависит его вид. Импульс возбуждения (например, щипок, удар молоточка) воздействует на колебательную систему (струну, пластинку, мембрану), механические колебания которой должны преобразоваться в акустические колебания. Иногда для этого нужен «резонатор» – например, корпус гитары. Формализуем наши рассуж-

дения, запишем выражение для комплексного спектра результирующего звукового давления:

$$S(j\omega) = S_{\text{возб}}(j\omega) \cdot H_{\text{КС}}(j\omega) \cdot H_{\text{рез}}(j\omega), \text{ где} \quad (9)$$

$S(j\omega)$ – комплексный спектр звукового давления;

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

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

$H_{\text{рез}}(j\omega)$ – частотная характеристика резонатора.

Какие из перечисленных множителей в формуле (9) постоянны и какие меняются? Очевидно, постоянными остаются параметры резонатора. Слабо меняющейся можно считать форму импульса возбуждения – меняется его размах. А вот параметры колебательной системы меняются. Большинство музыкальных инструментов способно воспроизводить звуковой ряд (в отличие от камертона). В некоторых случаях это достигается использованием набора колебательных систем, настроенных на разные частоты. Очевидно, что при неизменных АЧХ резонатора и импульсе возбуждения, возбуждая разные колебательные системы, получим колебания с совершенно различными амплитудны-

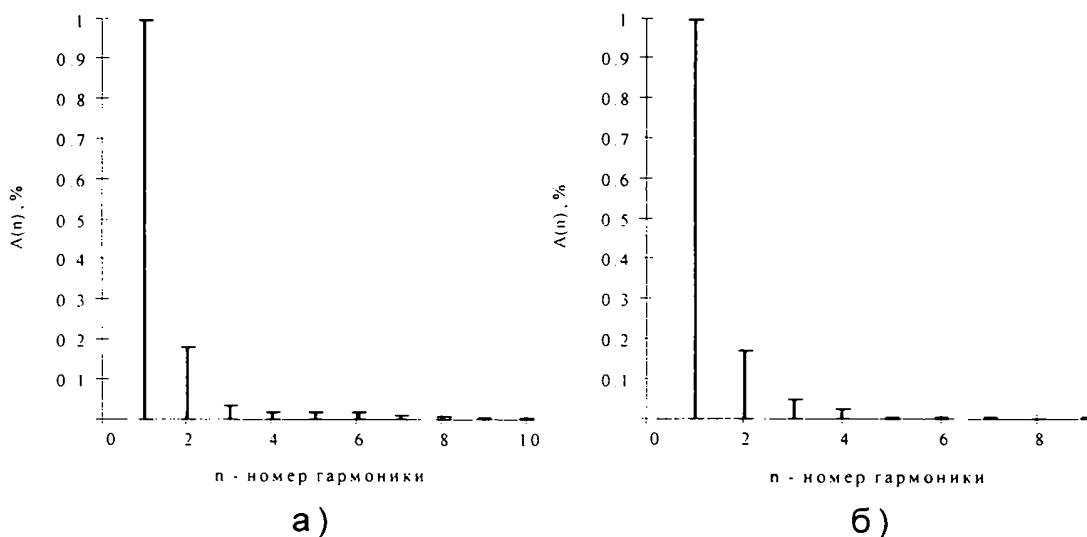


Рис. 30 Соотношение усреднённых амплитуд обертонов: а — фортепиано (для второй октавы), б — кларнет (ми третьей октавы)

ми спектрами, при этом звучать они будут похоже. Таким образом, и это подтверждается экспериментальными данными, при рассмотрении амплитудных спектров звуков одного и того же инструмента, сохраняются формантные области и характер затухания обертонов, соотношение же усредненных амплитуд обертонов для узнавания тембра может оказаться неинформативным. Более того, звуки совершенно разных инструментов могут иметь одинаковое соотношение усреднённых амплитуд обертонов (рис. 30). Неоспорим следующий факт: мы узнаём инструмент независимо от высоты тона, манеры исполнения, приёмов игры — мы узнаём колебательную систему, т.е. звукопорождающее тело (струна, столб воздуха). Обобщённый образ звучания знакомых нам колебательных систем хранится у нас в памяти.

2. *Траектории амплитуд обертонов.* Важны не только для узнавания тембра, но и для определения высоты тона, разделения звуков созвучий. Для струнных инструментов, как мы выяснили, свойственна малая взаимосвязь отдельных обертонов. Возможно, поэтому именно для струнных инструментов характерно «раздвоение» ощущения высоты тона, духовым это не свойственно. Кроме того, свойства колебательных систем определяют характерные признаки траекторий амплитуд обертонов: например, для струнных инструментов характерны периодические угасание и возбуждение обертонов. С другой стороны, как мы выяснили, в инструментах с вынужденными колебаниями соотношение амплитуд обертонов мало изменяется на установившейся стадии колебаний, все обертоны ведут себя почти одинаково, и изменении их амплитуд в основном определяется интенсивностью возбуждения. Иными словами, в траекториях амплитуд обертонов может быть не заложено информации о тембре.

3. *Переходные процессы (атака).* Складывается общее убеждение в том, что наличие атаки, по сути, является «необходимым и достаточным условием опознания тембра». Достаточным — да, но необходимым ли? Процесс установления колебаний всегда сопровождается возбуждением колебательной системы на наибольшем количестве резонансных частот, т.е. самый богатый спектр — у атаки, далее тембр только беднеет, но всегда ли исключение «атаки» делает невозможным узнавание инструмента? Для инструментов со свободными колеба-

ниями — да, обертоны быстро затухают, тембр быстро беднеет, кроме того, наше ухо привыкло к тому, что звучание струны, например гитары, сопровождается резким «всплеском» и быстрым затуханием, флейта так не звучит. Возьмите фрагмент звучания тромбона и «отрежьте» атаку — тромбон останется тромбоном.

4. *Негармонические призвуки.* Речь идёт о сопровождающих колебания шумах: стук клапанов (вентилей) духовых инструментов, звук выдуваемого воздуха, звук, возникающий при движении пальцев по гитарным струнам с обмоткой. Безусловно, все эти призвуки помогают опознать тембр, но от них, как правило, стремятся избавиться и при записи и при концертном исполнении, так что слышим мы их редко и они мало знакомы.

Модель (9) описывает звуковой объект в частотной области. Её целесообразно использовать при анализе, звуковых объектов, идентификации музыкальных инструментов: по множеству реализаций звуков музыкального инструмента прежде всего можно установить положение формант амплитудного спектра и примерно описать $H_{рез}(j\omega)$, что упростит подбор $H_{КС}(j\omega)$ и $S_{возб}(j\omega)$ для идентификации.

Основной идеей данного исследования, уже высказанной при обсуждении классификации музыкальных инструментов, является то, что при узнавании тембра мы пытаемся узнать именно колебательную систему, а не конкретную реализацию колебаний. Автору представляется, что узнавание тембра происходит при сопоставлении слышимого звука и предполагаемого способа возбуждения. Так поступают при определении частотной характеристики устройства: подают на вход заранее известный сигнал (например, единичный импульс) и регистрируют выходной сигнал. В терминах формулы (9) можем записать:

$$H_{КС}(j\omega) \cdot H_{рез}(j\omega) = \frac{S(j\omega)}{S_{возб}(j\omega)} \quad \text{— это и есть формализация сопоставления слышимого звука и способа возбуждения (левая часть равенства характеризует собственно музыкальный инструмент).}$$

Возможно, в мозгу проис-

ходят подобные операции: из уже известных человеку моделей колебательных систем он выбирает наиболее подходящее звено между $S(j\omega)$ и $S_{\text{своб}}(j\omega)$. С опытом у человека накапливаются модели поведения различных физических тел, т.е. характерные издаваемые ими звуки: скрип, звон, свист, колебания струны.

5.4. 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.
3. Во временной области музыкальный звуковой объект с достаточной точностью описывается суммой модулированных гармонических колебаний.

6. РАЗРАБОТКА АЛГОРИТМА РАСПОЗНАВАНИЯ МНОГОГОЛОСНЫХ МУЗЫКАЛЬНЫХ СИГНАЛОВ

6.1. РАЗРАБОТКА АЛГОРИТМА СЕГМЕНТАЦИИ

Автоматическая сегментация является составной частью задачи распознавания звуковых сигналов. Сегментация всегда предшествует собственно распознаванию, т.е. принятию решения о принадлежности реализации объекта к тому или иному классу. Соответственно, достоверность распознавания зависит, в том числе, и от достоверности сегментации.

Из-за различий в свойствах звуковых сигналов, подлежащих распознаванию в современной технике, — а это не только речевые и музыкальные сиг-

налы, но также шумы всевозможных машин, узлов машин (например, двигателей), — специфичны и подходы к сегментации сигнала. Выбор подхода к сегментации определяется, в первую очередь, выбором *элемента распознавания*, т.е. элементарного с точки зрения системы распознавания звукового объекта, подлежащего распознаванию. При распознавании слитной речи элементом распознавания может быть и фонема, и слог и целое слово. Задача алгоритма сегментации — определение границ элементов распознавания и при необходимости их разделение. В случае, если элементы распознавания перекрываются во временной области, необходимо определение границ в частотной области.

Основная сложность сегментации многоголосных музыкальных сигналов определяется тем, что звуковые объекты перекрываются. Перекрываются звуковые объекты во времени, т.е. перекрываются интервалы существования звуковых объектов, моменты их возникновения могут совпадать, за время существования одного звукового объекта могут возникнуть и затухнуть ещё несколько. Перекрываются звуковые объекты и в частотной области, т.е. звуковые объекты могут иметь спектральные компоненты с примерно совпадающими частотами. Сложнейший вариант — т.н. исполнение в унисон, когда несколько инструментов одновременно исполняют одну и ту же мелодию. Перекрытие почти стопроцентное. Тем не менее, в случае, если тембры инструментов различны, человеческое ухо в некоторых случаях способно определить их количество.

После выполнения задачи сегментации образуется ритмическая структура распознаваемой мелодии. Ритмическая структура мелодии не менее важна, чем её высотная структура. В этом состоит одно из отличий задачи распознавания музыкального сигнала от распознавания речевого сигнала. При распознавании речи ошибки сегментации могут привести лишь к снижению достоверности распознавания фонемы, возникшие ошибки можно исправить автоматически, имея словарь языка. Ошибки сегментации музыкального сигнала повлекут и неверное определение высоты тона (если на вход алгоритма определения частоты основного тона попадут фрагменты последовательных звуко-

вых объектов с разной высотой тона), и искажение ритмической структуры, в результате чего мелодия может стать неузнаваемой.

6.1.1. Современные подходы к сегментации музыкальных сигналов

Перечисленные отличия полифонических музыкальных сигналов от одноголосных не позволяют применять для сегментации алгоритмы, разработанные для одноголосных мелодий. Для сегментации линейных последовательностей звуковых объектов первоначально применялся *последовательный* спектральный анализ. Т.е. распознаваемый сигнал считывается короткими выборками, на которых выполняется спектральный анализ. Анализируемые выборки могут перекрываться, но анализатор спектра один, и анализ производится во всей полосе частот. Современные алгоритмы сегментации музыкальных сигналов строятся на использовании *параллельного* спектрального анализатора. Как показывает обзор публикаций зарубежных исследователей за последние 5-7 лет, работы по созданию систем распознавания музыкальных сигналов ведутся в рамках решения более общей задачи – компьютеризированного анализа звуковых сцен (CASA – Computational Auditory Scene Analysis). К решаемым CASA задачам относятся пространственная локализация и идентификация источников звука. Бурное развитие CASA, видимо, вызвано развитием многоканальных систем звукозаписи, потребовавшее новых исследований психофизиологии слухового восприятия. По сути, CASA строится и в первую очередь занимается созданием модели слухового восприятия человека. По современным представлениям слуховой аппарат человека снабжён именно *параллельным* анализатором спектра – волокна покровной мембраны восприимчивы лишь к колебаниям своей узкой полосы частот, при этом информация от каждого волокна передаются в мозг отдельным нейроном. Представляется очевидным разделение звуковых объектов, перекрывающихся во времени, именно в частотной области. Работа волокон покровной мембраны моделируется с помощью набора фильтров. Сами по себе отклики фильтров, как правило, не используются при вынесении решения о появлении или затухании звуковых объектов. Для принятия решения сопоставляются различные параметры откликов фильтров,

различные алгоритмы сегментации отличаются выбором этих параметров и схемой принятия решения.

Существует, правда, подход к сегментации звуковых сигналов, явно не использующий набор фильтров, но основанный на параллельном спектральном анализе. Речь идёт об использовании для распознавания и цифровой обработки сигнала *сонограммы (спектрограммы)*, т.е. временной зависимости коэффициентов кратковременного дискретного преобразования Фурье (ДПФ). Сонограмма используется для визуализации результатов спектрального анализа звуковых фрагментов большой продолжительности.

Сегментация с использованием сонограммы не нашла пока широкого применения. Большого внимания заслуживает другой подход, позволивший добиться существенных результатов. Этот подход, уже упомянутый, основан на модели восприятия громкости человеческим ухом, подробное описание которой дано в [73]. Последовательность процессов обработки и анализа сигнала представлена на рисунке 33.

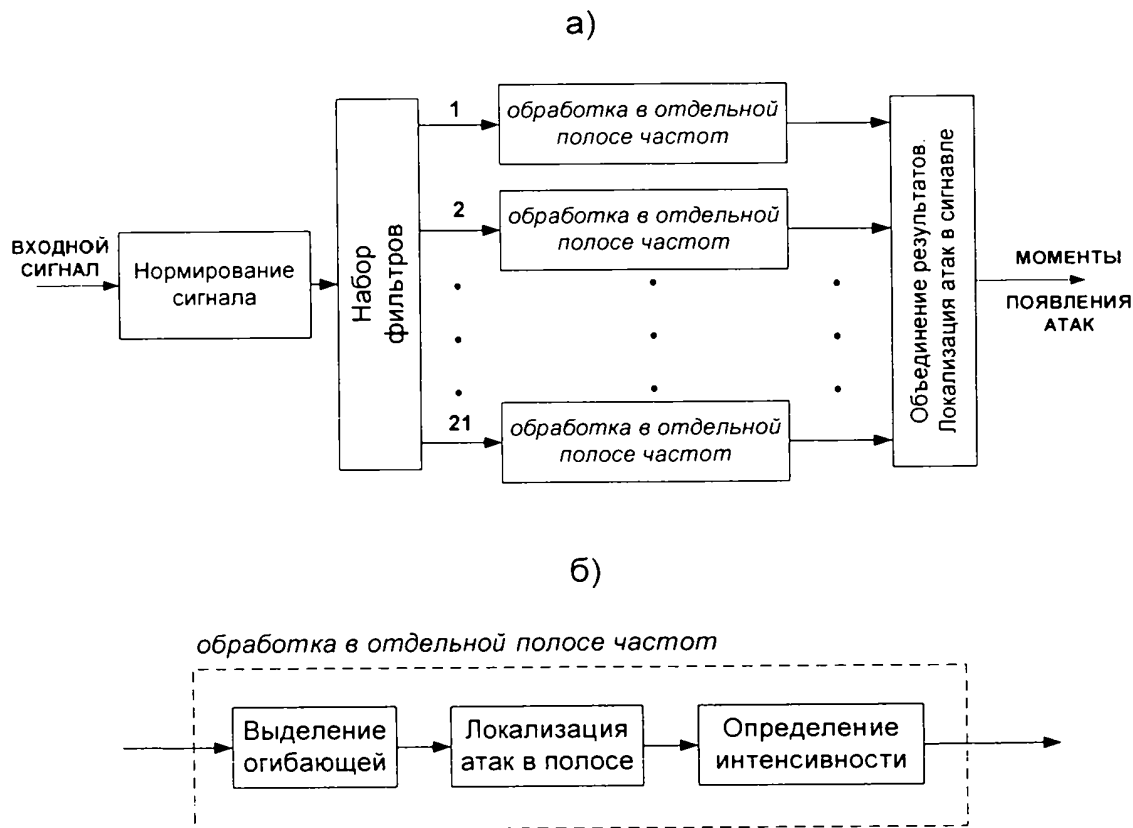


Рис. 31 Схема сегментации звукового сигнала моделированием слухового восприятия

Все входные сигналы должны быть приведены к одному определённом значению относительной громкости, для этого перед обработкой сигнал нормируется. Нормированный сигнал поступает на набор полосовых фильтров. Частотный диапазон от 40 Гц до 18 кГц делится на полосы таким образом, чтобы ширина каждой полосы была примерно равна ширине критической полосы слуховой системы. Первые три фильтра (в самой низкочастотной области) — октавные, остальные — третьоктавные. Для перекрытия указанного диапазона частот достаточно 21 фильтра. В системе сегментации, описанной в [73], применены рекурсивные фильтры 8-го порядка, в системе, описанной в [50] — цифровые аналоги эллиптических фильтров 6-го порядка, весь диапазон частот делится на 6 полос. После разделения все узкополосные сигналы (компоненты) одинаковым образом обрабатываются (см. рис. 31 б)). Прежде всего, выделяется огибающая узкополосного сигнала, сглаживаемая свёрткой с временной оконной функцией Хэмминга («косинус на пьедестале») длительностью 100 мс, чем моделируется интеграция энергии слуховой системой человека. Далее для каждого компонента определяются моменты появления атак. Моменты появления атак компонентов в дальнейшем рассматриваются как кандидаты при выборе моментов появления атак сигнала как единого целого.

В алгоритме сегментации, описанном в [73], как и во многих других алгоритмах, за моменты появления атак в узкополосных сигналах принимаются моменты времени, соответствующие максимальной скорости нарастания огибающей. Находя локальные максимумы первой производной огибающей $A(t)$, находят моменты появления атак. Однако, у такого подхода есть важный недостаток. Как указывается в [73], функцию $\frac{dA(t)}{dt}$ удобно использовать для оценки интенсивности атаки, но локальные максимумы этой функции не всегда точно соответствуют истинному моменту атаки. Основных причины две:

- низкочастотные компоненты нарастают медленнее, чем высокочастотные, в результате момент времени, соответствующий наибольшей скорости нарастания огибающей узкополосного сигнала, сильно отстаёт от субъективно воспринимаемого момента возникновения звукового объекта;

— не все узкополосные компоненты имеют монотонно нарастающую и спадающую огибающую, вблизи наибольшего локального максимума функции $\frac{dA(t)}{dt}$ имеются другие локальные максимумы (см. рис. 32), что приводит к ошибочным решениям при объединении промежуточных результатов всех компонентов.

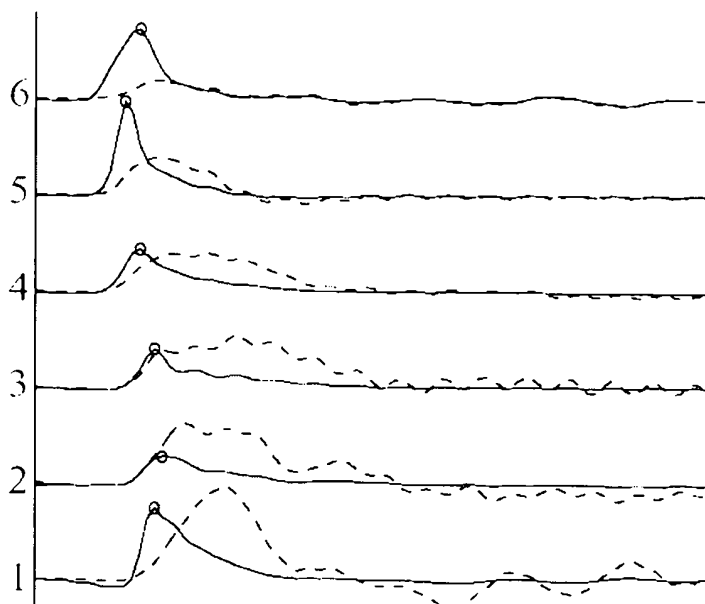


Рис. 32 Атака звука фортепиано: штриховые линии — производные огибающих узкополосных компонентов, сплошные линии — производные логарифмов огибающих

Для избежания последствий указанного недостатка в [73] предлагается

искать локальные максимумы функции $\frac{dA(t)}{A(t)}$, т.е. относительного приращения огибающей. Такой подход согласуется с установленным свойством и слуховой, и зрительной систем человека реагировать именно на относительные приращения интенсивностей воздействия. Известно, что

$\frac{d}{dx}(\ln(F(x))) = \frac{1}{F(x)} \cdot \frac{dF(x)}{dx}$, поэтому $\frac{dA(t)}{A(t)} = \frac{d(\ln(A(t)))}{dt} = W(t)$. Функция

$W(t)$ используется и при обнаружении атак узкополосных компонентов и при определении моментов появления этих атак, при этом достаточно простого на-

хождения локальных максимумов, поскольку функция $W(t)$ более гладкая, чем $A'(t)$. Рисунок 32 иллюстрирует преимущество использования при сегментации функции $W(t)$ перед использованием $\frac{dA(t)}{dt}$.

Определив моменты появления атак в компонентах можно переходить к локализации атак обрабатываемого сигнала. Атаки компонентов являются компонентами атаки сигнала, на основе соотношений моментов появления атак компонентов и их интенсивностей принимается решение о наличии атаки в сигнале. Интенсивности компонентов определяется следующим образом:

$$I_i(t) = \frac{dA_i(t)}{dt} \cdot \Delta f_i = D_i(t) \cdot \Delta f_i, \quad (10)$$

т.е. производная огибающей i -того компонента $D_i(t)$ умножается на ширину спектра компонента Δf_i (т.е. ширину полосы пропускания i -того фильтра). Интенсивности атак, появляющихся одновременно в разных компонентах суммируются. Далее атаки всех компонентов сортируются по времени появления и рассматриваются в дальнейшем как кандидаты при выборе момента появления атаки в сигнале. Каждому такому кандидату ставится в соответствие величина *относительной громкости*. Напомним, что перед разделением на узкие частотные полосы входной сигнал нормировался, т.е. приводился к определённому значению относительной громкости. Такое нормирование необходимо по одной причине: при сегментации необходимо иметь точку отсчёта энергетических параметров сигнала и его компонентов, например для игнорирования атак с интенсивностями ниже условного порога слышимости. Вычисление относительной громкости кандидата производится в соответствии с моделью слухового восприятия, описанной в [73]. При вычислении громкости кандидата учитываются интенсивности атак компонентов, находящихся во временном отрезке 50 мс с серединой в моменте появления данной атаки-кандидата. Последним этапом является устранение «лишних» атак, т.е. атак с громкостями ниже условного порога слышимости и условно маскируемых атак. За маскируемые атаки принимаются атаки с меньшей относительной громкостью, находящиеся к атаке с большей относительной громкостью ближе, чем 50

мс. При рассмотрении близко расположенных атак-кандидатов, имеющих равные относительные громкости, за истинную атаку принимается средний по времени кандидат, остальные игнорируются.

Описанный подход можно назвать «сегментацией по громкости». Если быть точным, то рассмотренные в [73],[74], [50], [49] алгоритмы выполняют не сегментацию в том смысле, в котором договорились понимать её мы. В англоязычной литературе описанная операция называется *sound onset detection*, что дословно означает *определение момента начала звучания звука*. Т.е. Определяются лишь моменты появления звуковых объектов, но не моменты их затухания. Необходимо заметить, что описанный алгоритм, как и многие другие, разрабатывался в рамках построения систем выделения ритма, ритмической структуры сигнала. Действительно, для этого достаточно определения лишь моментов начала звучания звуковых объектов, однако при распознавании музыкального сигнала необходимо определение не только моментов появления нот, но и их длительностей, что требует определения моментов затухания звуковых объектов, выделения и определения длительностей пауз. Важной особенностью описанного подхода является его универсальность. Определение моментов появления звуковых объектов может проводиться в сигналах любого типа звучания: инструментальной, камерной, симфонической, эстрадной или рок-музыке, речевом сигнале. Однако типом звучания определяется и достоверность определения ритма. По данным автора [73], достоверность работы его системы колеблется от 95% для фортепианной музыки (Шопен) до 7% для симфонической музыки (Бетховен). Такой разброс и есть плата за отсутствие специализации, построение алгоритма, не учитывающего особенности конкретного типа звучания.

6.1.2. Сегментация музыкального сигнала с использованием априорной информации

Развитие систем автоматической сегментации утвердило два принципа, на которых и строятся все современные системы:

— необходимым и основным параметром сигнала, по которому осуществляется сегментация, является его огибающая;

— для сегментации полифонических мелодий необходимо разделение сигнала на узкополосные компоненты.

Как правило, огибающая не является единственным информативным параметром при сегментации. Очевидно, что при переходе от одного звукового объекта изменяются и энергетические, и спектральные параметры сигнала, нестабильны эти параметры и на длительности отдельного звукового объекта. Выделение огибающей сигнала не требует сложных операций и вычислений, но при этом огибающая не всегда позволяет однозначно принять решение о появлении, а тем более затухании звукового объекта. Огибающая широкополосного многокомпонентного сигнала может вообще оказаться непригодной для сегментации: например, тихие пассажи в высоком диапазоне высот тонов могут совершенно не оставить в огибающей следа на фоне мощного длительного низкочастотного звука или его затухающего отзвука. Огибающие узкополосных составляющих сигнала оказываются более информативными при сопоставлении и объединении результатов разных частотных полос. При разделении спектра сигнала на узкие полосы количество звуковых объектов и их составляющих, модулирующих энергию колебаний в данной полосе, уменьшается с уменьшением ширины полосы, поэтому становится легче обнаружить «следы» отдельных звуковых объектов, перекрывающихся во времени.

При сегментации одноголосных мелодий, как мы говорили, достаточно слежения за изменениями основного тона за исключением случая следования друг за другом двух одинаковых нот. Выделение основного тона в полифонических мелодиях значительно усложняется частичным перекрытием обертонов звуковых объектов. При сегментации музыкального сигнала желательно разделение не по частоте, а *по периоду*. Иными словами, желательно создание такой системы фильтров, чтобы интенсивность отклика каждого фильтра была тем больше, чем ближе период входного сигнала к значению периода, на которое настроен этот фильтр. Конечно, это идеализация, но подумаем, как к этой идеализации можно приблизиться. Для музыкального сигнала справедливы два утверждения:

— созвучия, состоящие из нот, отстоящих на целое число октав, появляются в сигнале реже, чем терции, кварты или квинты, при этом пока не разработаны алгоритмы, позволяющие достоверно разделять ноты, отстоящие на октаву;

— музыкальный сигнал имеет более строгую, чем речевой сигнала, организацию спектральной структуры: понятие музыкального строя подразумевает определённое соотношение высот тонов звуков, выраженность ощущения высоты тона означает определённое соотношение частоты основного тона и частот обертонов.

Втрое утверждение означает, что при распознавании любого музыкального сигнала мы всегда обладаем априорной информацией о нём. Например, если известен музыкальный строй инструментов, участвующих в исполнении, а также основной тон эталона, по которому ансамбль (инструмент) настраивался, автоматически становятся известны примерные значения частот основных тонов и обертонов, соответствующие любой ноте, которую может исполнить данный ансамбль (инструмент).

У звуков, чьи высоты тонов отстоят на целое число октав k (т.е. частоты их основных тонов соотносятся как 2^k) в идеальном случае все 100% гармоник перекрываются. В случае реальных звуков это означает, что частоты основного тона и обертонов, например ноты «ля» третьей октавы примерно равны частотам обертонов ноты «ля» второй октавы с чётными номерами, или частотам обертонов ноты «ля» первой октавы с номерами, кратными 4.

Известно, что если частоты основных тонов звуков Z_1 и Z_2 соотносятся как $\frac{m}{n}$, $m, n \in Z$, $m, n \geq 1$, то гармоники звука Z_1 с номерами $P(k) = n \cdot k$, где $k = 1, 2, 3, \dots$ перекрывают гармоники звука Z_2 с номерами $Q(k) = m \cdot k$, где $k = 1, 2, 3, \dots$. Для звуков с основными тонами, отличающимися на целое число октав, это означает, что большинство их гармоник, имеющих одинаковые частоты будут иметь номера, являющиеся степенями 2. Т.е. система полосовых фильтров с центральными частотами, образующими ряд $f_{ii} = f_0 \cdot 2^i$

(частоту f_0 будем называть *основной частотой системы*), будет иметь мощность отклика на тональный звук тем большую, чем ближе будет основной тон этого звука к центральной частоте любого из фильтров системы. Иными словами, можно построить систему фильтров, настроенную, например, на ноты «ля» определённого числа октав. Для этого центральные частоты полосовых фильтров системы должны быть равны частотам основных тонов звуков «ля» этих октав.

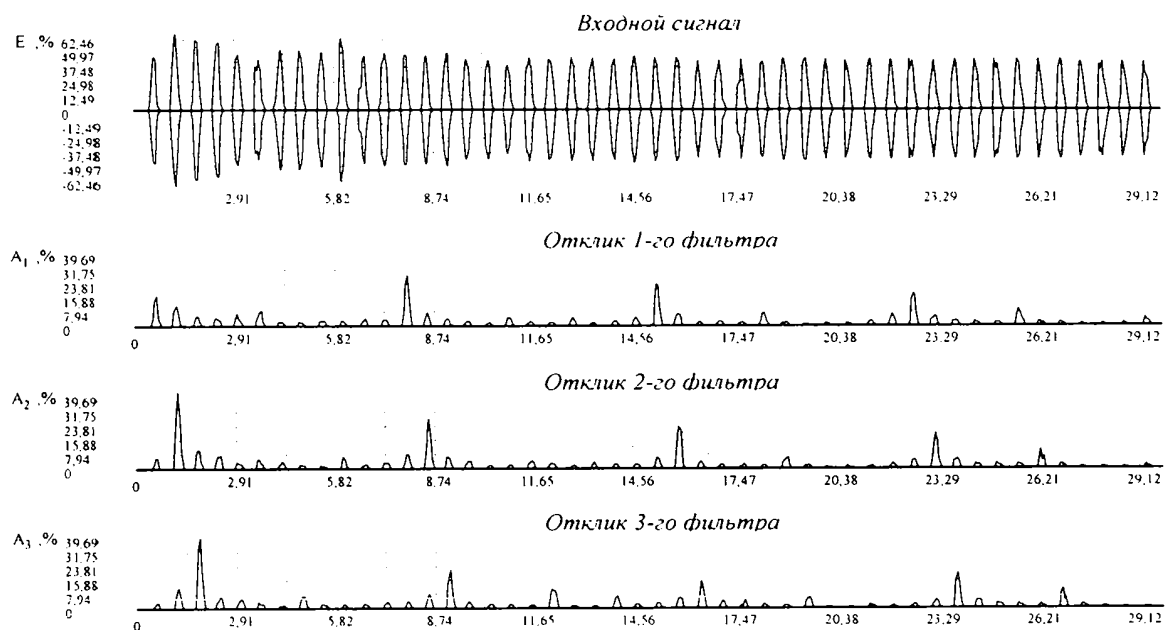


Рис. 33 Испытательный сигнал и отклики на него трёх систем фильтров

На рисунке 33 приведены огибающие испытательного сигнала и отклики трёх систем фильтров, настроенных на разные ноты. В качестве испытательного сигнала использована запись хроматической гаммы от «до» малой октавы до «си» третьей октавы, исполненной на скрипке. Три системы фильтров настроены соответственно на ноты «до», «до-диез» и «ре». Как и ожидалось, максимальные амплитуды откликов соответствуют именно тем нотам, на которые фильтры настроены, значительные интенсивности откликов соответствуют малым секундам, квартам и квинтам от нот, на которые настроены системы фильтров (соответственно, 1, 5 и 7 отклики после каждого максимального). Такое поведение системы фильтров обсудим позже, а пока рассмотрим, что она из себя представляет.

Каждая система фильтров состоит из 8 полосовых фильтров. В качестве полосовых фильтров используются рекурсивные фильтры 2-го порядка, разностные уравнения которых получены методом билинейного z -преобразования из коэффициентов передачи аналоговых прототипов. Фильтры-прототипы построены по схеме на рис. 34, имеют одинаковую добротность и отличаются центральной частотой.

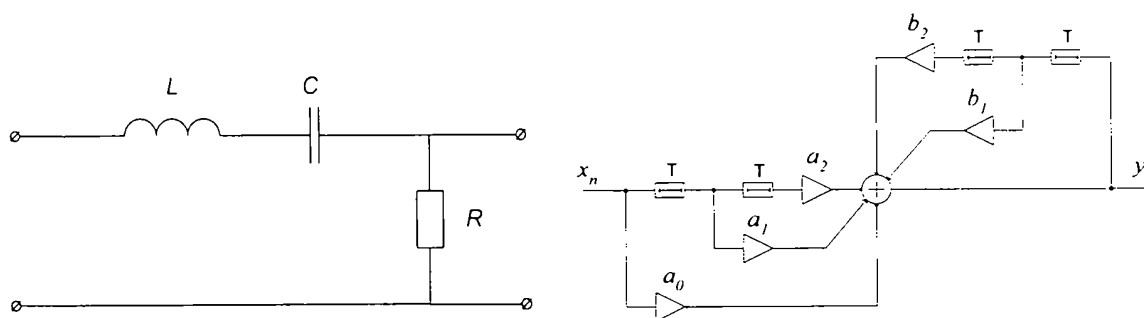


Рис. 34 Схема аналогового фильтра-прототипа и рекурсивного фильтра 2-го порядка

На рисунке 35 приведены АЧХ, ФЧХ и характеристик групповой задержки системы с основной частотой 100 Гц.

Обсудим поведение построенной системы фильтров. В действительности система откликается не только на звуки с основным тоном, отличающимся от основной частоты системы на целое число октав. Звуки, отличающиеся по высоте тона на кварту и квинту, имеют примерное соотношение основных частот соответственно $4/3$ и $3/2$. В таком соотношении находятся «до» и «фа» (кварты) и «до» и «соль» (квинта). Тогда каждая 3-я гармоника «фа» имеет ту же частоту, что каждая 4-я гармоника «до», т.е. совпадают частоты гармоник: 3-ей и 4-ой, 6-ой и 8-ой, 12-ой и 16-ой. Это означает, что при звучании ноты «фа» её 3-я, 6-я и 12-я гармоники будут подавлены системой фильтров (будут пропущены 1-я, 2-я, 4-я, 8-я и т.д.), настроенных на «фа», но будут пропущены системой, настроенной на «до». Очевидно, что это приведёт к ошибочной сегментации в том смысле, что будет принято решение о появлении двух звуковых объектов вместо одного. Следовательно, необходим дополнительный параметр для принятия решения о появлении нового звукового объекта. В качестве такого параметра предлагается использовать основной тон отклика системы фильтров.

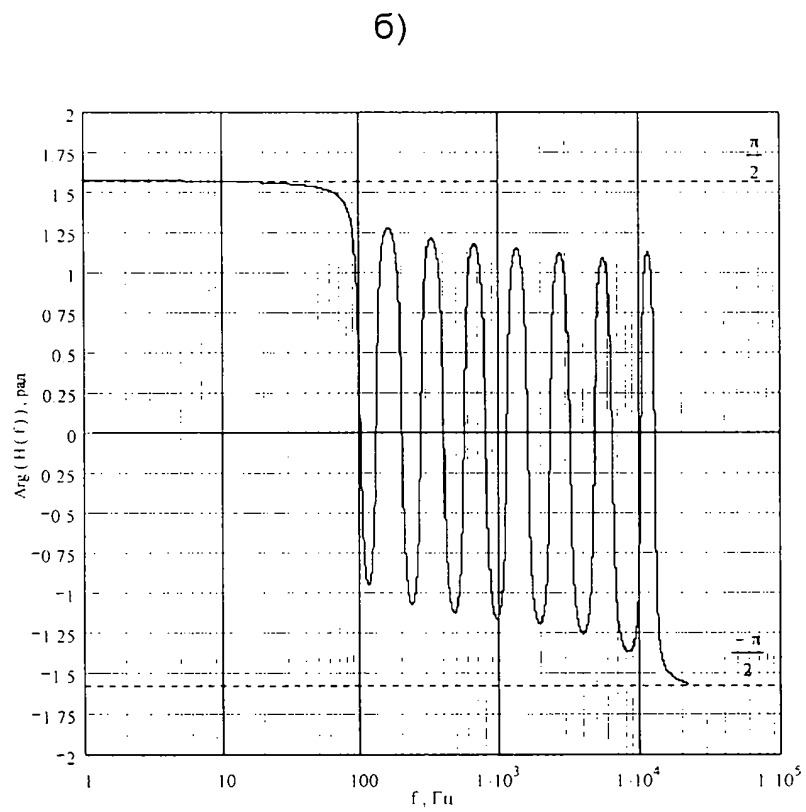
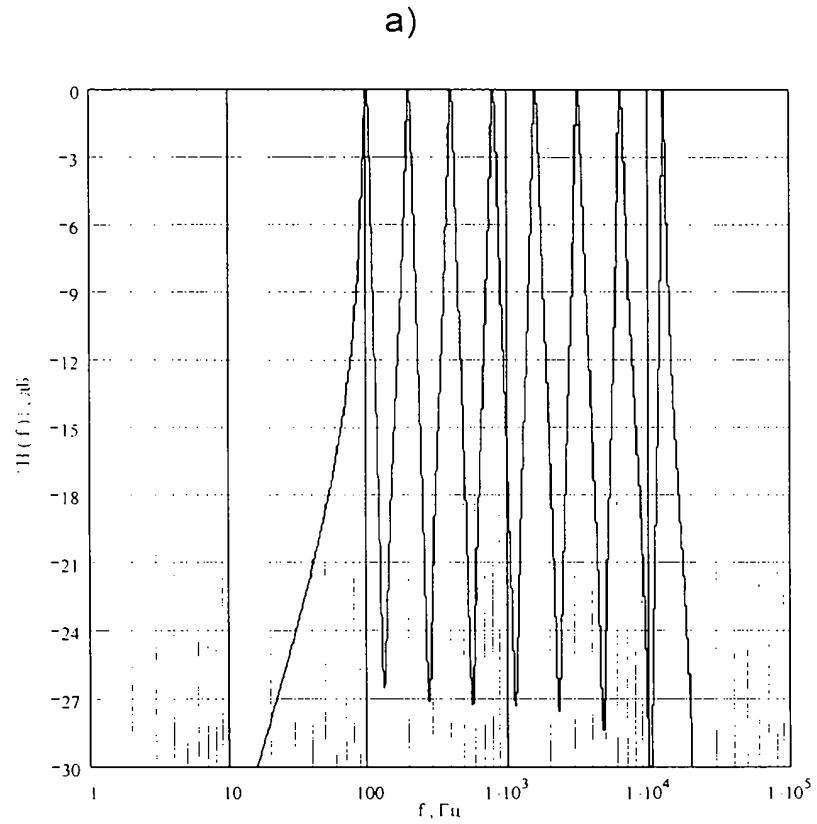


Рис. 35 Характеристики системы с основной частотой 100 Гц: а) АЧХ, б) ФЧХ

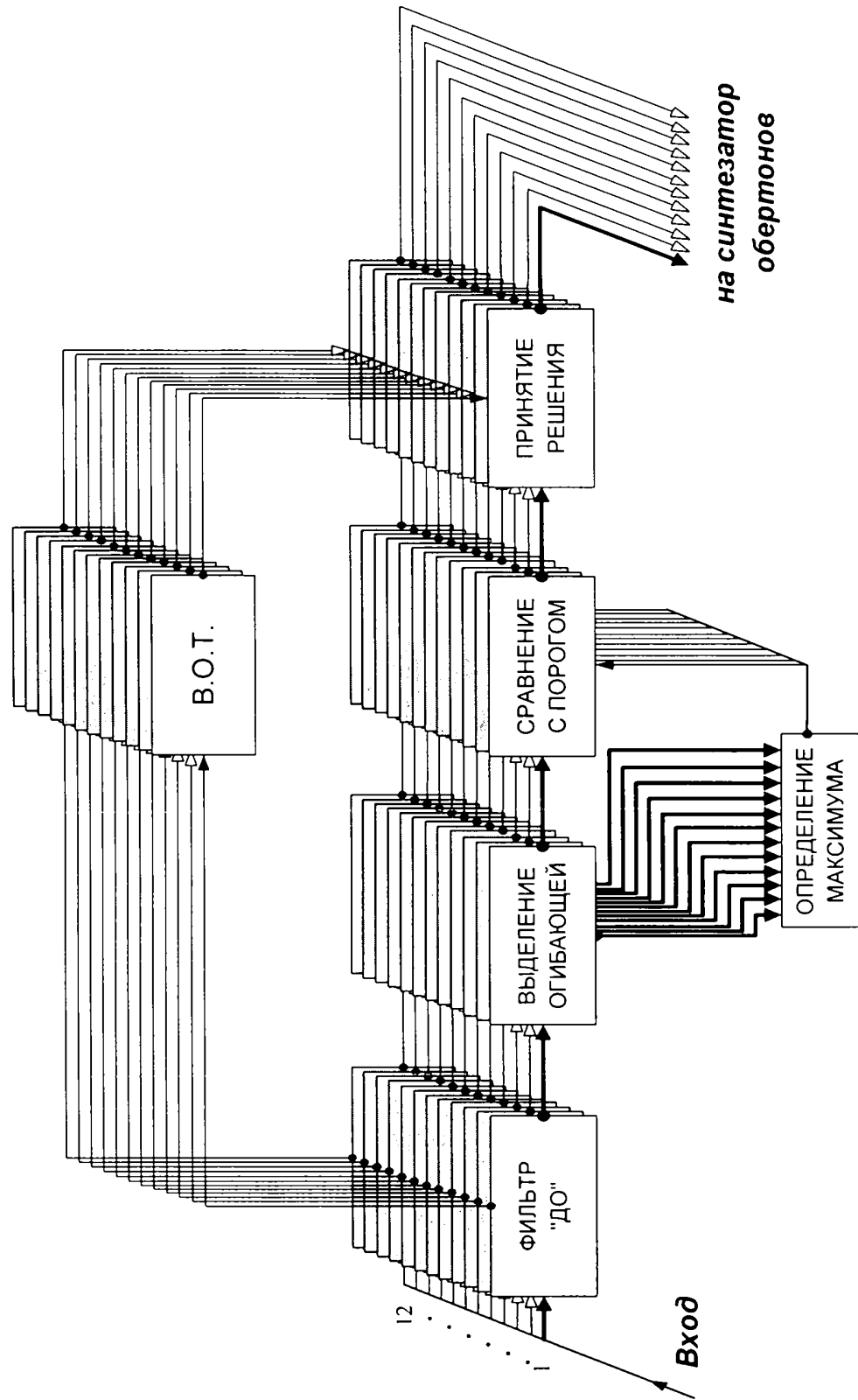


Рис. 36 Схема сегментации на основе фильтрации по периоду

Если основной тон выходного сигнала системы фильтров отличается на целое число октав от основной частоты системы, а значение огибающей превышает некий порог, принимается решение о появлении нового звукового объекта. Для иллюстрации работы алгоритма будем устанавливать значение огибающей отклика системы равным нулю, если не принято решение о появлении нового звукового объекта:

$$E_i^c(t) = \begin{cases} E_i(t) & , f_{OT}(t) = f_0 \cdot 2^k , E_i(t) > E_{ПОР}(t) \\ 0 & \text{иначе} \end{cases} , \text{ где (11)}$$

f_0 — основная частота системы фильтров,

$E_i(t)$ — огибающая выходного сигнала (отклика) системы фильтров,

$E_{ПОР}(t)$ — пороговое значение огибающей.

Схема системы автоматической сегментации на основе априорной информации и «фильтрации по периоду» приведена на рисунке 36. В основе системы — 12 фильтров с АЧХ вида рис. 35 а), разделяющих входной сигнал на сигналы широкополосные в отличие от ранее описанных систем. Схема обработки откликов всех 12-ти фильтров одинакова: последовательное подавление сигналов с интенсивностями ниже пороговой и сигналов с ОТ (ВОТ — выделитель основного тона), не соответствующими ожидаемым для данного фильтра значениям. Более жирными линиями на схеме показано прохождение основных (*информационных*) сигналов, более тонкими линиями — сигналов, управляющих работой схемы сегментации. Интенсивности откликов сравниваются для определения максимального для данной выборки значения. В блоке «сравнение с порогом» сигнал с интенсивностью меньше пороговой приравнивается нулю. В блоке «принятие решения» нулю приравниваются сигналы с несоответствующим ОТ.

Выходных сигналов 12 — по количеству фильтров. Сигналы эти названы *информационными*, поскольку несут информацию об обнаруженных звуковых объектах. И информация эта почти полная: мощности откликов пропорциональны мощностям звуковых объектов, номер отклика (по номеру фильтра)

позволяет с точностью до октавы определить основной тон обнаруженного звукового объекта (т.е. утверждать, например, что звучит «ля», но не известно, какой октавы).

Важно отметить, что точное определение высоты тона звуковых объектов по откликам фильтров некорректно – информация в них неполная. Более того, при определении ОТ анализом входного сигнала можно исправить и ошибки сегментации. В то же время, предлагаемый алгоритм сегментации не только локализует во времени звуковые объекты, но и даёт алгоритму определения высоты тона набор *гипотез*, т.е. значений ОТ служащих наиболее вероятными кандидатами. Выделение ОТ многоголосных сигналов намного сложнее, чем одноголосных, и наличие гипотез, полученных ещё до начала анализа (на предмет определения ОТ) повышает достоверность распознавания. Поэтому предлагаемый алгоритм сегментации можно назвать *ориентированным на распознавание*.

На рисунке 37 приведены отклики 12 систем фильтров с основными частотами, составляющими хроматическую гамму от «до» до «си» малой октавы. Огибающая входного музыкального сигнала и соответствующая нотная запись приведены на том же рисунке. Музыкальный сигнал синтезирован программным синтезатором с помощью звуков фортепиано. Отклики фильтров «почищены» по алгоритму (11), в качестве $E_{пор}(t)$ использовано значение $0,4 \cdot E_{МАКС}(t)$, где $E_{МАКС}(t)$ – максимальное по всем 12-ти откликам значение огибающей на данном интервале времени. Результаты, можно сказать, близки к идеальным: правильно определено количество звучащих нот, моменты их появления и затухания, однако при синтезе той же мелодии с помощью звуков трубы правильно локализованы лишь 50% звуковых объектов. Такое различие вызвано различиями свойств музыкальных инструментов. Звуки трубы богаче обертонами, при этом наиболее интенсивны обертоны среднего диапазона спектра звука. Звуки фортепиано, напротив имеют наиболее интенсивные первые несколько обертонов, амплитуда обертонов спадает с частотой. В результате поведение системы сегментации сильно зависит от свойств сигнала.

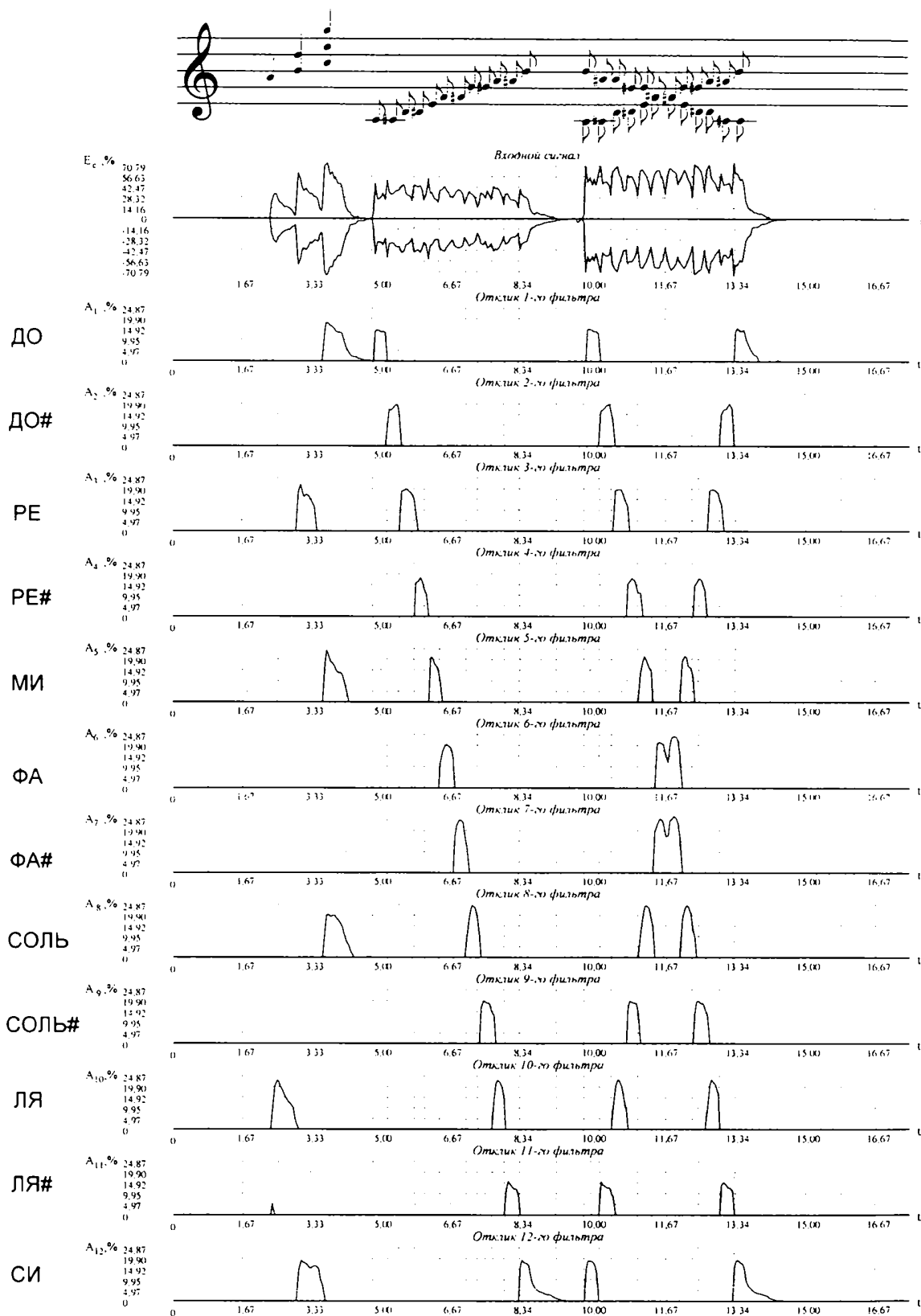


Рис. 37 Пример работы алгоритма сегментации

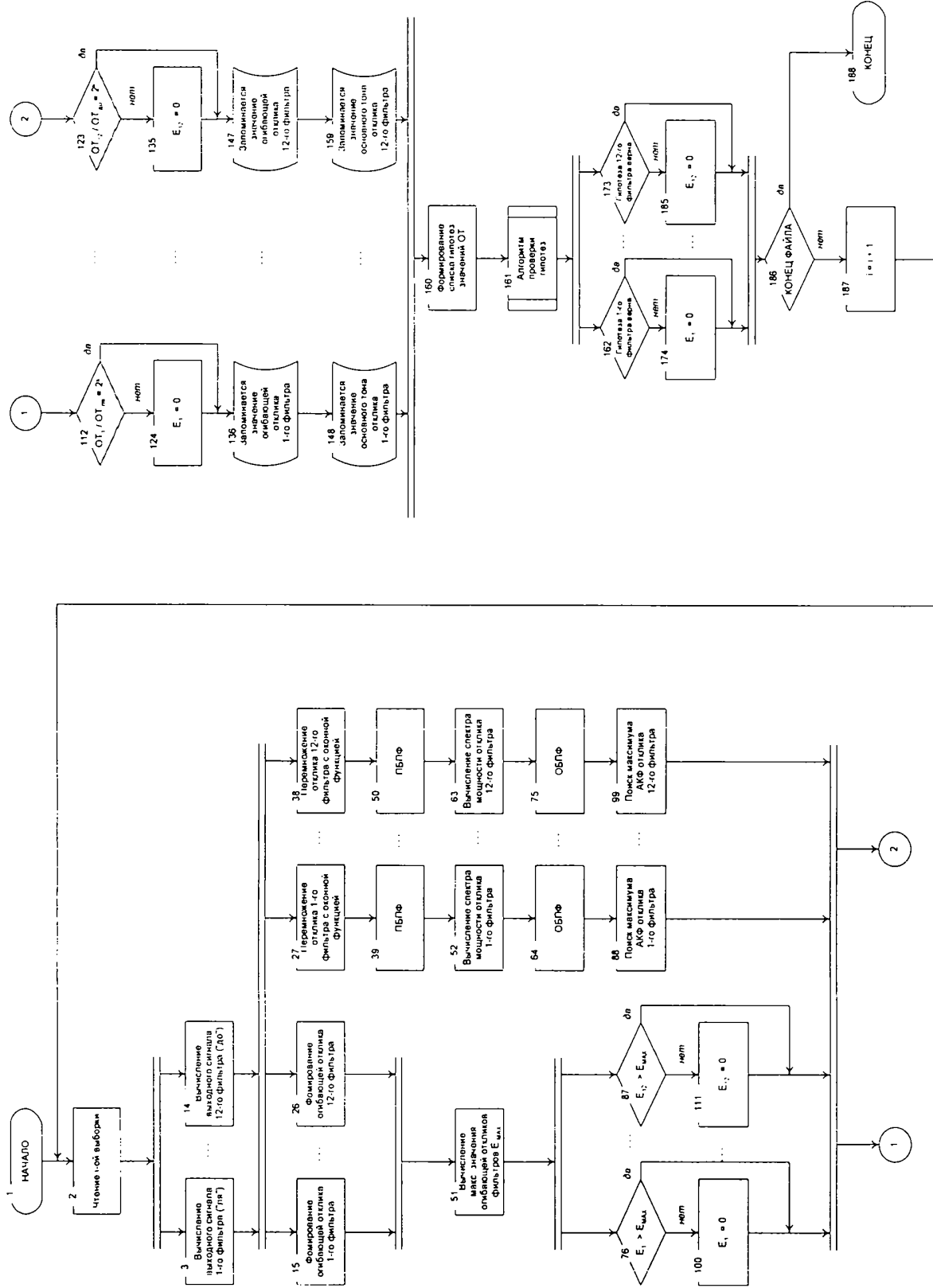


Рис. 38 Разработанный алгоритм сегментации музыкальных сигналов

Скажем несколько слов о добротности системы фильтров. Уменьшение добротности ухудшает избирательность, увеличение добротности ведёт к «затягиванию» фронтов выходного сигнала. Стремиться к максимальному увеличению избирательности нецелесообразно: в реальных сигналах всегда имеет место отклонение частот обертонов и основных тонов от точных значений, определяемых музыкальным строем, эталоном для настройки и номером гармонии. Выбор добротности системы — отдельная проблема, требующая большого количества экспериментов. Целесообразным представляется построение *настраиваемых* систем фильтров для точной подстройки центральных частот полосовых фильтров при отклонении высот тонов от эталонных.

Подведём итоги. Прежде всего, предлагаемый алгоритм сегментации в определённых условиях оказался работоспособным. Основные достоинства и недостатки алгоритма определяются использованием априорной информации. Априорная информация может быть не всегда точной: в реальных сигналах возможны следующие отклонения:

- отклонения от ожидаемого строя (соотношения между высотами тонов)
- отклонения от ожидаемых значений высот тонов (отклонения от эталонных значений)
- отклонения частот обертонов от значений, кратных основной частоте.

С другой стороны, именно использование априорной информации позволяет ожидать и фиксировать появление звуковых объектов с определёнными параметрами: каждый из 12-ти фильтров настроен на свою ноту, хотя истинное значение высоты тона не определяется. Именно использование априорной информации позволяет ещё при сегментации составить набор гипотез для алгоритма определения ОТ звуковых объектов.

Исследования требует вопрос о виде АЧХ разделяющих фильтров. Прежде всего, очевидна необходимость высокой прямоугольности АЧХ фильтров, составляющих системы. Рассмотренный вариант — наиболее простой, использованы фильтры 2-го порядка, минимизированы объём требуемой памяти

и вычислительная сложность. Неочевидно, что области пропускания систем фильтров, настроенных на разные ноты не должны перекрываться, возможно, такое перекрытие повысит устойчивость системы к описанным выше отклонениям частот компонентов звуковых объектов.

6.2. РАЗРАБОТКА АЛГОРИТМА ОПРЕДЕЛЕНИЯ ОСНОВНЫХ ТОНОВ ЗВУКОВЫХ ОБЪЕКТОВ ПЕРЕКРЫВАЮЩИХСЯ ВО ВРЕМЕНИ

Как было установлено в разделе 3.6, процедура принятия решения о *высоте тона ноты по значению частоты основного тона звукового объекта* (именно так и формулируется часть стоящей задача распознавания музыкального сигнала) — задача тривиальная. Гораздо более сложной процедурой, являющейся источником ошибок, является *получение значения признака*, по которому принимается решение. И признаком этим является основной тон звукового объекта.

Разработанный алгоритм сегментации во многом облегчает получение значения признака — список возможных значений-кандидатов (гипотез) частоты основного тона сформирован. Остаётся проверить достоверность этих гипотез, оставив подтверждённые и отбросив ложные. Кроме того, гипотезы частот основных тонов сформированы *с точностью до октавы*, т.е. необходимо ещё и уточнение значения признака.

Известна причина возникновения ложных гипотез при сегментации сигнала — совпадение частот гармоник и основных тонов звуковых объектов, подлежащих распознаванию. При определённом соотношении частот ОТ звуков созвучия формируются ложные гипотезы в откликах фильтров, не настроенных на частоты ОТ этих звуков. Даже при содержании в сигнале всего одного звукового объекта могут возникнуть более одной гипотезы частоты основного тона. Таким образом, при проверке гипотезы $F_{гип}$ частоты ОТ следует проверить, была она сформирована компонентами звукового объекта с частотой ОТ $F_{гип}$ или какого-либо другого объекта. Иными словами, необходимо убедиться, что в спектре распознаваемого сигнала содержатся основной тон с частотой $F_{гип}$ или хотя бы его обертоны с достаточной мощностью. О наличии гармонических составляющих сигнала можно судить по амплитудному спектру, полу-

чаемому из ДПФ выборки сигнала. Для сигналов, содержащих гармонические составляющие характерно наличие локальных максимумов на частотах коэффициентов ДПФ, близких к частотам этих гармонических составляющих.

При проверке гипотезы ОТ $F_{гип}$ целесообразно убедиться не просто в наличии гармонических составляющих с частотой $F_{гип}$ и её гармоник, но убедиться в наличии составляющих, не совпадающих по частоте с основными тонами, соответствующими остальным гипотезам ОТ, а также с их обертонами. Факт совпадения частот *всех* обнаруженных гармоник $F_{гип}$ и самой $F_{гип}$ с частотами основных тонов и гармоник остальных гипотез означает либо что $F_{гип}$ находится в октавном соотношении с частотой одной из остальных гипотез, либо обертоны звуковых объектов, сформировавших остальные гипотезы, сформировали и гипотезу $F_{гип}$, и, соответственно она является *ложной*. Разработанный алгоритм сегментации не может сформировать гипотезы ОТ, отличающихся на октаву: каждый гребенчатый фильтр настроен на периоды ОТ, отличающихся на октаву, т.е. разделение звуковых объектов, с ОТ, отличающимися на целое число октав *не происходит*.

Итак, при проверке гипотезы частоты ОТ $F_{гип}$ необходимо убедиться либо в наличии самого основного тона с частотой $F_{гип}$, либо в наличии его обертонов, не совпадающих по частоте с основными тонами и их обертонами, соответствующими остальным гипотезам. Очевидно, для этого

при проверке гипотезы основного тона F'_1 из набора гипотез $\{F'_1, F'_2, \dots, F'_N\}$ необходимо в спектре сигнала полностью подавить составляющие с частотами $\{F'_2, F'_3, \dots, F'_N\}$, а также их гармоники и убедиться, что оставшиеся гармонические составляющие могут являться основным тоном с частотой F'_1 , либо его обертонами.

Убедившись, что гипотеза не является ложной, можно утверждать, что звуковой объект существует, и остаётся уточнить значение его основного тона. Как было установлено ранее, методы определения ОТ, применимые к речевым и одноголосным музыкальным сигналам не применимы к многоголосным звучаниям. Однако, подавив все спектральные составляющие сигнала, не относя-

щиеся к звуковому объекту, чей основной тон уточняется, можно применить и уже известные методы. Таким методом **выбран метод определения ОТ по кратковременной АКФ**, как просто реализуемый и достаточно точный для распознавания музыкальных сигналов.

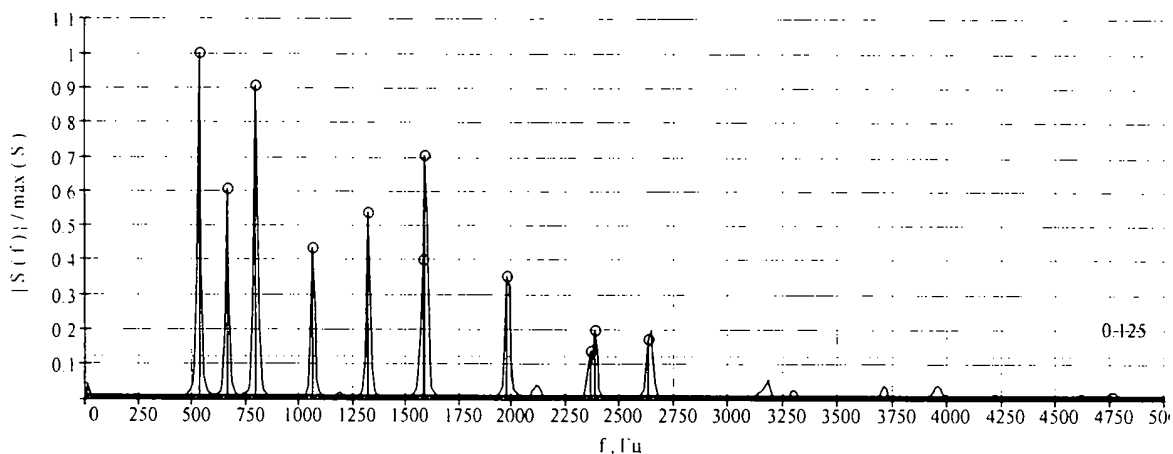


Рис. 39 Амплитудный спектр реализации аккорда до-мажор («до»+«ми»+«соль»)

Рассмотрим конкретный пример. Допустим, анализируется выборка сигнала, содержащая 3 звуковых объекта с частотами ОТ, соответствующими нотам «до», «ми» и «соль». Амплитудный спектр сигнала этой выборки представлен на рисунке 39.

Выделим локальные максимумы, превышающие некий порог. Опытным путём было выбрано оптимальное по критерию минимума вероятности ошибочного решения значение порога $0,125 \cdot A_{MAX}$, где A_{MAX} — максимальное значе-

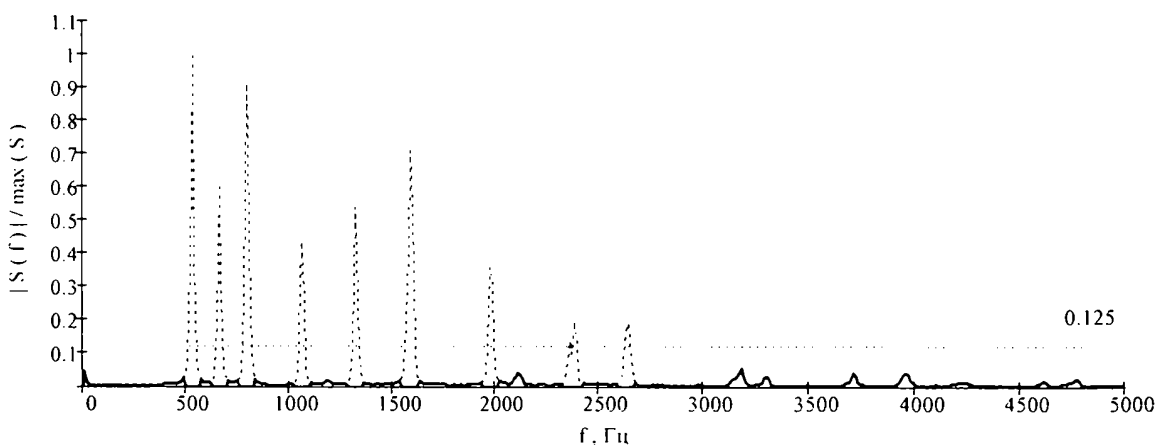


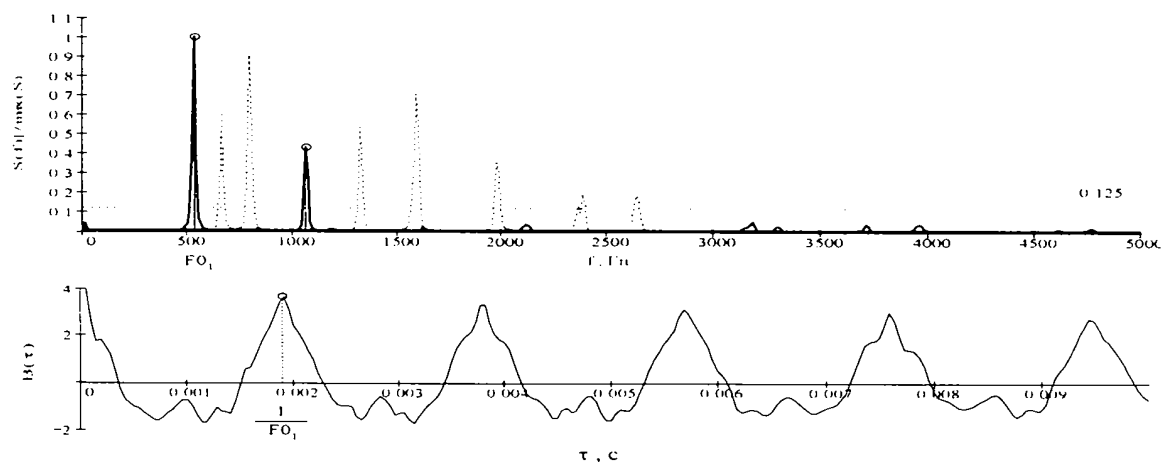
Рис. 40 Амплитудный спектр при проверке ложной гипотезы частоты основного тона

ние амплитудного спектра на данной выборке. Набор гипотез частот основных тонов выглядит следующим образом: {494,7 Гц; 523,2 Гц; 660,1 Гц; 782,5 Гц}. Первая гипотеза соответствует «си» первой октавы и является ложной, остальные три гипотезы соответствуют нотам «до», «ми» и «соль» второй октавы.

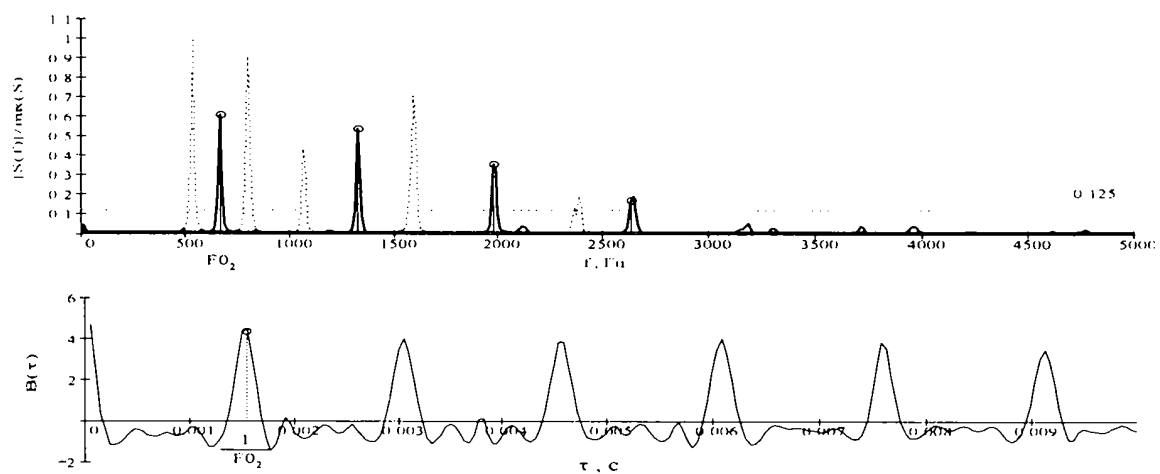
При проверке гипотезы «си» первой октавы в спектре исходного сигнала будут подавлены все составляющие с частотами основных тонов, соответствующих «до», «ми» и «соль» второй октавы, а также их гармоники. Результат подавления этих составляющих представлен на рисунке 40. Сплошной жирной линией изображён амплитудный спектр после подавления составляющих, не относящихся к проверяемой гипотезе, тонкой пунктирной – спектр исходного сигнала. Очевидно, ни одного локального максимума, превышающего выбранный порог, после подавления не обнаружено, следовательно, предположение о наличии звукового объекта с частотой ОТ 494,7 Гц ложно.

На примере проверки оставшихся трёх гипотез рассмотрим и предложенный метод уточнения частот основных тонов звуковых объектов, гипотезы о существовании которых подтвердились. На рисунке 41 изображены амплитудные спектры и соответствующие им автокорреляционные функции при проверке гипотез о существовании звуковых объектов с основными тонами 523,2 Гц (рис. 41 а), 660,1 Гц (рис. 41 б) и 782,5 Гц (рис. 41 в). АКФ вычисляется обратным преобразованием Фурье квадрата амплитудного спектра, в соответствии с теоремой Винера-Хинчина. Во всех трёх случаях обнаруживаются гармонические составляющие, которые могли бы быть компонентами звукового объекта, гипотеза о наличии которого проверяется. Во всех трёх случаях наибольший локальный максимум кратковременной АКФ соответствует уточнённым значениям периодов основных тонов проверяемых звуковых объектов.

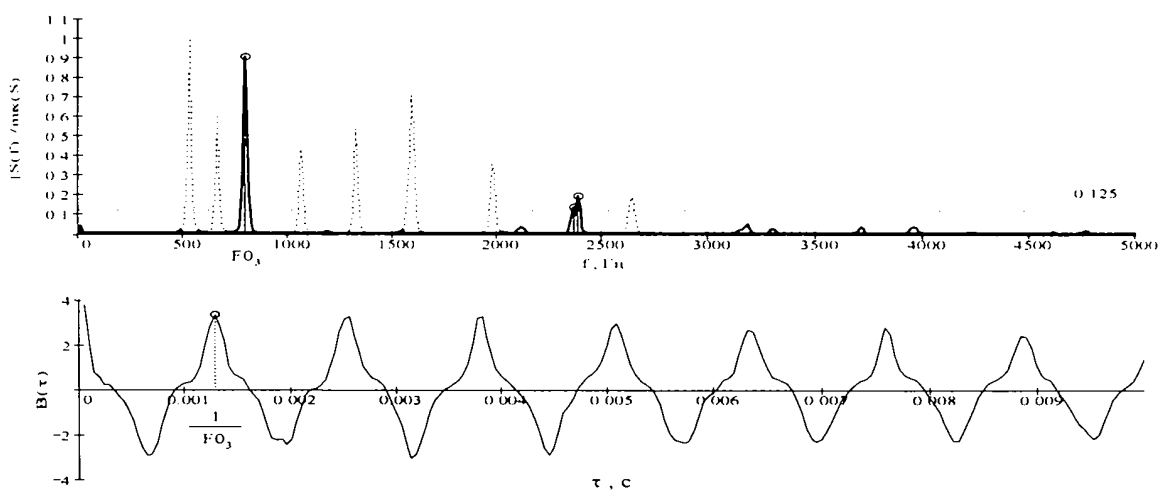
Подтверждённые гипотезы используются для классификации звуковых объектов по высоте тона: по уточнённым значениям частоты основного тона принимается решение о высоте тона. Решение о длительности ноты, соответствующей звуковому объекту также принимается лишь после подтверждения гипотезы.



a)



б)



в)

Рис. 41 Нормированные амплитудные спектры и соответствующие АКФ при проверке гипотез частот основного тона

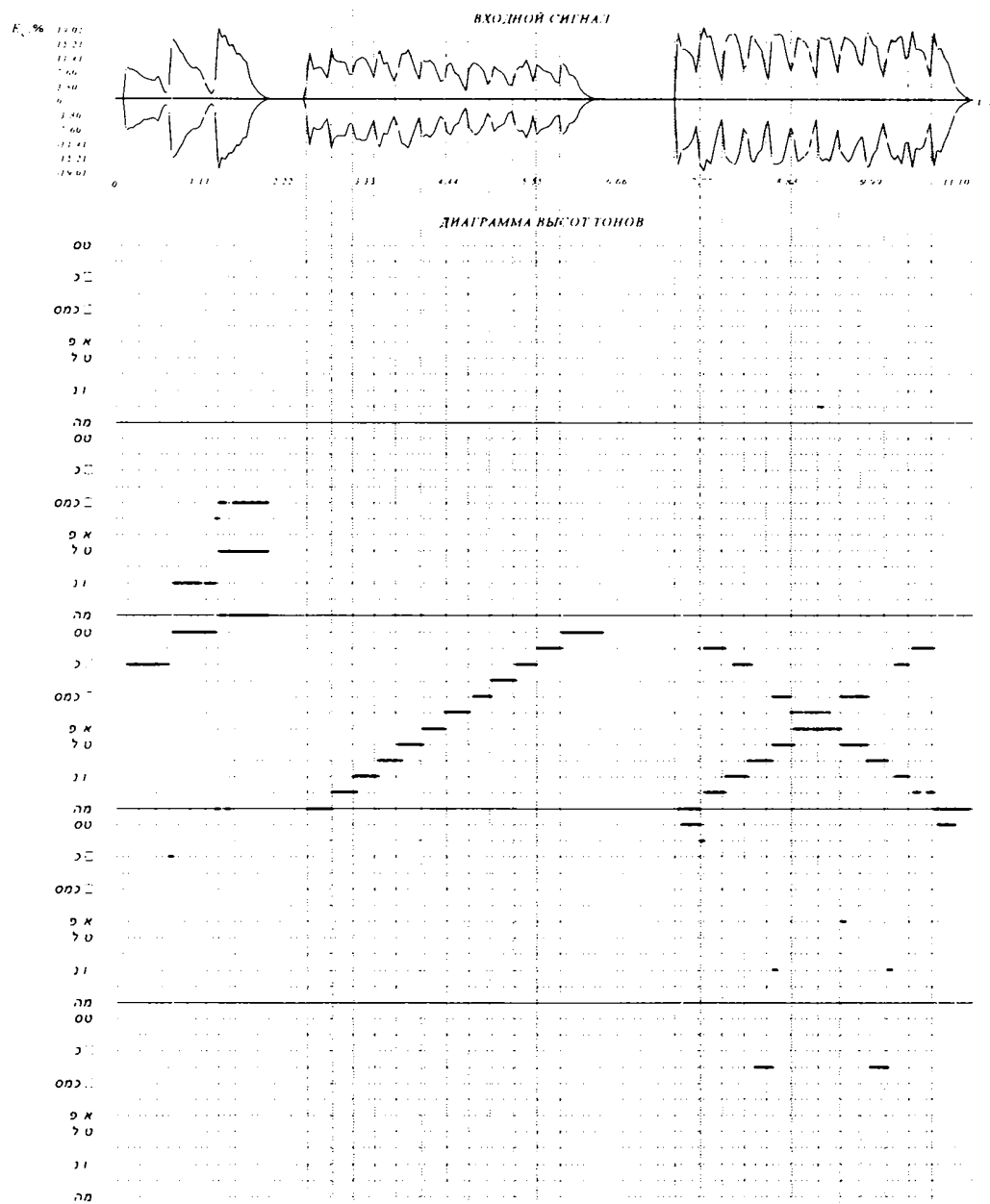


Рис. 42 Пример работы системы определения высоты тона

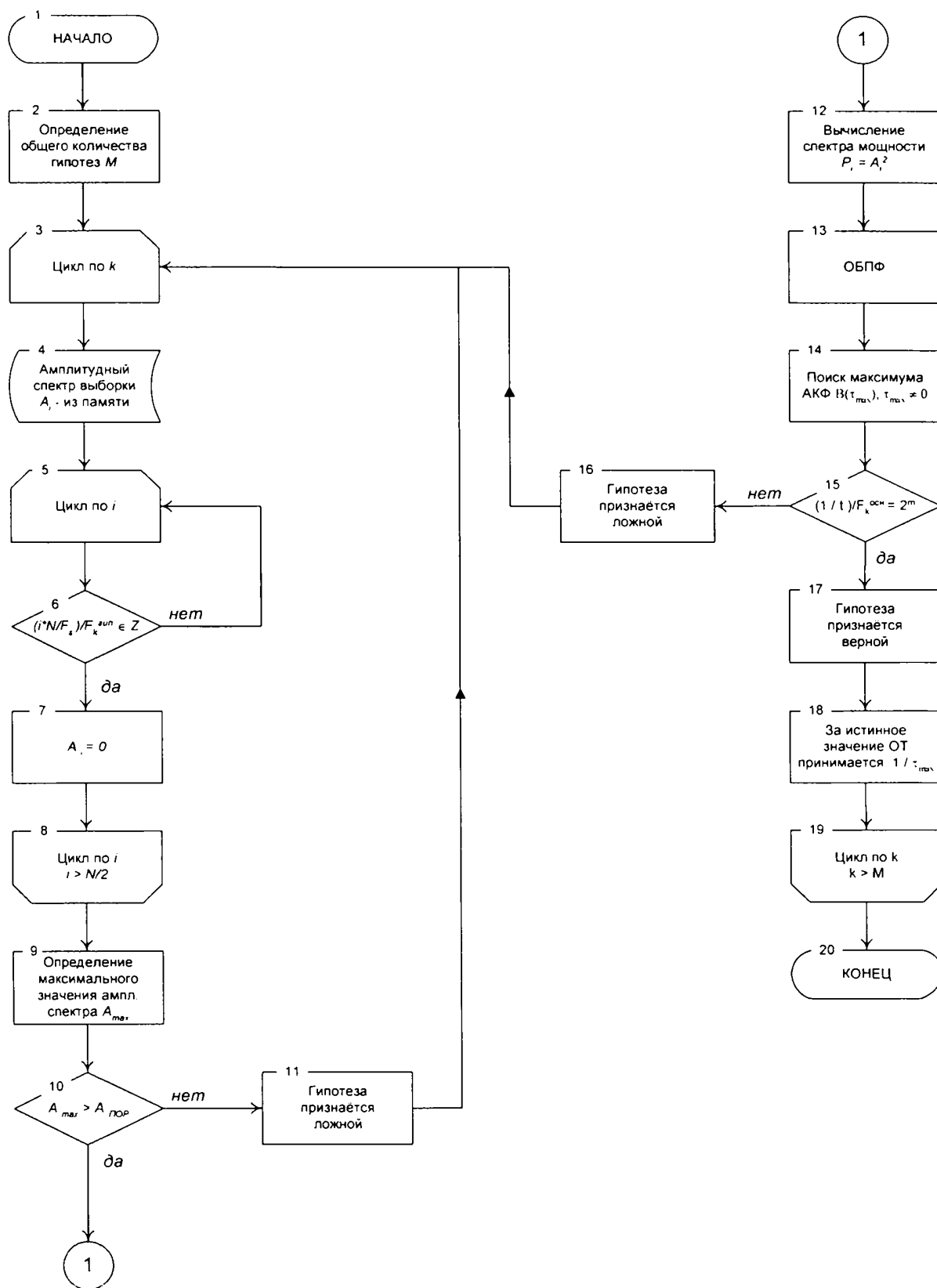


Рис. 43 Разработанный алгоритм проверки гипотез частот основного тона

За абсолютное значение длительности ноты принимается интервал времени, на котором огибающая отклика фильтра системы сегментации, сформировавшего подтверждённую гипотезу, отлична от нуля.

Если гипотеза оказалась ложной, это означает, что звукового объекта с такой частотой основного тона не существует. Следовательно, необходимо исправить и ошибку работы системы сегментации — приравнять нулю огибающую фильтра, сформировавшего ложную гипотезу на интервале времени, соответствующем этой самой гипотезе.

На рисунке 42 приведён пример работы алгоритма определения высоты тона — временная диаграмма высот тонов нот, соответствующих обнаруженным звуковым объектам. Этот пример соотносится с примером на рис. 37 — в качестве испытательного взят тот же сигнал.

Отметим особенности работы разработанной системы распознавания многоголосных музыкальных сигналов:

1. При наличии в звучании сигнала реверберации длительности нот несколько отклоняются от длительностей, предписанных нотной записью, по которой музыкальный сигнал синтезирован.

2. В моменты времени, соответствующие атакам звуковых объектов формируются и подтверждаются гипотезы об одновременном звучании звуковых объектов с основными тонами, отличающимися на полутон. Такое поведение системы можно объяснить лишь значительными отклонениями частоты ОТ на интервале нестационарности звукового сигнала от частоты ОТ на стадии установившихся колебаний. Наличие таких отклонений подтверждается и результатами проведённых в главе 5 исследований свойств звуковых объектов.

3. Наиболее часто встречающаяся ошибка определения частоты ОТ (и, соответственно, высоты тона ноты) — ошибка на целое число октав.

6.3. Выводы

1. Подход к сегментации, основанный на модели слухового восприятия (сегментация «по громкости») может быть использо-

ван для любого типа звучания, но ориентирован на определение моментов начала звучания, поэтому, в основном, используется для выделения ритма.

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

7. РАЗРАБОТКА АЛГОРИТМА ОБРАБОТКИ СИГНАЛА НА ОСНОВЕ РАСПОЗНАВАНИЯ

В соответствии с предложенной во «Введении» общей концепцией восстановления музыкальных сигналов в процессе обработки сигнала необходимо провести фильтрацию (подавление шумовых компонентов) и синтез утраченных компонентов звуковых объектов.

Разработанная системы распознавания способна предоставить все необходимые сведения для управления этими процессами. Во-первых, определены длительности звуковых объектов и их основные тона, следовательно, для сигнала каждой временной выборки в частотной области могут быть локализованы компоненты звуковых объектов, т.е. полезные компоненты восстанавливаемого сигнала. Полезными при разрабатываемом подходе признаются гармонические составляющие сигнала с частотами основных тонов обнаруженных звуковых объектов, а также составляющие с частотами гармоник этих основных

тонов. Все остальные гармонические составляющие признаются «шумовыми» — это могут быть, например, продукты нелинейных искажений, тональные помехи.

Очевиден принцип формирования АЧХ фильтра шумоподавления на основе данных, полученных при распознавании. Известны приблизительные значения частот обертонов звуковых объектов, амплитуды которых оставим без изменений, т.е. модуль коэффициента передачи фильтра на каждой из этих частот и в некоторой области (справа и слева) каждой из этих частот должен быть равен 1. Ширина полосы частот Δf , содержащей обертонов полезного сигнала, может и должна быть регулируемой. Проведённый (глава «Свойства звуковых объектов») анализ результатов исследования структуры звуковых объектов показывает, что отклонения частот обертонов от среднего за длительность звукового объекта 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. Кроме того отклонения частот обертонов от средних значений не одинаковы для разных музыкальных инструментов (см. рис. 29).

В остальной области частот коэффициент передачи должен быть меньше единицы: $|\dot{H}(f)| = H_{под} < 1$ для подавления шумов (см. рис. 45). Соответственно, при таком подходе считается, что область частот, не занятая полезным сигналом, занята шумом.

Сразу можно отметить недостаток такого подхода: музыкальная фонограмма всегда содержит, кроме шумов носителя записи и шумов тракта формирование-запись-воспроизведение, «полезные шумы» — акустический фон помещения, в котором производилась запись, негармонические призвуки (в том числе и шумовые), возникающие в ходе исполнения или обусловленные особенностями звукообразования в том или ином инструменте. Очевидно, что при $H_{под} = 0$ фильтрация губительно скажется на впечатлении от прослушивания восстановленного сигнала. Пока не разработаны алгоритмы распознавания «полезных» шумов, разумно сделать $H_{под}$ регулируемым вручную или ав-

томатически, например, исходя из условия поддержания заданного минимального соотношения сигнал/шум.

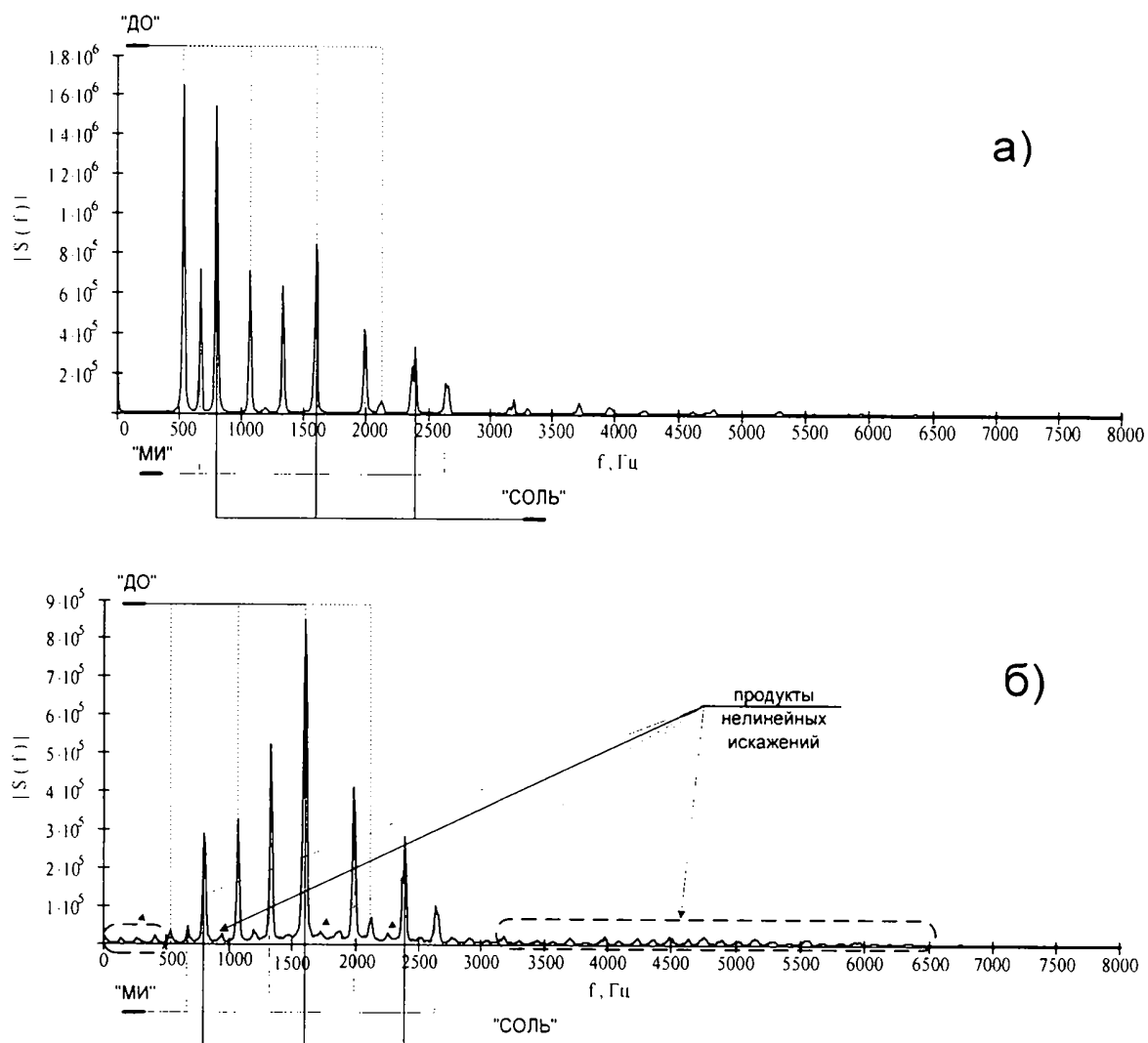


Рис. 44 Амплитудные спектры фрагментов музыкального сигнала: а) исходного, б) искажённого, подлежащего восстановлению

Рассмотрим конкретный пример — фильтрацию искажённой реализации созвучия нот «до», «ми» и «соль» второй октавы (сигнал, на котором демонстрировалась работа системы распознавания). Амплитудный спектр исходного (неискажённого) и искажённого фрагментов изображён на рисунке 44. В исходный сигнал внесены следующие изменения:

1. ограничена полоса частот фильтром с граничными частотами 500 Гц и 5,1 кГц;

2. добавлен шум с равномерным в указанной полосе спектром плотности мощности;
3. внесены сильные нелинейные искажения (КНИ около 10 %) ограничением мгновенных значений.

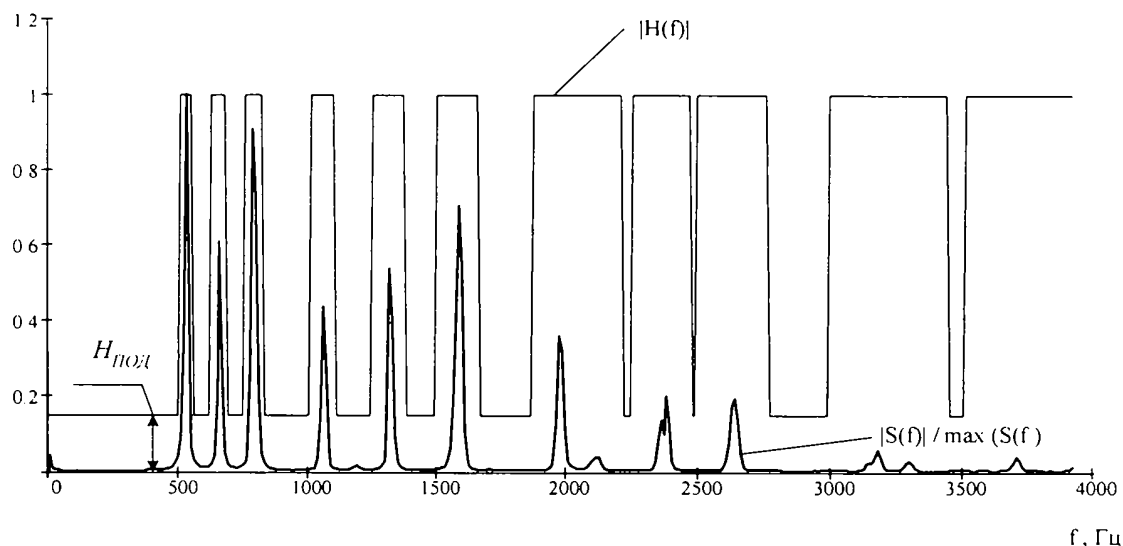


Рис. 45 Формирование АЧХ фильтра шумоподавления

На рисунке 45 изображена АЧХ фильтра для подавления шума в сигнале выборки, соответствующей рис. 44 а). На рисунке 46 приведён спектр восстанавливаемого фрагмента до и после прохождения через фильтр.

Наиболее простым способом сформировать фильтр с АЧХ сложного вида и линейной ФЧХ при цифровой обработке сигнала является БПФ фильтрация. Именно такой способ и выбран. АЧХ фильтра $H(f_k)$ формируется поэтапно, количество этапов зависит от количества обнаруженных звуковых объектов. Первоначально $H(f_k) = 0$, $k = 0 \dots N - 1$ (N — количество отсчётов выборки). Запишем выражение для АЧХ БПФ фильтра на этапе подавления компонентов, не относящихся к i -ому звуковому объекту:

$$H(f_k) = \begin{cases} 1, \text{ если } \left(\left[\frac{f_k}{F_i^{OT}} \right] = m > 1 \text{ и } \left| \frac{m \cdot F_i^{OT}}{f_k - m \cdot F_i^{OT}} \right| \leq Q \right) \text{ или } H(f_k) = 1 \\ H_{\text{под}}, \text{ в противном случае} \end{cases} \quad (12)$$

$$f_k = \frac{k \cdot F_{\text{д}}}{N} \text{ — частота } i\text{-ого коэффициента ряда Фурье,}$$

$F_{\text{д}}$ — частота дискретизации восстанавливаемого сигнала,

N — длина выборки,

F_i^{OT} — частота основного тона i -того звукового объекта,

$$\left[\frac{f_k}{F_i^{\text{OT}}} \right] = m \text{ — целая часть } \frac{f_k}{F_i^{\text{OT}}},$$

Q — величина эквивалентной добротности, характеризующей ширину полосы пропускания фильтра в окрестности m -ого обертона F_i^{OT} .

$$\left| \frac{f_k - m \cdot F_i^{\text{OT}}}{m \cdot F_i^{\text{OT}}} \right| \text{ — относительное отклонение частоты } k\text{-ого коэффициента}$$

ряда Фурье от частоты m -ого обертона F_i^{OT} ,

$H_{\text{под}}$ — модуль коэффициента передачи фильтра в области подавления.

Поясним выражение (12). Если частота k -ого коэффициента ряда Фурье отклоняется от частоты ближайшего обертона на величину, меньшую, чем задано полосой пропускания в окрестности этого обертона, то модуль коэффициента передачи фильтра шумоподавления на частоте k -ого коэффициента устанавливается равной 1. Если $H(f_k) = 1$, т.е. частота k -ого коэффициента ряда Фурье попала в область пропускания обертона другого звукового объекта, существующего на обрабатываемой выборке, то $H(f_k)$ остаётся без изменений. На частотах всех остальных коэффициентов $H(f_k)$ устанавливается равным $H_{\text{под}}$. Как было сказано выше, величины $H_{\text{под}}$ и Q устанавливаются пользователем вручную.

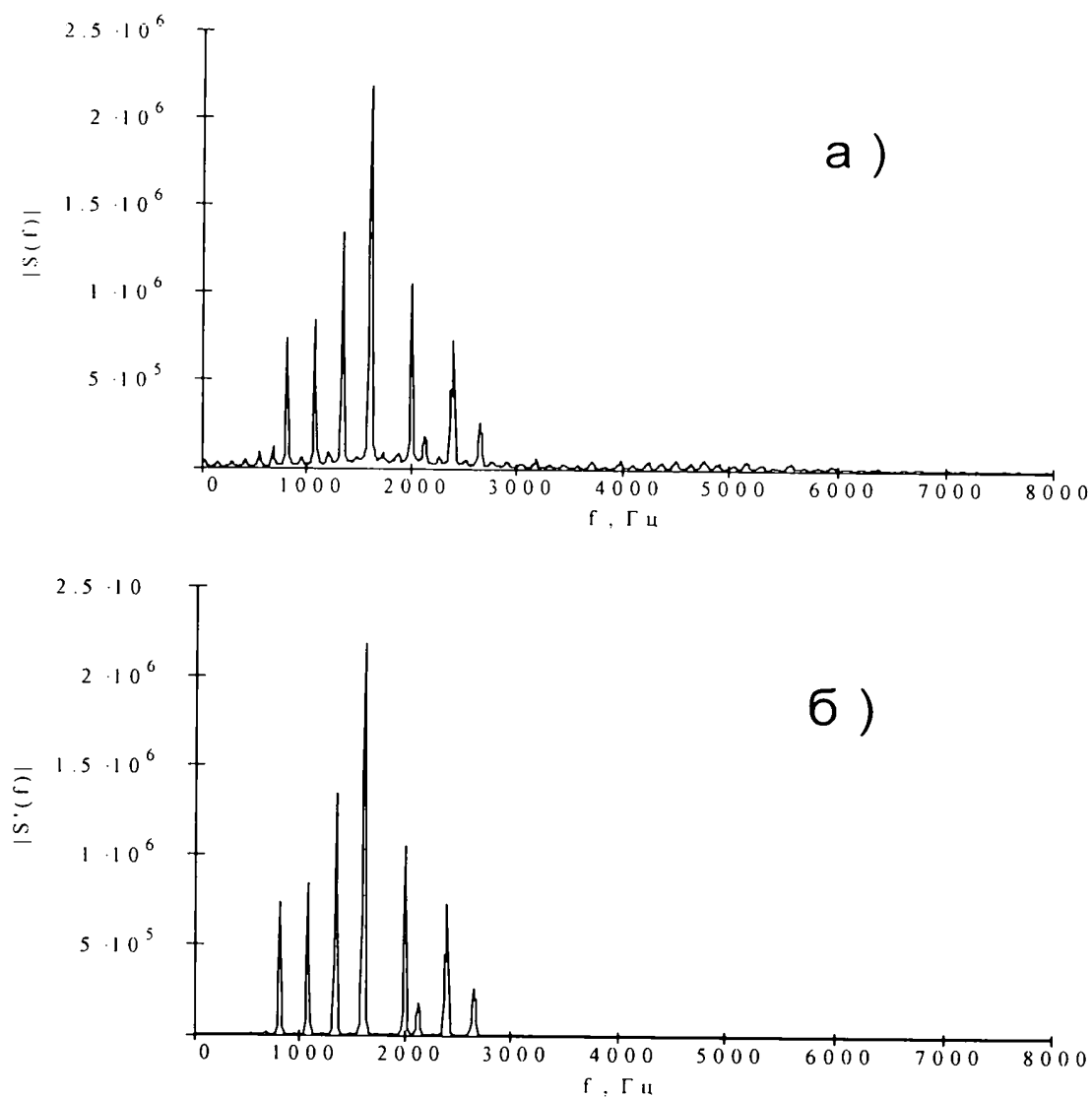


Рис. 46 Спектр фрагмента искажённого сигнала а) до фильтрации, б) после фильтрации

Как было отмечено ещё во «Введении» достоверность восстановления музыкального сигнала зависит от достоверности его распознавания. Можно назвать наиболее вероятные последствия неверного распознавания для фильтрации. Очевидно, могут быть подавлены компоненты нераспознанных звуковых объектов, в том числе и достаточно мощные, определяющие субъективное восприятие сигнала в целом. Могут быть пропущены без изменений продукты нелинейных искажений.

Обсудим процесс восстановления утраченных обертонов на основе данных, полученных при распознавании. Как установить факт утраты обертонов? Отсутствие обертонов может быть и следствием физических особенностей звукообразования в музыкальном инструменте, и следствием линейных искажений, например ограничения полосы частот при формировании, записи, воспроизведении или передаче сигнала. Очевидно, не имея дополнительных сведений (нам известны лишь частоты основных тонов звуковых объектов), мы можем ограничиться лишь сравнением мощности спектра в областях, в которых находятся гармонические составляющие, наиболее близкие по частоте к гармоникам основных тонов распознанных звуковых объектов (и, соответственно, принимаемых за обертоны), с некоторым порогом. По результатам сравнения можно констатировать «отсутствие» или наличие «обертон». Итак

для каждого обнаруженного звукового объекта становится возможным составить список недостающих обертонов.

Соответственно, отсутствующие обертоны необходимо синтезировать. Для этого нужно определиться с амплитудой, частотой и начальной фазой гармонического колебания, которое заменит утраченный обертон.

Для формирования амплитудной огибающей синтезируемых обертонов предлагается использовать выходные сигналы системы сегментации. Эти сигналы были специально сформированы таким образом, чтобы быть использованными при синтезе обертонов. С другой стороны, целесообразно и легко реализуемо при выбранном подходе

предоставить пользователю возможность вручную регулировать амплитуды отдельных обертонов, если для синтеза каждого обертона использовать отдельный генератор.

Обратим внимание на то, что интенсивность выходного сигнала системы сегментации пропорциональна интенсивности колебаний, имеющих определённый период, будем считать — интенсивности звукового объекта. Количество же обертонов, которые необходимо синтезировать может быть больше одного. Самый простой вариант — использовать одну и ту же огибающую для всех синтезируемых обертонов данного основного тона. Положительный эф-

фekt от такого подхода может быть вызван корреляцией интенсивности синтезированных обертонов с интенсивностью сохранившихся компонентов звукового объекта. Таким образом, синтезируемые обертоны будут иметь одинаковые по виду амплитудные огибающие, отличающиеся множителями, установленными пользователем.

Частота синтезируемого обертона, очевидно, определяется частотой основного тона того звукового объекта, к которому обертон относится, и целочисленным множителем — порядком обертона. Важно отметить, что хотя частота синтезируемого «обертона» получается целочисленным умножением частоты основного тона распознанного звукового объекта, такая спектральная составляющая не может быть названа гармоникой — всегда имеется погрешность в определении частоты ОТ, т.е. частота синтезированного обертона не обязательно будет кратна истинной частоте основного тона. Как и в случае амплитуд обертонов, есть возможность ручного регулирования частоты обертона, например, её отклонения от значения, кратного частоте ОТ, полученной при распознавании.

Наиболее интересный вопрос состоит в определении начальной фазы обертона. Учитывая психофизиологические особенности восприятия, целесообразно синтезировать разные обертоны с разными начальными фазами — для улучшения натуральности звучания. Т.е. необходим генератор случайных чисел, вносящий дополнительный сдвиг фазы отдельных обертонов. Есть возможность и ручного регулирования начальной фазы отдельного обертона. Очень важно обеспечить непрерывность мгновенной фазы обертона. Для этого при переходе от одной временной выборки к другой необходимо запоминать значение мгновенной фазы на последнем отсчёте выборки и использовать его в качестве опорного при начале синтеза на следующей выборке.

На рис. 47 представлена разработанная в соответствии с предыдущими рассуждениями схема синтеза обертонов.

Входными сигналами синтезатора обертонов являются выходные сигналы системы сегментации (см. рис. 37), выходной сигнала — сумма сигналов генераторов, модулированных по амплитуде и фазе.

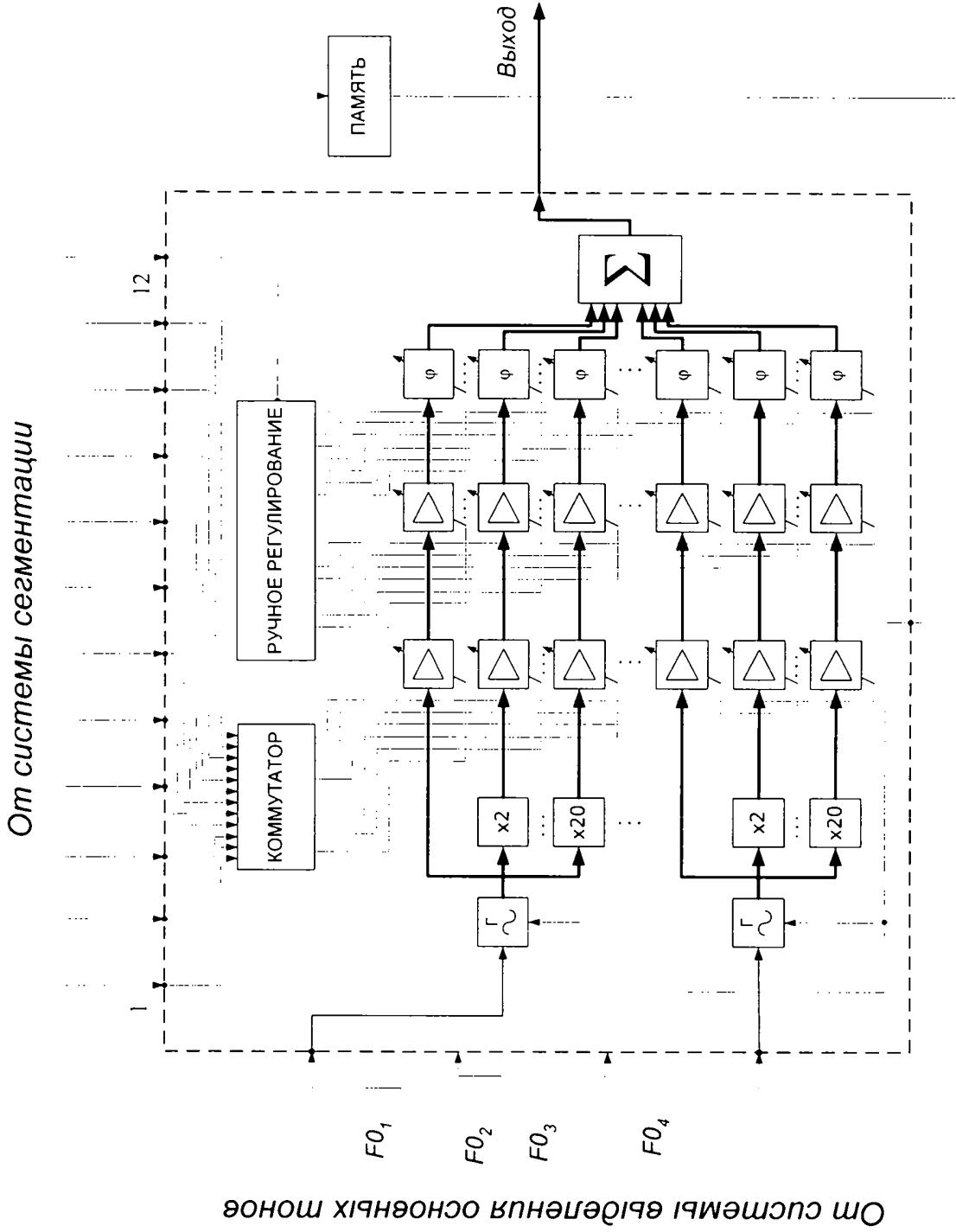


Рис.47 Схема синтеза обертонов на основе распознавания сигнала.

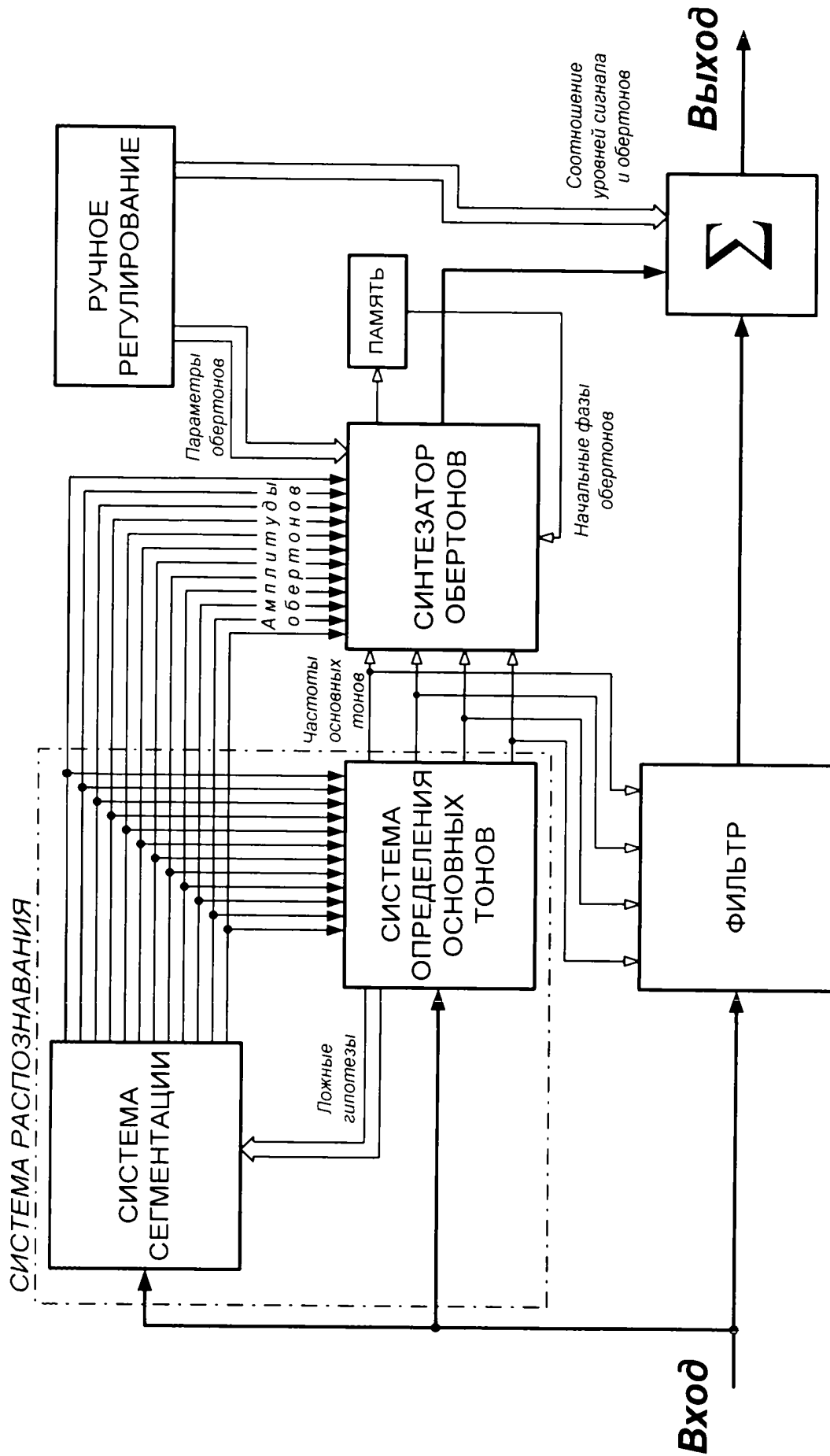


Рис. 48 Разработанная схема обработки сигнала, управляемой разработанной системой распознавания

На рис. 48 представлена разработанная схема системы обработки музыкального сигнала, управляемой системой распознавания. Отображены все взаимодействия между разработанными системами сегментации и определения основных тонов звуковых объектов, фильтром шумоподавления и синтезатором обертонов. Цепь прохождения обрабатываемого сигнала изображена тонкими стрелками, управляющие воздействия отображены широкими стрелками.

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.

Данная глава является наиболее важной главой диссертации, все результаты которой являются следствием результатов, полученных в предыдущих главах. По сути, итоги и результаты данной главы являются итогами и результатами всей работы.

8 . ЗАКЛЮЧЕНИЕ

Разработанные алгоритмы сегментации и определения основных тонов реализованы в программном обеспечении. С помощью этого программного обеспечения и получены примеры работы разработанных алгоритмов. Система обработки сигнала реализована лишь в части подавления шумов. Реализация синтезатора обертонов по разработанному алгоритму составляет предмет дальнейших исследований.

Значительным недостатком разработанной системы является её непригодность к детонации. При детонации происходит отклонение основ-

ных тонов звуковых объектов от того строя, на который настроена система сегментации. Происходят серьёзные ошибки при определении высот тонов, и, как следствие, снижается эффективность шумоподавления.

Проведённые субъективно-статистические экспертизы показали, что более половины слушателей предпочли звучания, обработанные с помощью разработанной системы фильтрации, звучаниям, обработанным с помощью программного средства восстановления сигналов DART, а также подключаемых модулей (Plug In) программы Steinberg WaveLab. Экспертам были предъявлены различные несложные (количество голосов не более двух-трёх) звучания, средний процент экспертов, отдавших предпочтение звучаниям, полученным с помощью разработанной системы составляет 65%. Таким образом, достигнута цель исследования — повышено качество (по субъективным критериям) восстановления сигналов.

Полностью выполнены и задачи исследования: разработана система распознавания музыкальных сигналов, способная автоматически управлять восстановлением сигнала. Разработан и алгоритм обработки музыкального сигнала на уровне структуры отдельного звукового объекта, полностью управляемый и контролируемый пользователем. Алгоритм фильтрации реализован в программном обеспечении.

Сформулируем **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. Сокращено время обучения неподготовленных или неопытных пользователей систем восстановления музыкальных сигналов за счёт интуитивно более понятных функций, выполняемых системой, например, регулировка амплитуд отдельных обертонов и мощности отдельных звуковых объектов вместо регуляторов АЧХ эквалайзера.

ЗАЩИЩАЕМЫЕ ПОЛОЖЕНИЯ:

1. Целью реставрации является повышение качества сигнала по субъективным критериям 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 обучения и самообучения без предоставления этой информации.
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. Естественность звучания сигналов, получаемых при воссоздании звуков музыкальных инструментов и вокализованных звуков речи аддитивным синтезом, достигается при наличии информации о каждом полном цикле колебаний.

Обсудим и ПРАКТИЧЕСКУЮ ЗНАЧИМОСТЬ полученных в работе результатов:

— 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, а также аранжировщиками и композиторами для сокращения рутинной работы по написанию партитур. Соответственно, практическую значимость представляет разработанный алгоритм распознавания музыкальных сигналов.

— 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.

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

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

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

В работе предложен новый подход к решению задачи восстановления сигнала по неполной или искажённой информации. Выявлены недостатки подхода, предложены пути устранения этих недостатков, а также расширения возможностей обработки музыкального сигнала и использованием распознавания образов.

9. СПИСОК ЛИТЕРАТУРЫ

1. **Алдошина И.** Основы психоакустики часть 14. Тембр часть 1 // Звукорежиссёр № 2, 2001.
2. **Алдошина И.** Основы психоакустики часть 14. Тембр часть 2 // Звукорежиссёр № 3, 2001.
3. **Баскаков С. И.** Радиотехнические цепи и сигналы: Учеб для вузов по спец. «Радиотехника» – 4-е изд., перераб. и доп. – М.: Высш. шк., 2003. – 462 с.
4. **Вахромеев В. А.** Элементарная теория музыки. – М.: Музыка, 1999. – 173 с.
5. **Волков А.Л.** Адаптивный алгоритм цифровой обработки звуковых сигналов для реставрации фонограмм: Автореферат диссертации на соискание ученой степени канд. техн. наук: 05.09.08. -СПб., 2000. -23 с.
6. **Гольденберг Л. М., Матюшкин Б. Д., Поляк М. Н.** Цифровая обработка сигналов: справочник. – М.: Радио и связь, 1985. – 312 с.
7. **Горон Е. И.** Радиовещание: Учебник для вузов. – М.: Связь, 1979. – 368 с.
8. **Дворянкин С.** Взаимосвязь цифры и графики, звука и изображения // Открытые системы №3, 2000.
9. **Иванченко Г. В.** Психология восприятия музыки: подходы, проблемы, перспективы. – М.: Смысл, 2001. – 264 с.
10. **Ковалгин Ю. А.** Стереофония. – М.: Радио и связь, 1989. – 272 с.
11. **Лейтес Р. Д., Соболев В. Н.** Цифровое моделирование систем синтетической телефонии. – М.: Связь, 1969. – 208 с.
12. **Маркел Дж. Д., Грэй А. Х.** Линейное предсказание речи, пер. с англ. / Под ред. Ю. Н. Прохорова и В. С. Звездина. – М.: Связь, 1980. – 308 с.
13. **Моль А.** Теория информации и эстетическое восприятие – М.: Мир, 1966. – 312 с.
14. **Мясников Л. Л., Мясникова Е. М.** Распознавание звуковых образов. – М.: Наука, 1984. – 158 с.
15. **Назаров М. В., Прохоров Ю. Н.** Методы цифровой обработки и передачи речевых сигналов. – М.: Радио и связь, 1985. – 148 с.
16. **Некоторые проблемы обнаружения сигнала, маскируемого флюктуационной помехой.** Сборник статей: Пер с англ. Под ред. Н. И. Шнер. – М.: Советское радио, 1965. – 263 с.
17. **Никамин В. А.** Форматы цифровой звукозаписи. – С.-Пб.: ЗАО «Элби», 1998. – 264 с.
18. **Павленко А.** Реставрация фонограмм с помощью программы DART Pro. // Мир ПК, №11, 1997.

19. **Потапова Р. К.** Речь: коммуникация, информация, кибернетика: Учеб. пособие для вузов. — М.: Радио и связь, 1997. — 528 с.
20. **Применение цифровой обработки сигналов** под ред. Э. Оппенгейма, пер. с англ. — М.: Мир, 1980. — 552 с.
21. **Рабинер Л., Гоулд Б.** Теория и применение цифровой обработки сигналов, пер. с англ. — М.: Мир, 1978. — 848 с.
22. **Рабинер Р., Шафер Р.** Цифровая обработка речевых сигналов, пер. с англ. — М.: Мир, 1981. — 496 с.
23. **Радиовещание и электроакустика: Учебное пособие для вузов / С. И. Алябьев, А. В. Выходец, Р. Гермер и др.; Под ред. Ю. А. Ковалгина.** — М.: Радио и связь, 1999. — 792 с.
24. **Римский-Корсаков А. В.** Электроакустика. — М.: Связь, 1973. — 272 с.
25. **Сапожков М. А.** Электроакустика. Учебник для вузов. — М.: Связь, 1978. — 272 с.
26. **Система цифровой реставрации и ремастеринга фонограмм «Канонь»**
<http://ru.ecomstation.ru/showarticle.php?id=68>
27. **Станции ТРЕК** от фирмы «Тракт» // 625, №7, 1998.
28. **Фомин Я. А., Тарловский Г. Р.** Статистическая теория распознавания образов. — М.: Радио и связь, 1986. — 264 с.
29. **Фу К.** Последовательные методы в распознавании образов и обучении машин, пер. с англ. — М.: Наука, 1971. — 256 с.
30. **Харкевич А. А.** Автоколебания. — М.: Государственное издательство технико-теоретической литературы, 1954. — 212 с.
31. **Харкевич А. А.** Спектры и анализ. — М.: Государственное издательство технико-теоретической литературы, 1962. — 236 с.
32. **Чернецкий М.** Психоакустические процессоры - что это такое? // Звуко-режиссёр, №8, 1999.
33. **Чулаки М.** Инструменты симфонического оркестра. — Л.: Союз советских композиторов СССР Ленинградское отделение Музфонда, 1950. — 94 с.
34. **Шелухин О. И., Лукьянцев Н. Ф.** Цифровая обработка и передача речи / Под ред. О. И. Шелухина. — М.: Радио и связь, 2000. — 456 с.
35. **Blauert J.** Spatial Hearing. MIT Press, 1983.
36. **Bregman A.** Auditory Scene Analysis. Bradford Books MIT Press, 1990.
37. **Cooke M.** Modelling auditory processing and organization. // PhD thesis, University of Sheffield Dept of Computer Science, 1991.
38. **Dolson M.** The phase vocoder : a tutorial. // Computer Music Journal №10(4), 1986.

39. **Dove Webster P.** Knowledge-based pitch detection. // PhD thesis, EECS dept, MIT, 1986.
40. **Duda R., Lyon R., Slaney M.** Correlograms and the separation of sounds. // Proc Asilomar Conf on Sigs, Sys & Computers, 1990.
41. **Ellis D.** Mid-level representations for Computational Auditory Scene Analysis. // Proceedings of the International Joint Conference on AI, Workshop on Computational Auditory Scene Analysis, Aug. 1995.
42. **Ellis. D.** Prediction-driven computational auditory scene analysis. // Ph.D. thesis, Massachusetts Institute of Technology, 1996.
43. **Forsberg J.** Automatic Conversion of Sound to the MIDI Forma // Computer Music Journal №12(10), 1999.
44. **Fujinaga I.** Machine recognition of Timbre Using Steady-State tone of Acoustic musical Instruments // International Computer Music Conference, 2003, P. 89-96.
45. **Fujinaga I., Moore S., Davad S. Sullivan Jr.** Implementation of Exemplar-based Learning Model for Music Cognition.
46. **Ghitza O.** Auditory nerve representation criteria for speech analysis/synthesis. // IEEE Tr ASSP 35(6), 1987.
47. **Glasberg B., Moore B.** Derivation of auditory filter shapes from notched-noise data. // Hearing Research №4(7), 1990.
48. **Goto M., Muraoka Y.** A Real-time Beat Tracking System for Audio Signals. // Proceedings of the 1995 International Computer Music Conference, pp.171-174, September 1995.
49. **Goto M., Muraoka Y.** Beat Tracking based on Multiple-agent Architecture - A Real-time Beat Tracking System for Audio Signals. // Proceedings of The Second International Conference on Multiagent Systems, pp.103-110, 1996.
50. **Greimas A. -J.** Semantique structural. Recherche de methode. Paris: Seuil, 1977.
51. **Gromko J.-E.** Perceptual differences between expert and novice music listeners: A multidimensional scaling analysis//Psychology of Music. 1993. Vol. 21 (1). P. 34 – 47.
52. **Gurney E.** The Power of Sound. London, 1880.
53. **Habermas J.** Identitat // Zur Rekonstruktion des Historischen Materialismus. Frankfurt, 1976.
54. **Halam S.** The development of memorisation strategies in musicians. Paper presented at the VII European Conference on Developmental Psychology, August. Krakow, 1995.

55. **Halpern A.R.** Organization in memory for familiar songs // *Journal of Experimental Psychology. Learning, Memory and Cognition*. 1984 a. Vol. 10 (3). P. 496–512.
56. **Halpern A.R.** Perception of structure in novel music // *Memory and Cognition*. 1984 b. Vol. 12(2). P. 163-170.
57. **Handel St.** *Listening*. // Bradford Books MIT Press, 1989.
58. **Handel St.** *Listening: An introduction to the perception of auditory events*. Cambridge (Ma): MIT Press, 1993.
59. **Hansen Ch.-H., Hansen R.-D.** Constructing personality and social reality through music: Individual differences among fans of punk and heavy metal music // *Journal of Broadcasting and Electronic Media*, Sum. 1991. Vol. 35 (3). P. 335–350.
60. **Hargreaves D.J.** Verbal and behavioral responses to familiar and unfamiliar music // *Current Psychology Research and Reviews*, Winter 1987–1988. Vol. 6 (4). P. 323-330.
61. **Hargreaves D.J., Colman A.M.** The dimensions of aesthetic reactions to music // *Psychology of Music*. 1981. Vol. 9 (1). P. 15–20.
62. **Hawley S.** *Structure out of Sound*. // Ph.D. thesis, Massachusetts Institute of Technology, 1998.
63. **Hebert S., Peretz I.** Recognition of music in long-term memory: Are melodic and temporal patterns equal partners? // *Memory & Cognition*, Jul. 1997. Vol. 25 (4). P. 518–533.
64. **Hester V.** *The Meaning of Poetic Metaphor*. London: The Hague, 1967.
65. **Hindemith P.** *A Composer's World. Horizons and Limitations*. Cambridge, 1953.
66. **Hirsch E.D.** *Cultural literacy*. Boston: Houghton Mifflin, 1987.
67. **Holleran S., Jones M.-R., Butler D.** Perceiving implied harmony: The influence of melodic and harmonic context // *Journal of Experimental Psychology. Learning, Memory, and Cognition*, May. 1995. Vol. 21(3). P. 737–753.
68. **Hoover D.-M., Cullari S.** Perception of loudness and musical preference: Comparison of musicians and nonmusicians // *Perceptual and Motor Skills*, Jun. 1992. Vol. 74(3). P. 1149-1150.
69. **Howell P., Cross I., West R.** (Eds.). *Musical Structure and Cognition*. London: Academic Press, 1985.
70. **Introduction** to Audio Restoration. <http://www.cedar-audio.com/>

71. **Kashino G., Tanaka T.** A sound source separation system with the ability of automatic tone modeling. // Proceedings of the International Computer Music Conference, 1993.
72. **Katayose Y., Inokuchi D.** The Kansei music system. // Computer Music Journal, 13(4), 1989.
73. **Klapuri A.** Automatic transcription of music. // Master of science's thesis. Tampere University of Technology, Department of Information Technology, 1998.
74. **Klapuri A.** Conventional and periodic N-grams in the transcription of drum sequences. // IEEE International Conferences on Multimedia and Expo, Baltimore, USA, 2003.
75. **Klapuri A.** Multipitch estimation and sound separation by the spectral smoothness principle. // IEEE International Conference on Acoustics, Speech and Signal Processing, Salt Lake City, USA, 2001.
76. **Klapuri A.** Number theoretical means of resolving a mixture of several harmonic sounds. // Proc. EUSIPCO, 1998.
77. **Klapuri A.** Pitch estimation using multiple independent time-frequency windows. // Proc. EUSIPCO 1999.
78. **Klapuri A.** Sound onset detection by applying psychoacoustic knowledge IEEE // International Conference on Acoustics, Speech and Signal Processing, ICASSP 1999.
79. **Klapuri A., Eronen, A.** Musical instrument recognition using cepstral coefficients and temporal features. // Proc. of the ICASSP. 2000.
80. **Klapuri A., Peltonen V., Eronen A., Parviainen A.** Recognition of everyday auditory scenes: Potential, latencies and cues. // Proc. of 110th AES Convention, May 2001.
81. **Klapuri A., Sillanp J., Seppnen J., Virtanen T.** Recognition of Acoustic Noise Mixtures by Combined Bottom-up and Top-down Processing. European Signal Processing Conference, Tampere, Finland 2000.
82. **Klapuri A., Virtanen T.** Separation of Harmonic Sounds Using Multipitch Analysis and Iterative Parameter Estimation. // Proc. of IEEE WASPAA, 2001.
83. **Maher R.** An approach to the separation of voices in composite musical signals. // PhD thesis, University of Illinois, Urbana-Champaign, 1989.
84. **Martin K. D.** A Blackboard System for Automatic Transcription of Simple Polyphonic Music // International Computer Music Conference, 1993, P. 100-112.

85. **Martin K. D.** Automatic Transcription of Simple Polyphonic Music // Computer Music Journal №1(7), 2002.
86. **McAdams S.** Audition: Cognitive Psychology of Music // Proc. of the ICASSP. 2001.
87. **McAdams S., Marin C.** Auditory Processing of Frequency Modulation Coherence/ J. Acous. Soc. Am. №86(12), 2001, P. 776-823.
88. **McAdams S., Marin C.** Segregation of Concurrent Sounds II: Effects of Spectral envelope Tracing, Frequency Modulation Coherence and Frequency Modulation Width // International Computer Music Conference, 2002, P. 145-168.
89. **McAdams S., Bigand E.** Introduction a la cognition auditive // International Computer Music Conference, 2001, P. 19-67.
90. **McAdams S., Bregman A.** Hearing Musical Streams. // Computer Music Journal №3 (4), 1979 : P. 26-43.
91. **McAdams S., Cunibile J.-C.** Perception Of Timbral Analogies // J. Acous. Soc. Am. №94(3), 2000, P. 918-954.
92. **McAdams S., Donnadieu S., Winsberg S.** Context Effects in Timbre Space // Proc. of the ICASSP. 2000.
93. **McAdams S.,** Spectral Fusion, Spectral parsing and the Formation of Auditory Images. // Ph.D. dissertation, Stanford University, CCRMA, 1984.
94. **McAdams S., Winsberg S.** A meta-analysis of timbre space. I: Multidimensional scaling of group data with common dimensions, specificities, and latent subject classes. // J. Acous. Soc. Am. №18(1), 1996, P. 332-342.
95. **McAdams. S.** Segregation of concurrent sounds. I: Effects of frequency modulation coherence. // J. Acoust. Soc. Am., Dec. 1989.
96. **McAulay R., Quatieri J.** Speech analysis/synthesis based on a sinusoidal representation. // IEEE Trans. on Acoustics, Speech and Signal Processing, 34(4), pp. 744-754, 1986.
97. **McAulay R., Quatieri J.** Speech analysis/resynthesis based on a sinusoidal representation. // IEEE Tr ASSP 34, 1986.
98. **Mellinger D.** Event formation and separation in musical sound. // PhD thesis, CCRMA, Stanford University, 1991.
99. **Moore B., Glasberg B., Baer T.** A model for the prediction of thresholds, loudness and partial loudness. // Journal AES, Vol. 45, No. 4, pp. 224-240. April 1997.
100. **Redden S.K.** Musik listening responses of groups differing in listening achievement // Psychomusicology. 1982. N° 1 – 2. P. 52 – 59.

101. **Saleh F.** A la découverte de l'ancienne gamme musicale égyptienne. – <http://egyptsound.free.fr/fathiVF.htm#>
102. **Scheirer Eric D.** Music-Listening Systems. // PhD Thesis, Massachusetts Institute of Technology, 2000.
103. **Sheldon D.A., Gregory D.** Perception of tempo modulation by listeners of different levels of educational experience // Journal of Research in Music Education, Fall 1997. Vol. 45 (3). P. 367–379.
104. **Shoen M.** The Psychology of MUsic. New York: Free Press, 1940.
105. **Siegel J.-A., Siegel W.** Categorical perception of tonal intervals: Musicians can't tell sharp from flat // Perception and Psychophysics, May. 1977. Vol. 21 (5). P. 399–407.
106. **Simonton O.K.** Thematic fame and melodic originality: A multivariate computer content analysis//Journal of Personality. 1980 a. Vol. 48. P. 206–219.
107. **Slaney B.** A critique of pure audition. // Joint International Conference on AI, CASA workshop, Aug. 1995.
108. **Slaney B.** An efficient implementation of the Patterson-Holdsworth auditory filter bank. // Apple Computer Technical Report 35, 1993.
109. **Sloboda J.A.** Parker D.H. Immediate recall of melodies // Musical Structure and Cognition. London: Academic press, 1985. P. 143–167.
110. **Sloboda J.A.** The music mind: a cognitive psychology of music. Oxford: Clarendon Press, 1985.
111. **Smith E., Serra A.** PARSHL: an analysis/synthesis program for non-harmonic sounds based on a sinusoidal representation. // International Computer Music Conference, 1987.
112. **Smith P.S., Cumow R.** Arousal Hypotheses and the Effects of Music on Purchasing Behavior//Journal of Applied Psychology. 1986. Vol. 50. № 3. P. 255–256.
113. **Smith, L. S.** Sound segmentation using onsets and offsets. // Interface Journal of New Music Research, №18(6), 2000.
114. **Soede, W., Berkhout, A. J., Bilsen, F. A.** Development of a directional hearing instrument based on array technology. // J. Acous. Soc. Am. 94(2), P. 785-798.
115. **SonicStudio 1.8 NoNoise Guide.**
ftp://ftp.sonicstudio.com/PDFs/NoNoise_Guide_1.8.pdf
116. **Stadler, R. W., Rabinowitz, W. M.** On the potential of fixed arrays for hearing aids. // J. Acous. Soc. Am. 94(3), P. 1332-1342.
117. **Stautner A.** The auditory transform. // MSc. thesis, Massachusetts Institute of Technology, 1995.

118. **Stautner, J. P.** Analysis and synthesis of music using the auditory transform. // S.M. thesis, Dept. of EECS, Massachusetts Institute of Technology.
119. **Steiger, H., Bregman, A. S.** Competition among auditory streaming, dichotic fusion and diotic fusion. *Perception & Psychophysics* 32, P. 153-162.
120. **Stubbs, R. J., Summerfield, Q.** Effects of signal-to-noise ratio, signal periodicity, and degree of hearing impairment on the performance of voice-separation algorithms. // *J. Acous. Soc. Am.* №89(3), 1998, P. 1383-1393.
121. **Stumpf C.** *Tonpsychologie*. 2 Bände. Leipzig, 1890.
122. **Suga, N.** Cortical computational maps for auditory imaging. // *Neural Networks* №3, 1997, P 3-21.
123. **Sundberg J.** *The Science of Singing Voice*. Decalb (111): North Illinois University Press, 1987.
124. **Taylor K., Greenhough L.** An object oriented ARTMAP system for classifying pitch. // *International Computer Music Conference*, 1993, pp. 244-247.
125. **Terhardt E.** Gestalt principles and music perception // *Auditory processing of complex sounds* / W.A.Yost, C.S.Watson (Eds.). Hillsdale (NJ): Lawrence Erlbaum Associates, Inc., 1987. P. 157 – 166.
126. **Thomassen R.** Melodic Accent: Experiments and a Tentative Model. // *J. Acoust. Soc. Am.*, №71(6), 1982.
127. **Tillmann B., Bigand E.** Influence of global structure on musical target detection and recognition // *International Journal of Psychology*, Apr. 1998. Vol. 33 (2). P. 107 – 122.
128. **Trehub S.E.** Auditory pattern perception in infancy // *Auditory development in infancy* / S.E.Trehub & BASchneider (Eds.). New York: Plenum Press, 1985. P. 185-195.
129. **Wendt C., Athina P. Petropulu.** Pitch Determination and Speech Segmentation Using the Discrete Wavelet Transform. // *Computer Music Journal* №5(8), 1996.
130. **Wessel D.** Timbre as a Musical Control Structure // *International Computer Music Conference*, 1998, P. 211-242.

10. СПИСОК ПУБЛИКАЦИЙ СОИСКАТЕЛЯ

1. Восстановление музыкальных сигналов с использованием распознавания образов // «Научноёмкие технологии» №3, 2003, том 4.
2. Реставрация без иллюзий // Broadcasting. Телевидение и радиовещание, №4, 2003.
3. Распознавание музыкальных сигналов // Депонирована в ЦНТИ «Информсвязь», №2177св. 2000 от 04.05.00, с. 108-121.
4. Подавление взаимного влияния сигналов, соответствующих разным группам инструментов, при многомикрофонной звукозаписи // Депонирована в ЦНТИ "Информсвязь", № 2200св. 2001 от 22.05.01, с. 2-9.
5. Исследование структуры звуковых объектов // Депонирована в ЦНТИ «Информсвязь», № 2208св. 2002 от 10.06.02, с. 95-117.
6. Распознавание музыкальных образов// 55-ая студенческая НТК МТУСИ. – Москва, 2000. – Тезисы докладов. с.65-67.
7. Автоматическая сегментация речевых и музыкальных сигналов// 55-ая студенческая НТК МТУСИ. – Москва, 2000. – Тезисы докладов. с.67-69
8. Основной тон неречевых сигналов // 10-ая Межрегиональная конференция МНТОРЭС им. А.С.Попова «Обработка сигналов в системах двусторонней телефонной связи». М. – : МТУСИ, 2000. Тезисы докладов, с.127-128.
9. От вокодера к музкодеру // 4-я международная НТК «Перспективные технологии в средствах передачи информации». – Владимир, 2001. Тезисы докладов, стр. 112-113.
10. Обработка звукового сигнала на основе распознавания// 4-я международная НТК «Перспективные технологии в средствах передачи информации». – Владимир, 2001. Тезисы докладов, стр. 113-115.
11. Узнавание тембра: что важно и что – нет? // НТК ППС и ИТС М.: МТУСИ, 2002. - Тезисы докладов, с. 213-214.
12. Классификация музыкальных инструментов // НТК ППС и ИТС М.: МТУСИ, 2002. - Тезисы докладов, с. 214-215.
13. Аддитивный синтез вокализованных звуков речи // НТК ППС и ИТС М.: МТУСИ, 2003. - Тезисы докладов, с. 196-197.

ПРИЛОЖЕНИЕ 1. ПРОТОКОЛ ИСПЫТАНИЙ ПО ОЦЕНКЕ
ЭФФЕКТИВНОСТИ РАЗРАБОТАННОГО АЛГОРИТМА ПОДАВЛЕНИЯ ШУМОВ

ПРОТОКОЛ № 1

ПРОСЛУШИВАНИЯ (СУБЪЕКТИВНО-СТАТИСТИЧЕСКИХ ИСПЫТАНИЙ)
МУЗЫКАЛЬНЫХ СИГНАЛОВ, ОБРАБОТАННЫХ ПРОГРАММНОЙ СИСТЕМОЙ
ШУМОПОДАВЛЕНИЯ

• 1. Цель прослушивания.

Прослушивание музыкальных сигналов проводилось с целью определения бальной оценки качества звучания искажённых сигналов после их обработки с целью восстановления и подавления шумов.

• 2. Место и время прослушивания.

Прослушивания проводились в марте 2003 г. в помещении прослушивания № 5 лаборатории Радиовещания и электроакустики МТУСИ.

• 3. Звуковой материал.

В качестве звукового материала использовались записи полифонических музыкальных произведений, полученные с помощью программного синтезатора Gigasampler 2.5.

Звуковой материал был сформирован следующим образом: синтезированные музыкальные сигналы с выхода звуковой карты компьютера записывались на компакт-кассету с помощью бытового магнитофона PHILIPS AW7550, затем магнитные записи воспроизводились и записывались в цифровом виде на жёсткий диск компьютера с помощью звуковой карты. Использованный бытовой магнитофон не оснащён системой шумоподавления, с его помощью моделировались искажения, вносимые трактами записи и воспроизведения устройств аналоговой магнитной записи. Далее зашумлённые музыкальные сигналы обрабатывались с целью подавления шумов программой Dart PRO и с помощью разработанного программного обеспечения.

Все фонограммы были записаны в цифровой форме с частотой дискретизации 44100 Гц, при 16-ти разрядной линейной ИКМ. В качестве носителей использовались CD-R диски.

• **4. Состав оборудования:**

1. Проигрыватель CD-дисков Technics № VT 7LA06512;

2. Стойка усилителей № 126;

3. Акустическая система HZK-12061 №27980, №27981;

4. Комната прослушивания № 5 с практически равномерным временем реверберации $T_p = 0,4$ с. Основные параметры комнаты прослушивания (время реверберации и степень однородности структуры звукового поля) соответствуют рекомендациям ОИРТ №86 от 1977 г. и AES-20-96 от 1996 г.

• **5. Методика прослушивания:**

1. Оценка качества звучания проводилась с помощью субъективно-статистических экспертиз методом сравнительного прослушивания музыкального сигнала в соответствии с документом МЭК 286-5 от 1985 г. Использовалась балльная оценка (шкала пятибалльная).

2. Объекты прослушивания. Каждая предъявляемая экспертам пара звучаний содержала исходный сигнал и сигнал, обработанный программным средством шумоподавления.

3. Порядок прослушивания. При проведении прослушивания использовались цифровые записи, что предлагается последними рекомендациями, в силу того, что они сохраняют свои характеристики во времени и могут точно копироваться. В качестве источника сигнала использовался проигрыватель компакт-дисков Technics, качественные характеристики которого соответствуют требованиям ГОСТ 12107-74.

Перед проведением прослушивания воспроизводящий тракт подстраивался для соблюдения баланса и уровня громкости, а обработанные программы выравнивались с исходными по средней мощности.

В роли экспертов выступали студенты, не имеющие патологий слуха. Эксперты в количестве 19-21 человек располагались внутри площади прослушивания (в соответствии с рекомендациями AES-20-96), границы которой находились на расстоянии 1,8 м от фронтальных поверхностей акустических систем, 1,5 м от боковых стен и 2,5 м от задней стены помещения прослушивания. Ось симметрии площади прослушивания совпадала с направлением перпендикуляра, восстановленного из середины линии, соединяющей акустические системы.

Согласно рекомендации МЭК 268-5, рабочая ось АС была направлена горизонтально, расположена на высоте 1,25 м от пола и сориентирована на точку, находящуюся в центре зоны размещения экспертов. Величина базы акустических систем равнялась трем метрам. Размещение экспертов внутри площади прослушивания проводилось на расстоянии 0,7 м друг от друга амфитеатром.

Для проверки степени достоверности данных, получаемых от отдельно взятого эксперта, был использован метод «самосовпадений». В процессе проведения прослушивания периодически вводились проверочные пары, состоящие из одинаковых исходных фрагментов. О времени появления в процессе ССИ этих пар и степени их регулярности эксперты ничего не знали. При проведении обработки результатов ССИ и формировании обобщенных оценок заметности искажений, «вклад» каждого эксперта производился с учетом его весового коэффициента, найденного согласно «правильности ответов» на проверочные пары.

- **6. Методика математической обработки результатов экспертиз.**

1. Для обработки результатов массовых экспертиз естественно применение математического аппарата статистики, который позволяет получить на основе результатов многих экспериментов с большим числом экспертов некоторые общие усреднённые характеристики, не зависящие от случайных ошибок. Кроме того, математическая статистика позволяет решить вопрос о необходимом числе экспертов для обеспечения заданной точности и определения точности уже полученных результатов.

2. Закон распределения. Для применения аппарата математической статистики к результатам экспертиз необходимо выяснить статистический смысл полученных данных.

Искомой величиной является бальная оценка звучания некоторого сигнала, которая определяется типовыми свойствами человеческого уха и, следовательно, должна иметь какую-то определённую величину. Однако, на показания экспертов будут оказывать влияние и всякие побочные факторы: тренировка в прослушивании искажений, музыкальный слух, психическое состояние и т. д. Вследствие этих факторов показания экспертов не постоянны, а колеблются вокруг какой-то средней величины \bar{p} , которая при очень большом числе экспертов уже не зависит от разнообразных побочных причин, а определяется только свойствами слуха. Поэтому среднюю величину \bar{p} можно считать бальной оценкой качества данного сигнала. Величина бальной оценки p_i , которую покажет i -тый человек, в общем случае не совпадает с \bar{p} . Отклонение p_i от \bar{p} зависит от того, каковы количество, состав, психическое состояние экспертов.

При проведении измерений эксперты подбирались из разных групп без глубокого предварительного их обследования. Состав экспертов был случайным, а следовательно, и результат p_i и оценка \tilde{p}_i , также есть случайная величина. Математическое ожидание этой случайной величины есть средняя бальная оценка качества сигнала.

Поскольку величина \tilde{p}_i является случайной величиной, и в том случае, когда число экспертов в группе n достаточно велико, закон распределения величин \tilde{p}_i , можно считать нормальным.

При нормальном законе распределения вероятность того, что оценка заметности будет равна \tilde{p}_i , определяется по формуле:

$$f(\tilde{p}_i) = \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-\frac{1}{2} \left(\frac{\tilde{p}_i - a}{\sigma} \right)^2} \quad (\text{П-1})$$

здесь \tilde{p}_i – оценка заметности искажения,

n – число экспертов,

a – математическое ожидание случайной величины,

σ – среднеквадратическое отклонение случайной величины,

$$\frac{\tilde{p}_i - a}{\sigma} = t \quad - \text{нормированное отклонение случайной величины.} \quad (\text{П-2})$$

Функция распределения величины нормированного отклонения при $n \rightarrow \infty$ имеет вид

$$P(t < x) = F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt \quad (\text{П-3})$$

3. Проверка функции распределения. После проведения экспертиз необходимо, прежде всего, проверить функцию распределения, чтобы убедиться в правильности сделанных предположений о функции распределения, а, следовательно, и в правильности полученных результатов. Это нужно также и для того, чтобы, пользуясь найденным законом распределения, обработать полученные данные и найти доверительные интервалы, которые определяют точность результатов. Проверка закона распределения производится по критерию согласия Колмогорова. С этой целью по полученным данным для какого-либо искажения необходимо построить эмпирическую функцию распределения $F_N(x)$. Эмпирическая функция распределения строится следующим образом: группа экспертов на одном отрывке даст определённую величину оценки p . Если провести с этой группой N экспериментов, то получим N значений заметности: $p_1, p_2, \dots, p_i, \dots, p_N$ и N значений оценок заметности $\tilde{p}_1, \tilde{p}_2, \dots, \tilde{p}_i, \dots, \tilde{p}_N$.

Расположим их в порядке возрастания в ряд (вариационный ряд). При повторении эксперимента N раз не обязательно значение \tilde{p}_i , будет принимать каждый раз новую величину. Некоторые величины могут повторяться по несколько раз. Число повторений какой-либо величины \tilde{p} называется её частотой m , а отношение частоты m к общему числу экспериментов $m/n = \varpi$ частотой этой величины. Таким образом, если проведено N экспериментов, то в ре-

зультате получится M различных оценок, причём $M \leq N$. Обозначим эти значения через x_i . Между значениями существует соотношение

$$x_1 < x_2 < x_3 < \dots < x_i < \dots < x_m.$$

Каждому значению соответствует определённая частота m_i и частость $\varpi_i = m_i / n$.

Сумма частот $\sum_{i=1}^n m_i = N$, а сумма частостей $\sum_{i=1}^n \varpi_i = 1$. Теперь составим функцию накопленных частостей $F_N(x)$. Эта функция строится таким образом: значению x_1 соответствует частость $\varpi_1 = m_1/n$, значению x_2 соответствует частость $\varpi_2 = m_2/N$ и накопленная частость $F_N(x_2) = \varpi_1 + \varpi_2$. Значению заметности x_3 соответствует частость $\varpi_3 = m_3/N$ и накопленная частость $F_N(x_3) = \varpi_1 + \varpi_2 + \varpi_3$ и т. д. Значению заметности x_m соответствует накопленная частость $F_N(x_m) = \varpi_1 + \varpi_2 + \varpi_3 + \dots + \varpi_m = 1$.

Функция накопленных частостей $F_N(x)$ и есть эмпирическая функция распределения случайной величины \tilde{p}_i . После того как построена эмпирическая функция распределения, можно определить среднее значение случайной величины \bar{p} и её среднеквадратичное отклонение.

Среднее значение определяется по формуле

$$\bar{p} = \frac{\sum_{i=1}^N \tilde{p}_i}{N} = \frac{\sum_{i=1}^M x_i m_i}{N} \quad (\text{П-4})$$

Среднеквадратичное отклонение случайной величины определяется по формуле

$$\sigma^2 = \frac{\sum_{i=1}^N (\tilde{p}_i - \bar{p})^2}{N} = \frac{\sum_{i=1}^M (x_i - \tilde{p}_i)^2 m_i}{N} \quad (\text{П-6})$$

Иногда эту величину удобнее вычислять по другой формуле:

$$\sigma^2 = \frac{\sum_{i=1}^M x_i^2 m_i}{N} - \left(\frac{\sum_{i=1}^M x_i m_i}{N} \right)^2 \quad (\text{П-7})$$

Теперь можно приступить к проверке функции распределения по критерию Колмогорова.

Критерий согласия Колмогорова гласит, что если число экспериментов N достаточно велико (не менее нескольких десятков) и теоретическая функция распределения $F(x)$ (с которой сравнивается полученная эмпирическая) непрерывна, то вероятность того, что наибольшее отклонение $F_N(x)$ от $F(x)$ превысит заданное число $\frac{\lambda}{\sqrt{N}}$, может быть приближённо вычислена по формуле:

$$P(\lambda) = P\left\{ \max[F_N(x) - F(x)] \geq \frac{\lambda}{\sqrt{N}} \right\} = 1 - \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2 \lambda^2} \quad (\text{П-8})$$

Значения функции $1 - \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2 \lambda^2}$ табулируются.

Сравнивая значения накопленных частостей $F_N(x)$ с расчётными значениями функции $F(x)$, находим наибольшее отклонение $D = \max[F_N(x) - F(x)]$. Умножая это наибольшее значение на $\lambda = \sqrt{N}$, определяем величину $\lambda = D\sqrt{N}$. По этой λ находим в таблице величину $\kappa(\lambda)$ и затем вероятность

$$P(\lambda) = 1 - \kappa(\lambda).$$

Если окажется, что вероятность $P(\lambda)$ наблюдаемого расхождения между $F_N(x)$ и $F(x)$ очень мала (менее 0,01), то по принципу практической невозможности маловероятных событий заключаем, что предполагаемое теоретическое распределение не соответствует исследуемому эмпирическому. Если же $P(\lambda)$ имеет достаточно большую величину, то это значит, что наблюдаемое значение D вполне возможно и, следовательно, можно считать, что эмпирическое распределение соответствует теоретическому.

В нашем случае сравнивается эмпирическое распределение с нормальным

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{t^2}{2}} dt \quad (\text{П-9})$$

При этом считаем, что теоретическое распределение имеет математическое ожидание, равное среднему значению $a = \bar{p}$ и дисперсию σ^2 . Значение \bar{p} и σ^2 найдено из данных экспериментов

Таким образом, проверка закона распределения проводится следующим образом: по результатам экспериментов строим эмпирическую функцию распределения $F_N(x)$ и определяем среднее значение \bar{p}_2 и дисперсию σ^2 . Затем строим по этим значениям $\tilde{p} = \bar{p}_2$ и σ^2 теоретическую функцию распределения $F(x)$. Значения $F(x)$ находим из формулы

$$F(x) = \frac{1}{2} + \Phi\left(\frac{x_i - \tilde{p}}{\sigma}\right) \quad (\text{П-10})$$

Для каждого из x_i находим разность

$$D_i = [F_N(x_i) - F(x_i)]$$

и берём максимальную из них $D_{\text{макс}} = \text{макс}[F_N(x_i) - F(x_i)]$.

После этого находим $\lambda = D\sqrt{N}$ и по λ определяем вероятность $P(\lambda) = 1 - K(\lambda)$.

Если $P(\lambda)$ имеет достаточно большую величину, можно считать, что эмпирическое распределение совпадает с теоретическим. Если $P(\lambda)$ мала (меньше 0,01), то эмпирическое распределение не совпадает с теоретическим.

Величина \tilde{p} определяется из формулы

$$\tilde{p} = \bar{p} = \sum_{i=1}^M \frac{x_i m_i}{N} \quad (\text{П-11})$$

Величина σ — из формулы

$$\sigma = \sqrt{\sum_{i=1}^M \frac{x_i^2 m_i}{N} - \tilde{p}^2} \quad (\text{П-12})$$

4. Определение точности результатов. Так как проведенная проверка показала, что закон распределения оценок \tilde{p}_i нормальный, можно определить точность полученных результатов. Точность результата определяется разбросом случайной величины вокруг её истинного значения. При расчёте необходимого значения n мы задавались допустимым разбросом и вероятностью, с которой случайная величина будет попадать в заданный интервал при повторении опытов. Имея данные о заметности искажения, и зная, сколько экспертов участвовало в эксперименте по определению бальной оценки того или иного сигнала, можно построить интервал, в который с заданной вероятностью будут попадать значения оценок, если бы мы стали повторять эксперимент снова. Таким образом, этот интервал определит с заданной достоверностью точность полученных результатов (такой интервал называется доверительным).

Определение доверительного интервала производится следующим образом. По полученной бальной оценке для данного сигнала определяем величину S_x , которая в теоретическом распределении соответствует дисперсии оценки

$$S_x = \frac{S}{\sqrt{n}} \quad , \text{ где} \quad (\text{П-13})$$

$$S^2 = \frac{\tilde{p}_i(1 - \tilde{p}_i)n}{n - 1} \quad (\text{П-14})$$

Здесь через \tilde{p}_i обозначена относительная бальная оценка качества,

n — число экспертов.

Так как $n \gg 1$, то можно считать $n/(n-1) \approx 1$ и тогда $S = S_x \approx \sqrt{\tilde{p}_i(1 - \tilde{p}_i)}$.

Нормированное отклонение оценки величины заметности будет определяться как:

$$t = \frac{\tilde{p}_i - \tilde{p}}{S_x^-} \quad (\text{П-15})$$

Здесь \tilde{p} есть истинное значение оценки для данного сигнала (искомое нами). Вероятность того, что это значение нормированного отклонения лежит в интервале $t_1 < t < t_2$, равна

$$p(t_1 < t < t_2) = \int_{t_1}^{t_2} S(t, k) dt \quad (\text{П-16})$$

Здесь $S(t, k)$ означает функцию от величин t и $k = n - 1$; $S(t, k)$ — функция распределения величины t .

Зададим определённую вероятность α и найдём такой интервал $-t_\alpha < t < t_\alpha$, такой, что вероятность того, что t будет лежать в этом интервале, равна α

$$\overline{p}(-t_\alpha < t < t_\alpha) = \int_{-t_\alpha}^{t_\alpha} S(t, k) dt = 2 \int_0^{t_\alpha} S(t, k) dt = \alpha \quad (\text{П-17})$$

Таблица значений t_α , определяемых этим равенством, приводится в литературе. Зададимся величиной α и по таблице найдём соответствующее t . Это значит, что с достоверностью α нормированное отклонение оценки лежит в интервале: $-t_\alpha < t < t_\alpha$, или

$$-t_\alpha < \frac{\tilde{p}_i - \tilde{p}}{S_x^-} < t_\alpha \quad (\text{П-18})$$

Отсюда следует, что с заданной достоверностью α .

$$-t_\alpha S_x^- < \tilde{p}_i - \tilde{p} < t_\alpha S_x^- \quad \text{или}$$

$$\tilde{p}_i - t_\alpha S_x^- < \tilde{p} < \tilde{p}_i + t_\alpha S_x^-$$

истинная оценка величины оценки качества звучания сигнала по бальной шкале с достоверностью α лежит в доверительном интервале $(\tilde{p}_i - \varepsilon) < \tilde{p} < (\tilde{p}_i + \varepsilon)$,

где $\varepsilon = t_{\alpha} S_{\bar{x}}$ называется точностью приближённого равенства $\tilde{p} = \tilde{p}_i$.

Вычисления доверительных интервалов проводится с достоверностью $\alpha=0,95$. Результаты вычислений сведены в табл. П-1.

Таблица П-1

Фрагмент звучания		Средняя оценка по бальной шкале	Число экспертопока- заний	Исправленное сред- нее квадратическое отклонение	Дисперсия	Аргумент функции Стьюдента	Точность
№	Обра- ботка	\tilde{p}_i	n	$S = \sqrt{\tilde{p}_i(1 - \tilde{p}_i)}$	$S_{\bar{x}} = \frac{S}{\sqrt{n}}$	$t_{\alpha} = \varphi(\alpha, n)$	$\varepsilon = t_{\alpha} S_{\bar{x}}$
1	DART	3,02	145	2,55	0,21	1,92	0,4158
	Prog	3,82	145	2,36	0,17	1,92	0,3366
2	DART	3,54	145	2,42	0,15	1,92	0,297
	Prog	4,18	145	2,48	0,15	1,92	0,297
3	DART	3,33	164	2,76	0,24	1,94	0,4752
	Prog	4,01	164	2,53	0,29	1,94	0,5742

В таблице П-1 фрагменты звучания обозначены следующим образом:

№1 – «Как пойду...», русская народная песня, количество голосов – два, синтезирована с помощью звуков флейты и кларнета, длительность 23 с.

№2 – «Канон», неизвестный автор XVII века, количество голосов – три, синтезирована с помощью звуков органа, длительность 18 с.

№3 – «Эль Чокло», аргентинское танго, автор – А. Вильельдо, количество голосов – три (мелодия + аккомпанемент), синтезирована с помощью звуков гитары с нейлоновыми струнами, длительность 25 с.

В колонке «Обработка»: DART Pro – программный продукт компании Dartech Inc., используемый при реставрации звукозаписей для подавления различных шумов и помех, Prog – условное название программного обеспечения, разработанного в диссертационной работе.

Таблица П-2

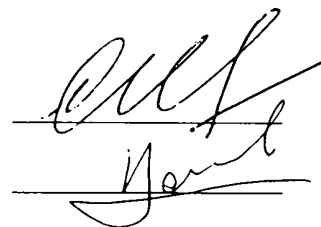
Фрагмент звучания		Процент экспертов, поставивших более высокую оценку
№	Обработка	P
1	DART	32
	Prog	68
2	DART	41
	Prog	59
3	DART	31
	Prog	69

Из таблицы (П-1) видно, что усреднённые по всем экспертопоказаниям оценки звучаний, полученных с помощью разработанного программного обеспечения выше, чем оценки звучаний, полученных с помощью одного из широко применяемых сегодня средств шумоподавления. Кроме того, из таблицы П-2 видно, что в среднем 65% экспертов предпочли (т.е. поставили более высокую оценку) звучания, обработанного с помощью предложенной системы подавления шумов.

При измерениях присутствовали:

профессор кафедры РВ и ЭА, д.т.н. С. Л. Мишенков

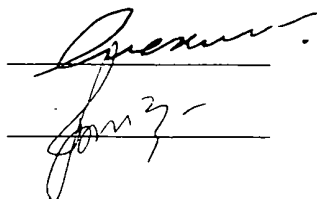
доцент кафедры РВ и ЭА кафедры, к.т.н. О.Б. Попов



Измерения проводили аспиранты кафедры РВ и ЭА:

А.А. Кудинов

Е.А. Хрянин





ОБЩЕСТВО С ОГРАНИЧЕННОЙ ОТВЕТСТВЕННОСТЬЮ

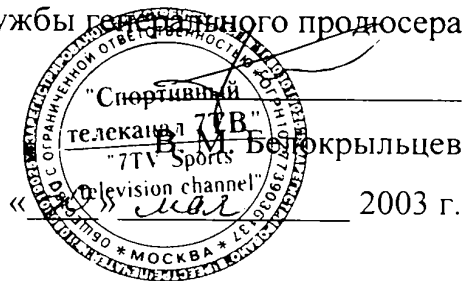
"СПОРТИВНЫЙ ТЕЛЕКАНАЛ 7ТВ"

127427, Москва, ул. Академика Королёва, 19, офис 3-38, тел 217-9940, 217-8549

на № _____ № _____
от _____

УТВЕРЖДАЮ.

Директор производственной дирекции
службы генерального продюсера




А К Т

о применении результатов диссертационной работы

Настоящий акт составлен в том, что разработанное А. А. Кудиновым программное обеспечение, реализующее алгоритм обработки музыкальных сигналов, предложенный в его диссертационной работе, используется в аппаратной озвучивания ООО «Спортивный телеканал 7 ТВ» при подготовке музыкальных заставок программ и спортивных трансляций.

Акт составлен для представления в диссертационный совет К 219.001.02 при Московском техническом университете связи и информатики.

Главный звукорежиссёр

ООО «Спортивный телеканал 7 ТВ»  Заборин А. И.