

Восстановление музыкальных сигналов с использованием распознавания образов

УДК 681.84.087

Кудинов А. А.

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, нередко являются толчком к развитию методов цифровой обработки сигналов, а иногда и поводом к вопросу о целесообразности реставрации фонограмм.

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. Задачу реставрации фонограмм можно сформулировать следующим образом:

1. устранение «лишних» спектральных составляющих, т.е. компонентов спектра, не содержавшихся 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 записи-воспроизведения, 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. В данной статье не обсуждается усовершенствование применяющихся методов восстановления звуковых сигналов, но предлагается новый подход к решению этой задачи.

Реставрация — понятие универсальное, используемое различными отраслями 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. Например, если на листе бумаги изобразить окружность, а затем разорвать этот лист на части так, чтобы каждая из частей содержала фрагмент окружности, то в зависимости от длины дуги, содержащейся на той или иной части листа, мы сможем с большей или меньшей уверенностью утверждать, что эта дуга является фрагментом изображения окружности. Можно рассуждать иначе: когда нам дают клочок бумаги с изображением дуги и просят сказать, фрагмент какого изображения мы видим, нас просят решить задачу реставрации. Чтобы решить эту 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. 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.

Зависимость достоверности распознавания от количества информации иллюстрируется рисунком 1. Априорная информация состоит в том, что объекты, подлежащие восстановлению — фрагменты симметричного изображения, количество информации, содержащейся в восстанавливаемых фрагментах различно.

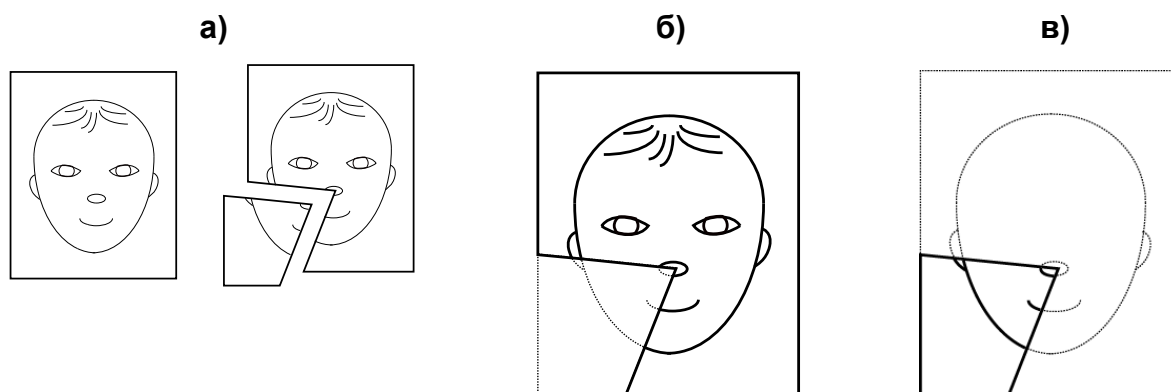


Рисунок 1. Пример восстановления образа при наличии априорной информации: а) исходный образ и фрагменты, подлежащие восстановлению; б) восстановление с большой достоверностью; в) восстановление с малой достоверностью.

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. Это доводы умозрительные. Теоретически обоснованный довод состоит в следующем: из теории передачи сигналов известно, что оптимальным (т.е. обеспечивающим наибольшее выходное отношение сигнал/шум) является фильтр, имеющий частотную характеристику, совпадающую со спектром сигнала. Достоверное распознавание образов позволяет разделить спектр сигнала и спектр шумов за исключением перекрывающихся областей. Значит, если мы с помощью распознавания образов выделим спектр сигнала из спектра суммы сигнала и шумов, то сможем приблизиться к оптимальной фильтрации ровно настолько, насколько точно проведём распознавание.

Рассмотрим возможную последовательность действий при восстановлении музыкального сигнала. Отметим, что при восстановлении музыкального сигнала кроме той информации, которую можно получить, проанализировав сам сигнал, мы обладаем и априорной информацией: не учитывая незначительные отклонения, основные тона звуков составляют дискретное множество, зная один из членов которого, можно восстановить все остальные. Например, для равномерного строя:

$f_i^{OT} = f_{\text{э}} \cdot \sqrt[12]{2^i}$, где $f_{\text{э}}$ — некая эталонная частота, для натурального строя $\frac{f_i^{OT}}{f_{\text{э}}} = \frac{m}{n}$,

где $m, n \in \mathbb{Z}, m, n \geq 1$, для африканской или индийской музыки закономерности иные. Кроме того, выраженность высоты музыкального звука означает определённое (конкретно — целочисленное, в приближении, допустимом в задаче распознавания) соотношение частот основного тона и обертонов, иными словами, квазипериодичность сигнала означает дискретность спектра (и наоборот). Отметим также, что далее обсуждается цифровая обработка сигнала. Алгоритмы цифровой обработки, основанные на анализе сигнала, работают с короткими выборками сигнала.

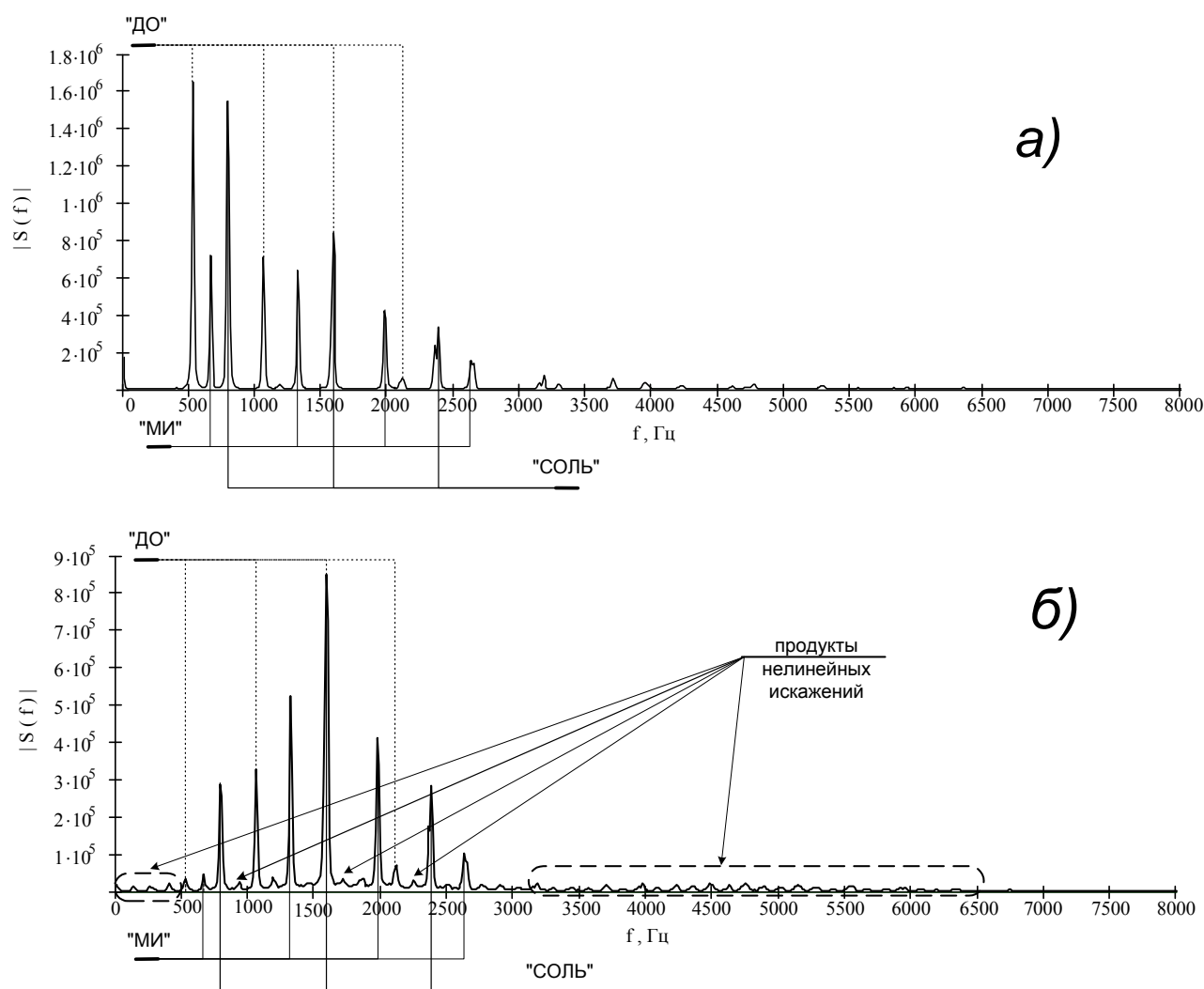


Рисунок 2. Амплитудные спектры фрагментов музыкального сигнала: а) исходного, б) искажённого, подлежащего восстановлению

Рассмотрим последовательность восстановления музыкального сигнала на примере выборки, содержащей три звуковых объекта, соответствующие нотам «до», «ми» и «соль» второй октавы, исполненных флейтой. На рисунке 2 приведены ам-

плитудные спектры исходного и искажённого фрагментов. Искажённый фрагмент получен из исходного моделированием. 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.

Не будем касаться обсуждения автоматического распознавания многоголосных музыкальных сигналов. Отметим лишь, что в случае, приведённом на рисунке 2 б) достаточно информации для достоверного распознавания, т.е. для принятия решения о том, что на данной выборке сигнала существуют три звуковых объекта (перекрываются лишь третья гармоника «до» со второй гармоникой «соль»), и определения высот тонов (двух-трёх гармоник для этого достаточно). Какую информацию мы получаем, проводя распознавание? Количество существующих на анализируемом сегменте звуковых объектов (нот) и частоты их основных тонов. Этой информации достаточно для проведения фильтрации, поскольку локализованы спектральные компоненты полезного сигнала. Для восстановления структуры звуковых объектов (т.е. соотношения амплитуд и фаз обертонов) этой информации недостаточно, поскольку исходное соотношение амплитуд обертонов утрачено безвозвратно (вероятность угадать его, очевидно, ничтожна). Поэтому для восстановления спектральной структуры сигнала придётся пользоваться какими-то эвристическими правилами. Об этом ниже.

Первым этапом восстановления сигнала является фильтрация, поскольку проще сначала подавить шум, а затем провести синтез спектральных составляющих, чем наоборот. АЧХ фильтра сформируем 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. Отметим, что ширина полосы частот Δf ,

содержащей обертоны полезного сигнала, может и должна быть регулируемой. В остальной области частот коэффициент передачи должен быть меньше единицы: $|\dot{H}(f)| = H_{\text{под}} < 1$ для подавления шумов (см. рис. 3). Соответственно, при таком подходе считается, что область частот, не занятая полезным сигналом, занята шумом.

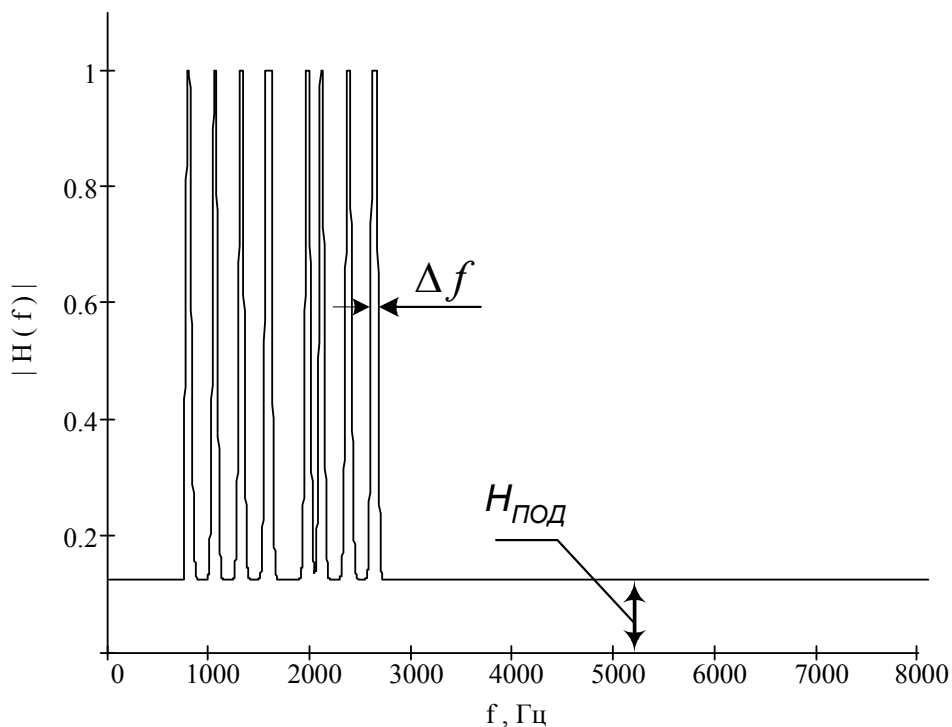


Рисунок 3. АЧХ фильтра для подавления шума выборки, соответствующей рисунку 2 б)

Сразу можно отметить недостаток такого подхода: музыкальная фонограмма всегда содержит, кроме шумов носителя записи и шумов тракта формирования-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 $H_{\text{под}} = 0$ фильтрация губительно скажется на впечатлении от прослушивания реставрированной фонограммы. Пока не разработаны алгоритмы распознавания «полезных» шумов, разумно сделать $H_{\text{под}}$ регулируемым вручную или автоматически, например, исходя из условия поддержания заданного минимального соотношения сигнал/шум. На рисунке 4 приведён спектр восстанавливаемого фрагмента до и после прохождения через фильтр с АЧХ, изображённой на рисунке 3.

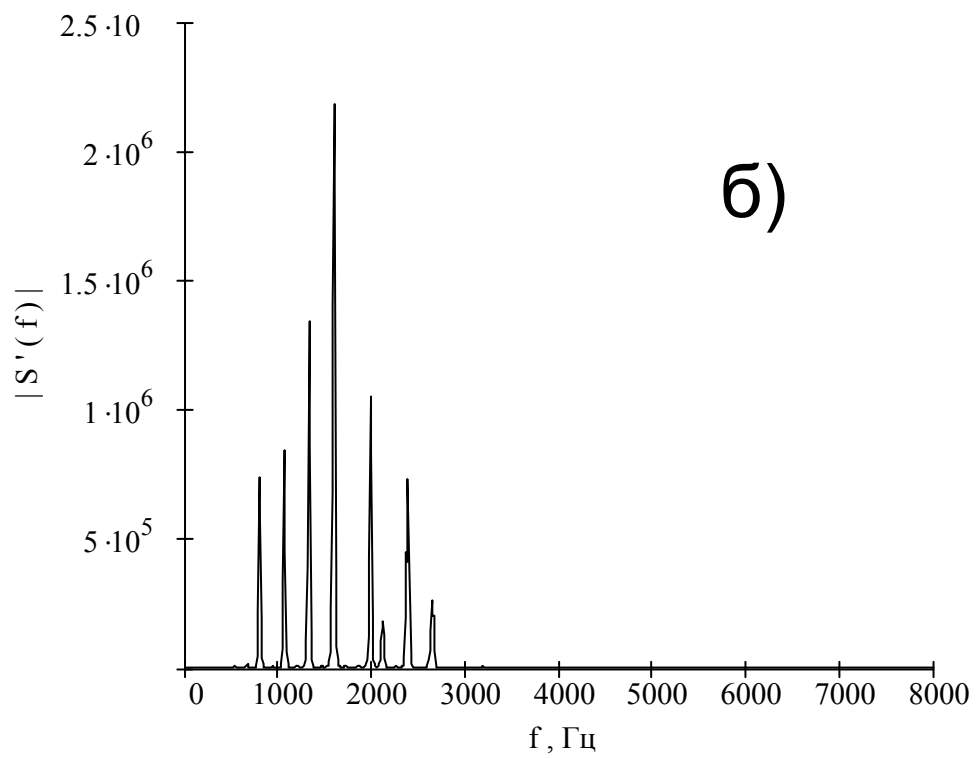
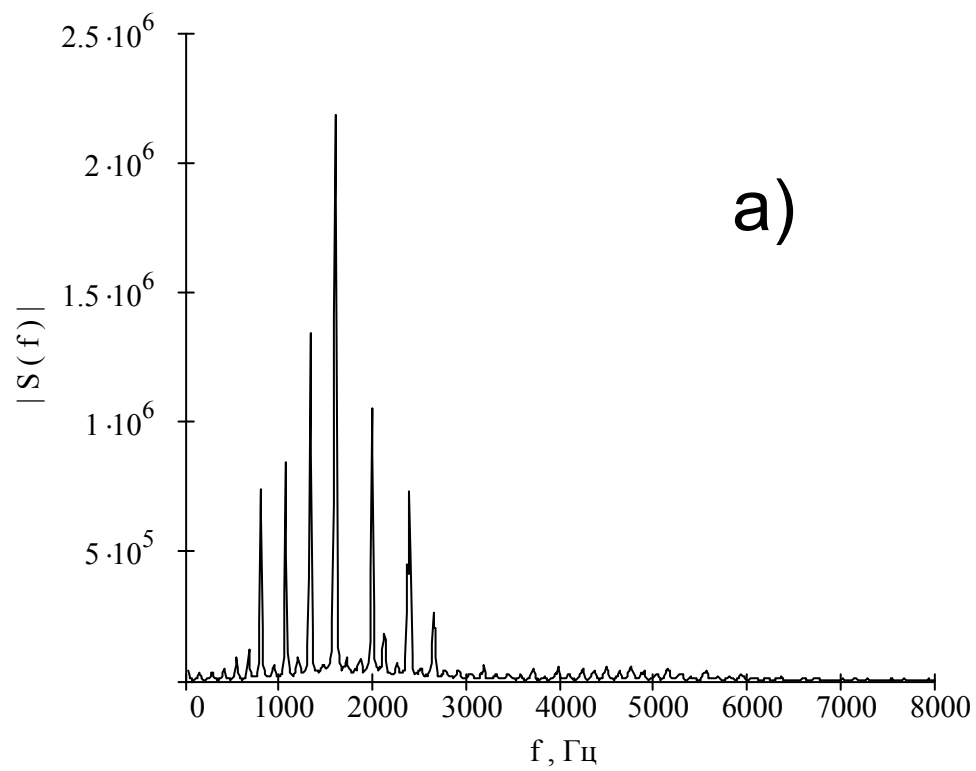


Рисунок 4. Спектр фрагмента искажённого сигнала а) до фильтрации, б) после фильтрации

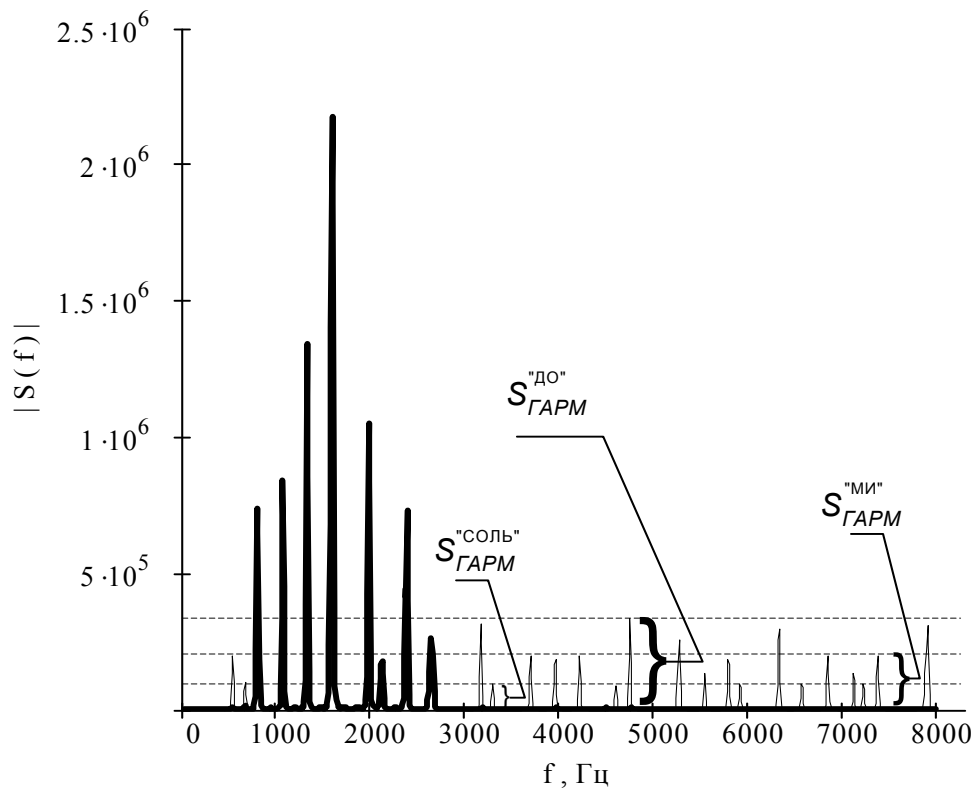


Рисунок 5. Синтез обертонов по простому алгоритму (жирной линией изображён спектр сигнала после фильтрации)

После фильтрации можно переходить непосредственно к восстановлению утраченных компонент спектра – синтезу обертонов. Как мы уже говорили, известны лишь приблизительные частоты недостающих обертонов, но не их амплитуды. Недостающими можно считать обертоны, мощность которых ниже определённого порога. В такой ситуации можно предложить два возможных подхода (конечно, не единственно возможных):

1. синтез недостающих обертонов по простому алгоритму — одинаковой амплитуды, например, зависящей от амплитуды максимального обертона данного звукового объекта $S_{ГАРМ} = k \cdot S_{МАКС\ ЗО}$ (см. рис. 5), или, с амплитудами, спадающими с частотой, например, пропорциональными $\frac{1}{f}, e^{-b \cdot f}, (1 - a \cdot f)$.
2. синтез недостающих обертонов по заранее заданной *структуре звукового объекта*.

Под структурой любого объекта понимается состав и взаимосвязь составляющих его элементов. Соответственно, структура звукового объекта описывается соотношениями параметров обертонов, изменяющихся во времени. Упрощённой структурой звукового объекта может служить соотношение амплитуд и фаз обертонов в зависимости от частоты основного тона. 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.

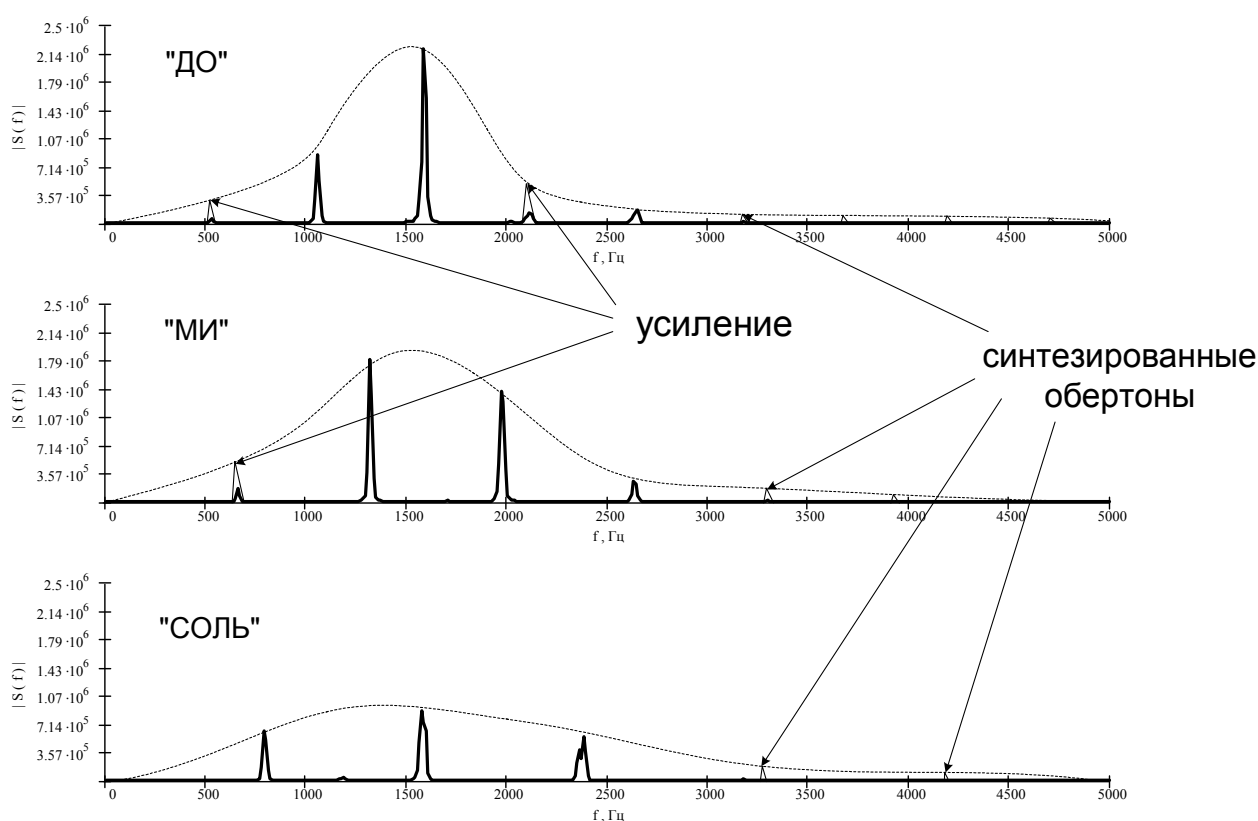


Рисунок 6. Пример приведения структур звуковых объектов к заданному виду

Возвращаясь к алгоритму синтеза обертонов, обратимся к рисунку 6. На нём изображены примеры приведения структуры отдельного звукового объекта к заданному виду. В восстанавливаемом звуковом фрагменте содержатся три звуковых объекта — реализации нот «до», «ми» «соль». Их амплитудные спектры для наглядности «разнесены» на разные графики. Пунктирными линиями изображены огибающие амплитудных спектров звуковых объектов, определяющие заранее заданное соотношение амплитуд обертонов. Чтобы привести соотношение амплитуд обертонов

в обрабатываемом сигнале к требуемому, необходимо изменить амплитуды (уменьшить или увеличить) существующих обертонов и синтезировать недостающие, что схематично и изображено на рисунке 6. Структура звукового объекта, как мы уже говорили, может быть задана заранее, но может быть сформирована и на основе анализа сигнала в процессе обработки. Например, огибающая амплитудного спектра может быть сформирована интерполяцией и экстраполяцией по двум-трём обертонам, имеющим наибольшие амплитуды (в случае, приведённом на рисунке 6, использован именно такой метод).

Для подведения итогов, во-первых, изобразим схему (см. рис. 7) действий и управления процессами при восстановлении музыкального сигнала, а во-вторых, обсудим достоинства и недостатки предлагаемого подхода.

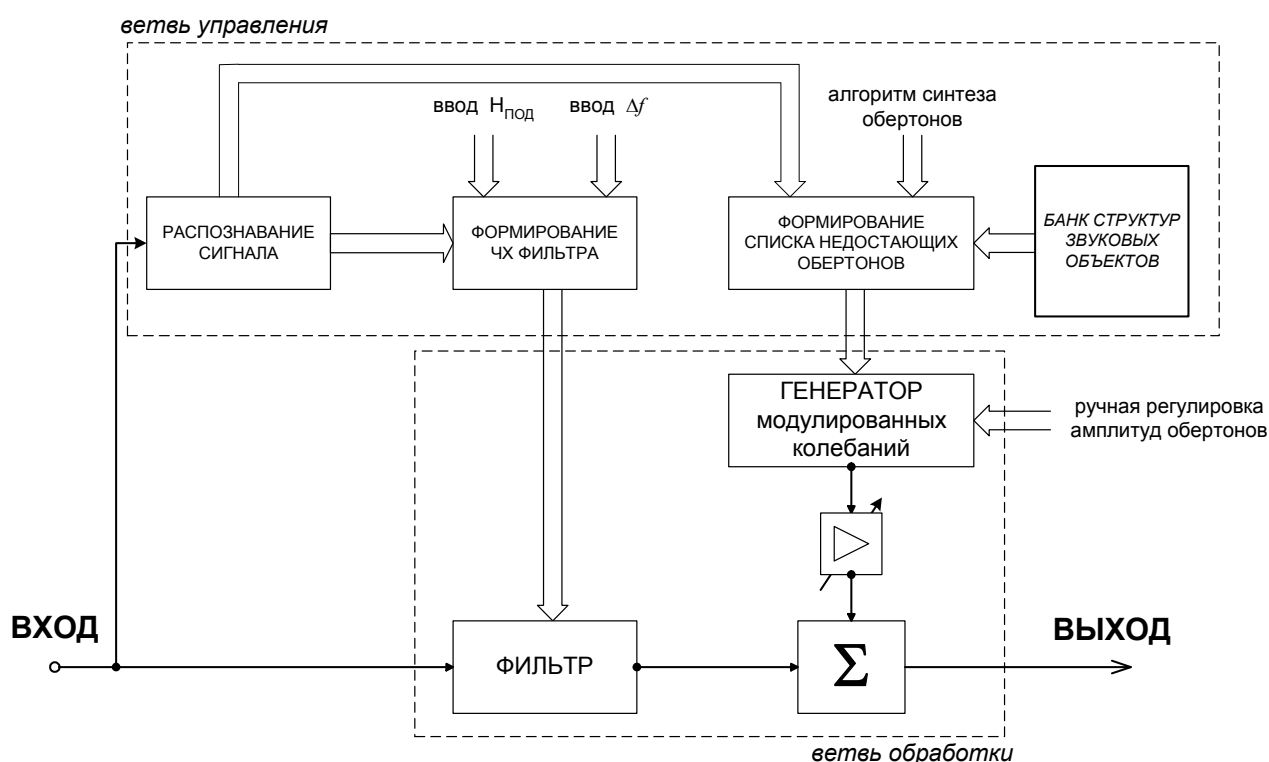


Рисунок 7. Схема управления и последовательность процессов обработки при восстановлении

В схеме сразу 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. Выражаясь образно, остальные процессы управления, действуя по намного более простым алгоритмам, механически исполняют указания процесса распознавания. Процесс распознавания формирует информацию, необходимую для работы остальных процессов. Это, как

мы уже говорили, информация о спектре сигнала, о содержании и структуре звуковых объектов. Формирование частотной характеристики фильтра — процесс управляемый. При этом локализация полезных составляющих спектра по времени и частоте (сегментация и определение высоты тона) производится распознаванием, а коэффициент ослабления шума ($H_{под}$) и «степень свободы» (Δf) отдельного обертона может и должна регулироваться как вручную, так и автоматически.

Синтез обертонов — также процесс регулируемый и управляемый. Алгоритм синтеза выбирается пользователем, в случае синтеза по заданной структуре звукового объекта информация о параметрах синтезируемых обертонов поступает из *банка структур звуковых объектов*. Результатом формирования списка недостающих обертонов является передаваемый в генератор модулированных колебаний набор параметров, содержащий амплитуды, частоты, начальные фазы, параметры модуляции частоты и амплитуды генерируемых колебаний. На данной стадии доступно ручное регулирование амплитуд **отдельных обертонов**, т.е., например, пятого или восьмого, или всех чётных или нечётных. На выходе генератора — сумма всех синтезированных обертонов, поступающая на регулируемый усилитель. Этот усилитель необходим для регулирования (опять же ручного или автоматического) мощности добавляемых (в сумматоре) к прошедшему фильтр сигналу спектральных компонент.

И достоинства, и недостатки предлагаемого подхода обусловлены применением распознавания образов. Два основных достоинства:

1. Возможность работать с сигналом на «элементарном уровне», т.е. с отдельным звуковым объектом, что даёт беспрецедентные возможности по изменению тембра, подавлению шумов и продуктов нелинейных искажений. Важно отметить, что для синтеза утраченных компонент спектра используется генератор модулированных колебаний, а не синтезатор гармоник, как во многих психоакустических процессорах. Принципиальное отличие состоит в том, что синтезатор гармоник генерирует гармоники исходного сигнала, т.е. смеси полезного сигнала, шумов, продуктов нелинейных искажений. Фактически, спектр кроме гармоник сигнала обогащается новыми посторонними компонентами. Использование же распознавания

позволяет произвести «адресный» синтез обертонов, т.е. добавить желаемое количество компонент с желаемыми параметрами (частотами, амплитудами и т.д.).

2. Реализация предлагаемого подхода позволяет создавать устройства обработки, полностью контролируемые и управляемые пользователем. Это очень важно, поскольку вместо чёрного ящика, снабжённого единственной ручкой «глубина введения эффекта», «улучшающего», «расширяющего», «добавляющего воздушности и прозрачности» и т.д., пользователь получает рабочий инструмент, научившись владеть которым, он сможет решать многие задачи, возникающие не только в процессе реставрации фонограмм.

Основных недостатка также можно выделить два:

1. Распознавание, как мы говорили, — задача статистическая, и достоверность распознавания современных систем не превышает 95%. 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.

2. Современные системы распознавания звуковых сигналов не универсальны и классифицируются по типам распознаваемых сигналов: музыкальных или речевых. Соответственно, создание универсальных устройств обработки на основе распознавания пока невозможно, к сожалению, область их применения будет очень узка. Создание и использование же универсальных систем распознавания, отличающих речь от музыки и, тем более, распознающих смесь речевого и музыкального сигналов значительно удорожат устройства обработки. При этом универсальная система распознавания, очевидно, будет обеспечивать меньшую достоверность распознавания, чем специализированная система.

В заключении 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. Целью реставрации является не только улучшение субъективной оценки качества, но и повышение соотношения уровня полезного сигнала

к уровню шума. Архивные и фондовые записи используются при формировании программ теле- и радиовещания наравне с современными, и при реставрации фонограммы необходимо подготовить её к этому использованию, сделать возможными дальнейшую обработку, передачу и хранение сигнала. Если забытую картину доставать раз в 15 лет, стирать с неё пыль, и делать репродукцию для выставки, то что будет дальше? Репродукция попадёт в помойку, а картина рано или поздно рассыпется. Не так ли мы поступаем с архивными записями? Можно рассуждать о целесообразности восстановления и сохранении «духа времени» или «оригинального звучания». Конечно, время вспять не повернуть, и как слышал Шаляпина звукорежиссёр, мы не узнаем, но есть предел деградации качества, который делает невозможным восприятие информации, каким бы аудиофилом вы не были...

Литература

1. **Павленко А.** Реставрация фонограмм с помощью программы DART Pro. Мир ПК, №11/1997
2. Станции ТРЕК от фирмы "Тракт". 625, №7/1998
3. SonicStudio 1.8 NoNoise Guide
ftp://ftp.sonicstudio.com/PDFs/NoNoise_Guide_1.8.pdf
4. Introduction to Audio Restoration <http://www.cedar-audio.com/>
5. Система цифровой реставрации и ремастеринга фонограмм «Канонь»
<http://ru.ecomstation.ru/showarticle.php?id=68>