## Введение

Окружающий акустический шум является мощным мешающим фактором, снижающим качество работы систем цифровой речевой связи. При этом падает как разборчивость речи, так и ее качество, выражаемое в терминах естественности звучания, узнаваемости голоса и т.д. Помимо эффекта маскирования речи шумом, одной из основных причин такого снижения является сильный рост искажений при прохождении зашумленной речью преобразования в устройстве низкоскоростной компрессии речи (вокодере). В частности, для низкоскоростных (0.6-4 кбит/с) параметрических вокодеров, снижение качества становится заметным для отношений сигнал/шум (ОСШ) на входе менее +15…+20 дБ, вследствие возрастания числа ошибок в оценке параметров модели синтеза речевого сигнала на основе анализа текущей речи. Нижний предел для ОСШ, при котором речь становиться неразборчивой, лежит около 0…+3 дБ, в зависимости от типа вокодера, вида шума и способа измерения ОСШ.

Так как частотные спектры акустического шума и речи перекрываются, набор линейных методов фильтрации, приводящих к повышению ОСШ для речи на входе вокодера, ограничен. К ним можно отнести пространственную фильтрацию с помощью направленных микрофонных систем и предварительную частотную фильтрацию входного сигнала, формирующую АЧХ для узкополосных (0.3-3.4 кГц) систем связи. Первые методы имеют самостоятельное значение, они лишь отодвигают наступление порога понижения ОСШ на входе вокодера в конкретной ситуации, но на сам порог не влияют. Вторые методы не эффективны, так как имеют постоянные среднестатистические параметры фильтрации, не учитывающие динамики спектральных характеристик конкретных шумовых и речевых сигналов.

В условиях квазистационарных аддитивных статистически независимых от речи акустических шумов для улучшения ОСШ речевых сигналов широко применяются шумопонижающие устройства (ШПУ), построенные на основе нелинейных адаптивных алгоритмов очистки речи от шумов. Чаще всего они строятся с использованием различных кратковременных преобразований сигнала, оценки компонент шума и речи в преобразованной области, подавления компонент шума с последующим обратным преобразованием очищенного сигнала во временную область. Обработка ведется по кадрам длительностью 10-30 мс. Используются следующие преобразования: дискретное преобразование Фурье (ДПФ), дискретное косинусное преобразование, вейвлет-преобразование, преобразование Карунена-Лоева и др.

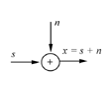
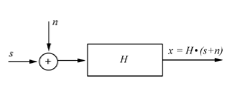
Наиболее распространены алгоритмы на основе ДПФ с обработкой кратковременного спектра амплитуд сигнала в частотной области следующими методами: спектрального вычитания; Винеровской фильтрации; статистических оценок амплитуд речевого сигнала по критерию максимального правдоподобия или минимума среднеквадратической ошибки. В настоящее время, наиболее развит метод статистических оценок значений логарифма спектра амплитуд речевого сигнала по критерию минимума среднеквадратической ошибки (MMSE-LSA), дополненный статистической оценкой вероятности присутствия речи в шуме.

Считается, что для людей с нормальным слухом адаптивные алгоритмы очистки речи от шумов на основе методов спектрального вычитания увеличивают субъективное качество очищенной речи, повышают ее ОСШ, снижают утомляемость при длительном прослушивании, но не способны повысить ее разборчивость по сравнению с разборчивостью исходной речи в шумах.

# Общие сведения о шумах и адаптивной фильтрации речевого сигнала

Регистрация, анализ и обработка аудиоинформации являются одним из важнейших факторов при проведении мероприятий по организации информационной безопасности. При этом зачастую возникает необходимость обработки аудиосигнала с целью повышения его качества и разборчивости. При проведении слухового контроля или получении магнитофонных записей речевого сигнала в реальных условиях на этот сигнал воздействуют различные помехи, которые снижают качество полезного (речевого) сигнала, в том числе и его разборчивость, вплоть до срыва связи. Задача снижения уровня помех с целью восстановления смысла сообщения для ряда практических ситуаций крайне актуальна.

Воздействие помех на полезный сигнал упрощенно может быть представлено на следующих моделях. На рис. 1.1 представлена модель воздействия аддитивных шумов на речевой сигнал, т.е. шум складывается с полезным сигналом. Эта модель соответствует ситуации, когда запись производится на открытом пространстве и в качестве помех могут быть шум ветра, уличные и строительные шумы и т.д.



На рис. 1.2 представлена модель воздействия аддитивных и мультипликативных помех. В этом случае до поступления в приемник (ухо человека) информации аддитивная смесь (речевой сигнал плюс акустический шум) проходит по тракту передачи, имеющему частотнозависимую передаточную характеристику. Таким образом, аддитивная смесь претерпевает дополнительные мультипликативные искажения: смесь домножается на резонансы передаточной характеристики тракта (сворачивается с импульсной характеристикой тракта «Н»). Эта модель соответствует записи сигнала в помещении или передаче сигналов по радио и телефонным трактам.

Задача устранения или снижения уровня аддитивных и мультипликативных помех осложняется вариативностью характеристик акустических помех (шумы ветра, листвы, проходящего мимо транспорта, музыки и т.п.) и трактов передачи (говорящий человек ходит по комнате, поворачивает голову и т.д.). Таким образом, для эффективного устранения искажений речевого сигнала необходимо, чтобы устройство, выполняющее эту функцию, постоянно отслеживало изменения характеристик помех во времени и постоянно корректировало свою импульсную характеристику в соответствии с этими изменениями. Такими возможностями обладают устройства, использующие адаптивную фильтрацию с целью выделения помехи, точнее ее оценки, с последующей ее компенсацией в смеси полезного сигнала и помехи.

Искаженный сигнал может быть представлен в виде одноканального сигнала, т.е. в виде смеси полезного сигнала и помехи (зашумленный речевой сигнал - ЗРС), или в виде двухканального сигнала, когда помимо основного канала - ЗРС, присутствует и опорный канал, сигнал в котором максимально близок к помехе, присутствующей в ЗРС.

По виду представления входного сигнала различают одноканальные и двухканальные устройства адаптивной фильтрации. Упрощенные блок-схемы одно- и двухканальных устройств представлены на рис. 1.3 и 1.4, соответственно, где представлены адаптивный фильтр, или процессор, состоящий из двух блоков: трансверсального фильтра (для вычисления оценки шума ) и процессора КЛП (для вычисления импульсной характеристики фильтра или вектора коэффициентов линейного предсказания ) и отдельного сумматора для вычисления результата компенсации .

В процессоре КЛП значения  вычисляются таким образом, чтобы предсказанное на момент времени  значение  компенсировало шумовую составляющую  с минимальным остатком. Значения ,  и  вычисляются на каждом периоде дискретизации. Настройка на полную компенсацию шумовой составляющей осуществляется не мгновенно, а за определенное время (время адаптации), которое регулируется с помощью коэффициента адаптации .

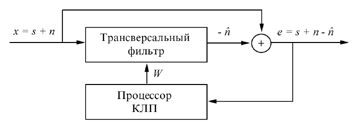


Рис. 1.3

При наличии только одноканального сигнала компенсация производится по схеме на рис. 1.3. В этом случае опорный сигнал формируется из ЗРС. По этой схеме могут быть уменьшены аддитивные шумы, имеющие периодические составляющие (например, шумы различных моторов, двигателей, музыки и т.п.), а также может быть уменьшено влияние мультипликативных помех, в том числе и реверберационных искажений.

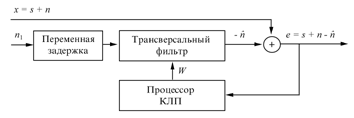
Для осуществления компенсации шума в двухканальном сигнале используется схема адаптивной фильтрации, представленная на рис. 1.4, где по основному каналу поступает ЗРС, а по опорному - только шум , коррелированный с шумом  в ЗРС. Регулируемая задержка предназначена для компенсации акустической задержки сигнала, возникающей в одном из каналов (на рис. 1.4 показана компенсация задержки в основном канале). При наличии соответствующего сигнала в опорном канале по этой схеме можно с той или иной эффективностью скомпенсировать практически любые аддитивные шумы.

Рис. 1.4

В обоих вариантах представления входного сигнала адаптивная фильтрация осуществляется по одной и той же процедуре. В цифровом адаптивном фильтре на каждом периоде дискретизации производится вычисление  проекций  вектора  и вычисление свертки  с входным сигналом. В результате этого на -ый момент времени для исходного сигнала  определяется значение выходного сигнала , где скомпенсирована помеховая составляющая.

Подстройка (адаптация к внешним условиям) вектора  осуществляется на принципах оптимизации по критерию минимума среднего значения модуля выходного сигнала. При получении вычислительного алгоритма адаптации используется математический аппарат оптимальной фильтрации. Сходимость алгоритма осуществляется по методу наискорейшего спуска, причем, для упрощения вычислений используется стохастическая аппроксимация градиента по Уидроу-Хопфу.

В одноканальном варианте для обработки ЗРС используется алгоритм адаптивной dеcоnvоlution («развертки»), а в двухканальном - адаптивной компенсации. Принципиальное отличие вариантов обработки заключается в формировании входных сигналов, которые используются в последующей вычислительной процедуре. В одноканальном варианте оба входных сигнала (основной и опорный) формируются из одного входного сигнала, при этом исходный входной сигнал является основным, а опорный формируется из исходного с помощью единичной задержки. В двухканальном варианте основной и опорный сигналы реально существуют и непосредственно используются в последующей вычислительной процедуре.

Следует отметить некоторые моменты, полезные для практической работы с одноканальным сигналом. В пределе, адаптация  происходит до полной декорреляции входного сигнала, т.е. до получения на выходе «белого» шума. При этом не имеет значения за счет каких помех спектральная огибающая входного сигнала имеет неравномерности: за счет аддитивных помех с «окрашенным» спектром или за счет свертки с резонансами тракта передачи. Это касается и самого речевого сигнала, который является продуктом свертки голосового и шумового источников возбуждения с импульсной характеристикой артикуляторного тракта, т.е. при неудачном выборе скорости адаптации, которая регулируется с помощью коэффициента адаптации m, можно не только скомпенсировать помеху, но и значительно исказить речевой сигнал.

# Цифровые мето ды фильтрации речевых сигналов

Речевые сигналы, с которыми приходится иметь дело на практике, всегда в той или иной степени зашумлены. В тех случаях, когда шум имеет значительную интенсивность, его наличие может существенно исказить результаты обработки, анализа или распознавания речи. В целом ряде других случаев, например, при анализе зашумленных записей в криминалистических целях или восстановлении аудиозаписей в архивах, задача очистки сигнала от шума носит самостоятельный характер и является единственной целью работы. Поэтому разработка методов очистки сигнала от шума является весьма актуальным направлением исследований. К настоящему времени разработано очень большое количество различных методов цифровой обработки зашумленных речевых сигналов.

Основным типом шумов, для представленных методов, является аддитивный шум.

В целях упорядочения рассмотрения методов очистки сигнала от шума целесообразно произвести их классификацию. Основным признаком, по которому будут классифицироваться алгоритмы, является характер или тип тех закономерностей, которые служат основой для выделения речевого сигнала из смеси с шумом. В качестве вспомогательного признака будет использоваться классификация по типу того математического или алгоритмического аппарата, который использован для фильтрации. Подобная классификация, конечно, весьма условна, так как многие из рассматриваемых методов нельзя безоговорочно отнести к какой-либо одной категории. Как правило, одни и те же методы используют одновременно различные принципы, и в этом случае можно говорить лишь о преимущественном влиянии какой-либо концепции.

С учетом сделанного замечания можно выделить следующие группы методов цифровой обработки зашумленных речевых сигналов:

- методы адаптивной компенсации помех;

- методы, основанные на использовании математических моделей речевых сигналов во временной области (например, авторегресионная модель речевого сигнала и рекуррентные алгоритмы оценки параметров и речевого сигнала);

- методы, основанные на использовании математических моделей речевых сигналов в частотной области (оценивание минимальной среднеквадратической ошибки, марковские модели сигнала и шума);

- методы, основанные на использовании спектральных характеристик шума (вычитание амплитудных спектров, Винеровская фильтрация);

- методы, основанные на использовании моделей искусственных нейронных сетей;

- методы, основанные на моделях восприятия речи человеком.

## Адаптивные компенсаторы помех

Этот класс методов цифровой обработки зашумленных сигналов основан на использовании, помимо собственно зашумленного сигнала, который подлежит очистке, также одного или нескольких опорных сигналов - сигналов, которые коррелированны с шумовым сигналом и некоррелированные (или слабо коррелированные) с полезным сигналом, подлежащим выделению. С помощью опорных сигналов формируется сигнал, который является оценкой помехи. Этот сигнал затем вычитается из зашумленного сигнала и результат этой операции рассматривается как оценка не зашумленного сигнала.

На рис. 2.1 представлена схема адаптивного компенсатора помех, который использует один опорный сигнал.

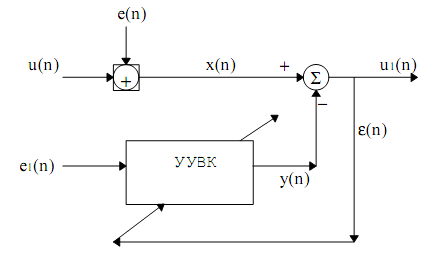


Рис. 2.1. Схема адаптивного компенсатора помех

Здесь  - дискретный отсчет полезного сигнала в момент времени ;  - шумовой сигнал;  - опорный сигнал;  - сигнал ошибки;  - выходной сигнал компенсатора; УУВК - устройство управления весовыми коэффициентами.

Наиболее важной частью адаптивного компенсатора помех является устройство управления весовыми коэффициентами - линейный фильтр, через который пропускается опорный сигнал . Задача адаптивной компенсации помехи  сводится к подбору коэффициентов фильтра таким образом, чтобы минимизировать энергию сигнала на выходе компенсатора . В этом случае будет максимизировано выходное отношение сигнал/шум. Минимизация энергии обычно осуществляется на основе градиентных методов поиска экстремума функций многих переменных.

Известно, что адаптивные компенсаторы помех позволяют значительно улучшить качество зашумленных сигналов - на несколько десятков децибел, но требование наличия опорного сигнала существенно сужает их область применения. Во многих приложениях цифровой обработки речевых сигналов (например, при реставрации архивных записей или в криминалистике), опорного сигнала, по крайней мере, в явном виде, не имеется. Поэтому для применения методов адаптивной компенсации помех опорный сигнал в таких случаях приходится получать на основе косвенных соображений, связанных с особенностями речевого сигнала, а сам адаптивный компенсатор в этом случае будет являться одной из составных частей более сложного алгоритма выделения речевого сигнала.

# Методы, основанные на использовании статистических моделей речевых сигналов во временной области

Класс методов цифровой обработки зашумленных речевых сигналов, который основан на построении математических моделей речевых сигналов и обработке речевых сигналов с использованием этих моделей быстро развивается и в настоящее время эти методы приводят к самым успешным результатам. Задача выделения речевого сигнала из смеси с шумом в случае использования достаточно адекватной модели сводится к оценке каким-либо образом параметров этой модели и последующим синтезом или фильтрации речевого сигнала фильтром, построенным на основе или с помощью оцененных параметров.

Одними из наиболее перспективных методов в этом классе являются методы статистической фильтрации во временной области. Фильтрация речевого сигнала, моделируемого авторегрессией, осуществляется при этом методами теории оптимального оценивания, например, с помощью построения оптимального линейного фильтра (фильтра Калмана).

Вычислительно эффективная (но с менее удачным результатом обработки) реализация алгоритма фильтрации речевого сигнала, моделируемого авторегрессионной моделью с параметрами, связанными в марковскую цепь. Совместная оценка сигнала и параметров марковской цепи вычисляются рекуррентным способом с помощью алгоритма максимизации математического ожидания (expectation maximization - EM), причем для вычисления условного ожидания (expectation step) сигнала относительно наблюдений использован фильтр Калмана-Бьюси.

Экспериментальные испытания на речевом сигнале в смеси с некоррелированным аддитивным белом шумом с отношениями сигнал/шум 0, 10 и 20 дБ показали увеличение отношения сигнал/шум в среднем на 4 дБ.

Авторегрессионная модель речевого сигнала, как показывает практика, не имеет такого выраженного дефекта как музыкальные тона, однако, артефакты обработки также имеют место. В целом фильтрация высокоэнергетических частотных диапазонов звуков выполняется точнее, в то время как моделирование нулей и минимумов спектра затруднено. Этот эффект непосредственно связан с порядком  используемой модели авторегрессии.

Высокие (предварительные) результаты продемонстрировал комбинированный метод анализа и фильтрации речевого сигнала, основанный на частотной декомпозиции сигнала c последующим раздельным моделированием сигнала в каждой частотной полосе моделью авторегрессии.

Речевой сигнал декомпозируется в несколько частотных полос, в каждой из которых оценивается отношение сигнал/шум, на основании которого выбирается порядок  модели сигнала в этой частотной полосе. В общем случае правило выбора величины  таково, что сигнал в частотном диапазоне с высоким отношением сигнал/шум (как правило, это соответствует формантным максимумам речевого сигнала) моделируется авторегрессией большего порядка, а сигнал для частотных полос с низким отношением сигнал/шум (нули спектра сигнала) моделируется авторегрессией более низкого порядка.

Собственно фильтрация осуществляется модифицированным фильтром Винера в частотной области. Предварительные измерения (смесь речи с белым шумом) показали значительное увеличение отношения сигнал/шум: на +15 дБ при начальном отношении −5 дБ (соответственно, при начальном ОСШ +5 дБ улучшение составило 11 дБ).

## Методы, основанные на обработке речевого сигнала с использованием аппарата скрытых марковских моделей

Другим классом методов обработки зашумленных речевых сигналов основанных на использовании статистических моделей речевого сигнала являются методы, в которых речевой сигнал моделируется скрытой Марковской цепью. То есть для моделирования речевого сигнала использован наиболее эффективный для распознавания речи подход.

Известно, что традиционно используемые методы фильтрации (вычитание спектров или фильтр Винера) не используют фонетическую информацию, переносимую речевым сигналом. Недавние исследования показали, что знание и применение в процессе обработки фонетической структуры сигнала приводит к улучшению качества фильтрации. Поэтому вполне естественным является применение в процессе очистки речевого сигнала от шумов его статистической модели в виде скрытой марковской цепи, которая связана с фонетической структурой сигнала.

Идея реализации такого подхода заключается в том, что первоначально, по записям не зашумленного речевого сигнала строятся статистические модели единиц речевого потока (фонов либо более широких классов звуков). После того, как статистическая модель для множества состояний сигнала построена, по ней можно рассчитать оптимальный фильтр Винера.

При обработке зашумленного речевого сигнала сначала оценивается (по отфильтрованному на предыдущем шаге сигналу) текущее состояние марковской модели, в соответствии с которым выбирается оптимальный фильтр, который затем используется для фильтрации сигнала и получения очередной оценки.

Алгоритм фильтрации выглядит следующим образом:

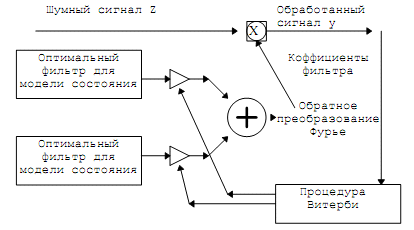


Рис. 2.2. Алгоритм фильтрации речевого сигнала с использованием скрытой марковской модели

Для эффективной обработки нестационарных сегментов отдельно оценивалась Марковская модель шума. В отличие от простых моделей состояний полезного сигнала шум моделировался набором состояний, каждое из которых содержало несколько гауссовских компонент.

Во время обработки зашумленного сигнала при определении отсутствия полезного сигнала выполнялось декодирование сегмента паузы процедурой Витерби для выбора оптимальной модели шума. Модель шума, обеспечивающая максимальное правдоподобие наблюдаемой последовательности использовалась далее для обработки сигнала.

Для сохранения «преемственности» между итерациями применялись инерционная схема фильтра Винера.

В реальных экспериментах по фильтрации речевого сигнала описанными методами число состояний марковской модели речевого сигнала выбиралось равным 5, то есть фактически речевой сигнал моделировался как последовательность фонов, соответствующих широким фонетическим категориям (звонкий - шумный - глухой и т.п.).

Сравнительные исследования выполнялись на различных типах шумов (белый, симуляция шума вертолете, одновременный разговор нескольких мешающих дикторов). Объективные измерения (изменение отношения сигнал/шум на входе и выходе системы) показали очевидное превосходство описанной методики (в среднем улучшение +2.5 дБ, в диапазоне от 0 до +20 дБ, причем при отношении сигнал/шум > 10 дБ, превосходство составило в среднем +7 дБ) над стандартной системой фильтрации, построенной по методу вычитания амплитудных спектров. Субъективные тесты (оценка качества звучания обработанного сигнала на слух по пятибалльной шкале) также показали превосходство марковских моделей над стандартными методиками. По мнению авторов, разборчивость речи в результате обработки также повысилась, вероятно вследствие того, что низкоэнергетические звуки в данном случае обрабатываются существенно аккуратнее.

Последовательное развитие этого подхода привело к использованию алгоритма минимизации среднеквадратической ошибки вместо процедуры Витерби с улучшенными возможностями.

Следует отметить, что очевидным недостатком подхода является необходимость иметь априорную информацию о возможных типах шумов (в виде предварительно обученных марковских моделей состояний). Типов возможных шумов для различных практически важных условий много и требование наличия заранее вычисленных моделей представляется мало выполнимым.

Кроме того, известно, что качество обработки сигнала ухудшается в тех случаях, когда помеха имеет существенно нестационарный характер.

В связи с этим дальнейший прогресс в этом направлении может быть, достигнут за счет использования более гибких способов моделирования помех. В работе предлагается эффективный алгоритм для переоценки параметров вероятностной модели шума, основанный на адаптивной подстройке элементов кодовой книги (которая состоит из матрицы корреляционных коэффициентов для авторегрессионной скрытой марковской модели) по методу скользящего среднего. Показано, что подобная методика приводит к дополнительному выигрышу примерно 2.3 дБ (по сравнению с исходным алгоритмом фильтрации, основанным на марковской модели) на нестационарных шумах и не ухудшает качества обработки для стационарных помех.

Использование марковских моделей речевого сигнала оказывается выигрышным при наличии априорной информации синтаксического характера о зашумленном речевом сигнале. Довольно часто (например, при анализе черных ящиков с борта самолета или реставрации аудиозаписей) время обработки сигнала не играет определяющей роли.

В тех случаях, когда аудитор может приблизительно указать (дать гипотезы), что именно было произнесено, или на что похоже зашумленное высказывание, можно улучшить качество сигнала с помощью использования аппарата марковских моделей, построенных для предстазанных звуков.

Анализ метода фильтрации речевых сигналов с помощью марковских моделей при наличии гипотез о произносимом тексте выполнен в работе. Было продемонстрировало улучшение, в результате обработки, качества звучания сигнала в условиях различных типов помех и в широком диапазоне отношений сигнал/шум.

## Методы, основанные на использовании, отдельных характерных свойств речевого сигнала

К методам этого типа относятся, прежде всего, класс методов обработки зашумленных речевых сигналов, которые используют квазипериодичность речевого сигнала. Первая группа методов использует периодичность речевых сигналов для построения адаптивного компенсатора помех, с помощью которого обрабатывается зашумленный речевой сигнал. Предполагается, что исходный речевой сигнал  строго периодичен с периодом , кратным частоте дискретизации, а случайный аддитивный шум  некоррелирован с . В качестве опорного сигнала для адаптивной компенсации помехи используется:

, где .

Отношение сигнал/шум может быть увеличено на 7-10 дБ, однако разборчивость отфильтрованной речи при этом несколько понижается.

Вторая группа методов, использующих периодичность звонких звуков основана на представлении сигнала в кепстральной области. В этом случае периодический характер речевого сигнала используется для синтеза адаптивной гребенки фильтров.

Периодичность звонких звуков выражается в частотной области в том, что их спектр имеет линейчатый характер, причем соседние пики (спектральные максимумы) отстоят друг от друга на интервал (в частотной области) равный частоте основного тона. Поэтому, если гребенка фильтров такова, что гармоники основного тона (спектральные пики) попадают в полосы пропускания, то можно рассчитывать на повышение качества речевого сигнала.

Исследования, проведенные на синтетических гласных звуках показали, что при надлежащем выборе взвешивающих коэффициентов можно добиться значительного эффекта для улучшения восприятия речи в тех случаях, когда помеха или шум являются структурированными.

Обобщение метода является также известный адаптивный фильтр Фрезиера. В этом случае учитывается изменение частоты основного тона на интервале времени, равном длине импульсной характеристики гребенки фильтров. Исследование характеристик такого фильтра показали, что он может дать выигрыш в отношении сигнал/шум до 10 дБ, однако при этом несколько снижается разборчивость речи.

Описанные методы имеют существенные недостатки. Помимо того, что все методики, кроме фильтра Фрезиера, не учитывают изменений частоты основного тона, эти методы непригодны для фильтрации глухих звуков. Например, существенные для разборчивости речевого сигнала звуки «п», «т», «к» не могут быть успешно обработаны такими методами. Наконец, качество обработки сигнала зависит от точности оценки частоты основного тона в зашумленном речевом сигнале, что само по себе не всегда возможно.

То обстоятельство, что на разборчивость речи существенно влияет правильное восприятие согласных звуков, в частности, взрывных «п», «т», «к» и шумного «с», было использовано в системе. Фильтрация речевого сигнала заключалась в том, что перед взрывными звуками вставлялась короткая пауза, - смычка, а согласный «с» фильтровался специально подобранным фильтром.

## Методы, основанные на оценке спектральных характеристик шума

Наиболее часто используемыми методами, основанными на использовании спектральных характеристик шума, являются методы, реализующие различные модификации алгоритма вычитания амплитудных спектров.

В качестве обоснования этих методов приводятся следующие соображения. Если стационарный сигнал ,  со спектральной плотностью мощности  искажен аддитивным стационарным шумом  со спектральной плотностью мощности , который предполагается некоррелированным с , то спектральная плотность мощности зашумленного сигнала  равна , следовательно спектральная плотность мощности полезного сигнала  может быть оценена как: .

В силу нестационарности речевых сигналов использовать соотношение непосредственно нельзя. На практике, при обработке речи на достаточно коротких участках, например квазистационарных участках гласных звуков, величины ,  аппроксимируют с помощью усредненных квадратов кратковременных амплитудных спектров наблюдаемого сигнала и шума. Спектр шума при этом должен оцениваться в моменты пауз. Полученная таким образом оценка соответствует квадрату амплитудного спектра сигнала. Восстановление речевого сигнала во временной области осуществляется с помощью обратного преобразования Фурье, причем фазовый спектр для восстановленного сигнала берется таким же, как и у наблюдаемого сигнала.

В наиболее общем виде операция спектрального вычитания может быть выражена соотношением:



Здесь коэффициент  (фактор переоценивания), вообще говоря, зависит от соотношения сигнал/шум на сегменте анализа, И имеет типичные значения близкие к 0.7-0.95, а коэффициент  (спектральный порог) - выбирается в диапазоне 0.01-0.1.

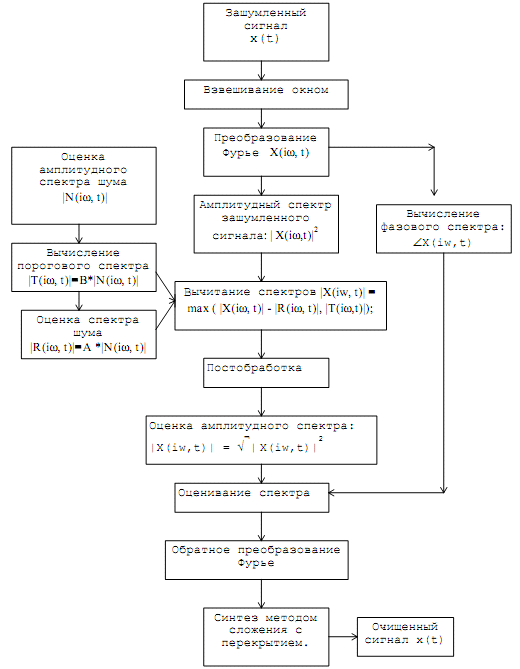
Блок-схема алгоритма вычитания амплитудных спектров приведена на следующем рисунке.

Рис. 2.3. Блок-схема алгоритма вычитания амплитудных спектров

Исследования качества и разборчивости речи, получаемой в результате применения описанной методики, показали, что в тех случаях, когда шум или помеха имеют стационарный (или квазистационарный) характер и их спектр имеет гармоническую структуру, достигается значительное на слух повышение, как качества, так и разборчивости речи. Однако, в случае шумов с быстроизменяющимися спектральными характеристиками такая обработка малоэффективна.

По мнению аудиторов, такая речь звучит чище и приятнее (несмотря на наличие характерных эффектов обработки - так называемых «музыкальных тонов», заключающихся в случайных кратковременных выбросах в спектре обработанного сигнала) чем до обработки, однако заметного повышения разборчивости в случае аддитивных широкополосных шумов не происходит, хотя отношение сигнал/шум повышается на 3-6 дБ.

В целом, методы, основанные на вычитании спектров считаются одними из лучших - они приводят к удовлетворительным результатам обработки и не требуют больших вычислительных ресурсов. Что же касается музыкальных тонов, которые существенно ухудшают восприятие обработанного сигнала, то для их подавления разработаны различные алгоритмы, основанные на эмпирических и эвристических соображениях.

К классу методов, основанных на оценке спектральных характеристик шума, относятся также методы коррекции спектра речевого сигнала, основанные на Винеровской фильтрации.

В этих методах зашумленный речевой сигнал  фильтруется фильтром с частотной характеристикой, рассчитанной из условия минимизации среднеквадратической ошибки фильтрации, то есть, если  - частотная характеристика такого фильтра, то

,

где  - спектральная плотность мощности сигнала,  - спектральная плотность мощности шума

В реальных условиях применения частотную характеристику (6.2) аппроксимируют как:

,

где  и  - усредненные квадраты амплитудных спектров сигнала  и шума , соответственно, причем оценка величин  и  осуществляется так же как и в методе вычитания амплитудных спектров.

Одним из основных недостатков спектрального вычитания является наличие артефактов в обработанном сигнале. Музыкальные тона существенно ухудшают качество сигнала, поэтому неудивительно, что одним из приоритетных направлений в развитии этого подхода стало создание постпроцессоров, снижающих эффект музыкальных тонов без дальнейших искажений в сигнале.

В большинстве случаев речь идет о пост-обработке сигнала в спектрально-временной области. Идея состоит в том, что музыкальные тона в спектрально-временном представлении представляют собой локальные (во времени и по частоте) спектральные максимумы, которые, как правило, можно найти из полуэвристических соображений. В работе поиск спектральных максимумов, соответствующих музыкальным тонам осуществляется методами обработки изображений.

Анализируя проводимые авторами упомянутых работ результаты, характеризующие эффективность очистки сигнала от шума, следует отметить, что практически во всех случаях при использовании методик типа вычитания спектров в качестве характеристики работоспособности алгоритма используются меры качества звучания сигнала или объективные меры типа усредненных евклидовых расстояний между отфильтрованным сигналом и незашумленным сигналом (предполагается, что он доступен).

Разборчивость в результате применения описанных методов к сигналу, содержащему случайный аддитивный шум (типа белого или розового, или нестационарные помехи с широкополосным спектром) по видимому, не изменяется.

Исключения к только что сформулированному тезису составили измерения, проведенные на методе вычитания кепстров (то есть вместо спектра сигнала использовался кепстр, а вместо логарифма при оценке спектральных амплитуд использовали кубический корень), когда для речевого сигнала с аддитивным шумом (двигатели самолета F-16) наблюдалась не только улучшение качества, но и разборчивости речи.

При сравнительных испытаниях методов фильтрации речевых сигналов реализованных как препроцессоры на входе ЛПК-вокодера, в результате артикуляционных испытаний для метода вычитания амплитудных спектров было зарегистрировано повышение разборчивости передаваемой речи на 22.8% при исходном отношении сигнал/шум 2 дБ.

Интересная модификация вычитания спектров - вычитание сигналов во временной области. Этот метод, в отличие от вычитания амплитудных спектров и Винеровской фильтрации дает хорошие результаты (в том числе, повышает как качество, так и разборчивость сигнала при исходных соотношениях сигнал/шум -10 дБ и −30 дБ) на нестационарных и коррелированных с речевым сигналов шумах типа мешающего диктора, гармонических помехах, одновременном фоновом разговоре (бормотании) нескольких дикторов.

Метод предполагает доступ к передаваемому незашумленному сигналу на передающей стороне и основан на технике добавления нулевых отсчетов в передаваемый речевой сигнал (то есть сигнал квантуется с удвоенной частотой, причем каждый второй отсчет - нулевой). На принимающей стороне характеристики шума оцениваются исходя из величин дополнительных (нулевых в начале передачи) отсчетов. Поскольку большинство практически встречающихся шумов (например, все речеподобные сигналы) коррелированны на интервалах между соседними отсчетами, оценка шума, выполненная для дополнительных отсчетов вполне пригодна для фильтрации сигнала.

Очевидными недостатками предлагаемой методики является условие доступа к данным на передающей стороне и удвоение скорости передачи данных, поэтому область практического применения алгоритма существенно сужена.

### Алгоритм амплитудного спектрального вычитания

Для повышения разборчивости речи применяется алгоритм спектрального вычитания. Спектральное вычитание оценивает спектр мощности очищенного сигнала путем вычитания спектра шума из зашумленного сигнала.

Алгоритм спектрального вычитания состоит из следующих этапов

Этап 1. Исходный зашумленный сигнал , состоящий из чистого речевого сигнала  и некоррелированного аддитивного шума .

, где  - индекс дискретизации.

Этап 2. Деление сигнала на перекрывающиеся кадры длиной .

Применение оконной функции  для точного вычисления коэффициентов Фурье. Окна подавляют просачивание спектральных составляющих, которое может привести к смещению оценок амплитуд и положений гармонических составляющих сигнала.

,

где  - индекс кадра,  - сдвиг от кадра к кадру,  - длина кадра.

Этап 3. Дискретное преобразование Фурье.

,

где  - значение *-*й спектральной компоненты -го кадра зашумленного сигнала,  - фазовый спектр -го кадра зашумленного сигнала,  - амплитудный спектр -го кадра зашумленного сигнала, .

Этап 4. Сглаживание спектра мощности -го кадра зашумленного сигнала

,

где  - значение *k-*й спектральной компоненты спектра мощности -го кадра зашумленного сигнала,  - коэффициент сглаживания или забывания, выбираемый для предотвращения как музыкального шума, так и слишком большого искажения сигнала. Обычно  лежит в пределах .

Этап 5. Оценка шума на основе отслеживания минимумов в области спеткра мощности от кадра к кадру:

,

где  - *k-*й локальный минимум спектра мощности -го кадра зашумленного сигнала

 - коэффициент адаптации к локальному минимуму.

Этап 6. Спектральное вычитание шума в области амплитудного спектра:

,

где  - передаточная функция фильтра подавления шума,

 - улучшенный кратковременный амплитудный спектр m-го кадра, *subf* - постоянная спектрального минимального уровня для ограничения максимального вычитания, - коэффициент избыточного спектрального вычитания как функция отношения сигнал-шум кадра и частотного индекса.

Этап 7. Аппроксимационные ОСШ или частотное расстояние в качестве критерия качества фильтрации зашумленного сигнала для оценки параметров алгоритма спектрального вычитания:

,

.

## Метод оценивания минимальной среднеквадратической ошибки

Описываемый алгоритм (оригинальное название Minimum Mean-Square Error estimation) как и вычитание спектров алгоритм основан на оценке амплитудного спектра сигнала и общая блок-схема алгоритма в целом соответствует рис. 2.3. Среди других методов фильтрации, предполагающих наличие только одного микрофона, алгоритмы, основанные на минимуме среднеквадратической ошибки являются одними из наиболее полезных. Их использование приводит к значительному сокращению уровня шума в сигнале без внесения остаточных искажений типа музыкальных тонов. В недавно проведенных исследованиях утверждается, что в значительной мере превосходство метода оценивания минимальной среднеквадратической ошибки над методиками типа Винеровской фильтрации или вычитания амплитудных спектров связано именно с введением априорной оценки сигнал/шум в каждой спектральной полосе. В связи с этим, были предложены модификации стандартных подходов (Винеровской фильтрации, вычитания амплитудных спектров и оценок максимального правдоподобия) использующие априорные отношения сигнал/шум, что привело к существенному улучшению результатов фильтрации.

## Подавление аддитивного квазистационарного шума методом вычитания амплитудных спектров

В качестве типичного примера можно привести шумы кондиционеров, видеокамеры, автотрансформаторов и усилителей. Поведение алгоритма контролируется набором параметров, включая предварительно измеренные характеристики шума. Для того чтобы пользователь смог полностью использовать все возможности, заложенные в алгоритме, предусмотрен аналоговый режим работы, когда выбор и изменение параметров и режимов обработки выполняется в ходе обработки, причем оператор контролирует качество работы метода прослушиванием обработанного сигнала.

Допустим, что  и  обозначают, соответственно, речевой сигнал и аддитивный шум, а  - наблюдаемый сигнал, то есть. Пусть также ,  и  обозначают соответственно спектральные компоненты речевого сигнала, шума и зашумленного сигнала, оцененные на интервале анализа, на котором предполагается квазистационарность речевого сигнала.

Оцениватель амплитудного спектра сигнала по минимуму среднеквадратичной ошибки (MMSE) определяется из двух следующих (апостериорного и априорного) локальных отношений сигнал/шум:

 и .

Передаточная функция шумоподавителя определяется формулой:

,где  - это обобщенное отношение правдоподобия, которое принимает во внимание величину неопределенности присутствия полезного сигнала (речи) в зашумленном сигнале. Отношение правдоподобия оценивается как

,

а «обычный» коэффициент подавления шума  на частоте  равен

,

где  - обозначают модифицированные функции Бесселя нулевого и первого порядка, а   - вероятность отсутствия полезного сигнала в соответствующей спектральной компоненте.

Приведенные формулы выведены при неявном предположении, что априорное отношение сигнал/шум известно. В реальных условиях, однако, этот параметр априори неизвестен, при этом предлагается его оценивать соотношением:

,

где  - индекс времени и  - обозначает операцию клиппирования полуволны. Параметр  выбирается из эмпирических соображений и обычно .

Превосходство метода оценивания минимальной среднеквадратической ошибки над методиками типа Винеровской фильтрации или вычитания амплитудных спектров в значительной мере связано именно с введением априорной оценки сигнал/шум в каждой спектральной полосе. В связи с этим, были предложены модификации стандартных подходов (Винеровской фильтрации, вычитания амплитудных спектров и оценок максимального правдоподобия) использующие априорные отношения сигнал/шум типа, что привело к существенному улучшению результатов фильтрации.

## Методы, основанные на искусственных нейронных сетях

Разработка аппарата искусственных нейронных сетей привела к появлению нового типа алгоритмов для обработки зашумленных речевых сигналов, основанных на использовании моделей нейронных сетей. Подобных работ пока еще немного и получаемые в этом направлении результаты пока хуже, чем достигаемые более традиционными методами, однако, поскольку нейронные сети обладают потенциально огромными возможностями по непараметрическому моделированию различных типов плотностей, можно ожидать в этом направлении появления мощных алгоритмов фильтрации.

Исследование свойств многослойного персептрона как нелинейного фильтра во временной области выполнено. В качестве помехи рассматривались аддитивный гауссовский шум и нелинейный шум, моделирующий артефакты низкоскоростного CELP кодера. Персептрон обучался на зашумленном сигнала, роль сигнала-учителя выполнял чистый сигнал (доступный на этапе обучения). В качестве речевого материала использовались записи гласного «e» (120 записей от 40 дикторов).

Наилучшие результаты были продемонстрированы на трехслойном персептроне (на каждом слое по 60 нейронов): улучшение отношения сигнал/шум составило 3 дБ для шума кодека и 6 дБ для белого шума при начальном уровне сигнал/шум = 7.4 дБ. Эти результаты, показывают, что пока выигрыша от использования многослойного персептрона по сравнению со многими стандартными методиками нет.

## Методы, основанные на использовании закономерностей восприятия речи человеком

Некоторые алгоритмы анализа речевых сигналов основаны на использовании свойств слухового анализатора человека. В основе развития этого класса методов лежит утверждение, что анализ речи, основанный на модели слуха человека, будет возможно более успешным, чем анализ, основанный на довольно абстрактных моделях речеобразования или статистических марковских моделях. В частности, утверждается, что системы цифровой обработки речевых сигналов, построенные на таких принципах, будут устойчивы по отношению к мешающему шуму и дикторам.

Одним из наиболее специфических механизмов слуха является эффект маскировки в слуховой системе, который во многом определяет свойства помехоустойчивости, присущие слуху.

Поскольку традиционные методы фильтрации типа вычитания спектров или оптимального среднеквадратического оценивания сопровождаются наличием постпроцессорных искажений сигнала, основным направлением использования моделей эффекта маскировки является их использование в качестве дополнительных блоков, имитирующих маскировку слабых сигналов более сильными в критических полосах слуха.

Метода повышения разборчивости зашумленной речи основан на том, что речевой сигнал сначала подвергался высокочастотной фильтрации таким образом, чтобы ослабить первую форманту, тем самым, повышая удельный вес высших формант в спектре речевого сигнала. Далее отфильтрованный речевой сигнал подвергается клиппированию. Авторы утверждают, что операция клиппирования увеличивает амплитуду речевой волны на участках, которые соответствуют важным для восприятия согласным по отношению к амплитуде гласных звуков.

Одним из часто применяемых на практике методов обработки зашумленных речевых сигналов является фильтрация этих сигналов полосовыми или режекторными фильтрами. Если спектральные характеристики шума известны заранее, то этот метод скорее можно отнести к классу методов, использующих априорные сведения о спектральных характеристиках шума.

Потенциально весьма многообещающие результаты получены при испытаниях систем анализа и обработки зашумленных речевых сигналов, построенных на представлении речевых сигналов с помощью волновых функциях-вейвлетах.

Волновой (вейвлетный, wavelet) анализ речи широко применяется при анализе и обработке речевых сигналов последние 10-15 лет.

Семейство волновых функций описывается соотношением:

,

где  обозначает время, параметры  и  регулируют перенос (трансляцию) по времени и сжатие / растяжение.

Так же как и в других базисах, эти ортогональные функции используются для декомпозиции речевого сигнала в сумму элементарных сигналов.

Основная идея предложенного метода фильтрации навеяна свойствами робастности по отношению к помехам, которыми обладает человеческая аудиторная система и заключается в моделировании уже упомянутого эффекта маскировки, когда слуховая система суммирует сигналы в критических полосках и, при одновременном присутствии двух сигналов разного уровня, сигнал с более высоким уровнем подавляет сигнал меньшего уровня.

Характеристики метода оценивались субъективными методами, посредством прослушивания записей исходного и обработанного сигнала. При этом аудитор постепенно добавлял к обработанному речевому сигналу шум до тех пор, пока персептивно получаемый сигнал не становился таким же шумным как и исходный зашумленный сигнал. Уровень шума, который необходимо было добавить к обработанному сигналу (чтобы получить такой же по качеству сигнал, как и до обработки), являлся мерой качества обработки.

Повышение качества и комфортности звучания очищенного сигнала вовсе не означает улучшения его разборчивости. Одной из причин такого положения является то, что большинство из предложенных методов ориентированы на подавление шума, а не на выделение полезного речевого сигнала, что не одно и то же. Как результат, подавление шума в переходных сегментах речи и низкочастотных полосах, незначительно и не сопровождается улучшением разборчивости.

**Заключение**

Обзор методов повышения качества и разборчивости зашумленных речевых сигналов показывает, что существует много различных подходов к обработке зашумленной речи. Такое разнообразие методов обусловлено как важностью проблемы так и отсутствием достаточно надежных методов ее решения.

Объективное сравнение этих методов и выбор наиболее приемлемых сделать весьма затруднительно, так как перед системами коррекции речевых сигналов ставятся различные задачи. Например, можно в качестве главного критерия использовать повышение разборчивости речи, допуская при этом возможность искажений в тембре голоса или появление артефактов в виде структурированного шума. Можно поставить целью понижение утомляемости аудитора или сохранение натуральности голоса диктора, что достигается в основном за счет повышения качества речевого сигнала. Наконец, могут быть известны заранее важные априорные сведения, например тип или параметры шума, характеристики голоса диктора, наконец, гипотезы о произносимом тексте, что также может определяющим образом повлиять на выбор метода фильтрации.

Важно отметить, что универсальных методов обработки, которые одинаково хорошо боролись бы с существенно нестационарными и стационарными, аддитивными и мультипликативными шумами, существенно повышали бы качество и одновременно разборчивость речи, сейчас нет, и возможно не будет. Как типичная (за редкими, указанными в обзоре исключениями, наблюдается обратная тенденция: если сравнивать системы обработки зашумленной речи по двум показателям - повышению качества звучания речевых сигналов и повышению разборчивости, то системы, повышающие качество и натуральность звучания, скорее всего снижают разборчивость и наоборот, повышение разборчивости приводит к понижению качества и натуральности звучания. Поэтому, многие из названных методов фильтрации нужно рассматривать как взаимодополняющие.

Список использованной литературы

1. Журавлев Ю.И. Цифровая фильтрация зашумленных речевых сигналов. М.: ВЦ РАН, 1998.

2. Шелухин О.И., Лукьянцев Н.Ф. Цифровая обработка и передача речи. М.: Радио и связь, 2000.

. Рабинер Л.Р., Шафер Р.В. Цифровая обработка речевых сигналов. М.: Радио и связь, 2001.

. Золотарев В.И. Методы и аппаратура адаптивной фильтрации речевого сигнала. 2008.

. Назаров М.В., Прохоров Ю.Н. Методы цифровой обработки и передачи речевых сигналов. М.: Радио и связь, 2005.

. Овчинникова О.П. Повышение разборчивости речи путем цифровой фильтрации. М., 1997.

. Курицын С.А. Адаптивные методы обработки сигналов в цифровых и аналоговых системах передачи: Учебное пособие. СПб.: СПбГУТ, 2004.