

Программный модуль шумоочистки и повышения разборчивости фонограмм низкого качества

Лобанова М.А.,

Сазанов В.В.

Описание проблемы

В последние годы одним из основных направлений нашей деятельности является создание аппаратно-программных комплексов (АПК) и автоматизированных рабочих мест (АРМ) для регистрации и обработки речевой информации, поступающей по различным каналам связи, и для автоматизации производства фоноскопических и лингвистических экспертиз. В результате данных работ были созданы АПК «Ассамблея» (для регистрации и обработки речевой информации), АПК «Сапфир» (для производства фоноскопических экспертиз) и АРМ «Изумруд» (для производства лингвистических экспертиз). В настоящее время АПК «Ассамблея», АПК «Сапфир» и АРМ «Изумруд» приняты на вооружение и используются в подразделениях БСТМ и ЭКЦ МВД России.

Одной из задач, которая ставится перед экспертом при обработке и анализе речевых сигналов или производстве фоноскопической или лингвистической экспертизы, является установление дословного содержания фонограммы. Фонограммы сигналов, поступающих по каналам связи или записанных в реальных условиях, довольно часто имеют низкое качество. Плохое качество фонограммы выражается в низкой разборчивости речевого сигнала и/или в наличии в сигнале шумов, затрудняющих прослушивание фонограммы, понижающих разборчивость речи.

Для установления дословного содержания фонограмм низкого качества оператору (эксперту) требуются специальные средства обработки фонограммы, позволяющие повысить разборчивость (провести шумоочистку) речевого сигнала. Иногда без применения таких специальных средств установление дословного содержания фонограммы является практически невозможным.

Описываемый в данном докладе «Программный модуль шумоочистки и повышения разборчивости» был разработан для использования экспертами при установлении дословного содержания речевых фонограмм низкого качества.

Используемый подход к созданию программного модуля шумоочистки и повышения разборчивости

Условно речевые сигналы низкого качества можно разделить на две группы:

1. Речевой сигнал имеет достаточную разборчивость для установления дословного содержания. Низкое качество сигнала проявляется в наличии в нем шумов, понижающих комфортность звучания и вызывающих утомляемость у оператора.

2. Речевой сигнал имеет низкую разборчивость, сильно затрудняющую установление дословного содержания. Низкая разборчивость может быть обусловлена различными причинами. Например, неудачное расположение микрофона записывающего устройства по отношению к источнику речевого сигнала; наличие шумов, имеющих сложную спектрально-временную структуру; искажения сигнала в канале связи.

В зависимости от того, к какой группе (к 1-й или ко 2-й) относится обрабатываемый звуковой сигнал, задача шумоочистки и повышения разборчивости может решаться различными путями.

Для сигналов 1-й группы наиболее удобным для использования средством является возможность применения ко всему сигналу некоторого оптимального универсального метода обработки (фильтрации) сигнала. Критерием оптимальности (эффективности) метода является повышение комфортности звучания сигнала, что часто количественно можно оценить, как понижение уровня шума по отношению к уровню речевого сигнала (повышение «отношения сигнал/шум»). Обычно для сигналов 1-й группы удается найти такой оптимальный метод. Этим оптимальным методом может быть, например, обычный эквалайзер или адаптивный фильтр. Иногда после выбора оптимального метода сигнал фонограммы фильтруется в новый файл, который затем используется для установления дословного содержания.

Следует отметить, что работу с сигналами 1-й группы можно разделить между несколькими (двумя) операторами. Первый оператор выбирает оптимальный метод фильтрации сигнала и готовит файл, содержащий сигнал лучшего по сравнению с исходным сигналом качества. Дальнейшую работу по установлению дословного содержания выполняет другой оператор.

Повышение разборчивости сигналов 2-й группы является значительно более сложной задачей, для решения которой нечасто удается найти универсальный метод, после применения которого сигнал становится настолько «хорошим», что установление его дословного содержания больше не вызывает проблем у оператора.

С нашей точки зрения, существующие в настоящее время методы шумоочистки и повышения разборчивости одноканальных сигналов достигли некоего предела в эффективности своей работы. Так, например, широко известный метод, носящий название «спектральное вычитание», несмотря на большое количество модификаций, создаваемых различными разработчиками, далеко не всегда гарантирует выделение речевого сигнала из зашумленного сигнала и повышение его разборчивости.

Начиная с последнего десятилетия 20 века, в мире активно развивается новое научное направление, моделирующее принципы, используемые слуховой системой человека для выделения из сложного звука (представляющего собой смесь от различных источников звуков) целевого речевого сигнала. Данное направление получило название CASA (“Computational Auditory Scene Analysis” — «слуховой анализ сцен»). Математические модели CASA анализируют и обрабатывают сигнал, используя иную систему критериев и признаков, чем это делают «традиционные методы шумоочистки». Но и

эти новые методы не могут безошибочно, без потерь, выделить речевой сигнал и обеспечить его разборчивость.

Несмотря на достигнутый предел в разработке алгоритмов шумоочистки для одноканальных файлов, можно создать инструментарий, повышающий эффективность использования данных алгоритмов, если использовать еще один ресурс, который тяжело переоценить — это слух и интеллект человека. Как из научных экспериментов, так и из нашего повседневного опыта, известны способности человека (его слуховой системы) адаптироваться под шумы и настраиваться на целевой сигнал. После некоторой тренировки человек начинает слышать изначально малоразборчивые речевые сигналы.

От инструментария, предназначенного для установления дословного содержания фонограмм низкого качества, требуется, чтобы в нем были реализованы существующие эффективные методы шумоочистки, а также, чтобы данный инструментарий давал возможность оператору настроить («натренировать») свой слух под конкретный речевой сигнал, выделить в сигнале важные для понимания сказанного акустические, просодические (интонационные) и лингвистические «ключи».

Наш опыт показал, что одним из эффективных методов настройки слуха на целевой сигнал является многократное прослушивание (в режиме «кольцо») выделенного участка сигнала (например, реплики), к которому оперативно применяются различные методы фильтрации, с возможностью изменения границ этого участка. В результате такого многократного прослушивания исходного и преобразованного различными методами сигнала происходит настройка слуховой системы оператора на целевой речевой сигнал. Через некоторое время оператор начинает слышать и разбирать сказанное в исходной сильно зашумлённой фонограмме.

При создании программного модуля шумоочистки и повышения разборчивости мы выдвигали к нему следующие требования:

- Оператор должен иметь возможность оперативно, «в реальном времени», прослушивать сигнал, обработанный различными методами шумоочистки и повышения разборчивости.
- Изменение используемых методов шумоочистки или их параметров должно происходить «в реальном времени» с минимальной задержкой реагирования (менее 100 мсек). Большие задержки (более 0,5 секунд, а тем более, больше нескольких секунд) мешают оператору в выборе оптимальных параметров фильтрации и в восприятии сигнала.
- Одним из регулируемых параметров для каждого из методов фильтрации должен быть параметр, определяющий степень фильтрации (степень «искажения») сигнала. Оператор должен иметь возможность оперативно изменять данный параметр, влияя на звучание сигнала. Время реакции программы на изменения значений данного параметра должно быть минимальным (практически незаметным).
- Оператор должен иметь возможность оперативно изменять границы прослушиваемого участка сигнала для наиболее полного использования присутствующей в сигнале просодической информации.
- Оператор должен иметь возможность видеть изображение осциллограммы сигнала и «позицию плеера» в прослушиваемом сигнале для использования информации, заключенной в форме звуковой волны сигнала.

- Оператор должен иметь возможность видеть спектрограмму сигнала с «позицией плеера» в прослушиваемом сигнале для использования спектрально-временных характеристик при анализе речевого сигнала.
- Возможность использования видеоматериалов (в случае их наличия) для синхронного прослушивания/просмотра звукового файла (извлеченного аудиотрека) и видеотрека (видеофайла). Известно, что использование слушателем видеоинформации (динамики работы голосовых органов) в некоторых случаях может значительно повысить разборчивость речи.
- Для работы с сигналами 1-й группы (речевые сигналы с достаточной разборчивостью для установления дословного содержания, но не комфортные для прослушивания из-за наличия в них шумов) должны быть предусмотрены оперативные методы повышения комфортности звучания, которые могли бы использовать операторы, не обладающие специальными знаниями или навыками (например, эксперты-лингвисты).

Описание функциональных возможностей «Программного модуля шумоочистки и повышения разборчивости»

На рис. 1 показан внешний вид окна «Программного модуля шумоочистки и повышения разборчивости».

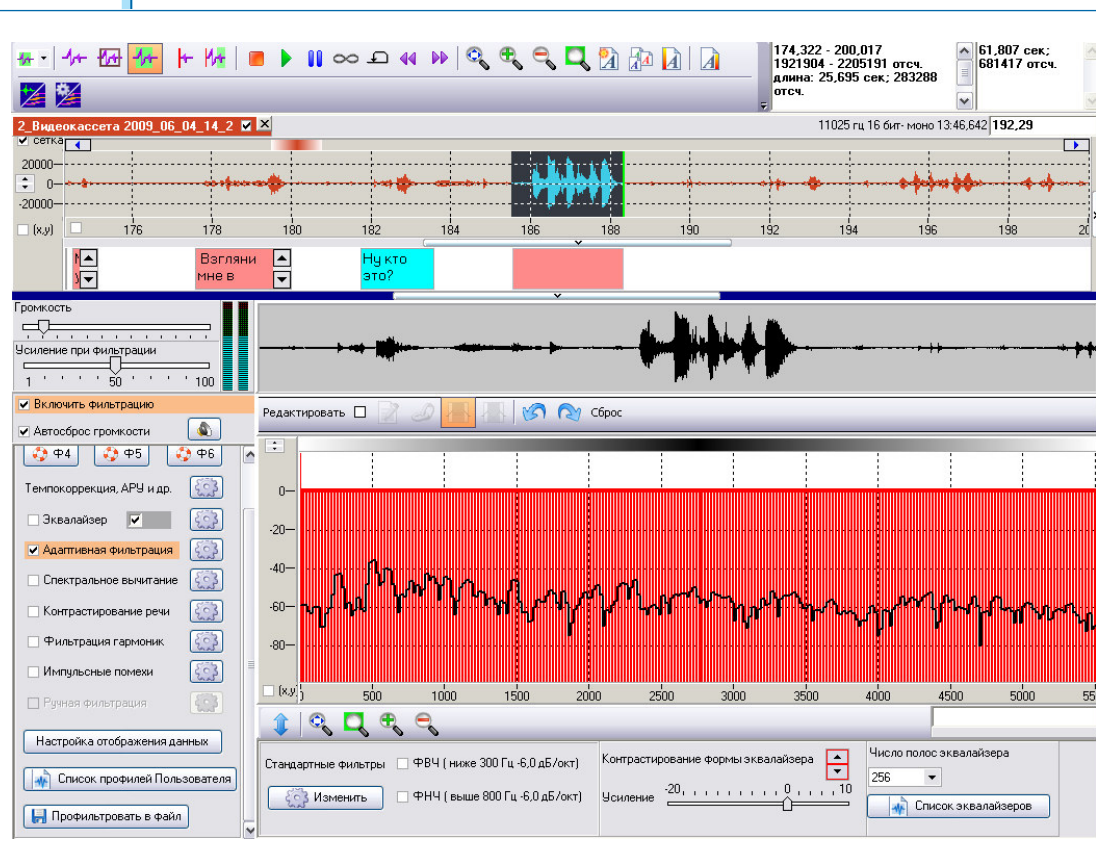


Рис. 1. Вид окна «Программного модуля шумоочистки и повышения разборчивости»

Ниже перечислены основные сведения о «Программном модуле шумоочистки и повышения разборчивости».

1. В программном модуле реализованы следующие методы преобразования (фильтрации) сигнала:

- Эквалайзер.
- Адаптивная фильтрация (одноканальная и двухканальная).
- Спектральное вычитание.
- Контрастирование речи.
- Фильтрация гармоник электросети.
- Импульсные помехи (фильтрация импульсных помех).

Дополнительно имеется возможность прослушивать сигнал:

- с изменением темпа с искажением тембра;
- с изменением темпа без искажения тембра;
- в режиме «псевдостерео»;
- в режиме «АРУ» (автоматическая регулировка усиления).

2. Оператор имеет возможность «в реальном времени» (при прослушивании сигнала):

- включать различные методы фильтрации в любой комбинации;
- изменять значения параметров методов фильтрации.

3. В программном модуле имеются возможности:

- сохранять в «файлы профилей» выбранную комбинацию методов фильтрации и их параметры;
- оперативно открывать «файл профиля», в том числе, не останавливая проигрывание;
- сохранять в «файлы эквалайзера» амплитудно-частотную характеристику (АЧХ) эквалайзера;
- оперативно открывать файл с АЧХ эквалайзера, в том числе, не останавливая проигрывание;
- проводить фильтрацию всего сигнала или его выбранного участка в новый звуковой файл.

4. Также оператору предлагаются 6 оперативных эффективных (предварительно настроенных) методов фильтрации (повышения комфортности звучания). Для выбора данных методов реализован простейший пользовательский интерфейс, что иногда позволяет оператору, не обладающему специальными знаниями в области шумоочистки сигнала, оперативно улучшить качество его звучания.

5. В модуле реализовано два метода автоматизированного построения эквалайзера – по инверсному спектру шума и по спектру речи. Участок сигнала, содержащий шум или содержащий речевой сигнал, задается оператором путем выделения соответствующего участка на осциллограмме.

6. Задание (редактирование) АЧХ эквалайзера может проводиться различными способами:

- Включение фильтров нижних и высоких частот с возможностью задания их параметров (частоты среза и наклона АЧХ);

- «Рисование» АЧХ в окне отображения эквалайзера с использованием указателя «мыши»;
 - Контрастирование (сглаживание) заданной формы эквалайзера;
 - Синхронная корректировка уровней выбранных соседних полос эквалайзера с помощью указателя «мыши» путем их параллельного сдвига или с использованием параболической интерполяции.
 - АЧХ эквалайзера редактируется в том же окне, где отображаются спектр прослушиваемого сигнала, спектр участка, заданного как образец шума, спектр вычисленного инверсного шума, спектр участка сигнала, заданного как образец речи. Благодаря этому при корректировке формы эквалайзера оператор может соотносить выбранную АЧХ эквалайзера со спектром сигнала.
7. В модуле реализован метод автоматизированного вычисления оценки спектра шума (по заданному образцу шума, выделенному на осциллограмме участку сигнала) для метода спектрального вычитания.
8. Метод контрастирования речи использует известные научные данные о восприятии речевого сигнала слуховой системой человека в затрудненных условиях. В данную группу методов включены методы, подчеркивающие формантную структуру сигнала и выраженность гармоник основного тона голоса диктора.
9. Методы фильтрации сигнала «фильтрация импульсных помех», «спектральное вычитание», «контрастирование речи» имеют в качестве одного из регулируемых параметров параметр, определяющий степень подавления шума. Регулируя данный параметр, оператор имеет возможность задавать степень искажения сигнала, подстраивая каждый из методов под конкретный сигнал. При изменении степени подавления шума меняется звучание преобразованного сигнала и создаются благоприятные условия для настраивания слуха оператора на сигнал.
10. Сигнал исследуемой фонограммы отображается в окне осциллограммы, в котором имеется возможность:
- выбирать участок сигнала для прослушивания с возможностью оперативного изменения границ прослушиваемого сигнала (не останавливая проигрывание);
 - выбирать участок сигнала для фильтрации в новый звуковой файл;
 - просматривать отображение сигнала в виде спектрограммы (с возможностью выбора параметров вычисления спектрограммы), используя спектрально-временное представление сигнала для выбора оптимального метода фильтрации сигнала;
 - открывать в окне осциллограммы любое количество звуковых файлов. Данная возможность удобна для открытия одновременно нескольких звуковых файлов, содержащих преобразованный различными методами сигнал исходной фонограммы;

- оперативно изменять прослушиваемый звуковой файл (не останавливая проигрывание), переключаясь между открытыми в окне звуковыми файлами.

11. В окне осциллограммы сигнала синхронно во времени отображается дословное содержание, представленное в виде атрибутированных по дикторам реплик. При работе со сложной фонограммой, устанавливая дословное содержание реплики, имеется возможность настроить свой слух на речевой сигнал путем многократного прослушивания нужной реплики с использованием «в реальном времени» различных методов фильтрации сигнала и возможностью изменения границ прослушиваемого участка сигнала.

12. Имеется возможность открывать в модуле видеофайлы, извлекать из них аудиотрек, синхронно во времени прослушивать сигнал звукового файла и просматривать видеотрек видеофайла.

Реализованные в представленном программном модуле подходы к проведению шумочистки и повышению разборчивости при установлении дословного содержания нашли успешное практическое применение при работе с фонограммами низкого качества и речевыми сигналами, имеющими низкую разборчивость.

Сведения об авторах

**Лобанова М.А.,
Сазанов В.В.**

ЗАО НПФ «ИСТА-СИСТЕМС», Санкт-Петербург, mal@ista.ru