*Н.В. Бобров*

**Исследование природы искажений, вносимых в речевой сигнал при его преобразовании методом амплитудной фильтрации**

В настоящем докладе представлены предварительные результаты исследования характера искажений, появляющихся в речевом сигнале после оптимизации для сжатия методом амплитудной фильтрации.

Метод амплитудной фильтрации (АФ) представляет собой исключительно экономную в отношении системных ресурсов альтернативу методам спектральной фильтрации – ставшим традиционными универсальным методам обработки речевого сигнала, основанным на вычислении спектров (в частности, при помощи преобразования Фурье).

Принцип действия данного метода заключается в выборочной модификации пиков речевой волны в зависимости от их магнитуды. Механизм этого преобразования был описан автором в предыдущих докладах.

Тогда в ходе экспериментов было установлено, что удаление пиков малой магнитуды приводит к упрощению формы речевой волны (см. рис. 1) без существенной потери разборчивости. Это свойство АФ позволяет использовать ее как средство оптимизации речевой волны для последующего сжатия [Бобров 2008]. В то же время удаление пиков большой магнитуды приводит к получению речевого сигнала минимальной (но не нулевой) разборчивости, аналогичного зашумленной шепотной речи. Это свойство АФ позволяет использовать ее в целях защиты речевой информации (ограничение распространения речевых волн, маскировка) [Бобров 2009]. Представление речевой волны в виде суммы ломаных линий с убывающей средней амплитудой, получающееся на одном из технологических этапов АФ, позволяет получить некоторую информацию о частотном спектре речевого сигнала путем измерения расстояний между пиками, что открывает возможность использования АФ для грубой оценки спектра акустического сигнала в условиях жестких ограничений на вычислительные ресурсы [Бобров 2010].

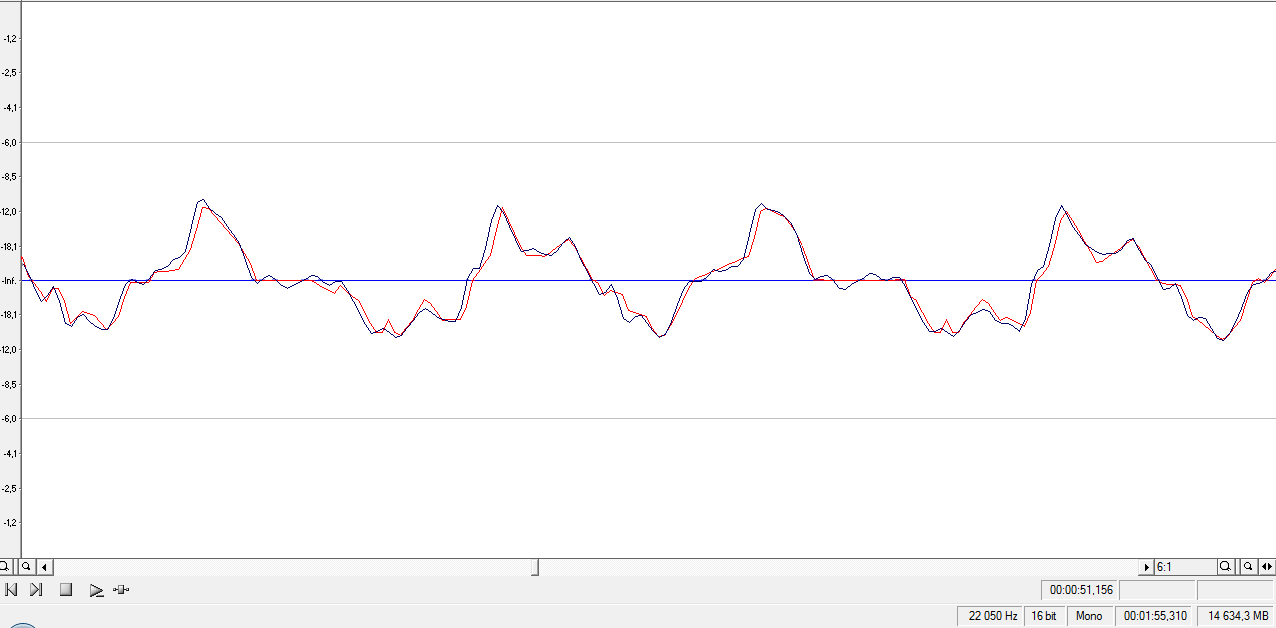


Рис. 1. Фрагмент речевой волны до и после АФ (гласный *а* в слове «так», диктор – мужчина). Синим (более темным при ч/б печати) показан исходный сигнал, красным (более светлым) – сигнал после АФ.

Нынешнее исследование имело целью определить характер искажений, вносимых АФ в речевой сигнал при оптимизации последнего для сжатия.

Для этого были проанализированы частотные спектры фонограмм, содержащих различные звуки речи, в исходном состоянии и после АФ, а также спектры разностных сигналов, полученных вычитанием измененных фонограмм из исходных. Для получения спектров была использована соответствующая функция аудиоредактора Sony Sound Forge 7.0. Во всех случаях применялись одни и те же параметры расчета спектров: длина окна анализа – 2048 отсчетов (при частоте дискретизации 22050 Гц), величина перекрытия окон – 75%, взвешивающее окно Хеннинга.

Результаты сопоставления спектров показали следующее.

1. Спектр речевого сигнала после АФ становится в целом более гладким (см. пример на рис. 2–3).

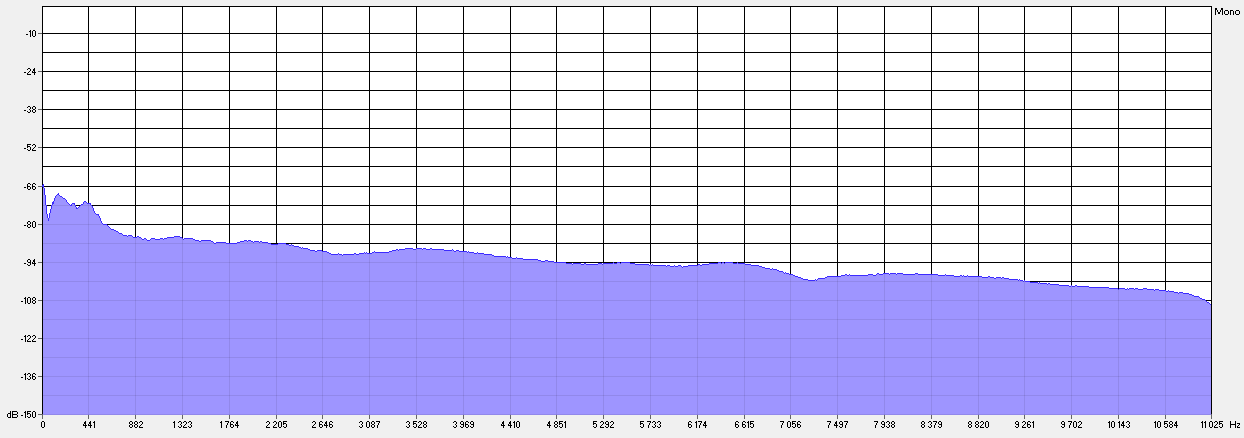


Рис. 2. Спектр исходной фонограммы спонтанного диалога (дикторы – мужчины). Длительность 1 мин 55 сек.



Рис. 3. Спектр той же фонограммы после АФ.

2. В наибольшей степени сглаживание затрагивает высокочастотную область сигнала (>6–7 кГц), не несущую важной для восприятия речи информации (см. рис. 4–5).

3. Спектры разностных сигналов представляют собой смесь белого и розового шума, причем розовая составляющая присутствует в низкочастотной области (<6–7 кГц; см. пример на рис. 6).

4. В некоторых случаях (это заметно на рис. 6) розовая составляющая разностного сигнала несет в себе явные (различимые также и на слух) следы исходного речевого сигнала. Это относится, прежде всего, к гласным звукам и объясняется регулярностью увеличения магнитуды удаляемых пиков, которая в этом случае связана с периодами основного тона и наиболее сильных резонансов. В то же время спектр розовой составляющей разностных сигналов лишен большинства особенностей, присутствующих в спектре исходного сигнала.

Вышеперечисленные наблюдения позволяют сделать следующие выводы:

1. Наиболее общей характеристикой искажений, вносимых АФ в спектр речевого сигнала, является увеличение его энтропии.

2. Увеличение энтропии спектра речевого сигнала в результате АФ происходит неравномерно, причем его степень находится в обратной зависимости от частоты.

3. Искажения, вносимые АФ, не могут быть описаны как добавление к исходному сигналу некоторой независимой шумовой составляющей (например, белого или розового шума).

В задачи дальнейших исследований входит поиск зависимости особенностей вносимых искажений (соотношения белой и розовой составляющих шума, крутизны спада энергии в зависимости от частоты, степени присутствия следов исходного сигнала и т.д.) от параметров АФ.

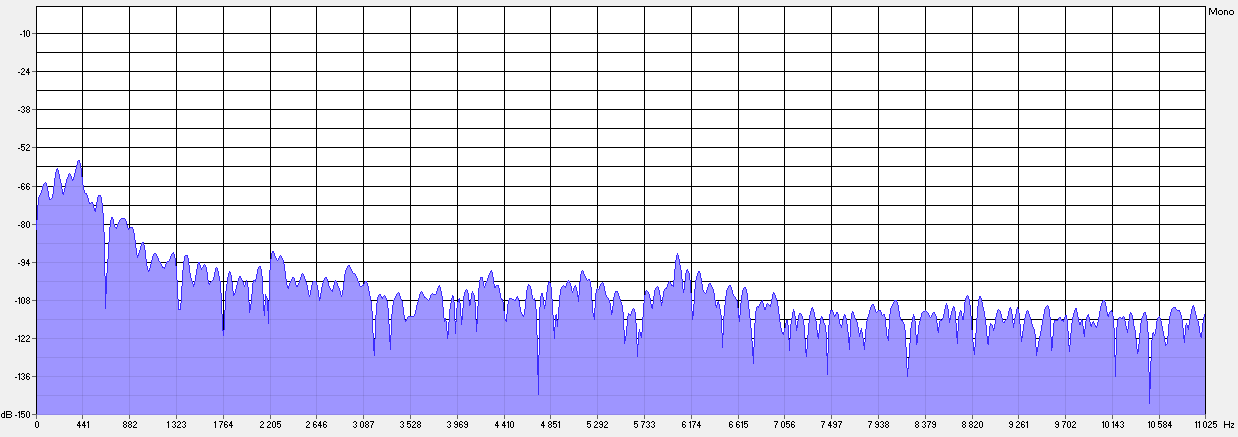


Рис. 4. Спектр фрагмента речевой фонограммы (квазистационарный участок гласного *у* в слове «лучше», мужской голос) в исходном состоянии.

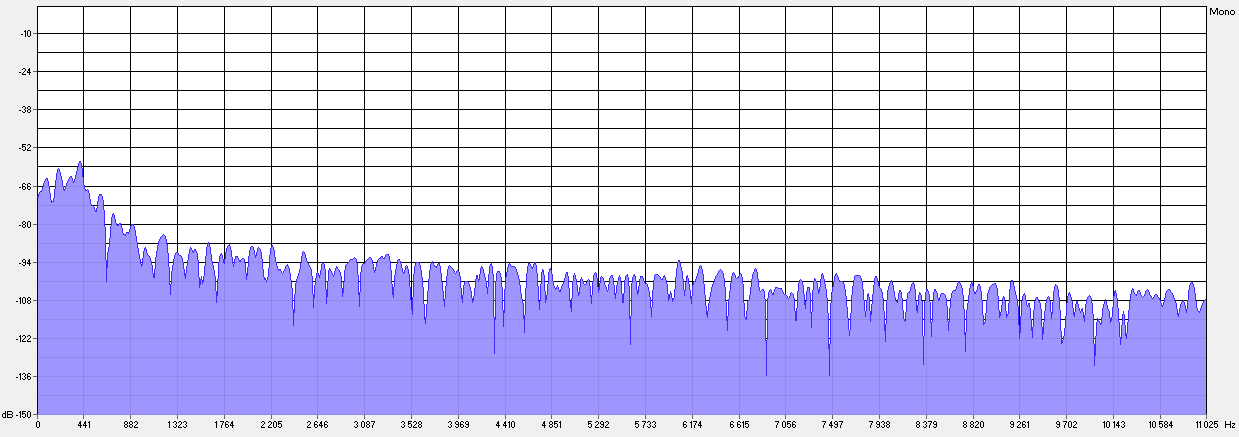


Рис. 5. Спектр того же фрагмента после АФ.

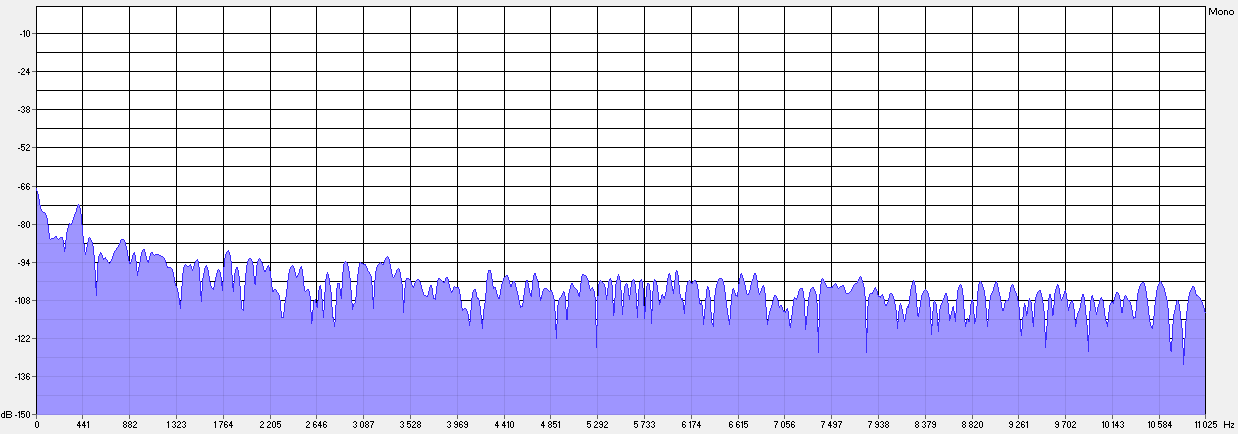


Рис. 6. Спектр сигнала, полученного вычитанием фонограммы после АФ из исходной, для этого же фрагмента.

**Список использованных источников**

1. Бобров Н.В. Преобразование речевого сигнала методом амплитудной фильтрации: возможности и перспективы. // Акустика речи. Медицинская и биологическая акустика. Сборник трудов XXII сессии Российского акустического общества и Сессии Научного совета РАН по акустике. Т. 3. – М.: ГЕОС, 2010.

2. Бобров Н.В. Возможности преобразования речевого сигнала методом амплитудной фильтрации. // Материалы Всероссийской конференции с элементами научной школы для молодежи «Проведение научных исследований в области обработки, хранения, передачи и защиты информации». – Ульяновск, 2009.

3. Бобров Н.В. Уменьшение информационной избыточности речевого сигнала методом амплитудной фильтрации. // Тезисы IX Всероссийской конференции молодых ученых по математическому моделированию и информационным технологиям. – Кемерово, 2008.