



Преобразование речевого сигнала методом амплитудной фильтрации: возможности и перспективы

Бобров Н.В.

Московский государственный лингвистический университет.
119034 Москва, ул. Остоженка, д. 38. E-mail: Arctangent@yandex.ru

В работе изложены результаты исследования возможностей преобразования речевого сигнала методом амплитудной фильтрации, намечены перспективы его применения. Данный метод отличается исключительно малой ресурсоемкостью и поэтому может использоваться в тех случаях, когда критичными оказываются не только быстродействие и объём памяти, но также размеры и энергетическая автономность устройства.

Лавинообразный рост производительности вычислительных машин во многом определил ход развития технологий анализа и преобразования речевого сигнала в последние десятилетия, создав крен в сторону ресурсоёмких методов, обеспечивающих более высокое качество результатов или требующих меньших вложений на этапе разработки. Вместе с тем, проблема создания алгоритмов преобразования речевого сигнала, не предъявляющих высоких требований к вычислительным ресурсам, сохраняет свою актуальность. Это становится очевидным при рассмотрении постановки задачи кодирования/декодирования речевого сигнала для сверхмалых устройств связи.

Метод амплитудной фильтрации (АФ), предлагаемый автором настоящей работы, предоставляет широкий диапазон возможностей для анализа и преобразования речевого сигнала, в том числе в режиме реального времени, при исключительно малой ресурсоемкости.

Принцип метода АФ заключается в разложении исходной дискретизированной речевой волны в сумму более простых ломаных линий с убывающей средней амплитудой (отсюда название), некоторые из которых (имеющие наименьшую амплитуду) могут быть отброшены без существенной потери разборчивости речи. Ход этого преобразования показан на рис. 1.

В результате суммирования оставшихся ломаных получается ломаная линия, похожая на исходную речевую волну, но имеющая более простую форму, т.е. меньшее количество вершин (рис. 2).

Одна из возможных технологических процедур, позволяющих осуществить такое преобразование, была реализована автором в виде программной библиотеки на языке FreePascal 2.0.4 и описана в работе [1].

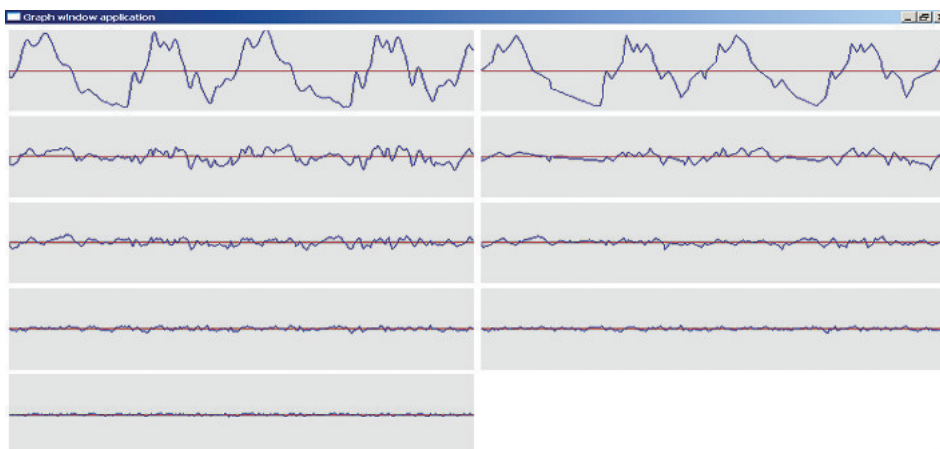


Рис 1. Преобразование речевого сигнала методом АФ. Гласный [и] (фраза «Посмотри сюда», мужской голос). Слева сверху — исходный сигнал. Справа — упрощённые ломаные, полученные в результате преобразования. Слева под исходным сигналом — остаточные сигналы, получаемые на каждом шаге преобразования. Если обозначить все графики слева a_1, a_2, a_3, \dots , а все графики справа — b_1, b_2, b_3, \dots , то $a_2 = b_1 - a_1$, $a_3 = b_2 - a_2$ и т.д. На каждом последующем шаге остаточный сигнал исполняет роль исходного для получения очередной упрощённой ломаной. Последний из остаточных сигналов, средняя амплитуда которого ниже некоторого заданного порогового значения, отбрасывается

Упрощение формы речевой волны обуславливает улучшение сжимаемости описывающих ее числовых данных. Изначально метод АФ разрабатывался именно с целью создания нового нересурсоёмкого алгоритма сжатия речевой волны, в связи с чем именно этот эффект АФ был исследован первым. Эксперименты, проведённые автором в 2008 году, показали, что сжимаемость речевой волны после преобразования методом АФ улучшается в 1,3–1,5 раза.

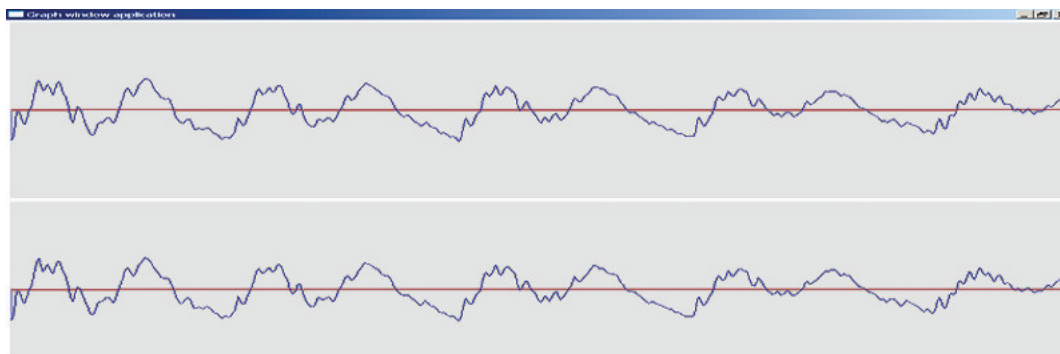


Рис. 2. Исходный сигнал (вверху) и сигнал, полученный в результате сложения упрощённых ломаных

Изучение спектральных характеристик фонограмм после АФ позволило определить, какая именно информация в частотной области оказывается отброшенной. В частности, для гласных таковой оказалась высокочастотная область ($F > 4,5$ кГц), а для шипящих и свистящих фрикативных — главным образом, низкие ($F < 2,5$ кГц) и очень высокие ($F > 10$ кГц) частоты. На спектральных срезах упомянутых звуков в указанных областях наблюдалось сглаживание огибающей вплоть до образования плато, т.е. интервалов белого шума (см. рис. 3). Следует отметить, что эта особенность работы метода АФ открывает некоторые возможности для автоматического распознавания по крайней мере классов звуков в потоке речи.

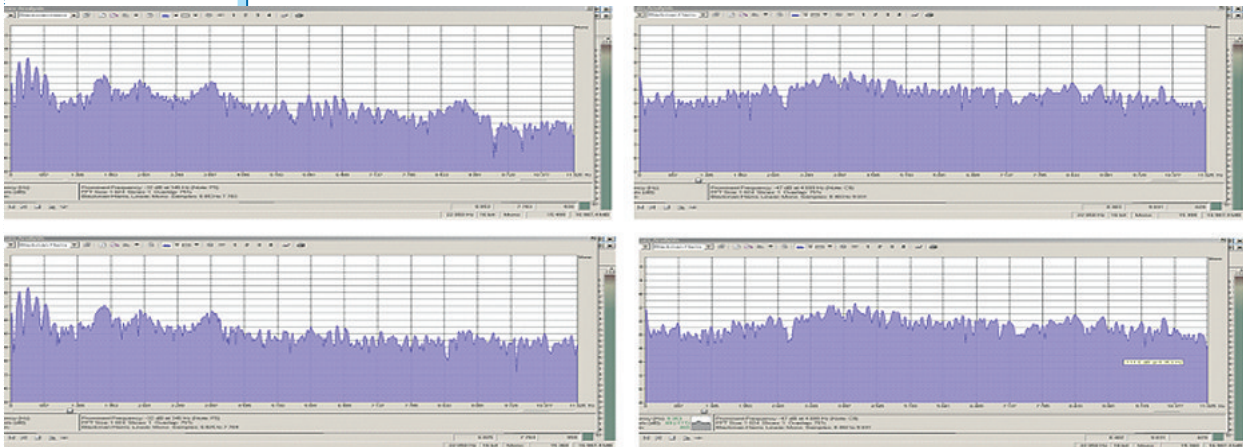
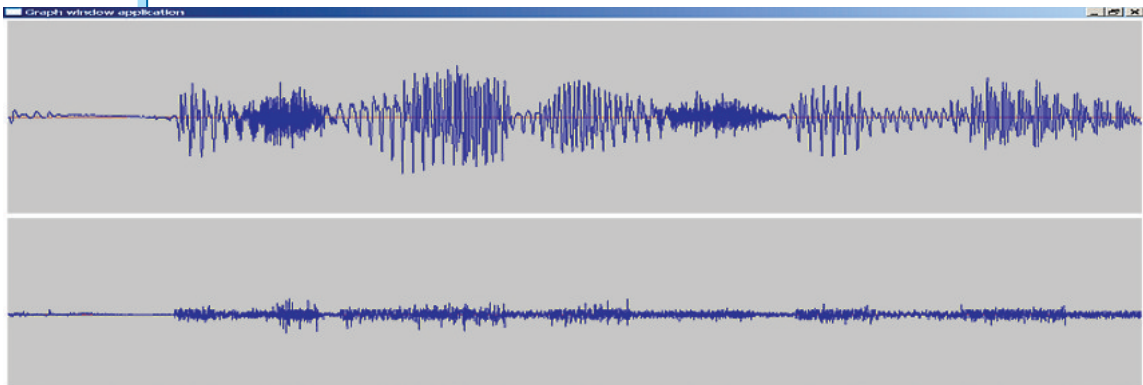


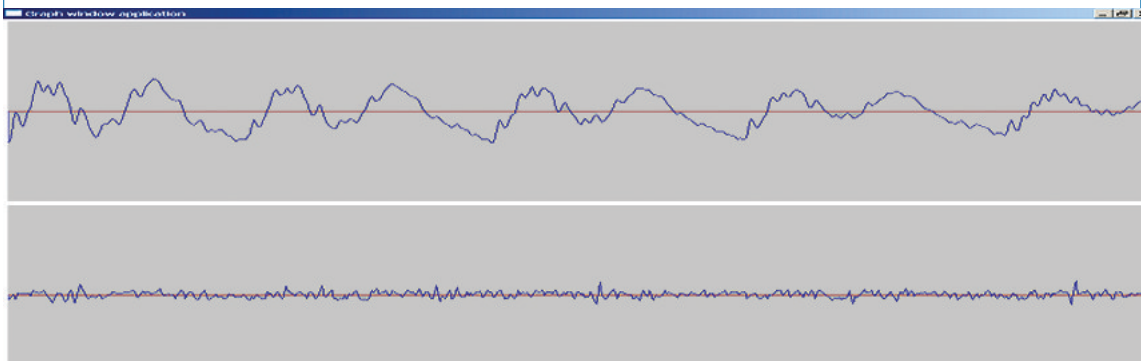
Рис. 3. Спектры гласного [и] (слева) и согласного [с] (справа) до (вверху) и после (внизу) амплитудной фильтрации. Изменения гораздо более заметны на гласном в силу изначального преобладания в его спектре гармонической составляющей по сравнению с шумовой

Помимо оптимизации формы речевой волны для сжатия, АФ открывает некоторые возможности для контролируемого снижения разборчивости речи (например, в целях предотвращения подслушивания или перехвата специальной распознающей аппаратурой, в том числе в каналах связи), а также для борьбы со скрытым несанкционированным встраиванием в речевой сигнал посторонней информации.

Контролируемое уменьшение разборчивости речи достигается путем выборочной обработки ломанных линий, полученных при разложении речевой волны методом АФ, и последующего ресинтеза. Проведенные в ноябре 2009 г. пилотные эксперименты показали, что результирующий сигнал, напоминающий на слух и по виду осциллограммы сильно зашумленную шепотную речь (см. рис. 4), оказывается достаточно разборчивым для понимания смысла сказанного на небольшом расстоянии (0,5–1 м от динамиков или в наушниках) и практически неразборчивым на большем расстоянии и, тем более, за преградой. В дальнейшем планируется получить точные количественные оценки описанного эффекта, позволяющие создать на его основе методику защиты речевой информации.



(a)



(б)

Рис. 4. Результат модификации речевого сигнала путём выборочного масштабирования ломаных линий, полученных при разложении исходной волны методом АФ: а — вся фраза («Посмотри сюда»), б — тот же участок, что был показан на предыдущих рисунках. Удивительно, но процентов 10 разборчивости здесь ещё осталось...

Что касается встраиваемой посторонней информации, то неустойчивыми к АФ оказываются, прежде всего, данные, введённые в младшие биты точек оцифрованной речевой волны, а также данные, размещённые на малозначимых для восприятия участках частотной области. В то же время данные, встроенные в существенные для восприятия медленно изменяющиеся характеристики речи (такие, как, например, частота основного тона вокалических звуков, длительность сегментов), АФ с большой вероятностью не затрагивает.

В 2009 году автором был проведён ряд экспериментов по получению «альтернативных», «амплитудных» спектров речевой волны путем её разложения на ломаные линии с убывающей амплитудой точно таким же образом, как это делалось и при преобразовании методом АФ [2]. Было установлено, что, измеряя расстояние между пиками ломаных линий, можно получить некоторые последовательности чисел, соотносящихся с максимумами мгновенных спектров Фурье, и даже нарисовать «альтернативную» спектрограмму, в своих основных чертах схожую с традиционной (см. рис. 5, 6). Вместе с тем, полученные результаты показали, что между «альтернативными» и «традиционными» спектрами существует немало различий, требующих объяснения. Определение причин этих различий и возможностей использования информации «амплитудных» спектров, является одной из первоочередных задач дальнейших исследований.

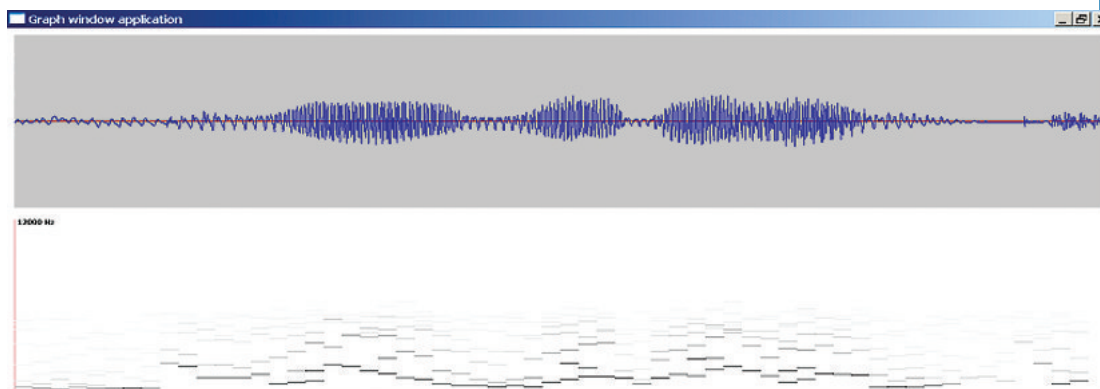


Рис. 5. Спектрограмма фразы «Два варианта» (мужской голос), построенная путем измерения расстояний между вершинами ломаных линий, полученных методом АФ

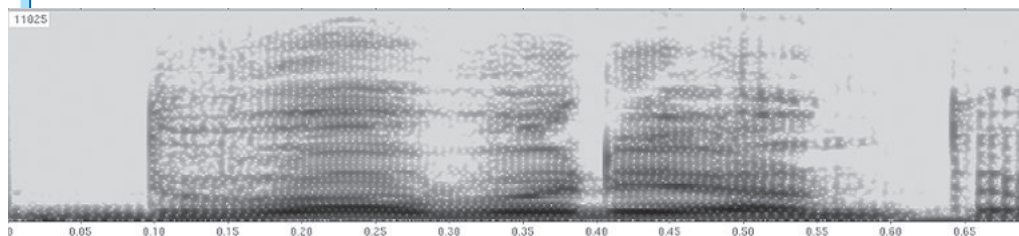


Рис. 6. «Традиционная» спектрограмма той же фразы, полученная при помощи преобразования Фурье

Выводы

На текущий момент можно констатировать, что преобразование речевого сигнала методом АФ позволяет оптимизировать дискретизированную речевую волну для сжатия без существенной потери разборчивости и производить контролируемое уменьшение разборчивости речевого сигнала в целях защиты речевой информации от подслушивания и перехвата с использованием специальной распознающей аппаратуры. АФ на данном этапе может иметь ограниченное применение в целях защиты речевого сигнала от встраивания в него посторонних данных с использованием некоторых простых распространенных методов. В ходе недавних исследований было установлено, что при помощи АФ также возможно получение информации о частотном спектре речевого сигнала, которая обычно извлекается путём использования таких ресурсоёмких методов, как преобразование Фурье, что открывает перспективы применения АФ в качестве экономной (по объёму вычислений) замены ныне используемых методов в ряде приложений реального времени. В задачи будущих исследований входит дальнейшее изучение вышеназванных возможностей, а также создание опытных образцов кодеков и многоцелевого программного инструментария для анализа речевого сигнала методом АФ.

ЛИТЕРАТУРА

1. Бобров Н.В. Уменьшение информационной избыточности речевого сигнала методом амплитудной фильтрации. // Тезисы IX Всероссийской конференции молодых учёных по математическому моделированию и информационным технологиям. Кемерово, 2008.
2. Бобров Н.В. Возможности преобразования речевого сигнала методом амплитудной фильтрации. // Материалы Всероссийской конференции с элементами научной школы для молодёжи «Проведение научных исследований в области обработки, хранения, передачи и защиты информации». Ульяновск, 2009.