

УДК 004.934.2

АЛГОРИТМ ОЧИСТКИ РЕЧЕВОГО СИГНАЛА ОТ СЛОЖНЫХ ПОМЕХ ПУТЕМ ФИЛЬТРАЦИИ В МОДУЛЯЦИОННОЙ ОБЛАСТИ

Азаров И.С., к.т.н., Белорусский государственный университет информатики и радиоэлектроники, e-mail: azarov@bsuir.by;

Вашкевич М.И., аспирант кафедры Электронных вычислительных средств Белорусского государственного университета информатики и радиоэлектроники, e-mail: vashkevich@bsuir.by;

Лихачев Д.С., Белорусский государственный университет информатики и радиоэлектроники;

Петровский А.А., д.т.н., профессор, зав. кафедрой Электронных вычислительных средств Белорусского государственного университета информатики и радиоэлектроники, e-mail: palex@bsuir.by.

Ключевые слова: речевой сигнал, фильтрация, модуляционная область, шумоподавление, синтез, оценка качества.

Введение

Обработка речи в модуляционной области появилась сравнительно недавно. Идеи, лежащие в ее основе, возникли в результате исследований, показавших важность амплитудных модуляций для восприятия звуков слуховой системой человека [1, 2, 3]. В последнее время проведено большое количество исследований, подтверждающих информационную значимость амплитудных модуляций речевого сигнала [3, 4]. На основании ряда наблюдений сделано предположение, что слуховая система человека распознает текст речевого сообщения при помощи частотно-временного анализа амплитудных модуляций речи. В работах [5, 6, 7] показано, что оценка амплитудно-модулированных тонов может служить методом исследования слухового восприятия. Кроме того, было установлено, что амплитудные модуляции на близких частотах маскируют друг друга. В работе [8] показана еще более тесная связь классических психоакустических принципов частотного и временного маскирования с аналогичными эффектами в модуляционном спектре. Исключительная роль модуляционного спектра в восприятии речи используется для оценки индекса разборчивости STMI (спектрально-временной модуляционный индекс), который определяет ухудшение декодирования спектральных и временных модуляций в условиях агрессивной шумовой обстановки [9].

В терминах модуляционного спектра речевой сигнал имеет целый ряд присущих ему характерных свойств, отличающих его от других сигналов. Одним из таких свойств является ограниченность модуляционного спектра [10], позволяющая подавлять различные акустические помехи путем применения узкополосной фильтрации в модуляционной области [11]. Преимуществом данного подхода является универсальность и простота реализации. В отличие, например, от широко применяемого метода спектрального вычитания шумоподавление происходит автоматически без вычисления спектральной плотности мощности шума, выделения пауз и использования детекторов голосовой активности. Выделение речевого сигнала выполняется на основе объективных

Предлагается способ очистки речевого сигнала от шума, основанный на фильтрации в модуляционной области. Способ является универсальным и позволяет подавлять широкий класс акустических помех. Приводятся результаты экспериментов, показывающие, что предлагаемый способ обеспечивает более высокий коэффициент ослабления шума и меньшую степень деградации полезного речевого сигнала в сравнении с известным алгоритмом на основе RASTA (RelAtive SpecTrA).

признаков, независимых от шумовой обстановки, что делает подход фильтрации в модуляционной области особенно надежным в условиях сильного зашумления. Особо высокое качество шумоподавления обеспечивается для полигармонических помех высокой интенсивности и реверберационных шумов. В работах [11, 12, 13] предложен ряд алгоритмов шумочистки, основанных на этой технике. Одним из основных недостатков данных алгоритмов является искажение речи, которое выражается в появлении артефактов (музыкальные тона), выравнивании спектральных огибающих (отбеливание) и подавлении полезного сигнала вместе с шумом.

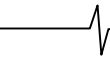
Настоящая работа направлена на поиск нового алгоритма очистки речевого сигнала от сложных помех путем фильтрации в модуляционной области, позволяющего минимизировать перечисленные недостатки.

Определение и особенности модуляционного спектра

Для определения «модуляционного спектра» и «модуляционной области» введем понятие «модуляционной частоты». Если «линейная частота» описывает частоту гармонических составляющих сигнала и используется для его декомпозиции при помощи преобразования Фурье, то «модуляционная частота» вводится как отдельное понятие, имеющее свой физический смысл. Рассмотрим амплитудно-модулированный периодический сигнал с фиксированной частотой ω_c и ограниченным спектром:

$$s(t) = m(t) \cos \omega_c t$$

где модулирующий сигнал $m(t)$ принимает неотрицательные значения и ограничен по частоте таким образом, чтобы обеспечить возможность его полного восстановления из $s(t)$. Тогда модуляционную частоту сигнала $s(t)$ можно получить, используя преобразование Фурье модулирующего сигнала:



$$M(e^{j\omega}) = F\{m(t)\} = \int_{-\infty}^{\infty} m(t)e^{-j\omega t} dt.$$

Модель имеет более простую интерпретацию через узкополосный синусоидальный модулятор:

$$s(t) = (1 + \cos(\omega_m t)) \cos(\omega_c t).$$

Для возможности восстановления модуляционного сигнала частота ω_m должна быть ниже несущей частоты ω_c . Для обеспечения неотрицательных значений модуляционного сигнала используется постоянное смещение. Без потери общности можно предположить, что модуляционный сигнал нормализован и принимает значения в диапазоне [-1, 1].

Степень влияния модуляционных частот хорошо исследована путем субъективной оценки разборчивости фильтрованной в модуляционной области речи [10]. Эксперименты проводились отдельно для модуляционных фильтров низких и высоких частот, что позволило выделить наиболее важную полосу. В результате были сформулированы следующие основные особенности обработки речи модуляционной области:

- амплитудный спектр речевого сигнала может быть ограничен сверху частотой 16 Гц без заметного снижения разборчивости для людей с нормальным слухом;

- амплитудный спектр речевого сигнала может быть ограничен снизу частотой 4 Гц без заметного снижения разборчивости для людей с нормальным слухом;

- слушатели лишь частично могут понимать речь в условиях идеальной акустической обстановки, если амплитудный спектр речевого сигнала обработан фильтром нижних частот с частотой среза 2 Гц либо фильтром верхних частот с частотой среза 32 Гц; при расширении полосы пропускания модуляционных фильтров разборчивость повышается;

- согласные звуки деградируют сильнее гласных при ограничении модуляционного спектра.

В работах [9] и [10] также отмечено, что основные модуляционные частоты, влияющие на разборчивость речи, находятся в диапазоне от 1 до 16 Гц с пиком около 3-5 Гц; кроме того, более 95% модуляционных компонент речевого сигнала сконцентрированы в данном диапазоне. Это предположительно обусловлено количеством слогов, произносимых диктором за одну секунду.

Шумоподавление на основе фильтрации в модуляционной области

На основе приведенных выше положений были предложены два наиболее известных метода модуляционной фильтрации речи для подавления аддитивного шума и реверберации: фильтрация модуляционного спектра – RASTA и перцепционное линейное предсказание – PLP (Perceptual Linear Prediction).

Основная идея данных алгоритмов заключается в том, что модуляционные компоненты сигнала, не входящие в важный для восприятия диапазон, можно удалить без существенной потери разборчивости. В результате шумовые составляющие сигнала (стохастические, полигармонические помехи и реверберация) будут подавлены, поскольку существенная часть их энергии находится

за пределами речевой полосы модуляционного спектра.

Алгоритм обработки речи на основе RASTA-PLP состоит из следующих основных шагов [12]:

- вычисление энергии сигнала в критических полосах;
- компрессия амплитудных значений путем статической нелинейной трансформации;

- фильтрация амплитудных огибающих каждого из субполосных сигналов;

- растяжение амплитудного спектра путем обратной статической нелинейной трансформации;

- умножение амплитудных значений на кривую равной громкости и возведение в степень 0.33 для выравнивания громкости.

Для нелинейного преобразования амплитуд (компрессии/растяжения) обычно используются логарифмические либо степенные функции. Логарифмическая трансформация приводит к тому, что такие искажения, которые представляют собой свертку во временной области (например реверберация), проявляются как аддитивный шум в логарифмическом амплитудном спектре. Однако в случае зашумления речи некоррелированным аддитивным сигналом возникают трудности, поскольку такой шум аддитивен в линейном спектре, но в логарифмическом становится зависим от самого сигнала, что делает невозможным его удаление путем фильтрации по частотным полосам. Решение данной проблемы заключается в использовании следующей функции преобразования:

$$Y(k, n\Delta t) = \ln(1 + JX(k, n\Delta t)),$$

где $X(k, n\Delta t)$ – амплитуда сигнала в k -ом канале, J – положительная константа, зависящая от типа сигнала. Такое преобразование линейно для малых значений амплитуды и является логарифмическим для больших величин. Формула обратного преобразования имеет вид

$$X(k, n\Delta t) = \frac{e^{Y(k, n\Delta t)} - 1}{J}.$$

Синтез модуляционного фильтра

В качестве фильтра, используемого для обработки сигнала в модуляционной области, авторами RASTA [0] был предложен БИХ фильтр с передаточной функцией

$$H(z) = 0.1z^4 \frac{2 + z^{-1} - z^{-3} - 2z^{-4}}{1 - 0.98z^{-1}}.$$

Нижняя граница полосы пропускания фильтра определяет максимальную скорость изменения логарифмического спектра, который отбрасывается в процессе фильтрации, в то время как верхняя граница полосы пропускания определяет максимальную скорость изменения сохраняемого спектра. На рис. 1 показана амплитудно-частотная характеристика модуляционного полосового фильтра RASTA. Как следует из представленных результатов, начиная с частоты 12 Гц коэффициент усиления уменьшается со скоростью 6 dB на октаву.

Основным недостатком RASTA-алгоритмов является то, что сигнал фильтруется без учета характера и степени зашумления. Поэтому системы на основе данных алгоритмов ориентированы на конкретную акустическую среду и не могут подстраиваться к изменяющимся усло-

виям шумовой обстановки. Более того, чистая речь после обработки RASTA-системой звучит неприятно из-за наличия музыкальных тонов.

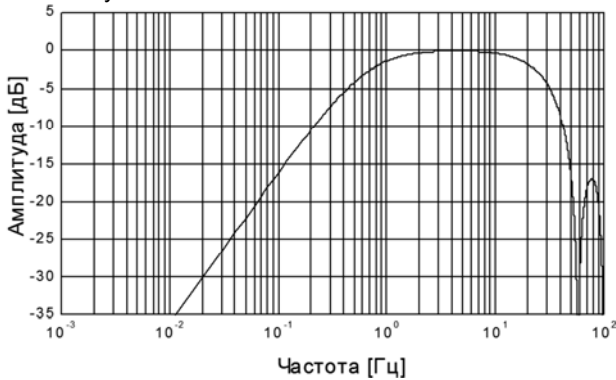


Рис. 1. Амплитудно-частотная характеристика модуляционного полосового фильтра RASTA

В работе [10] используются полосовые КИХ-фильтры (полоса пропускания 1-16 Гц) с различными амплитудно-частотными характеристиками. Известно применение модуляционного фильтра, который с целью улучшения разборчивости речи усиливал модуляционные компоненты в диапазоне 2-8 Гц [9].

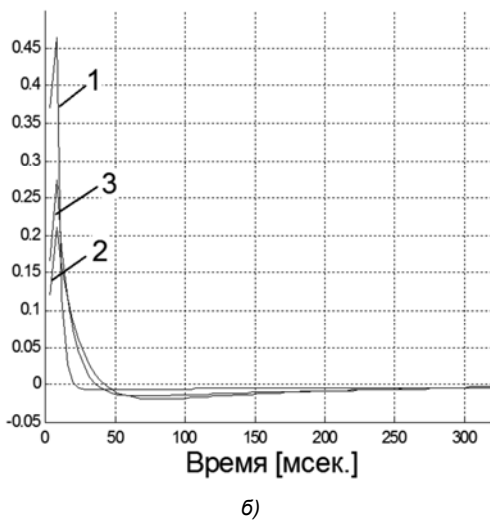
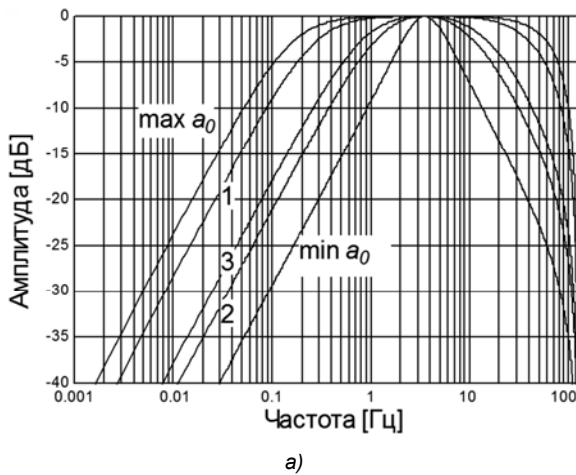


Рис. 2. Перестраиваемый фильтр для обработки речевого сигнала в модуляционной области:
а) – амплитудно-частотные характеристики;
б) – импульсные характеристики

Однако оба подхода не позволяют избавиться от вышеупомянутого недостатка. В работе [13] предпринимается попытка выбора модуляционного фильтра с учетом оценки условий окружающей акустической среды, и для повышения качества речи вместо одного модуляционного полосового фильтра с постоянной характеристикой используется их набор. Поскольку существует прямая зависимость между слоговой разборчивостью речи и энергией в полосе определенных частот модуляции, целесообразно выполнять слежение во времени за характером изменения модуляционного спектра и соответствующим образом менять полосу пропускания модуляционного фильтра, т.е. синтезировать модуляционный фильтр с изменяющимися во времени параметрами. В работе [11] в качестве прототипа используется параметрический перестраиваемый фильтр [14, 15], у которого в каждом частотном канале k во времени изменяется коэффициент a_{0k} , определяющий полосу пропускания

фильтра $\Delta\omega_k$. Центральная частота ω_0 сохраняется постоянной. Передаточная функция фильтра имеет вид

$$H_k(z) = a_{0k} \frac{1 - z^{-2}}{1 + (a_{0k} - 1)gz^{-1} + (1 - 2a_{0k})z^{-2}},$$

где $a_{0k} = \frac{\Delta\omega_k \Delta t}{2 + \Delta\omega_k \Delta t}$, $g = 2 \cos \omega_0 \Delta t$, $\Delta t = M / f_s$,

M – число полос банка фильтров, f_s – частота дискретизации. На рис. 2 изображены импульсные и частотные характеристики данного фильтра для разных значений коэффициента a_{0k} .

Алгоритм обработки сигнала в модуляционной области

Для субполосной декомпозиции речевого сигнала обычно используют банк фильтров [11, 12]. Фильтрация выполняется в фиксированных полосах, которые заранее определены на этапе расчета банка фильтров. Упрощенно схема обработки сигнала в модуляционной области может быть представлена следующим образом (рис. 3).

Согласно схеме, алгоритм состоит из следующих шагов [11]:

Шаг 1) речевой сигнал $x(n / f_s)$ разделяется на M частотных полос со следующими нормализованными центральными частотами: $(2k + 1)\pi / 2M$, причем каждый канал характеризуется своей импульсной характеристикой $p_k(n)$, а также частотной характеристикой $P_k(f)$ для $0 \leq k \leq M / 2$. Ширина каждой полосы не должна превышать частоту основного тона сигнала для того, чтобы исключить попадания нескольких гармоник основного тона в один канал.

Шаг 2) в каждом канале $k(0 \leq k \leq M / 2)$:

– вычисляется амплитудная огибающая сигнала $x_k(nM / f_s)$

$$y_k(nM / f_s) = \sqrt{\text{Re}^2[x_k(nM / f_s)] + \text{Im}^2[x_k(nM / f_s)]} \quad (1)$$

и фаза

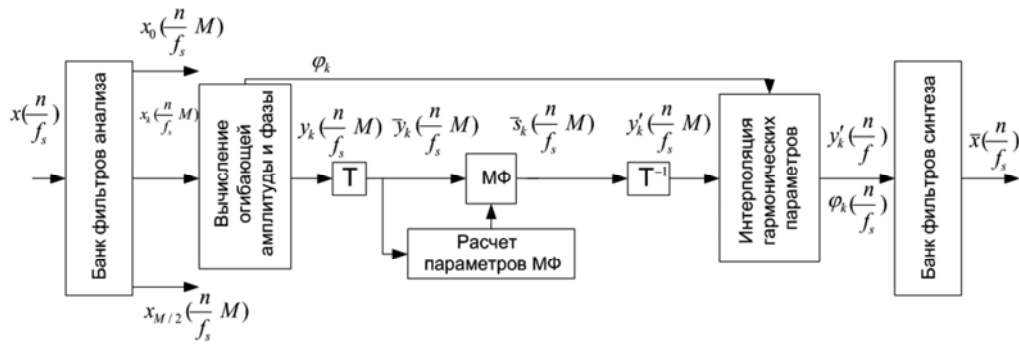


Рис. 3. Схема обработки сигнала в модуляционной области

$$\phi_k(nM / f_s) = -\arctg \frac{\text{Im}[x_k(nM / f_s)]}{\text{Re}[x_k(nM / f_s)]}; \quad (2)$$

– выполняется трансформация амплитуды огибающей спектра $y_k(nM / f_s)$ путем нелинейного статического сжатия

$$\bar{y}_k\left(\frac{nM}{f_s}\right) = \ln(1 + 1000|y_k(nM / f_s)|);$$

– производится фильтрация огибающей амплитуды спектра $\bar{y}_k(nM / f_s)$ модуляционным фильтром согласно схеме, приведенной на рис. 4;

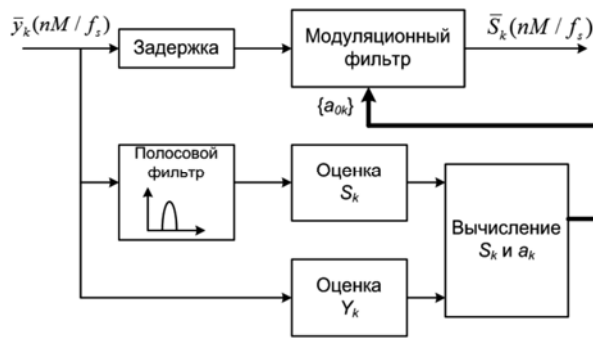


Рис. 4. Фильтрация амплитудной огибающей $\bar{y}_k(nM / f_s)$ перестраиваемым модуляционным фильтром

– изменяется параметр a_{0k} , задающий полосу пропускания модуляционного фильтра в каждом канале, согласно следующей процедуре:

1) вычисление оценки среднеквадратического значения S_k речевой (фильтрованной) компоненты;

2) вычисление оценки среднеквадратического значения Y_k сигнала $\bar{y}_k(nM / f_s)$ при помощи экспоненциального усреднения;

3) вычисление значения действительной весовой функции H_k и коэффициента модуляционного фильтра

$$a_{0k} = H_k^2.$$

– выполняется трансформация амплитуды огибающей спектра $\bar{s}_k(nM / f_s)$ в линейный масштаб путем обратного нелинейного преобразования

$$y'_k\left(\frac{nM}{f_s}\right) = \frac{e^{\bar{y}_k(nM/f_s)-1}}{1000}.$$

Шаг 3) восстановление речевого сигнала $\bar{x}_k(n / f_s)$ при помощи банка фильтров синтеза.

В приведенном алгоритме частотные полосы равно-

мерно перекрывают весь диапазон сигнала и не пересекаются. Причем, с одной стороны, число полос должно быть максимально большим, чтобы увеличить частотное разрешение алгоритма обработки, с другой стороны, каждая полоса должна быть достаточно широкой для сохранения эффективного диапазона модуляционного спектра. Если частота дискретизации составляет 8 кГц, то M следует выбирать не меньше 57 (что соответствует ширине полосы пропускания 70Гц), но и не более 130 (что соответствует ширине полосы пропускания 31Гц). Данные ограничения приводят к неточному разделению обрабатываемого сигнала на речь и шум в модуляционном спектре, что выражается в подавлении полезного сигнала и повышенном уровне остаточного шума, а также излишнего «отбеливания» речи. Таким образом, целесообразно использование иного способа расчета амплитудной огибающей, основанного на суммировании мгновенных амплитуд нескольких каналов.

Модификация алгоритма

Ниже приведен предлагаемый модифицированный алгоритм фильтрации в модуляционной области:

Шаг 1) речевой сигнал $x(n / f_s)$ разделяется на M частотных полос со следующими нормализованными центральными частотами: $(2k+1)\pi / 2M$. Ширина каждой полосы фиксирована (более 30Гц) и не зависит от числа каналов; M выбирается сколь угодно большим в зависимости от производительности вычислительной платформы (в экспериментальной реализации использовалось 2048 каналов).

Шаг 2) в каждом канале $k(0 \leq k \leq M / 2)$ амплитудная огибающая $y_k(nM / f_s)$ и фаза $\phi_k(nM / f_s)$ вычисляются по формулам (1) и (2).

Шаг 3) полученные амплитудные огибающие группируются в L каналов, причем $L < M$ (в экспериментальной реализации использовалось $L = 64$); Результирующие огибающие вычисляются как среднее квадратическое от мгновенных амплитудных значений субканальных сигналов, входящих в одну группу

$$z_l(nM / f_s) = \sqrt{\frac{\sum_{i \in \text{группа } l} y_i(nM / f_s)^2}{M / L}}, \quad 1 \leq l \leq L;$$

– выполняется трансформация амплитуды огибающей спектра $z_l(nM / f_s)$ путем нелинейного статического сжатия $\bar{z}_l(nM / f_s) = \ln(1 + 1000|z_l(nM / f_s)|)$;

– выполняется фильтрация огибающей амплитуды спектра $\bar{z}_l(nM / f_s)$ модуляционным фильтром согласно схеме, приведенной на рис. 4;

– параметр a_{0k} изменяется согласно процедуре, приведенной выше;

– производится трансформация амплитуды огибающей спектра $\bar{z}_l(nM / f_s)$ в линейный масштаб путем обратного нелинейного преобразования

$$\hat{z}_l(nM / f_s) = \frac{e^{\bar{z}_l(nM / f_s) - 1}}{1000}.$$

Шаг 4) Вычисление коэффициентов усиления G для каждого из L каналов:

$$G_l = \hat{z}_l(nM / f_s) / z_l(nM / f_s).$$

Амплитуда каждого из M исходных каналов умножается на соответствующий коэффициент G_l и выполняется восстановление речевого сигнала при помощи банка фильтров синтеза.

Результаты экспериментов

Сравнение полученного алгоритма (далее МФ2) с известным [11] (далее МФ1) проводилось с использованием специально подготовленных речевых сигналов, зашумленных помехами различных типов. Сравнение выполнено при помощи объективных оценок на основе анализа спектра сигналов и при помощи субъективных оценок на основе результатов прослушивания.

На рис. 5 представлен результат обработки сигнала, зашумленного белым шумом.

В обработанных сигналах энергия аддитивного шума снижена на 8дБ (МФ1) и на 10дБ (МФ2). Видно, что оба обработанных сигнала содержат музыкальные тона, однако интенсивность тонов заметно ниже для алгоритма МФ2.

На рис. 6 представлен результат обработки сигнала, зашумленного сетевой помехой. Сетевая помеха представляет собой совокупность гармонических компонент с медленно изменяющимися амплитудами и поэтому эффективно задерживается модуляционными фильтрами.

На участках, где отсутствует голос диктора, снижение энергии сетевой помехи составляет примерно 33дБ и 40дБ для алгоритмов МФ1 и МФ2 соответственно.

На рис. 7 представлен результат обработки сигнала, зашумленного шумом пылесоса, который состоит из периодических и непериодических составляющих. Особенностью данной помехи является ее нестационарность.

На участках, где отсутствует голос диктора, энергия помехи в среднем снижена на 20 дБ и 23 дБ для алгоритмов МФ1 и МФ2 соответственно.

На рис. 8 представлен результат обработки речи на фоне музыки. Автоматическое отделение речи от музыки в одноканальных системах представляет собой особенно сложную задачу. Тем не менее, использование модуляционных фильтров обеспечивает некоторое подавление тональных и переходных составляющих, характерных для музыкальных инструментов.

Снижение энергии музыки составляет примерно 4дБ для обоих алгоритмов модуляционной фильтрации.

На рис. 9 представлен результат обработки речи с реверберацией.

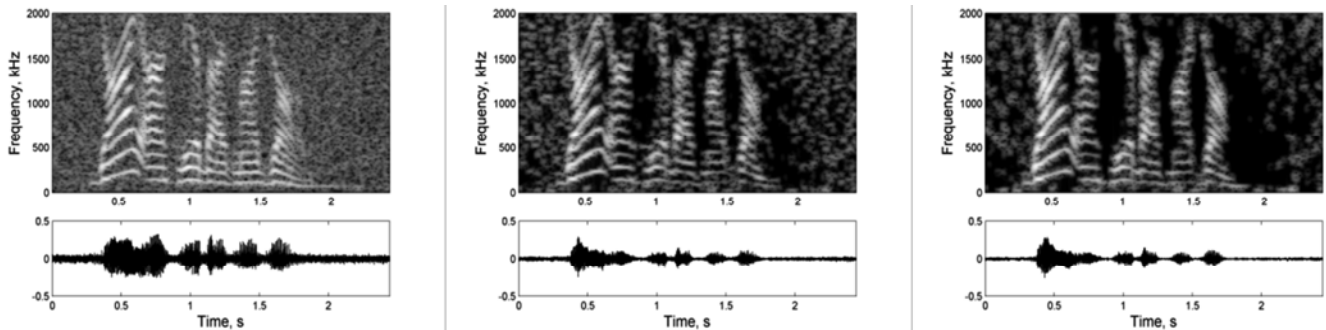


Рис. 5. Результат обработки сигнала, зашумленного белым шумом: а) – исходный речевой сигнал; б) – МФ1; в) – МФ2

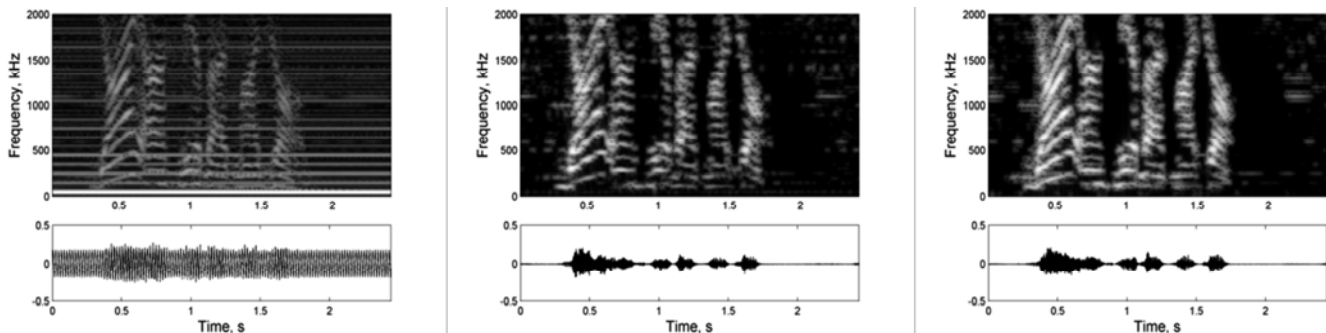


Рис. 6. Результат обработки сигнала, зашумленного сетевой помехой: а) – исходный речевой сигнал; б) – МФ1; в) – МФ2

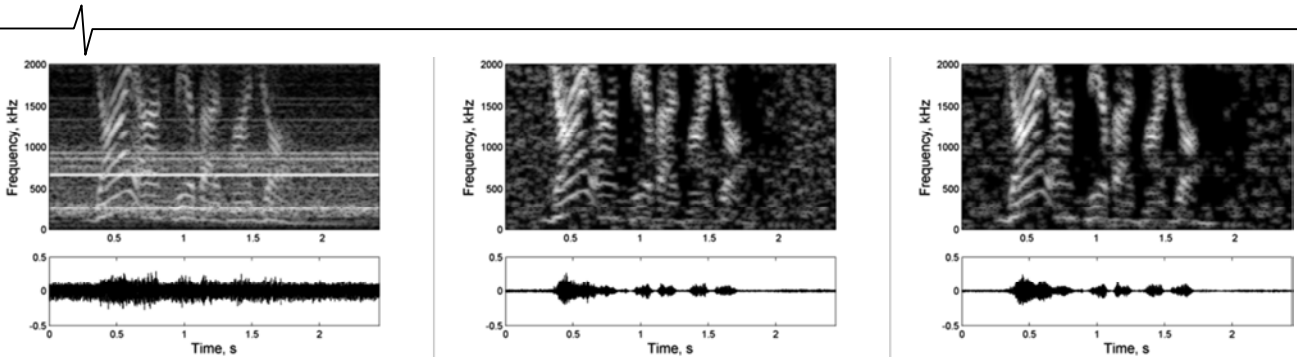


Рис. 7. Результат обработки сигнала, зашумленного шумом пылесоса: а) – исходный речевой сигнал; б) – МФ1; в) – МФ2

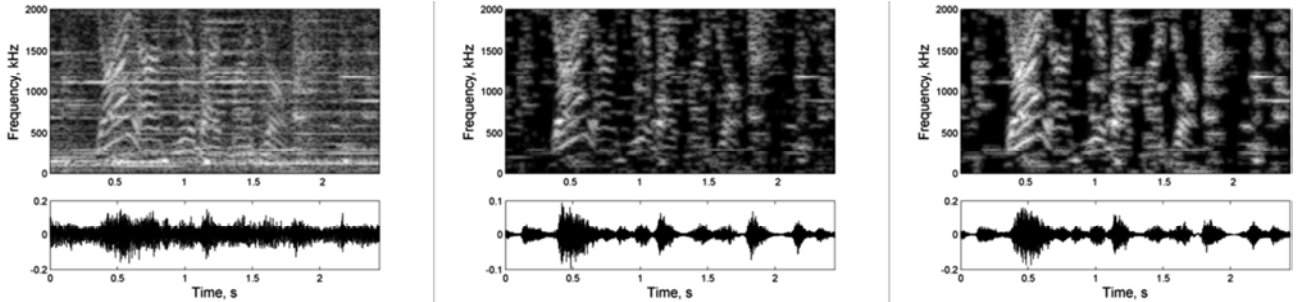


Рис. 8. Результат обработки сигнала, зашумленного фоновой музыкой: а) – исходный речевой сигнал; б) – МФ1; в) – МФ2

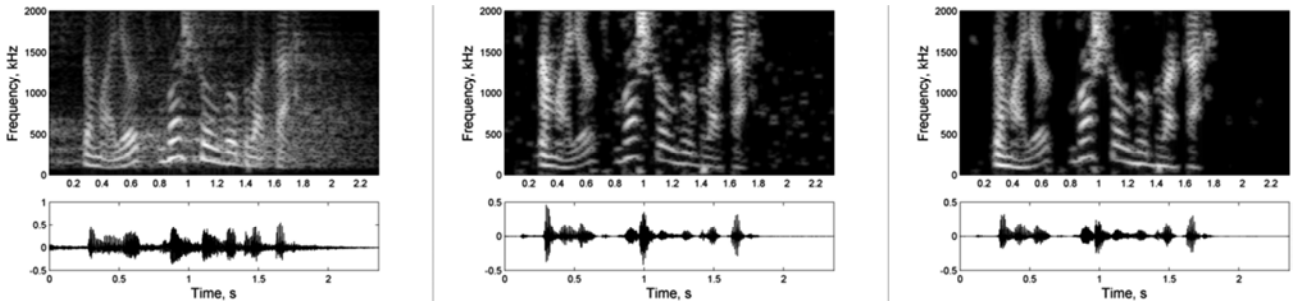


Рис. 9. Результат обработки сигнала с реверберацией: а) – исходный речевой сигнал; б) – МФ1; в) – МФ2

Во всех приведенных выше примерах предложенный алгоритм фильтрации обеспечивает более низкий уровень музыкальных тонов и других артефактов, более точно сохраняет форму спектральной огибающей и гармоническую структуру речевого сигнала. Различия особенно заметны в низкочастотной области спектра (ниже 500 Гц).

Результаты субъективной оценки качества очистки речевых сигналов приведены в табл. 1. В процессе эксперимента использовались записи мужского голоса продолжительностью 3 минуты, зашумленные помехами различных типов. Четверем слушателям была предложена субъективная оценка (по пятибалльной шкале) качества подавления шума в обработанных сигналах по следующим показателям:

– разборчивость (точность восприятия речевого сообщения: 5 – полная, 4 – почти полная, 3 – неполная, 2 – частичная, 1 – отсутствует);

– субъективное зашумление (степень зашумления: 5 – высокий уровень зашумления, ощущается сильный дискомфорт при прослушивании, 4 – умеренный дискомфорт при прослушивании, 3 – сохраняется средняя комфортность прослушивания, 2 – сохраняется высокая комфортность прослушивания, 1 – нет зашумления);

– качество реконструкции (естественность звучания обработанного речевого сигнала: 5 – абсолютно натуральное звучание, 4 – почти натуральное звучание с небольшим уровнем артефактов, 3 – средний уровень артефактов, 2 – высокий уровень артефактов, 1 – речь полностью неестественная).

Таблица 1 Результаты субъективной оценки качества очистки речевых сигналов

Тип помехи / оценка	Белый шум			Сетевая помеха			Шум пылесоса			Фоновая музыка			Реверберация		
	Исх	МФ1	МФ2	Исх	МФ1	МФ2	Исх	МФ1	МФ2	Исх	МФ1	МФ2	Исх	МФ1	МФ2
Разборчивость	4	4.25	4.25	3.25	4.0	4.25	2.75	3.5	3.75	2	2.5	2.75	4	4.25	4.25
Субъективное зашумление	4	3	2.75	4.5	2.5	2.25	4.75	3	2.5	4.5	4.25	4.25	3	1.5	1.25
Качество реконструкции	-	2.75	3	-	3.5	3.75	-	3	3.25	-	3.25	3.25	-	3.5	3.75

Полученные результаты свидетельствуют о том, что предложенный алгоритм фильтрации речи в модуляционной области обеспечивает более высокое субъективное качество шумоподавления. Это достигается как за счет снижения уровня слышимых артефактов, так и за счет более точного восстановления спектральных огибающих речевого сигнала.

Заключение

Исследования информационной значимости модуляционных частот речи являются одним из приоритетных современных направлений развития систем шумоподавления и распознавания речи в условиях агрессивной акустической обстановки. Высокий потенциал подхода модуляционной фильтрации для повышения разборчивости речи в шумах объясняется его физиологической мотивацией и тесной связью с психоакустикой. Практическим достоинством подхода является универсальность и автоматическая адаптивность к помехам разных типов.

При выполнении модуляционной фильтрации ключевыми являются вопросы выбора способа субполосной декомпозиции, функции преобразования амплитудной огибающей, модуляционного фильтра и способа синтеза выходного сигнала. В работе выполнен поиск наиболее удачных на наш взгляд решений каждой из перечисленных составляющих, и на их основе предложен оригинальный алгоритм для достижения максимально высокой степени очистки речевого сигнала.

Полученные результаты свидетельствуют о применимости предложенного алгоритма для обработки речевых сигналов, зарегистрированных в различных акустических условиях. Основным достоинством алгоритма является более низкий уровень слышимых артефактов и более высокое качество реконструкции речевого сигнала. Наиболее хорошо алгоритм применим для очистки речевого сигнала от тональных шумов высокой интенсивности (таких как сетевая помеха) и реверберации.

Литература

1. Schreiner and Urbas J., «Representation of amplitude modulation in the auditory cortex of the cat. I. The anterior auditory field (AAF)», *Hearing Research*, vol. 21, pp. 227-241, 1986.
2. Langner G., «Periodicity coding in the auditory system», *Hearing Research*, vol. 60, no. 2, pp. 115-142, 1992.
3. Kowalski N., Depireux D., and Shamma S., «Analysis of dynamic spectra in ferret primary auditory cortex: I. Characteristics of single unit responses to moving ripple spectra», *Journal of Neurophysiology*, vol. 76, no. 5, pp. 3503-3523, 1996
4. Drullman R., Festen J., Plomp R., «Effect of temporal envelope smearing on speech reception», *Journal of Acoustical Society of America*, vol. 95, no. 2, pp. 1053-1064, 1994.
5. Viemeister N., «Temporal factors in audition: A systems analysis approach», in *Psychophysics and Physiology*

of Hearing, E. Evans and J. Wilson, Eds., pp. 419-427, Academic Press, London, UK, 1977

6. Houtgast T. «Frequency selectivity in amplitude-modulation detection», *Journal of Acoustical Society of America*, vol. 85, pp. 1676-1680, 1989

7. Suga N. «Analysis of information-bearing elements in complex sounds by auditory neurons of bats», *Audiology*, vol. 11, pp. 58-72, 1972.

8. Bacon S. and Grantham D., «Modulation masking: Effects of modulation frequency, depth, and phase», *Journal of Acoustical Society of America*, vol. 85, pp. 2575-2580, 1989

9. Elhilali M., Chi T., Shamma S.A Spectro-temporal modulation index (STMI) for assessment of speech intelligibility // *Speech Communication – 2003 – 41 – pp. 331-348*

10. Arai T., Pavel M., Hermansky H., Avendano C. Syllable intelligibility for temporally filtered LPC cepstral trajectories // *J. Acoust. Soc. Am.* – 1999 – vol. 105 – pp 2783-2791

11. Bashun J., Petrovsky A. Speech enhancement for cochlear implants based on the reducing slow temporal modulations // *Proc. of The Acoustic Congress, Rom, Italy – 2000 – CD*

12. Hermansky H. and Morgan N., «RASTA processing of speech», *IEEE Transaction on Speech and Audio Processing*, vol. 2, no. 4, pp. 587–589, October 1994.

13. Avendano C. «Temporal processing of speech in a Time-Feature Space», Ph.D. thesis, Oregon Graduate Institute, April (1997)

14. Петровский А.А., Ганушкин Ю.А. Синтез полосовых цифровых фильтров с заданной добротностью // *Радиотехника*. 1985. № 10. С.24-25.

15. Петровский А.А., Башун Я.М. Пре-процессор повышения качества зашумленной и реверберирующей речи для систем улитковой имплантации. – *Цифровая обработка сигналов*. 2002. – № 2, Москва. С.48-61.

REDUCING COMPLEX NOISES IN SPEECH USING FILTERING IN MODULATION DOMAIN

Azarov I.S., Vashkevich M.I., Lihachev D.S., Petrovsky A.A.

The paper presents a method for noise reduction which is capable of handling noises of different nature such as white, colored quasiperiodic or reverberation. This all-purpose method is based on filtration in modulation domain. In order to obtain modulation signals source speech is decomposed into overlapping analytical subband signals using a filter bank. Then each analytical signal is parametrized in terms of instantaneous amplitude and frequency. Subchannel samples are grouped by instantaneous frequency values and their overall intensity envelopes are evaluated. Envelopes are filtered with a band-pass filter which retains speech modulations and attenuates noise modulations. Finally subband gains are calculated and applied to subchannel signals and the cleared speech signal is reconstructed. The paper contains experimental results that show that the proposed method has a better performance than its known counterpart based on RASTA (RelAtive SpecTrA).