

УДК 621.395

## ПОДАВЛЕНИЕ ШУМА В РЕЧЕВЫХ СИГНАЛАХ НА ОСНОВЕ МЕТОДА НЕЛОКАЛЬНОГО УСРЕДНЕНИЯ

*Новоселов С.А., аспирант кафедры динамики электронных систем Ярославского государственного университета им. П.Г. Демидова, e-mail: sergnovoselov@mail.ru;*

*Топников А.И., аспирант, ассистент кафедры динамики электронных систем Ярославского государственного университета им. П.Г. Демидова, e-mail: topnikov@gmail.com;*

*Савватин А.И., аспирант кафедры динамики электронных систем Ярославского государственного университета им. П.Г. Демидова, e-mail: dcslab@uniyar.ac.ru;*

*Приоров А.Л., д.т.н., доцент кафедры динамики электронных систем Ярославского государственного университета им. П.Г. Демидова, e-mail: andcat@yandex.ru*

**Ключевые слова:** подавления шума, речевые сигналы, гауссовский шум, нелокальное усреднение.

### Введение

Речевые сигналы, с которыми приходится иметь дело на практике, всегда в той или иной степени зашумлены. При наличии внешних шумов значительной интенсивности результаты анализа и распознавания речи существенно ухудшаются, поэтому разработка методов шумоподавления для речевых сигналов является актуальным направлением исследований.

На сегодняшний день в области цифровой обработки речевых сигналов существует множество различных алгоритмов шумоподавления. Наиболее распространенной моделью шума для рассматриваемых алгоритмов является аддитивный белый гауссовский шум (АБГШ). Классификацию методов повышения качества и разборчивости речи удобнее всего проводить согласно основной концепции и идее, лежащей в основе метода. На современном этапе наибольшее распространение получили следующие группы методов [1-4]:

- методы, основанные на использовании авторегрессионной модели речевого сигнала;
- методы, основанные на обработке речевого сигнала с использованием аппарата скрытых марковских моделей;
- методы, основанные на искусственных нейронных сетях;
- методы, основанные на оценке параметров шума, минимизации среднеквадратичной ошибки и пороговой обработке в области трансформант.

Существуют и другие подходы. Например, сравнительно недавно в области цифровой обработки речевых сигналов появилось ещё одно направление – слепое разделение речевых сигналов (Blind Speech Separation), объединяющее методы, позволяющие получить оценки сигналов источников, основываясь лишь на знании сигналов смесей и некоторых общих предположениях о свойствах сигналов источников и процессе смешивания [5-8]. В ряде случаев эти методы позволяют достичь значительного улучшения качества и разборчивости речи. В области цифровой обработки изображений (ЦОИ)

Описывается метод нелокального усреднения, анализируется возможность его применения в задаче подавления шума в речевых сигналах. Для демонстрации эффективности подхода представлены результаты моделирования, полученные с использованием алгоритма, основанного на нелокальном усреднении. Проведено сравнение полученных результатов с результатами, полученными для алгоритма, реализующего метод спектрального вычитания.

также существует множество методов шумоподавления [9-14], часть из которых можно распространить и на обработку речевых сигналов. Предложенный в статье алгоритм базируется на хорошо зарекомендовавшем себя в области ЦОИ [11-14] методе нелокального усреднения (Non-Local Means, NLM).

### Метод нелокального усреднения

Кратко опишем метод нелокального усреднения, используемый в задаче подавления шума в изображениях [14]. Пусть цифровое изображение  $x$  искажено аддитивным белым гауссовским шумом  $n$  с нулевым математическим ожиданием и дисперсией  $\sigma^2$ .

1. Для обрабатываемого пикселя  $i$  зашумленного изображения  $y = x + n$  описываем квадратную окрестность фиксированного размера, центрированную на этот пиксель.

2. Определяем подобность обрабатываемого пикселя  $i$  зашумленного изображения  $y$  с пикселем  $j$  того же изображения, используя взвешенное евклидово расстояние  $\|y(N_i) - y(N_j)\|_{2,a}^2$ , где  $N_i$  и  $N_j$  – квадратные окрестности, центрированные на пиксели  $i$  и  $j$  соответственно,  $a$  – положительное число, определяющее среднеквадратическое отклонение гауссова ядра, используемого для вычисления взвешенного евклидова расстояния.

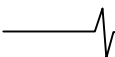
3. Определяем вес подобного к  $i$  пикселя  $j$  в итоговой оценке пикселя  $i$ :

$$w(i, j) = \frac{1}{Z(i)} \cdot e^{-\frac{\|y(N_i) - y(N_j)\|_{2,a}^2}{h^2}}, \quad Z(i) = \sum_j e^{-\frac{\|y(N_i) - y(N_j)\|_{2,a}^2}{h^2}},$$

где  $h$  – параметр, влияющий на степень фильтрации цифрового изображения.

4. Формируем итоговую оценку пикселя  $i$  на основе следующего выражения:

$$\hat{x}(i) = \sum_j w(i, j) y(j).$$



Так как спектрограмму звукового сигнала можно рассматривать как изображение, то это позволяет адаптировать алгоритмы шумоподавления, предназначенные для изображений, для обработки звуковых сигналов. Подобный подход позволяет, в частности, значительно снизить уровень «музыкального шума», который является основным артефактом работы алгоритмов, реализующих метод спектрального вычитания [15]. Также возможно применение метода нелокального усреднения как к реальной, так и к мнимой частям спектра звукового сигнала для шумоподавления с целью повышения эффективности системы распознавания [16].

Однако структура речевого сигнала во временной области позволяет сделать предположение о возможности применения метода нелокального усреднения непосредственно к одномерному сигналу. Действительно, в пределах интервала стационарности речевой сигнал состоит из последовательности похожих фрагментов, аналогично тому, как похожие фрагменты могут располагаться в различных частях изображения. Обоснование возможности применения метода нелокального усреднения к речевым сигналам, а также достижимость положительного эффекта в задаче шумоподавления были продемонстрированы в работе [17].

Настоящая работа направлена на дальнейшее развитие и исследование возможности подавления шума в речевых сигналах на основе применения метода нелокального усреднения во временной области. Особое внимание уделяется методикам поиска похожих фрагментов речевого сигнала, так как качество выполнения этого этапа в значительной степени определяет конечную эффективность алгоритма.

### Описание предлагаемого алгоритма

Структура отрезков речевого сигнала, приведенная на рис.1, позволяет сделать вывод о том, что речь по своей природе, в некотором смысле, является примером «помехоустойчивого кода». Как видно из осциллограмм, помехоустойчивость речи достигается за счет избыточного повторения «информационных блоков» (это свойство особенно характерно для гласных звуков). Возможно, звуковой анализатор человека использует подобную избыточность и благодаря этому способен выделять информацию на фоне достаточно интенсивных помех (-10 дБ). Эти положения позволяют рассматривать и применять следующую модель речевого сигнала: речь – периодический на интервалах стационарности процесс, состоящий из повторяющихся «информационных блоков» или фрагментов. В рамках этой модели для решения задачи шумоподавления в речевых сигналах эффективно использовать алгоритм нелокального усреднения. При адаптации алгоритма нелокального усреднения для задачи подавления АБГШ в речевом сигнале учитываются особенности речи, благодаря чему получившийся алгоритм значительно отличается от своего двумерного аналога. Например, усреднение производится непосредственно для фрагмента длиной несколько десятков отсчетов, а не для одного отсчета. Рассмотрим особенности работы предлагаемого алгоритма подавления АБГШ в речевых сигналах.

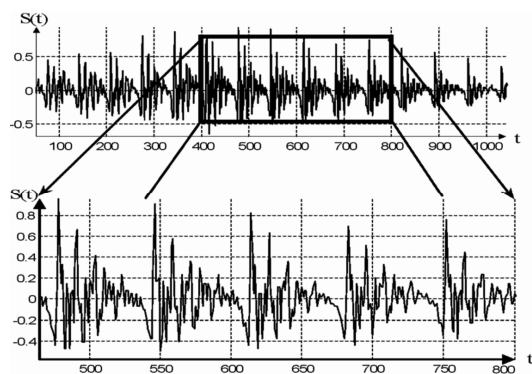


Рис. 1. Структура речевого сигнала

Обработка речевого сигнала осуществляется во временной области. Можно выделить два основных этапа: поиск похожих фрагментов и усреднение этих фрагментов, позволяющие значительно снизить дисперсию аддитивного шума. Длина фрагмента выбирается в пределах интервала 40-80 отсчетов (установлено экспериментально). Поиск похожих фрагментов ведется в пределах окна длиной 400 отсчетов, так как при частоте дискретизации 8000 отсчетов в секунду это соответствует 50 мс, то есть сравнимо с интервалом стационарности речевого сигнала, который примерно равен 15-30 мс. Для определения степени близости между фрагментами используется евклидово расстояние. Как будет показано ниже, локальные минимумы функции расстояния между выбранным блоком и остальными соответствуют положениям наиболее похожих на данный блок фрагментов. Теоретически чем больше число идентичных фрагментов усредняется, тем сильнее подавляется аддитивный шум, однако для речевых сигналов их число ограничено, к тому же их нельзя считать строго идентичными. На практике число таких фрагментов для речевого участка длиной 50 мс равно 4-12 (рис. 1), так как усреднение меньшего числа фрагментов малоэффективно, а поиск большего числа похожих фрагментов ограничен структурой речевого сигнала. После того как интересные фрагменты найдены, происходит их усреднение и замена отсчетов сигнала, соответствующих этим отрезкам, на отсчеты, полученные в результате процесса усреднения. При низких значениях отношения сигнал / шум (ОСШ) процесс поиска похожих фрагментов значительно осложнен. Для повышения эффективности алгоритма имеет смысл формировать опорный сигнал путем подавления шума в исходном сигнале при помощи метода спектрального вычитания. Этот сигнал необходим лишь для поиска похожих фрагментов.

**Выбор опорного сигнала.** В предлагаемом алгоритме шумоподавления определяющим фактором является выбор опорного сигнала, по которому проводится поиск похожих фрагментов на интервалах речи длительностью 50 мс. Исследования показывают, что если в качестве опорного сигнала взять сам исходный малошумленный сигнал (ОСШ > 30 дБ) и выбирать похожие фрагменты на основе простого порогового сравнения евклидова расстояния, то можно добиться значительного снижения уровня шума при сохранении почти всех особенностей речевого сигнала (рис. 2). Этот факт подтверждает адекватность выбранной модели сигнала.

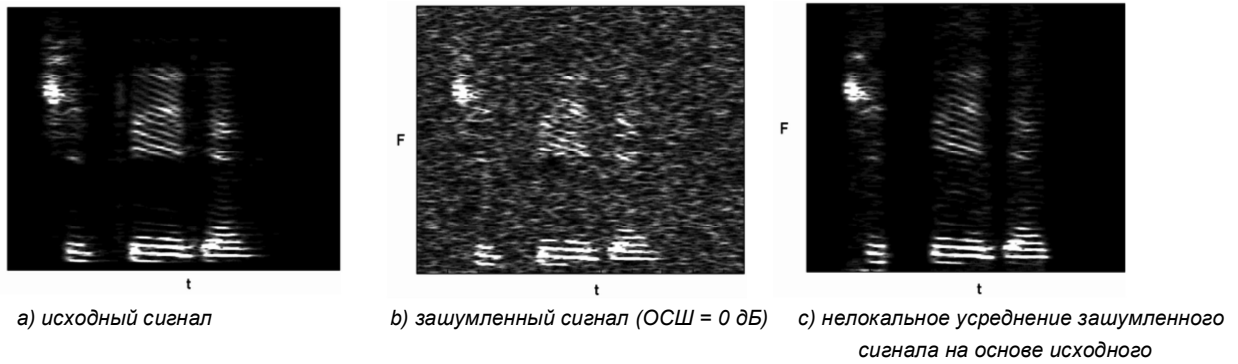


Рис. 2. Спектрограммы сигналов

На практике использовать исходный незашумленный сигнал не представляется возможным. В работе предлагается использовать в качестве опорного сигнала, полученный из зашумленного сигнала путем шумоподавления с помощью метода спектрального вычитания (Spectral Subtraction, SS), описанного в статье [2]. При этом коэффициент сглаживания  $\beta$  в этом алгоритме следует выбирать так, чтобы максимально сохранить особенности речевого сигнала (шипящие звуки, глухие согласные и др.) при вынужденном сохранении остаточного шума. Экспериментально установлено, что целесообразно выбирать значения  $\beta \approx 0.7-0.8$ . При таких значениях, ослабление исходного шума составляет 7-8 дБ. Остаточный шум почти равномерно распределен по сигналу и по свойствам схож с АБГШ (рис. 3).

**Поиск похожих фрагментов.** Фрагментом или блоком стационарного отрезка речевого сигнала будем называть участок речи, протяженностью в 5 мс, находящийся внутри этого отрезка. При частоте дискретизации 8000 отсчетов в секунду 5 мс соответствует 40 отсчетам. Фрагменты выбираются с максимальным перекрытием в 39 отсчетов. Как уже отмечалось, длина самого отрезка равна 50 мс, что соответствует 400 отсчетам. Таким образом, в одном отрезке содержится  $L = 361$  фрагментов.

Итак, опорный и зашумленный сигнал одинаковым образом разбиваются на перекрывающиеся отрезки длиной 50 мс. Величина перекрытия составляет 90% от размера окна.

Для общности обозначим фрагменты текущего отрезка речевого сигнала векторами  $\vec{S}_i$ , соответствующие им шумовые фрагменты текущего отрезка опорного сигнала – векторами  $\vec{n}_i$ , а соответствующие им шумовые фрагменты текущего отрезка исходного обрабатываемого сигнала – векторами  $\vec{N}_i$ .

Похожесть речевых фрагментов-векторов определяется евклидовым расстоянием между ними. Речевые фрагменты  $\vec{S}_i$  и  $\vec{S}_j$  будем считать похожими, если евклидово расстояние между ними не превышает некоторого порогового значения:

$$\|\vec{S}_j - \vec{S}_i\| \leq t_{уцх}.$$

Этот порог соответствует допустимым незначительным отклонениям похожих речевых блоков друг от друга.

Требуется для некоторого выбранного вектора  $\vec{S}_j + \vec{n}_{j_2}, j \in [1; L]$ , из всего набора фрагментов  $\{M_{\vec{n}} : \vec{S}_i + \vec{n}_i\}$  опорного сигнала найти такое множество векторов  $\{O_{\vec{n}} : \vec{S}_o + \vec{n}_o\}$ , что  $\forall \vec{S}_o : \|\vec{S}_j - \vec{S}_o\| \leq t_{уцх}$ , т.е. любой  $\vec{S}_o$  является похожим на  $\vec{S}_j$ . Рассмотрим евклидово расстояние между зашумленным фрагментом  $\vec{S}_j$  и произвольным вектором  $\vec{S}_i$ . Используя неравенство треугольника, получим:

$$\|(\vec{S}_j + \vec{n}_j) - (\vec{S}_i + \vec{n}_i)\| \leq \|\vec{S}_j - \vec{S}_i\| + \|\vec{n}_j - \vec{n}_i\|.$$

Если фрагменты  $\vec{S}_j$  и  $\vec{S}_i$  похожи, то положим, что  $\|\vec{S}_j - \vec{S}_i\| \leq t_{уцх} \ll \|\vec{n}_j - \vec{n}_i\|$ . Дополнительно, учитывая стационарность шума (рис. 3b) и используя решение детектора речевой активности, можно оценить максимальную амплитуду шумового вектора  $\vec{n}_i$  на интервалах, где отсутствует речевая активность  $\|\vec{n}_i\| \leq \|\vec{n}\|_{max}$ . Тогда имеем:

$$\|(\vec{S}_j + \vec{n}_j) - (\vec{S}_i + \vec{n}_i)\| \leq \|\vec{n}_j - \vec{n}_i\| \leq 2 \|\vec{n}\|_{max}.$$

Последнее неравенство определяет условия похожести двух зашумленных речевых фрагментов и задает порог, который необходимо применять для нахождения множества  $\{O_{\vec{n}} : \vec{S}_o + \vec{n}_o\}$ :

$$\vec{S}_i + \vec{n}_i \in \{O_{\vec{n}}\} : \|(\vec{S}_j + \vec{n}_j) - (\vec{S}_i + \vec{n}_i)\| \leq 2 \|\vec{n}\|_{max}.$$

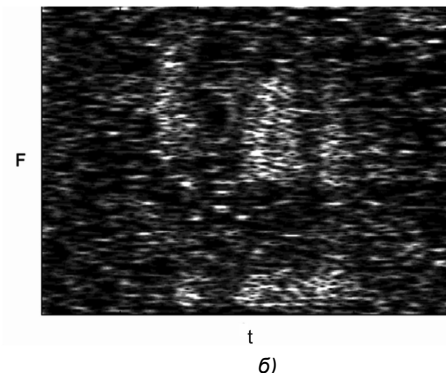
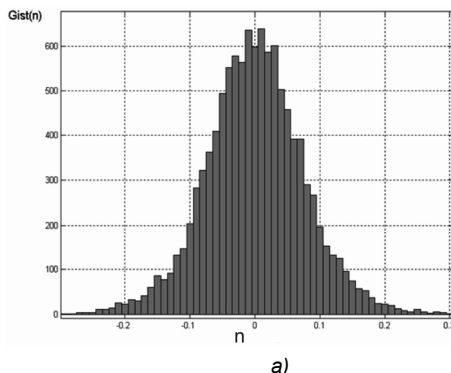
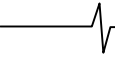


Рис. 3. Гистограмма а) и спектрограмма, б) остаточного шума



Такой метод поиска приводит к некоторому приближению множества похожих фрагментов. Однако следует иметь в виду, что смежные блоки выбранного фрагмента речевого сигнала являются «ложно» похожими на сам фрагмент из-за высокой корреляции соседних отчетов сигнала. Если их не исключить из рассмотрения и произвести оценку незашумленного блока путем нелокального усреднения по найденному множеству зашумленных векторов, то получим сглаженную копию искомого, потеряв при этом часть информации. Для исключения «ложных» фрагментов из множества  $\{O_n : \vec{S}_o + \vec{n}_o\}$  предлагается находить позиции локальных минимумов функции расстояния между векторами. Эти позиции соответствуют положениям пригодных для усреднения фрагментов (рис. 4). Таким образом, из рассмотрения исключаются окрестности минимумов, соответствующие «ложным» фрагментам.

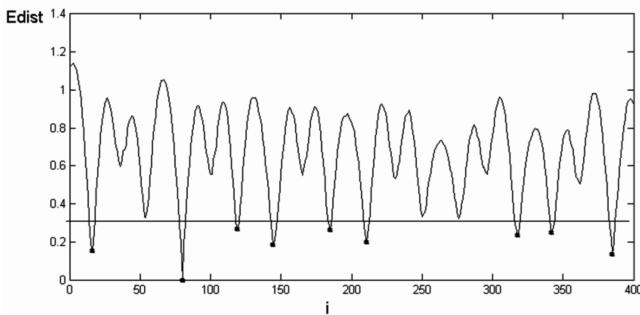


Рис. 4. Евклидово расстояние между текущим и другими фрагментами зашумленного отрезка речевого сигнала

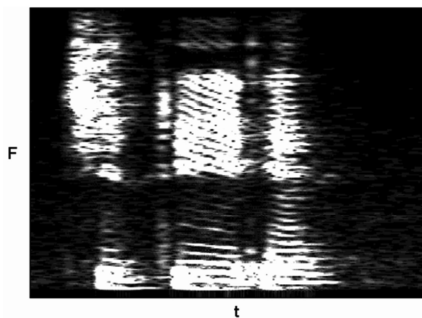
**Нелокальное усреднение.** После того как множество  $O_n$  для заданного фрагмента  $\vec{S}_j + \vec{n}_j$  найдено, с помощью гауссова ядра  $w_o$  реализуется процесс взвешенного усреднения множества векторов  $\{O_n : \vec{S}_o + \vec{N}_o\}$  отрезка обрабатываемого зашумленного сигнала, соответствующего векторам множества  $\{O_n : \vec{S}_o + \vec{n}_o\}$  опорного сигнала, с целью нахождения оценки  $\vec{S}_j'$ :

$$\vec{S}_j' = \sum_{o \in O_n} w_o (\vec{S}_o + \vec{N}_o),$$

$$\text{где } w_o = \frac{e^{-\frac{\|\vec{S}_j + \vec{n}_j - (\vec{S}_o + \vec{n}_o)\|^2}{h^2}}}{\sum_o e^{-\frac{\|\vec{S}_j + \vec{n}_j - (\vec{S}_o + \vec{n}_o)\|^2}{h^2}}},$$

$$\vec{S}_j' = \sum_{o \in O_n} w_o (\vec{S}_o + \vec{N}_o) \approx \sum_{o=1}^N \frac{1}{N} (\vec{S}_j + \vec{N}_o) \approx \vec{S}_j + \sum_{o=1}^N \frac{1}{N} \vec{N}_o.$$

Здесь  $N$  – количество векторов множества  $O_n$ ,  $h$  – параметр, влияющий на степень фильтрации (в работе  $h=1$ ). При этом  $h^2 \gg 2\|\vec{n}_{\max}\|^2$ . Полагая, что



а) исходный сигнал

$\{\vec{N}_o\}$  – набор случайных независимых векторов с постоянным вектором математического ожидания, равным  $\vec{\mu}$ , получим  $\vec{S}_j \approx \vec{S}_j + \vec{\mu}$ . Подобный вывод подтверждается тем, что отсчеты АБГШ независимы друг от друга, и математическое ожидание шума не зависит от времени. Таким образом, можно получить оценку  $\vec{S}_j$  незашумленного фрагмента  $\vec{S}_j$  с помощью выражения:

$$\hat{\vec{S}}_j = \vec{S}_j' - \vec{\mu} \approx \vec{S}_j.$$

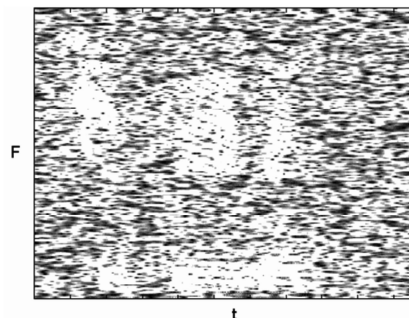
После проведения оценки всех фрагментов незашумленного отрезка сигнала, можно получить оценку всего отрезка. Затем по оценкам всех отрезков сигнала можно получить оценку и самого незашумленного речевого сигнала. Все это осуществляется с помощью процедуры восстановления сигнала с применением усреднения перекрывающихся данных.

### Результаты моделирования

Для проверки эффективности предложенного алгоритма проведены практические исследования. Алгоритм реализован в среде математического моделирования Matlab. Речевые сигналы, оцифрованные с частотой 8000 отсчетов в секунду и разрешением 8 бит на отсчет, зашумляются АБГШ. После этого сигнал поступает на вход алгоритма подавления шума.

Субъективная оценка позволяет сделать вывод, что предложенный алгоритм способен значительно снизить уровень шума, при этом улучшается разборчивость речи и не возникает артефактов, получивших название «музыкальный шум». Сравнение спектрограмм речевого сигнала (рис. 5) до добавления шума, после, а также при обработке с помощью предложенного алгоритма и алгоритма, реализующего метод спектральных вычитаний, подтверждают сделанные ранее выводы. В частности видно, что при использовании предложенного алгоритма (рис. 5d) хорошо сохраняются все значимые детали спектра и отсутствуют нежелательные всплески, приводящие к возникновению «музыкального шума», что характерно для методов фильтрации в спектральной области (рис. 5с).

Для объективной оценки речевого сигнала на выходе алгоритма шумоподавления используется сегментное отношение сигнал/шум (CeROCШ, segmental SNR, SSSNR). Результаты моделирования показывают, что при высоких уровнях шума предложенный алгоритм обладает большей эффективностью по сравнению с алгоритмом, реализующим метод спектральных вычитаний (рис. 6).



б) зашумленный сигнал (ОСШ = 0 дБ)

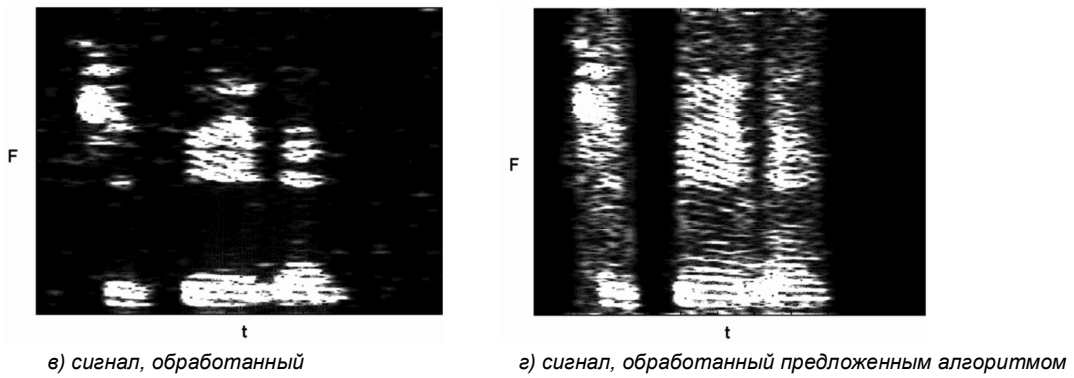


Рис. 5. Спектрограммы речевого сигнала

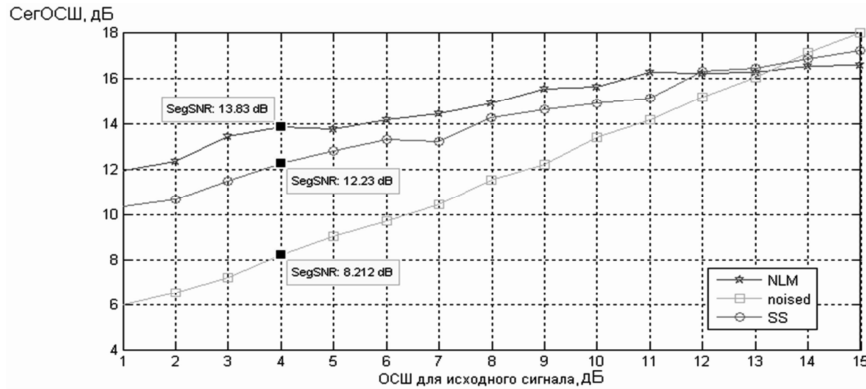


Рис. 6. Зависимость сегментного ОСШ от ОСШ исходного сигнала

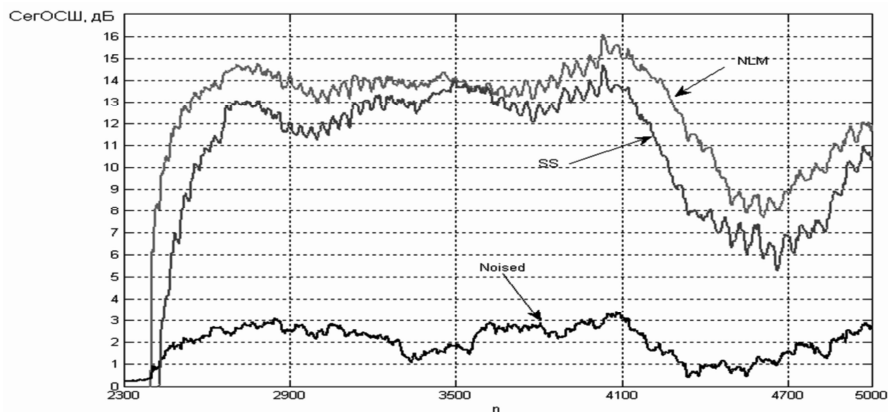


Рис. 7. Сегментное ОСШ для участка речевого сигнала до и после обработки алгоритмами шумоподавления

Наибольший выигрыш достигается при обработке участков речевого сигнала, отвечающих гласным звукам (рис. 7). Это объясняется особенностями алгоритма и структуры речевого сигнала: гласные звуки содержат больше похожих фрагментов (имеют структуру, близкую к периодической), на поиске которых и основан алгоритм нелокального усреднения. Предложенный алгоритм увеличивает сегментное ОСШ по сравнению с зашумленным сигналом на 10-12 дБ (рис. 7). Для сравнения, метод спектрального вычитания уступает предложенному методу на 1-1.5 дБ.

Ограничениями предложенного алгоритма можно считать то, что участки речевого сигнала, соответствующие шипящим звукам и глухим согласным, искажаются, что негативно влияет на качество звучания речи. Исходя из этого, актуальной является задача разработки специального алгоритма для фильтрации фриктивных звуков речи. Например, можно использовать комбинацию двух алгоритмов: гласные обрабатываются алго-

ритмом нелокального усреднения, а участки, отвечающие согласным, – методом спектрального вычитания.

### Заключение

В работе предложен метод, позволяющий решать задачу подавления аддитивного белого гауссовского шума в речевых сигналах. К достоинствам алгоритма можно отнести высокую эффективность при работе с сильнозашумленными сигналами (ОСШ ~ 0-7 дБ), а также отсутствие «музыкального шума» в обработанном сигнале. Основным недостатком алгоритма можно считать относительно высокую вычислительную сложность. Для повышения быстродействия алгоритма рекомендуется использовать параллельные вычисления.

### Литература

1. Чучупал В.Я., Чичагов А.С., Маковкин К.А. Цифровая фильтрация зашумленных речевых сигналов: Сообщения по программному обеспечению ЭВМ. М.:

Вычислительный центр РАН, 1998. 52 с.

2. Scalart P., Filho J.V. Speech enhancement based on a priori signal to noise estimation // IEEE international conference on acoustics, speech and signal processing (ICASSP-96). 1996. P. 629–632.

3. Ephraim Y., Malah D. Speech enhancement using a minimum mean square error short time spectral amplitude estimator // IEEE transactions acoustics, speech and signal processing. 1984. P. 1109–1121.

4. Ephraim Y., Malah D. Speech enhancement using a minimum mean square error log spectral amplitude estimator // IEEE transactions acoustics, speech and signal processing. 1985. P. 443–445.

5. Hyvärinen A., Karhunen J., Oja E. Independent component analysis. – Wiley-Interscience, 2001. 504 p.

6. Yilmaz O., Rickard S. Blind separation of speech mixtures via time-frequency masking // IEEE transactions on signal processing. 2004. V. 52, No. 7. P. 1830–1846.

7. Jourjine A., Rickard S., Yilmaz O. Blind separation of disjoint orthogonal signals: demixing N sources from 2 mixtures // IEEE international conference on acoustics, speech and signal processing (ICASSP2000). Istanbul, Turkey, 2000. V. 5. P. 2985–2988.

8. Araki S., Makino S., Blin A., Mukai R., Sawada H. Blind separation of more speech than sensors with less distortion by combining sparseness and ICA // International workshop on acoustic echo and noise control (IWAENC2003). 2003. P. 271–274.

9. Моисеев А.А., Кобелев В.Ю., Волохов В.А. Курвлет-преобразование в задаче подавления шума в изображениях // Цифровая обработка сигналов. 2008. № 1. С. 43–50.

10. Приоров А.Л., Хрящев В.В., Голубев М.Н. Удаление импульсного шума со случайными значениями импульсов из изображений // Радиотехника. 2010. № 5. С. 72–79.

11. Buades A., Coll B., Morel J.M. A review of image denoising algorithms, with a new one // Multiscale modeling and simulation. 2005. V. 4. P. 490–530.

12. Buades A., Coll B., Morel J.M. Nonlocal image and movie denoising // International journal of computer vision. 2008. V. 76, No. 2. P. 123–139.

13. Katkovnik V., Foi A., Egiazarian K., Dabov K. From local kernel to nonlocal multiple-model image denoising // International journal of computer vision. 2010. V. 86, No. 8. P. 1–32.

14. Волохов В.А., Сергеев Е.В., Мочалов И.С., Приоров А.Л. Модификация алгоритма нелокального усреднения для удаления шума из цифровых изображений // Доклады 13-й международной конференции и выставки «Цифровая обработка сигналов и ее применение» (DSPA-2011). Москва, 2011. Т. 1. С. 267–270.

15. Ткаченко М.С., Лукин А.С. Многомасштабный метод спектрального вычитания для подавления шумов в аудиосигналах // Доклады 12-й международной конференции и выставки «Цифровая обработка сигналов и ее применение» (DSPA-2010). Москва, 2010. Т. 1. С. 223–226.

16. Haitian X., Zheng-Hua T., Dalsgaard P., Lindberg B. Robust speech recognition by nonlocal means denoising

processing // IEEE signal processing letters. 2008. V. 15. P. 701–704.

17. Szlam A. Non-local means for audio denoising // Recent UCLA computational and applied mathematics reports. 2008. 5 p.

## **SPEECH DENOISING BY NON-LOCAL MEANS**

*Novoselov S., Topnikov A.,  
Savvatin A., Priorov A.*

A method of noisy speech enhancement based on non-local means algorithm is proposed. Non-local means algorithm is well-proven in the field of digital image processing. A description of the original two-dimensional method is performed and the possibility of its adaptation in one-dimensional case for noise suppression task is analyzed. Speech signals features takes into account in the method adapting so that the resulting algorithm is much differ from its two-dimensional analog. For example, the averaging is performed directly for a portion of a several tens of samples and not for a single one.

In this paper the suppression of additive white Gaussian noise in the speech signal is considered. There are two main steps in the proposed method of speech enhancement: the search for similar fragments and the average of these fragments to significantly reduce the variance of additive noise. The length of the fragment is chosen within the interval 40–80 samples. The similar fragments search is made in the window with 400 samples length. To determine the degree of similarity between segments Euclidean distance is used. In practice, the number of such fragments for 50 ms speech section is equal to 4–12, as averaging a smaller number of fragments is inefficient, and the search for a larger number of similar fragments is limited by the speech signal structure. After fragments of interest are found the procedure of its averaging is performed. The signal samples corresponding to these segments are changed to the samples obtained from the averaging process.

The article presents the simulation results of speech enhancement obtained using a non-local means algorithm. The effectiveness of the approach is shown. The results are compared with results obtained for the algorithm that implements the method of spectral subtraction. Segmental signal to noise ratio is used for an objective evaluation of enhanced speech quality. The proposed algorithm is more efficient than one that implements spectral subtraction method at high noise levels. The greatest benefit is achieved in vowel sounds processing that can be easily explained based on the characteristics of the algorithm and the speech signal structure.

The advantages of the proposed algorithm are high efficiency when working with low signal to noise ratios and the lack of "musical noise" in the processed signal. The main drawback of the algorithm is relatively high computational complexity. To improve the performance of the method parallel computing utilization is recommended.