



Шумоподавление на основе перцептуальных алгоритмов спектрального вычитания и обработки сигналов в подпространствах

*Петровский А.А., доктор технических наук, профессор,
Азаров И.С., кандидат технических наук,
Лихачёв Д.С., кандидат технических наук, с.н.с, доцент,
Ромашкин Ю.Н., кандидат технических наук,
Жигулёвцев Ю.Н., кандидат технических наук, доцент,
Харламов А.А., доктор технических наук, с.н.с.*

В данной статье оценивается эффективность алгоритмов шумоподавления, основанных на спектральном вычитании и обработке сигнала в подпространствах. Исследования проводятся с использованием аддитивных шумов различной природы и интенсивности. Алгоритмы оцениваются в контексте качества обработки речи. Применяются стандартные методики и показатели, такие как перцептуально-модифицированная оценка искажений спектра барков и индекс разборчивости речи.

• *фильтрация речи* • *метод спектрального вычитания* • *качество речи*

The paper assesses the effectiveness of the noise reduction algorithms, based on spectral subtraction and signal processing in subspaces. Studies are carried out with use of additive noise of different nature and intensity. The algorithms are evaluated in the context of the quality of speech processing. Standard methodologies and indicators such as the spectrum distortion and speech intelligibility index are used.

• *speech enhancement* • *the method of spectral subtraction* • *speech quality*

Введение

В настоящее время значительное внимание уделяется задаче выделения речи на фоне аддитивного шума [1–3]. Столь устойчивый интерес обусловлен широким кругом возможных применений и ограничениями существующих алгоритмов. Речевые сигналы, зарегистрированные в своей естественной обстановке, часто имеют невысокое качество звучания из-за мешающего воздействия. Это также значительно усложняет процесс автоматической обработки речи в задачах её распознавания, идентификации диктора, кодирования и т.д. Для достижения эффективности такой обработки зашумленная речь, как правило, должна быть очищена от посторонних звуков алгоритмами шумоподавления [4]. Оценка качества отфильтрованной речи

должна выполняться с использованием ряда объективных и субъективных показателей. На выбор алгоритма также влияет уровень присутствующего шума, который, в зависимости от приложения, может варьироваться от экстремального до умеренного (отношение сигнал/шум от -10 до 10 дБ).

Большинство существующих алгоритмов шумоподавления работают в частотной области, используя вариации метода спектрального вычитания [1, 2, 5, 6]. К сожалению, его негативной особенностью является появление в реконструированном речевом сигнале искажений, известных как «музыкальные тона». Было предложено много подходов, чтобы устранить этот феномен, включая перцептуально мотивированные [7,8], но их оптимальность в смысле линейной оценки неоднозначна. Подход обработки зашумленного речевого сигнала в подпространствах является интересным обобщением методов спектрального взвешивания. Данная техника первоначально была предложена в [9]. Оценка речи здесь рассматривается как задача оптимизации с ограничениями, где искажения речевого сигнала минимизируются с учётом остаточной мощности шума.

В данной статье рассматриваются оба указанных подхода. Для каждого из них исследуется зависимость объективных показателей качества фильтрации речи от интенсивности аддитивной помехи.

Описание алгоритмов

I. Перцептуальный алгоритм на основе спектрального вычитания

Методы спектрального вычитания были в числе первых средств, предложенных для фильтрации речи на фоне шума [1]. Достоинства таких средств — простота реализации и широкий спектр возможных применений. Основной идеей в них является вычитание спектральной плотности мощности (СПМ) шума из СПМ обрабатываемого сигнала. СПМ шума оценивается в паузах, где речевой сигнал отсутствует. Среди предложенных модификаций этого подхода можно выделить следующие: нелинейное спектральное вычитание (Non linear Spectral Subtraction) [10], субполосное спектральное вычитание (Multiband Spectral Subtraction) [11, 12], спектральное вычитание с оптимизацией параметров по минимуму среднеквадратичной ошибки (Minimum Mean Square Error Spectral Subtraction) [13,14,], избирательное спектральное вычитание (Selective Spectral Subtraction) [15], спектральное вычитание с перцептуальной оценкой вносимых искажений (Spectral Subtraction based on perceptual Properties) [7,16,17].

В алгоритмах этого класса речевой сигнал и аддитивная помеха считаются статистически независимыми и стационарными в широком смысле. В процессе обработки принятый зашумлённый сигнал разбивается на перекрывающиеся фреймы. Каждый входной фрейм умножается на оконную функцию и преобразуется в частотную область. Спектр выходного сигнала получается путём умножения текущего амплитудного спектра входного сигнала на действительные коэффициенты $H(\omega)$ фильтра. Следует заметить, что фаза сигнала не изменяется. После оценки спектра отфильтрованной речи сигнал преобразуется обратно во временную область. Слабым местом алгоритмов этого класса является остаточный шум, воспринимаемый на слух как случайные музыкальные тона (вследствие того, что $H(\omega)$ на некоторых частотах оказывается близкой к нулю).

Рассмотрим модификацию алгоритма спектрального вычитания на основе принципов психоакустики таким образом, чтобы оставить возникающие музыкальные тона ниже порога маскирования и так повысить субъективное качество обработанного сигнала. Для модификации используем преобразование Фурье с неравномерным частотным разрешением (warped discrete Fourier transform — WDFT) [8, 18, 19]. Обработка сигнала при этом выполняется в критических частотных полосах и более точна в контексте психоакустического моделирования по сравнению со схемами на базе ДПФ.

Структура такого алгоритма подавления шума показана на рисунке 1.

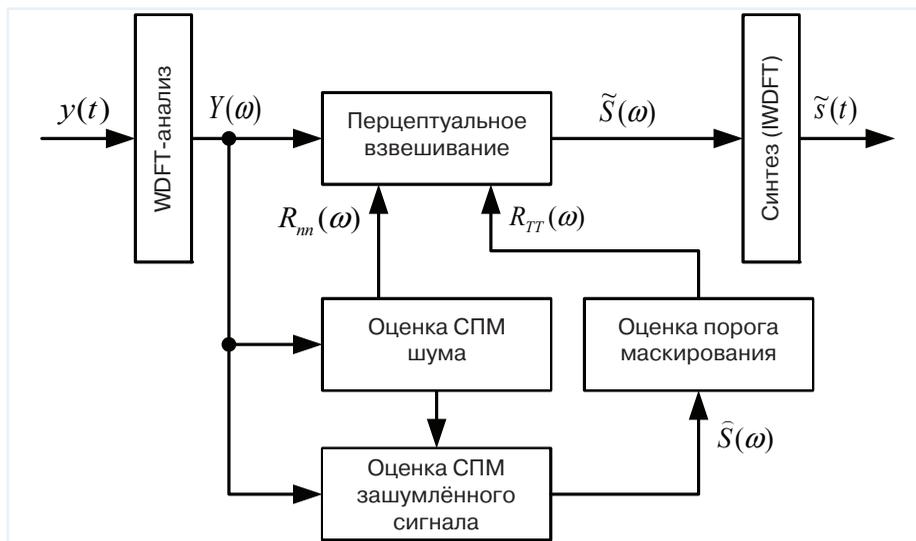


Рис. 1. Структура перцептуального алгоритма подавления шума на базе WDFFT

В ряде случаев, например, в системах телефонной связи, полное удаление шума нежелательно. Для того чтобы сохранить характеристики окружающего шума, необходимо определить комфортный уровень остаточного шума ζ_n [7]. Тогда разница между желаемым спектром отфильтрованного речевого сигнала и его оценкой может быть определена как

$$Q(\omega) = S(\omega) + \zeta_n N(\omega) - H(\omega) [s(\omega) + N(\omega)].$$

где $S(\omega)$ и $N(\omega)$ — спектры речевого сигнала и шума соответственно.

Так как речевой сигнал и шум статистически независимы, то СГМ этой разности $R_{qq}(\omega)$ может быть выражена следующим образом:

$$R_{qq}(\omega) = R\{|Q(\omega)|^2\} = [1 - H(\omega)]^2 R_{ss}(\omega) + [\zeta_n - H(\omega)]^2 R_{nn}(\omega),$$

или

$$R_{qq}(\omega) = R_{qsqs}(\omega) + R_{qnqn}(\omega),$$

где $E\{\}$ — операция вычисления математического ожидания, $R_{ss}(\omega)$ и $R_{nn}(\omega)$ — СГМ речевого сигнала и шума соответственно,

$R_{qsqs}(\omega)$ и $R_{qnqn}(\omega)$ — СГМ искажения речевого сигнала и шума соответственно.

Для минимизации $R_{qq}(\omega)$ в перцептуальном смысле необходимо, чтобы уровень искажений не воспринимался на слух. В идеальном случае все искажения должны быть замаскированы. Тем не менее, в большинстве реальных систем это требование не может быть удовлетворено, так как минимум $R_{qq}(\omega)$ может быть больше, чем СГМ порога маскирования $R_{TT}(\omega)$. Поэтому критерий минимизации формулируется так [7]:

$$R_{qnqn}(\omega) = [\zeta_n - H(\omega)]^2 R_{nn}(\omega) = R_{TT}(\omega).$$

Решая уравнение(4) относительно $H(\omega)$, находим весовую функцию

$$H^{IND}(\omega) = \min \left\{ 1, \sqrt{\frac{R_{TT}(\omega)}{R_{nn}(\omega)} + \zeta_n} \right\}.$$

где IND (Inaudible Noise Distortion) — неслышимое шумовое искажение.

Легко заметить, что если остаточный шум лежит ниже порога маскирования, выражение под корнем больше единицы и речь не искажается, так как

$H^{IND}(\omega) = 1$. В противном случае, окружающий шум оптимально ослабляется до уровня, не воспринимаемого на слух человеком. Заметим, что оценки порога маскирования и СПМ шума необходимы только для вычисления взвешивающих коэффициентов. Оценка порога маскирования вычисляется согласно психоакустическим моделям [19–21]. Обозначим далее этот алгоритм как NRS.

II. Психоакустически мотивированный алгоритм на основе обработки сигнала в подпространствах

Существует два основных способа определения линейного фильтра для обработки речевого сигнала в подпространствах [9]: во временной области (time-domain-constrained — TDC) и в спектральной области (spectral-domain-constrained — SDC). Входной зашумленный сигнал разделяется на подпространство речи и подпространство шума с использованием преобразования Карунена-Лозева (Karhunen-Loeve Transform — KLT), затем в подпространстве речи выполняется спектральное вычитание. Компоненты входного сигнала, которые проецируются на подпространство шума, просто обнуляются, что приводит к значительно более высокому качеству выделенной речи по сравнению с обычными методами, где обрабатывается спектр сигнала в полосе его пропускания.

К сожалению, эффективная реализация методов, основанных на KLT, является трудной задачей и часто на практике существенно упрощается. Например, в традиционных подходах [9], предполагается, что шум является белым. В случае же окрашенного шума, в первую очередь, предлагается выбеливать входной сигнал. Оптимальность фильтрации не гарантируется, поскольку к минимуму сводятся искажения скорректированной речи, а не исходной. Методы из [22, 23] решают проблему окрашенного шума с помощью аппроксимации его ковариационной матрицы, но фактически также сходятся к субоптимальным операторам.

Основная трудность в интеграции психоакустики и методов, основанных на KLT, заключается в том, что свойства слуха (т.е. маскирующие эффекты) необъяснимы в области разложения по собственным векторам. В [24] были предложены соответствующие преобразования, чтобы перейти к порогу маскирования в области KLT и наоборот. В этом способе используется психоакустически мотивированное правило взвешивания, но проблема окрашенного шума решается так же, как и в [22]. Подходы [25, 26] используют совместно диагонализацию матриц ковариации речи и шума, что позволяет сделать оптимальным оператор фильтрации для окрашенного шума. К сожалению, аналитические выражения вида как в [26] для этих операторов весьма непрактичны. На самом деле они связаны с множителями Лагранжа, которые должны быть заданы особым образом, чтобы получить требуемый фильтр. В общем случае аналитические выражения для этих множителей неизвестны. В [25] множителям Лагранжа было просто задано фиксированное значение, что привело к обычному фильтру Винера.

В работе [27] предложен перцептуально мотивированный алгоритм подавления шума на основе обработки сигнала в подпространствах (perceptually constrained signal subspace — PCSS), основанный на модифицированном SDC операторе. Решение представлено в новой форме, которое делает реализацию оператора более надёжной. В отличие от других подходов, такой способ использует перцептуально мотивированное построение огибающей остаточного шума и накладывает ограничения строго в частотной области, применяя базисные векторы ДПФ. Остаточные уровни шума устанавливаются чуть ниже порога маскирования для ослабления только слышимой компоненты шума. Так как множители Лагранжа используются в выражении для модифицированного SDC оператора, они должны быть точно установлены для данного набора остаточных уровней шума. Эти множители независимы друг от друга и могут быть вычислены численно. В [27] также предлагается версия способа PCSS с низкой вычислительной сложностью. Учитывая все вышеперечисленные особенности, в данной статье используется реализация именно этого способа обработки зашумленного сигнала в подпространствах.

Входной сигнал делится на фреймы длиной N с перекрытием N_0 отсчётов. Каждый фрейм разбивается на $m = N - k$ меньших перекрывающихся k -мерных векторов. Определим вектор x_l внутри фрейма следующим образом:

$$x_l = \begin{bmatrix} x(l(N - N_0) + t + 1) \\ \vdots \\ x(l(N - N_0) + t + k) \end{bmatrix},$$

где l — индекс фрейма. Последовательность этих векторов можно рассматривать как траектории в k -мерном евклидовом пространстве. Такая последовательность организована в так называемую матрицу траекторий размера $k \times m$:

$$X^{(l)} = [x_1, x_2, \dots, x_m].$$

Векторное произведение матрицы траекторий затем используется для вычисления значений ковариационной матрицы входного сигнала

$$C_x^{(l)} = \frac{1}{m} X^{(l)} (X^{(l)})^T.$$

Эта оценка является основой для расчёта структур собственных векторов шума (только в паузах речи) и KLT выделенного сигнала, соответственно:

$$C_n \approx U_n \Lambda_n U_n^*,$$

$$C_y \approx \sqrt{C_n C_x} \sqrt{C_n} - I \approx U \Lambda U^*.$$

В последней формуле опущен индекс фрейма l для краткости. Чтобы избежать численных проблем, квадратные корни из матриц рассчитываются с использованием структуры собственных векторов U_n , Λ_n , ковариационной матрицы шума.

Упрощённая схема алгоритма обработки показана на рисунке 2 [27]. Сначала вычисляется некоторый эффективный фильтр, а затем все векторы фрейма обрабатываются с помощью той же матрицы. Результат сохраняется в матрице траекторий отфильтрованного речевого сигнала. Обработанные векторы получают из этой матрицы, используя технику диагонального усреднения [28]. Выходной речевой сигнал синтезируется с помощью метода перекрытия с суммированием с использованием временного окна Хеннинга.

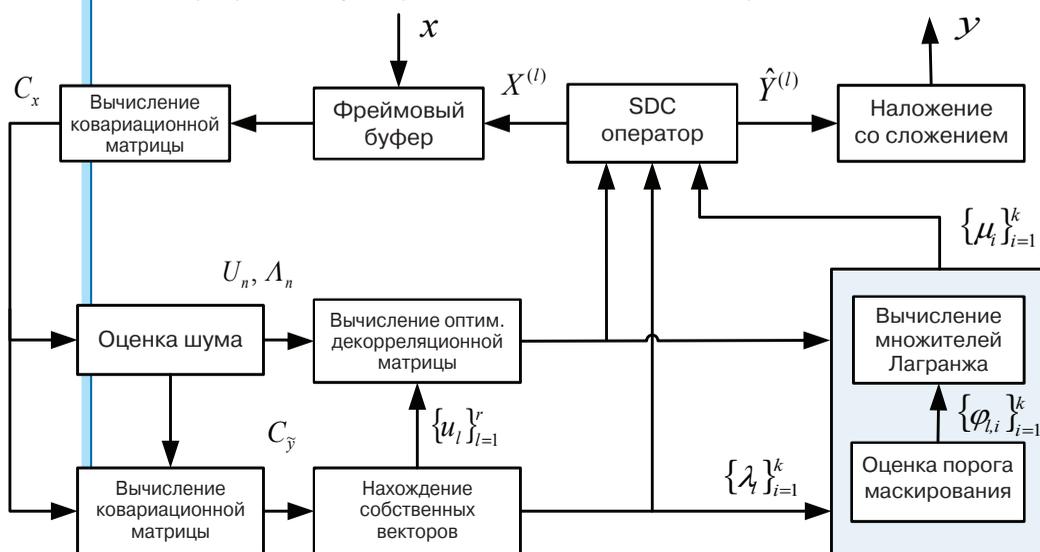


Рис. 2. Структура алгоритма по методу PCSS

Для вычисления эффективного фильтра необходимо множество неортогональных проекций, собственные значения выделенной чистой речи и множители Лагранжа. Множители рассчитываются итеративно по методу Ньютона. Поскольку ковариационная матрица шума считается положительно определённой, то максимальный уровень остаточного шума всегда больше нуля. Если это не так, данная матрица может быть реализована путём добавления маленькой положительной константы к оценённым собственным значениям.

Вычислительная сложность и эффективность метода PCSS зависят от нескольких факторов: модели данных (высокий/низкий ранг), статистики сигнала и схемы обработки. Наиболее затратная с точки зрения времени выполнения операция алгоритма — вычисление KLT и эффективная реализация фильтров SDC. Сложность других частей (оценка шума, психоакустическая модель и т.д.) относительно мала, и ею можно пренебречь. Матрица KLT обычно получается при помощи собственного разложения (ED — eigen decomposition) ковариационной матрицы. Вычислительная сложность ED равна $O(k^2)$, где k — размерность модели данных. Можно использовать любую процедуру отслеживания подпространств, т.е. алгоритмы аппроксимации проекций подпространств вместо ED. Такой подход зависит от данных, и в худшем случае его сложность такая же, как у процедуры ED, а общая производительность алгоритма, как правило, ухудшается. С другой стороны, в данном подходе KLT можно аппроксимировать дискретным косинусным преобразованием. Подобная структура, безусловно, является субоптимальной, но и требует меньше вычислительных затрат.

Условия экспериментов

Формирование тестовых аддитивных смесей, фильтрация сигнала и вычисление показателей качества проводились на персональном компьютере в среде моделирования Matlab. Обрабатываемые звуковые файлы были представлены в формате WAV при частоте дискретизации 8000 Гц и разрядности квантования 16 бит. В качестве речевого сигнала использовались фонограммы фразовых таблиц, начитанные тремя мужчинами и тремя женщинами, общей длительностью около 4 минут.

Акустические помехи представляли собой синтезированные стационарные белый и розовый шумы (БШ и РШ соответственно). Помеха добавлялась аддитивно с таким уровнем, чтобы обеспечить значения отношения «сигнал/шум» ($OS\mathcal{S}_x$), равные -10 , 0 , 10 и 20 дБ. При вычислении мощности речевого сигнала использовалось так называемое значение уровня активного речевого сигнала, определяемое в соответствии со стандартом ITU-T P. 56 «Объективное измерение уровня активных речевых сигналов»:

Для оценки эффективности фильтрации речи с помощью изложенных выше алгоритмов шумоподавления применялись следующие объективные показатели:

1) среднее отношение «сигнал/шум» на выходе:

$$SNR_y = \frac{1}{M} \sum_{m=1}^M 101g \left[\frac{\sum_{i=1}^N y_m^2(i)}{\sum_{i=1}^N [y_m(i) - S_m(i)]^2} \right];$$

2) среднее подавление помехи:

$$K_n = \frac{1}{M} \sum_{m=1}^M 101g \frac{P_{xm}}{P_{ym}} = \frac{1}{M} 101g \frac{\sum_{i=1}^N x_m^2(i)}{\sum_{i=1}^N y_m^2(i)},$$



где оценки текущей мощности входного и выходного сигналов рассчитываются на каждом m -ом отрезке длиной N отсчётов, не являющемся паузой (паузы определяются по чистому речевому сигналу с помощью детектора речи);

3) искажения спектра барков:

$$MBSD = \frac{1}{M} \sum_{m=1}^M \sum_{k=1}^K B(k) |G_{sm}(k) - G_{ym}(k)|^2,$$

где K — число полос;

$B(k)$ — показатель присутствия искажений в k -й полосе (равен 0, когда искажения в полосе не воспринимаются на слух, и равен 1 в противном случае); $G_{sm}(k)$ и $G_{ym}(k)$ — барк спектры m -го фрейма исходного речевого и выходного сигналов.

Данный показатель учитывает тот факт, что слух человека обладает неодинаковой чувствительностью на разных частотах [29–31];

4) перцептуальная оценка PESQ.

Алгоритм PESQ (Perceptual Evaluation of Speech Quality) представляет собой объективную методику определения качества речевых сообщений, передаваемых по каналам связи, которая прогнозирует результаты субъективной оценки качества слушателями-экспертами [32]. Для получения такой оценки в экспериментах использовалось специальное программное обеспечение, реализующее методику из [32]. Для определения качества фильтрации осуществляется сравнение входного и выходного сигналов. Результатом этого сравнения является оценка PESQ качества речевого сообщения, которая аналогична усреднённой субъективной оценке MOS (Mean Opinion Score);

5) индекс разборчивости речи SII.

Методика вычисления индекса разборчивости речи SII основана на подходах, которые изложены в [33–37]. В соответствии с ней весь частотный диапазон речевого сигнала разделяется на фиксированное количество полос. В каждой полосе рассчитывается отношение «сигнал/шум» и определённым образом взвешивается. Значение SII может изменяться в пределах от 0 до 1. Его интерпретируют следующим образом:

- ниже 0,3 соответствует плохой разборчивости;
- от 0,3 до 0,5 — удовлетворительная разборчивость;
- от 0,5 до 0,7 — хорошая разборчивость;
- выше 0,7 — очень хорошая разборчивость.

Результаты экспериментов

На рисунке 3 изображены экспериментальные оценки SNR_y , полученные для случаев воздействия БШ и РШ и усреднённые для мужских и женских голосов. При отрицательных значениях ОСШ на входе алгоритм NRS обеспечивает более высокий выигрыш, чем алгоритм PCSS. Однако при положительных ОСШ наблюдается противоположный эффект, обусловленный тем, что в алгоритме PCSS проводится более тонкая настройка на полезную часть сигнала, что сделать при интенсивной помехе затруднительно.

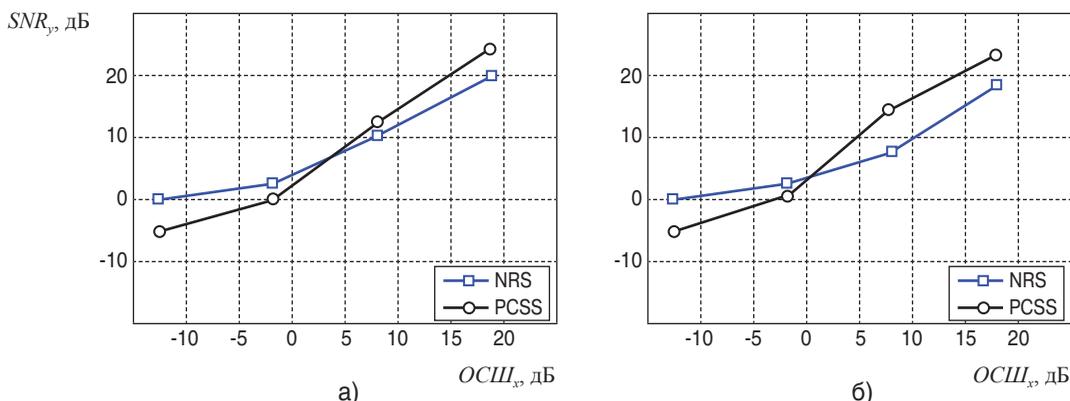


Рис. 3. Зависимости ОСШ на выходе алгоритмов от входного ОСШ: а) БШ; б) РШ

Величина K_n подавления шума в алгоритме PCSS значительно выше и достигает значения 30 дБ. Это объясняется тем, что в нём реализуется практически полное подавление помехи в паузах речи, чего не наблюдается в алгоритме NRS.

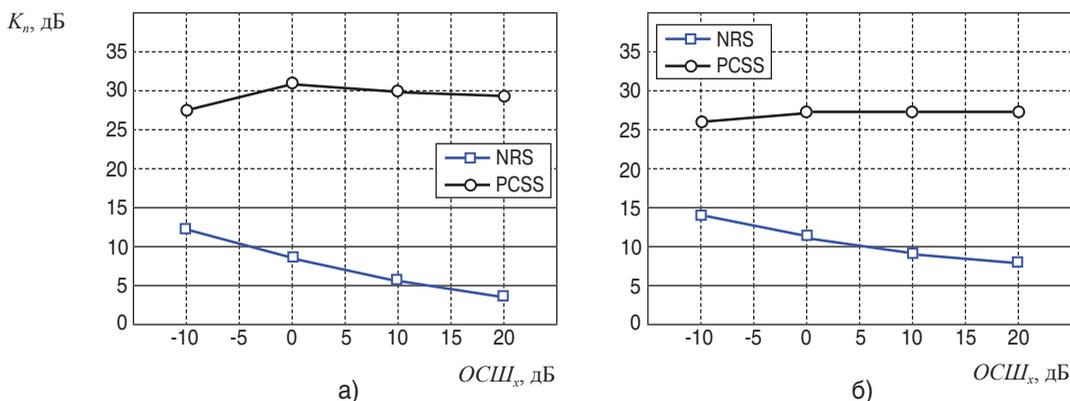


Рис. 4. Зависимости подавления помехи от входного ОСШ: а) БШ; б) РШ

Для иллюстрации особенностей работы метода PCSS на рисунке 5 изображены временные реализации отрезков входного и выходного сигналов. Из рисунка хорошо видно, что шумовая часть сигнала после обработки полностью обнулена. В алгоритме NRS после фильтрации практически всегда остаётся шум небольшого уровня, комфортного для восприятия слухом человека.

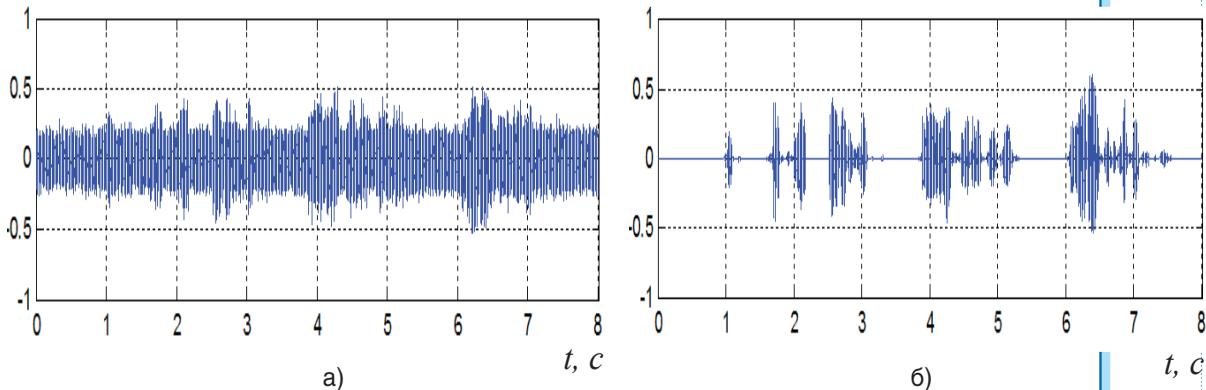


Рис. 5. Примеры реализации сигналов для алгоритма PCSS: а) входной сигнал; б) выходной сигнал

Полученные оценки искажения спектра барков для обоих алгоритмов показаны на рисунке 6. Из полученных результатов видно, что алгоритмы NRS и PCSS позволяют значительно уменьшить величину MBSD по сравнению с входным сигналом, особенно при отрицательных ОСШ.

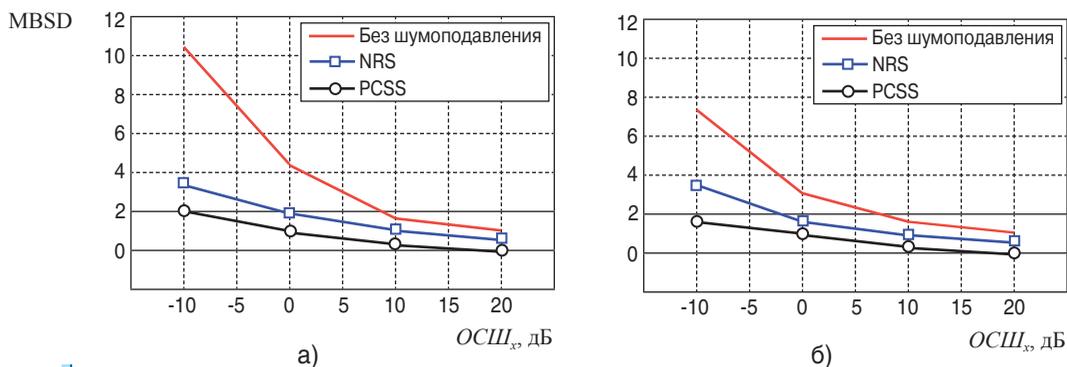


Рис. 6. Зависимости оценки MBSD от входного ОСШ: а) белый шум; б) розовый шум

Результаты оценки PESQ показаны на рисунке 7. Как видно, применение обоих алгоритмов шумоподавления при входном ОСШ выше — 5 дБ позволяет улучшить перцептуальное качество выходной речи.

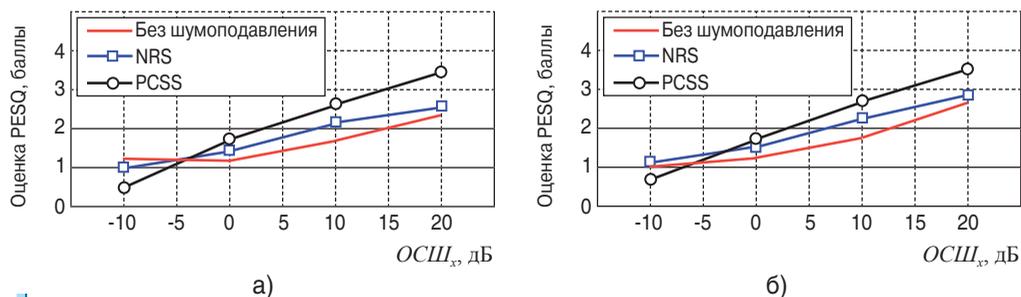


Рис. 7. Зависимости оценки PESQ от входного ОСШ: а) белый шум; б) розовый шум

На рисунке 8 показаны полученные оценки индекса разборчивости SII. Из них видно, что даже при очень интенсивном шуме (ОСШ= — 10 дБ) алгоритмы позволяют увеличить разборчивость с уровня «Очень плохая разборчивость» до уровня «Удовлетворительная разборчивость». Однако алгоритм NRS в целом оказывается немного более эффективным, чем PCSS.

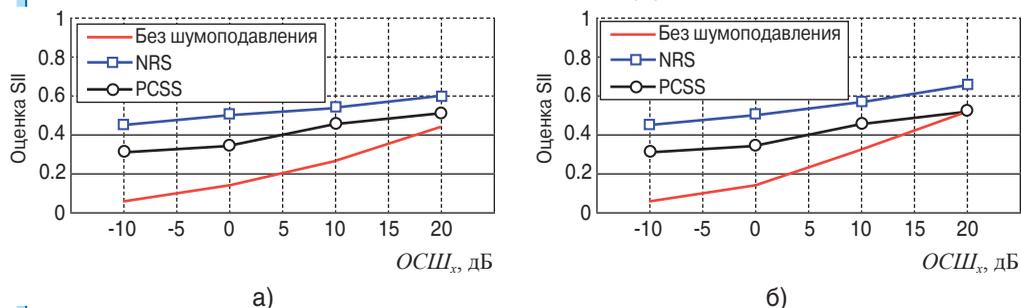


Рис. 8. Зависимости индекса разборчивости речи от входного ОСШ: а) белый шум; б) розовый шум

Заключение

Рассмотрено два алгоритма подавления аддитивного шума: алгоритм NRS на основе спектрального вычитания и алгоритм PCSS на основе обработки речевого сигнала в подпространствах. В обоих алгоритмах выполняется оценка шума исходя их перцептуально мотивированных критериев, что повышает субъективное качество отфильтрованной речи.

Полученные результаты обработки речи с аддитивными шумами различной интенсивности позволяют сделать вывод о приемлемой эффективности обоих алгоритмов. Показано, что их использование существенно увеличивает разборчивость речи в условиях шума повышенной интенсивности.

Эффективность работы алгоритмов различается в зависимости от уровня мешающего шума, но мало зависит от его окраски. Для сигналов с относительно слабым зашумлением ($ОСШ > 0$) целесообразно использовать алгоритм PCSS, при шумах большей мощности — алгоритм NRS.

Алгоритм PCSS обеспечивает значительно более высокое подавления шума вне зависимости от входного ОСШ, однако уступает по показателю разборчивости, а следовательно, сильнее искажает речь.

Список литературы

1. Phillips C Loizou. Speech enhancement theory and practice: 1st ed. Boca Raton, FL.: CRC, 2007. Releases Taylor & Francis.
2. J. Benesty, J. Chen, Y. Huang, I. Cohen Noise Reduction in Speech Processing // Springer-Verlag, 2009.
3. Анализаторы речевых и звуковых сигналов: методы, алгоритмы и практика (с MATLAB примерами); под редакцией д.т.н. профессора Петровского А.А. Минск: Бестпринт, 2009.
4. Gibak Kim, Phillips C Loizou. Why do speech-enhancement algorithms not improve speech intelligibility? // Processing of ICASSP-2010. Vol. 1. P. 397–400.
5. S.F. Boll. Suppression of acoustic noise in speech using spectral subtraction // IEEE Trans. on Acoustic, Speech, Signal Processing. 1979. Vol. 27. P. 113–120.
6. Sim B, Tong Y, Chang J., Tan C. A parametric formulation of the generalized spectral subtraction method // IEEE Trans. Speech Audio Process, 1998, 6(4). Pp.328–337.
7. Gustafson S., Jax P., Vary P. A novel psychoacoustically motivated audio enhancement algorithm preserving background noise characteristic. // Proc. of ICASSP-1998. Vol. 1. Pp. 397–400.
8. A. Petrovsky, M. Parfieniuk, A. Borowicz. Warped DFT based perceptual noise reduction system // AES116th Convention, 2004, May 8–11.
9. Ephraim Y., Van Trees H. A signal subspace approach for speech enhancement // IEEE Trans. Speech Audio Process, 1995. V.3 (4). Pp. 251–266.
10. Lockwood P., Boudy J. Experiments with a Nonlinear Spectral Subtractor (NSS), Hidden Markov Models and the projection, for robust speech recognition in cars // Speech Communication, 1992. V.11. Pp. 215–228.
11. Bielawski K., Petrovsky A.A. Speech enhancement system for hands-free telephone based on the psychoacoustically motivated filler bank with allpass frequency transformation // Proc. of EuroSpeech-1999. Pp. 2555–2558.
12. Петровский А.А., Белявский К., Петровский А.А. Перцептуальное кодирование аудио и речевых сигналов: Доклады БГУИР. 2004. № 1(5). С. 73–91.
13. Ephraim Y. and Malah D. Speech Enhancement Using minimum mean square error short-time spectral amplitude estimator // IEEE, Trans. on Audio, Speech, Signal Process. V. 6(4). Pp. 328–337.
14. Martin R. Speech Enhancement Using MMSE Short Time Spectral Estimation with Gamma Distributed Speech Priors. // Proc. of ICASSP-2002. Vol. 1. Pp. 253–256.



15. He C, and Zweig G. Adaptive two band spectral subtraction with multiwindow spectral estimation // Proc. of ICASSP-1999. V.2. P. 793–796.
16. Virag N. Single channel speech enhancement based on masking properties of the human auditory system // IEEE. Trans. Speech Audio Process, 1999. V. 7(3). P. 126–137.
17. A.A. Petrovsky, A.E. Anoshenko. Combined system for echo cancelation and noise reduction in frequency domain with psychoacoustic motivation // The 2nd International conference and exhibition on Digital Signal Processing and its Applications, DSPA-1999, Moscow, Russia. Pp. 166–169.
18. A. Borowicz, M. Parfieniuk, A.A. Petrovsky. An application of the warped discrete Fourier transform in the perceptual speech enhancement. // Speech Communication, 2006. V. 48. P. 1024–1036.
19. А.А. Петровский, А. Борович, М. Парфенюк. Обработка речи на основе дискретного преобразования Фурье с неравномерным частотным разрешением // Речевые технологии, 2008. № 3.
20. J.D. Johnston. Transform coding of audio signals using perceptual noise criteria // IEEE Transactions on Selected Areas Communication, 1988. V. 6. P. 314–323.
21. Parfieniuk M., Petrovsky A. Warped DFT as the basis for psychoacoustical model // Proc. of ICASSP-2004. V. 4. P. 185–188.
22. Mittal P., Phamdo N. Signal/noise KLT based approach for enhancing speech degraded by colored noise. // IEEE Trans. Speech, audio process, 2000. V.8(2). P. 159–167.
23. Rezayee A., Gazor S. An adaptive KLT approach for speech enhancement. // IEEE Trans. Speech, Audio Process, 2001. V.9 (2). P. 87–95.
24. Jablom F., Champagne B. Incorporating the human hearing properties in the signal subspace approach for speech enhancement // IEEE Trans. Speech, Audio Process, 2003. V. 11 (6). P. 700–708.
25. Hu Y., Loizou P. A generalized subspace approach for enhancing speech corrupted by colored noise. // IEEE Trans. Speech, Audio Process, 2003. V. 11 (4). P. 334–341.
26. Lev-Ari H., Ephraim Y. Extension of the signal subspace enhancement to colored noise // IEEE Sign. Process. Lett., 2003. № 10 (4). P. 104–106.
27. Borowicz A., Petrovsky A. Signal subspace approach for psychoacoustically motivated speech enhancement // Speech communication, 2011. P. 210–219.
28. Vetter R., Virag N., Renevey P., Vesin J. Signal channel speech enhancement using principal component analysis and MDL subspace selection // Proc. of EuroSpeech-1999. P. 2411–2414.
29. Yang W. Performance of the modified bark spectral distortion as an objective speech quality measure // Proc. of ICASSP-1998. P. 541–544.
30. Аношенко А.Е., Петровский А.А. Метод подавления эхо сигнала и шумов окружающей среды на основе спектрального вычитания с психоакустической мотивацией // 4-я Международная конференция «Цифровая обработка сигналов и её применение», 2002. С. 450–453.
31. Shihua Wang, Andrew Sekey, Allen Gersho. An Objective Measure for Predicting Subjective Quality of Speech Coders // IEEE Journal on Selected Areas in Communications, 1992. V.10, № 5.
32. ITU-T Recommendation P. 862, PESQ an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs, February 2001.
33. Programs for calculating the Speech Intelligibility Index (SII) — <http://www.sii.to/html/programs.html>

34. Продеус А.Н., Гавриленко А.В., Дидковский В.С. Сравнительный анализ некоторых методов оценки разборчивости речи. // Сб. трудов Акустического симпозиума «Консонанс-2007». С. 273–278.
35. Kamm C.A., Dirks D.D., Bell T.S. Speech recognition and the Articulation Index for normal and hearing-impaired listeners // J. Acoust. Soc. Am., 1985. V. 77. P. 281–288.
36. Алдошина И. Основы психоакустики. Субъективные и объективные методы оценки разборчивости речи // Звукорежиссёр, 2002. — www.kadva.ru/files/edu/AldoshinaPsychoacoustics.pdf
37. ANSI S3.5–1997 (R2007) Methods for Calculation of the Speech Intelligibility Index, American National Standards Institute, 1997.

Сведения об авторах:

Петровский Александр Александрович —

доктор технических наук, профессор. Работает в Учреждении образования «Белорусский государственный университет информатики и радиоэлектроники», кафедра «Электронные вычислительные средства». Окончил Белорусский государственный университет информатики и радиоэлектроники по специальности «Электронные вычислительные машины». Главные научные интересы лежат в области цифровой обработки сигналов речи и звука для целей компрессии, распознавания, редактирования шума, а также в области проектирования проблемно-ориентированных средств вычислительной техники реального времени для систем мультимедиа. Член НТО РЭС им. А.С. Попова, IEEE, EURASIP, AES.

Азаров Илья Сергеевич —

кандидат технических наук, Белорусский государственный университет информатики и радиоэлектроники (БГУИР), кафедра электронных вычислительных средств. Окончил Белорусский государственный университет, механико-математический факультет, отделение математической электроники. Область научных интересов: цифровая обработка речевых сигналов.

Лихачёв Денис Сергеевич —

кандидат технических наук, доцент, Белорусский государственный университет информатики и радиоэлектроники (БГУИР), кафедра электронных вычислительных средств, г. Минск. Закончил БГУИР, факультет компьютерного проектирования. Область научных интересов: цифровая обработка речевых сигналов, системы компрессии речи, антропоморфическая обработка речи, конверсия голоса.

Ромашкин Юрий Николаевич —

кандидат технических наук, Окончил Московский инженерно-физический институт, факультет «Автоматика и электроника». Область научных интересов: цифровая обработка речевых сигналов, фильтрация речи на фоне помех, автоматическое распознавание речи и языка, идентификация говорящего по голосу, низкоскоростное кодирование речи, оценка качества трактов речевой связи. E-mail: romayn@yandex.ru

Жигулёвцев Юрий Николаевич —

кандидат технических наук, старший научный сотрудник, доцент МГТУ им. Н.Э. Баумана. Окончил в 1969 г. МГТУ им. Н.Э. Баумана по специальности «Системы автоматического управления». Автор более 80 научных публикаций, 6 авторских свидетельств на изобретения, соавтор 2 монографий. Область научных интересов: методы и средства построения систем речевого взаимодействия.

Харламов Александр Александрович —

доктор технических наук, старший научный сотрудник Института высшей нервной деятельности и нейрофизиологии РАН. Область научных интересов: нейроинформатика, распознавание речи, анализ текстов, распознавание изображений, семантические представления, искусственные нейронные сети.