Психоакустически мотивированный алгоритм фильтрации шума окружающей среды на основе обработки речевого сигнала в подпространствах

Борович А., *доктор-инженер*

Петровский А.А., *доктор технических наук, профессор*

В данной работе предложен новый перцептуально мотивированный метод и алгоритм подавления шума окружающей среды на основе обработки речевого сигнала в подпространствах (PCSS), ядром которого является модифицированный оператор SDC.

• *фильтрация шума* • *преобразование Кархунена-Лоэва* • *речевой сигнал*

In this paper the perceptually motivated signal subspace method and algorithm for speech enhancement (perceptually constrained signal subspace (PCSS) based on the extended spectral-domain-constrained (SDC) estimator are proposed.

• *speech enhancement* • *Karhunen-Loeve transform (KLT)* • *speech signal*

Введение

Существует острая необходимость в разработке эффективных алгоритмов подавления шума в устройствах обработки речевых сигналов, работающих в шумах умеренной интенсивности (при отношениях сигнал-шум, близких к 0 дБ). С одной стороны, большинство существующих одноканальных алгоритмов подавления шума работают в частотной области и используют вариации метода спектрального взвешивания [1]. К недостаткам этих алгоритмов следует отнести появление в отфильтрованном речевом сигнале искажений, известных как «музыкальные тона». Много подходов было предложено, чтобы устранить этот недостаток, включая перцептуально мотивированные подходы [2–4], но их оптимальность в смысле линейной оценки не явна.

С другой стороны, подход обработки зашумленного речевого сигнала в подпространствах (signal subspace, SS) для фильтрации шума — это интересное обобщение методов спектрального взвешивания. Данная техника первоначально была предложена в [5]. Оценка речи здесь рассматривается как задача оптимизации с ограничениями, где искажения речевого сигнала минимизируются с учётом остаточной мощности шума, определяемой в соответствующем подпространстве. Было предложено два линейных оператора фильтрации: во временной (time-domain-constrained, TDC) и в спектральной областях (spectral-domain-constrained, SDC). В отличие от методов, основанных на дискретном преобразовании Фурье (ДПФ), SS подход разделяет зашумленный речевой сигнал на подпространство чистого речевого сигнала и подпространство шума, используя преобразование Кархунена-Лоэва (Karhunen-Loeve, KLT). При этом спектральное взвешивание выполняется только в подпространстве речевого сигнала, а компонента шума аддитивной смеси проецируется на подпространство шума, которое потом просто обнуляется. Это приводит к значительно более высокому качеству выделения речевого сигнала по сравнению с обычными методами, работающими в частотной области, где обрабатывается спектр сигнала во всём частотном диапазоне.

К сожалению, эффективная реализация методов, основанных на KLT, является трудной задачей и на практике часто существенно упрощается. Например, в традиционных подходах [5] предполагается, что шум является белым. В случае же цветного шума, в первую очередь, предлагается отбеливать зашумлённый речевой сигнал. В таком случае оптимальность оператора фильтрации не гарантируется, потому что к минимуму сводятся искажения отбелённого речевого сигнала, а не чистой речи. Другие методы [6,7] решают проблему цветного шума с помощью аппроксимации ковариационной матрицы шума, но фактически также сходятся к субоптимальным операторам.

Другие SS подходы так же, как и в [5], выполняют построение огибающей остаточного шума в области разложения по собственным векторам на основе обобщённого правила Винера. Такая методика зависит от ошибок в оценке отношения «сигнал-шум» и не является оптимальной с точки зрения перцетуально мотивированного подхода (в результате остаточный шум не может быть замаскирован корректно). Однако основная трудность в интеграции психоакустики и методов, основанных на KLT, состоит в том, что свойства слуха (т.е. маскирующие эффекты) необъяснимы в области разложения по собственным векторам. В [8] были предложены соответствующие преобразования, чтобы перейти к порогу маскирования в области KLT и наоборот. В этом методе используется психоакустически мотивированное правило взвешивания, но проблема цветного шума решается так же, как и в [7].

Расширенные подходы [9,10] используют совместно диагонализацию матриц ковариации речи и шума, что позволяет сделать оптимальный оператор фильтрации для цветного шума. К сожалению, аналитические выражения вида [10] для этих операторов весьма непрактичны. На самом деле они связаны с множителями Лагранжа, которые должны быть заданы деликатно, чтобы получить требуемый фильтр. Однако в общем случае аналитические выражения для этих множителей неизвестны. В [9] множителям Лагранжа было просто задано фиксированное значение, что привело к обычному правилу взвешивания Винера.

Основная задача данной работы заключается в использовании маскирующих свойств в независимом от шума SS подходе. Для повышения качества речевого сигнала здесь предлагается перцептуально мотивированный метод и алгоритм подавления шума окружающей среды на основе обработки речевого 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, предложенный метод использует перцепту-

ально мотивированное построение огибающей остаточного шума и накладывает ограничения строго в частотной области, применяя базисные векторы дискретного преобразования Фурье (ДПФ).

А именно 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 определены для данного набора остаточных уровней шума. Однако было установлено, что эти множители независимы друг от друга и могут быть вычислены численно. Кроме того, в работе в качестве альтернативного решения предлагается версия метода PCSS с низкой вычислительной сложностью.

1. Фильтрация шума на основе подхода обработки речевого сигнала в подпространствах

Модель зашумлённой речи, которая используется в SS методе, предполагает, что речь и шум являются аддитивными. Пусть $x = y + n$ обозначает k-мерный вектор зашумленной речи, где y и n — случайные векторы с нулевыми средними значениями, представляющие речевой сигнал и шум окружающей среды соответственно. Поскольку сигналы речи и шума считаются некоррелированными, ковариационная матрица аддитивной смеси R_x может быть записана в виде:

$$R_x = R_y + R_n , \tag{1}$$

где R_y и R_n — ковариационные матрицы речи и шума соответственно. Предполагается также, что матрица R_n положительно определена. Пусть $\hat{y} = Hx$ будет линейным оператором оценки речи. Эффективный фильтр H находится при минимизации средней мощности искажений речи и ограничении уровня мощности остаточного шума. Вектор ошибки определяется следующим образом:

$$\epsilon = \hat{y} - y = (H - I)y + Hn = \varepsilon_y + \varepsilon_n , \tag{2}$$

где ε_y и ε_n интерпретируются как векторы искажений речи и остаточного шума соответственно. Средний уровень искажения речи определяется по формуле:

$$\overline{\epsilon_y^2} = \frac{1}{k}\mathrm{tr}\, E\{\epsilon_y \epsilon_y^{\#}\} = \frac{1}{k}\mathrm{tr}\, E\{(\boldsymbol{H}-\boldsymbol{I})\boldsymbol{R}_y(\boldsymbol{H}-\boldsymbol{I})^{\#}\}, \tag{3}$$

где $E\{\cdot\}$ — оператор математического ожидания, $\mathrm{tr}\{\cdot\}$ — след матрицы, символ # обозначает транспонирование вещественной матрицы или сопряжённое транспонирование комплексной матрицы. Операторы фильтрации могут быть определены как во временной, так и спектральной областях, т.е. TDC и SDC операторы соответственно [10]. На самом деле TDC оператор является частным случаем SDC оператора. Ниже приводится краткое описание только оператора в частотной области. В этом случае задача оптимизации формулируется следующим образом:

$$\min_{\boldsymbol{H}} \overline{\epsilon_y^2} \quad \text{при условии: } E\{|v_i^{\#}\epsilon_y^{\#}|\} \le \alpha_i,\ i = 1, \ldots k, \tag{4}$$

где $\{v_i, i=1,..., k\}$) — множество k-мерных вещественных или комплексных векторов. Первоначально в [10] вид матрицы $V=[v_1, v_2, ..., v_k]$ был ограничен ортогональным или унитарным.

Решение (4) находится с использованием множителей Лагранжа следующим образом:

$$L(\boldsymbol{H}, \bar{\mu}) = \overline{\epsilon_y^2} + \sum_{i=1}^{k} \mu_i (\boldsymbol{v}_i^{\#} \boldsymbol{H} \boldsymbol{R}_n \boldsymbol{H}^{\#} \boldsymbol{v}_i - \alpha_i). \quad (5)$$

Полагая, что $M=k.diag\{\mu_1\mu_2,...\mu_k\}$ и $L=VMV^{\#}$. Из $\nabla_H L(H, \mu-)$ получим, что

$$\boldsymbol{L}\boldsymbol{H}\boldsymbol{R}_n + \boldsymbol{H}\boldsymbol{R}_y = \boldsymbol{R}_y. \quad (6)$$

Данное уравнение может быть решено итерационно, как это предлагается в [9]. Явное решение основано на факторизации матриц, которое преобразовывает совместно обе матрицы R_y и R_n в диагональную форму. Такое преобразование было показано в [9], где KLT сигнала было заменено неортогональным преобразованием. В [10] приводится эквивалентное решение с использованием отбеливающего подхода, а также приведены явные формы TDC и SDC операторов. А именно, KLT сигнала было заменено KLT отбеленной чистой речи. Поэтому собственное разложение ковариационной матрицы отбеленной чистой речи считается вместо матрицы R_y, т.е.

$$\boldsymbol{R}_{\tilde{y}} = E\{\tilde{\boldsymbol{y}}\tilde{\boldsymbol{y}}^T\} = \boldsymbol{R}_n^{-0.5} \boldsymbol{R}_y \boldsymbol{R}_n^{-0.5} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\#}, \quad (7)$$

где $\tilde{y}$ вектор отбеленной чистой речи, $U = [u_1,..., u_k]$ обозначает ортогональную матрицу собственных векторов, $\Lambda = diag\{\lambda_1,...\lambda_k\}$ — диагональная матрица соответствующих собственных значений (индекс $\tilde{y}$ опускается здесь для краткости). Пусть $Q=R_n^{-0.5}U$ и $Q^{\#}H(Q^{\#})^{-1}$. Подставляя эти соотношения в (6), получим

$$Q^{\#}L(Q^{\#})^{-1} G + G\Lambda = \Lambda. \quad (8)$$

В работе [10] предлагается следующая реализация:

$$H = R_n^{0.5} U \tilde{H} U^{\#} R_n^{-0.5}. \quad (9)$$

Столбцы матрицы $\tilde{H}$ определяются следующим образом:

$$h_l = T\lambda_l (R + \lambda_l I)^{-1} T^{-1} e_l, \quad l = 1,..., k, \quad (10)$$

где $T=U^{\#}R_n^{-0.5}V$ и e_l обозначает единичный вектор, для которого l-й элемент равен единице, а все остальные элементы равны нулю.

Отметим, что непосредственное применение фильтра (9) достаточно непрактично. Хотя подпространство сигнала небольшое, вычисление матрицы H всё ещё затратно, так как требуется знание полной матрицы U. Кроме того, H не диагональная матрица, подпространство разложения не является очевидным. Уравнение (9) является аналитическим выражением [10], но на самом деле речь идёт о наборе множителей Лагранжа, которые контролируют компромисс между искажением речи и остаточным шумом и должны быть тщательно подобраны для получения желаемого (возможно, психоакустически мотивированного) остаточного шума. Хотя установка фиксированных значений множителей даёт относительно хорошие результаты, но не может быть получен оптимальный результат с точки зрения акустического восприятия отфильтрованной речи. Как правило, ограничения остаточного шума определены в области собственных значений, в то время как маскирующие свойства вычисляются в частотной области.

В таком случае трудно использовать любое психоакустически мотивированное правило построения огибающей шума. Наконец, модифицированный SDC оператор, предполагающий операцию преотбеливания, которая является вычислительно затратной, может быть неэффективным для нестационарных шумов. Следует обратить внимание, что отбеливающие и неотбеливающие преобразования зависят от изменяющихся во времени характеристик шума. Обычно они могут просто вычисляться из ковариационной матрицы шума. Однако на практике эта матрица неизвестна и должна быть вычислена.

2. Перцептуально мотивированный метод фильтрации шума окружающей среды на основе обработки речевого сигнала в подпространствах

Принимая во внимание проблемы, кратко изложенные выше, предлагается новый перцептуально мотивированный метод и алгоритм подавления шума окружающей среды на основе обработки речевого сигнала в подпространствах (PCSS), ядром которого является оператор SDC. Модифицированный оператор SDC выбран потому, что он выполняет оптимальную декорреляцию в области преобразования и его эффективность не зависит от типа шума. Оптимальность преобразования KLT особенно важна для ослабления музыкального тона. Хотя основная обработка осуществляется в области KLT отбеленной речи, ограничения спектра остаточного шума могут быть определены в других областях, не обязательно связанных с KLT. Эта возможность была предложена в [10], но она не была рассмотрена на практике до сих пор.

2.1. Новая интерпретация SDC оператора

Как упоминалось ранее, прямая реализация фильтра (9) весьма непрактична, более того, разложение сигнала на подпространство речевого сигнала и подпространство шума не является очевидным. Однако, если матрица R_Y является положительной полуопределённой, то значения вектор-столбцов h_l, соответствующих нулевым собственным значениям, имеют все элементы равные нулю. Тогда (9) можно переписать следующим образом:

$$\boldsymbol{H} = \boldsymbol{R}_n^{0.5}\boldsymbol{U}\tilde{\boldsymbol{H}}\boldsymbol{U}^{\#}\boldsymbol{R}_n^{-0.5}. \qquad (11)$$

где r обозначает размерность SS. Параметр r обычно оценивается как число строго положительных собственных значений в соответствии со следующим правилом:

$$r = \mathrm{argmax}\{\lambda_l > \theta\}, \quad 1 \le l \le k. \qquad (12)$$

На практике порог θ обычно задаётся как некоторая малая положительная величина, чтобы избежать численных проблем. Большие значения θ приводят к снижению остаточного шума, однако, следует быть внимательным, поскольку сегменты речевого сигнала с малой амплитудой могут быть также убраны. В наших экспериментах просто устанавливается этот параметр в 3 раза больше, чем абсолютная величина меньшего собственного значения, но не меньше чем 2^{-52}.

С учётом формулы (10) выражение для эффективного фильтра можно упростить:

$$\boldsymbol{H} = \sum_{l=1}^{r} \boldsymbol{V}\lambda_l(\lambda_l \boldsymbol{I} + \boldsymbol{M})^{-1}\boldsymbol{V}^{\#}\bar{\boldsymbol{q}}_l \boldsymbol{q}_l^{\#}), \quad (13)$$

где $\bar{q}_l$ является l-м вектор-столбцом матрицы $(Q^{\#})^{-1}$ и q_l является l-м вектор-столбцом матрицы Q. Отметим, чтобы вычислить эти векторы, требуется только l-й собственный вектор матрицы U (процедуры отбеливания/неотбеливания). Проблема имеет единственное решение тогда и только тогда, когда матрица $\lambda_l I+M$ не вырождена.

Как видно из (13), предлагаемый подход не требует полного набора собственных векторов. Этот факт особенно важен, если собственные значения оцениваются с помощью любой итерационной техники, например, как PASTd алгоритм [11]. Кроме того, очевидно, что шумовые компоненты, которые проецируются на подпространство шума, обнуляются. Хотя оба решения являются эквивалентными, интерпретация, предлагаемая в этой работе, позволяет избежать многих числовых операций. А именно, вычислительная нагрузка предлагаемого метода зависит от данных. В наихудшем случае вычислительная сложность данного решения примерно такая же, как и в работе [10], но количество собственных значений меняется с течением времени. Как можно увидеть на ***рис. 1***, ситуация, когда $r<k$, является обычной для типовых образцов речи. Поэтому в общем случае предложенное здесь решение превосходит стандартный метод.

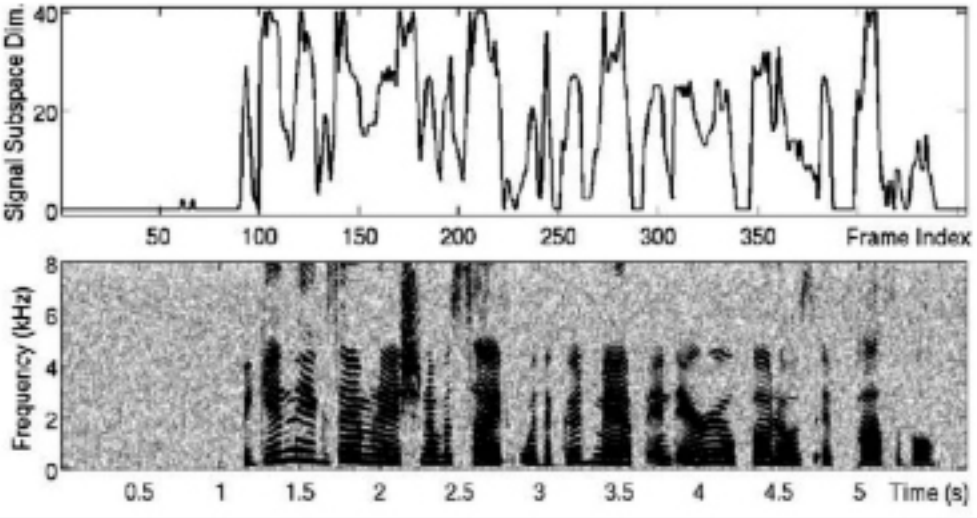

Рис. 1. Пример оценки размерности SS (вверху) для типового речевого сигнала (внизу)

Отметим также, что совокупность неортогональных подпространственных проекций можно интерпретировать как r-канальный банк фильтров. Такая интерпретация особенно полезна при параллельной обработке. Кроме того, если матрица является ДПФ подобной, внутриканальные фильтры могут быть эффективно реализованы с помощью быстрого преобразования Фурье. Такая прямая реализация модифицированного SDC оператора с помощью алгоритма БПФ представлена на рисунке 2. Следует обратить внимание, что матрицы могут быть рассмотрены как взвешивающие фильтры в частотной области.

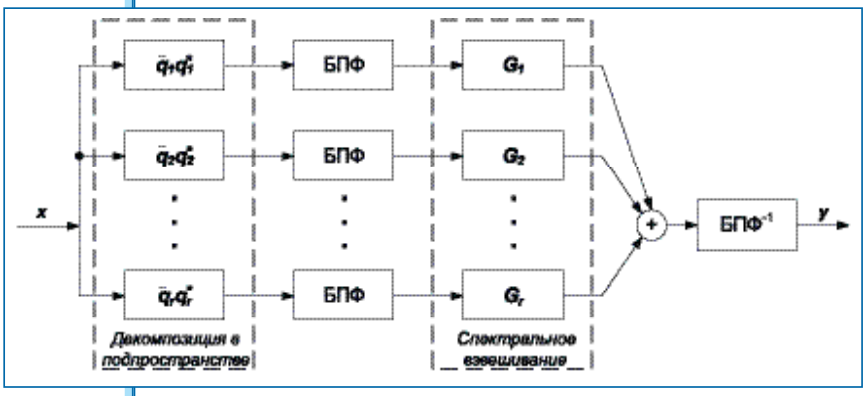

Рис. 2. Прямая реализация модифицированного SDC оператора на основе алгоритма БПФ

2.2. Перцептуально мотивированные ограничения

Эмпирически было проверено, что правило взвешивания наподобие правила Винера [1] делает обработанный спектр похожим на спектр чистой речи. К сожалению, такая техника слабо коррелирует со слуховым восприятием человека. Если ограничения определены в области собственных векторов, то трудно использовать психоакустически мотивированные правила взвешивания (они определяются обычно в частотной области) для формирования спектра остаточного шума. Согласно известному IND правилу [2], если какой-либо частотный компонент остаточного шума больше, чем порог маскирования, то он становится слышимым, и речевой сигнал искажается шумом. В противоположной ситуации, когда частотный компонент речи находится ниже порога маскирования, то получается ненужное ослабление чистой речи. Таким образом, в идеале, эти компоненты должны быть размещены ниже порога маскирования чистого речевого сигнала, чтобы сделать шум неслышимым и избежать ненужного ослабления речи.

Хотя представление спектра остаточного шума в частотной области можно получить с помощью соответствующего преобразования [8], здесь предлагается более простое решение. Одним из возможных выборов V является унитарная матрица. Тогда спектр остаточного шума $\{\alpha_i, i=1,..., k,\}$ может быть определён непосредственно в частотной области с помощью синусоидальных векторов:

$$vi\# = k-1/2[e-j\omega i.0,\ e-j\omega i.1,...,\ e-j\omega i.(k-1)\ ], \quad (14)$$

где $\omega_i = 2\pi(i-1)/k, \quad i = 1,2, ..., k$ (15)

Вектор $v_i^{\#}$ интерпретируется здесь как i-я строка нормированной ДПФ матрицы. Поскольку порог маскирования также определяется в частотной области, то преобразование «частотная область — область собственных значений» [8] выполнять не нужно. Принимая во внимание эти соображения, предлагается следующее правило, основанное на правиле IND, для формирования спектра мощности остаточного шума:

$$\alpha_i = \min(\phi_t(\omega_i)\ \alpha_{i,max_i}),\ \ i=1,2, ..., k\ , \quad (16)$$

где $(\phi_t(\omega_i)$ обозначает порог маскирования чистой речи и является максимально возможным остаточным уровнем шума для *i*-го спектрального отсчёта (не для случая ослабления шума).

$$\alpha_{i,\max} = \sum_{l=1}^{r} \left| v_i^{\#} q_l \right|^2, \quad (17)$$

2.3. Расчёт множителей Лагранжа

Интересным аспектом метода множителей Лагранжа является то, что значения множителей в точке решения обычно имеют некоторое значение. В данной задаче оптимизации они контролируют компромисс между остаточным шумом и искажениями речи, а следовательно, должны быть тщательно подобраны для получения требуемого фильтра. Как уже упоминалось ранее, в случае цветного шума, явный вывод множителей Лагранжа для определённого набора уровней остаточного шума является трудной задачей. Анализ литературы показывает, что такое выражение в настоящее время неизвестно. Однако в данной работе сделана попытка найти их численно.

Если требования в (4) выполнены с равенством, то уровни остаточного шума могут быть записаны следующим образом:

$$\alpha_i = v_i^{\#} R_n^{0,5} U G G^{\#} U^{\#} R_n^{0,5} v_i, \quad i = 1,2, ..., k \tag{18}$$

Можно заметить, что:

$$\boldsymbol{G}\boldsymbol{G}^{\#} = \sum_{l=1}^{r} \mathbf{g}_l \mathbf{g}_l^{\#}, \tag{19}$$

где g_i является l-м вектор-столбцом матрицы G. Таким образом, подставляя (19) в (18), имеем:

$$\alpha_i = \sum_{l=1}^{r} \left| \boldsymbol{v}_i^{\#} \frac{\lambda_l}{k\mu_i + \lambda_l} \overline{\boldsymbol{q}}_l \right|^2, \qquad i = 1,2,\ldots,k. \tag{20}$$

В общем, $v_i^{\#}\bar{q}_l \neq 0$ для всех i,l. Если предположить, что уровни остаточного шума определены в области собственных величин (т.е. $Y=U$) и шум является белым с дисперсией σ_n^2, т.е. $\bar{\boldsymbol{q}}_l = \sqrt{\sigma_n^2}\boldsymbol{u}$, то уравнения, представленные выше, могут быть упрощены до:

$$\alpha_i = \left(\frac{\lambda_l}{k\mu_i + \lambda_l} \right)^2 \sigma_n^2 . \tag{21}$$

Это приводит к следующему выражению для множителей Лагранжа:

$$\mu_i = \frac{\lambda_l}{k} \left[\left(\frac{\alpha_i}{\sigma_n^2} \right)^{-0.5} - 1 \right]. \tag{22}$$

Далее, подставляя эти соотношения в (9) и используя соответствующие правила для построения огибающей шума, можно получить обычные SDC операторы для белого шума [5]. Однако в данном случае не делается предположение о характере шуме, а также об области ограничений. Поэтому множители Лагранжа должны быть вычислены непосредственно. Принимая во внимание соотношение (20) легко видеть, что вычисление i-го множителя эквивалентно нахождению корня следующего уравнения:

$$\mathbf{g}_i(\mu_i) = \sum_{l=1}^{r} \left| \boldsymbol{v}_i^{\#} \frac{\lambda_l}{k\mu_i + \lambda_l} \overline{\boldsymbol{q}}_l \right|^2 - \alpha_i. \tag{23}$$

Как будет показано дальше, он может быть найден численно для определённого уровня остаточного шума.

3. Приближенное решение

Пусть $U_y \Lambda_y U_y^{\#}$ — разложение матрицы R_y по собственным векторам. В случае белого шума, т.е. $R_y = \sigma_n^2 I$, где σ_n^2 — дисперсия шума, обе матрицы R_y и R_n могут быть диагонализированы совместно с использованием матрицы U_y, что делает решение (6) тривиальным. В случае с цветным шумом было предложено следующее приближение [7]:

$$\boldsymbol{R}_n \approx \boldsymbol{U}_y \hat{\boldsymbol{\Lambda}}_y \boldsymbol{U}_y^{\#}, \tag{24}$$

где Λ_n — диагональная матрица с элементами, которые определяются следующим образом:

$$\hat{\lambda}_{n,i} = \boldsymbol{u}_{y,i}^{\#} \boldsymbol{R}_n \boldsymbol{u}_{y,i}, \qquad i = 1, \ldots, k. \tag{25}$$

Подставляя (24) в (6) и обозначая субоптимальный фильтр как H, получим:

$$\boldsymbol{L}\hat{\boldsymbol{H}}\boldsymbol{U}_y\hat{\boldsymbol{\Lambda}}_n\boldsymbol{U}_y^{\#} + \hat{\boldsymbol{H}}\boldsymbol{R}_y = \boldsymbol{R}_y. \tag{26}$$

Пусть $\hat{G} = U_y \# \hat{H} U_y$, тогда уравнение (26) может быть записано следующим образом:

$$\boldsymbol{U}_y^{\#}\boldsymbol{L}\boldsymbol{U}_y\hat{\boldsymbol{G}}\hat{\boldsymbol{\Lambda}}_n + \hat{\boldsymbol{G}}\boldsymbol{\Lambda}_y = \boldsymbol{\Lambda}_y. \tag{27}$$

Следует обратить внимание, что:

$$\boldsymbol{U}_y^{\#}\boldsymbol{L}\boldsymbol{U}_y\hat{\mathbf{g}}_l\hat{\lambda}_{n,l} + \hat{\mathbf{g}}_l\lambda_{y,l} = \lambda_{y,l}\boldsymbol{e}_l, \tag{28}$$

где $\hat{g}_i$ обозначает l-й вектор-столбец матрицы $\hat{G}$. Таким образом,

$$\hat{\mathbf{g}}_l = \boldsymbol{U}_y^{\#}\lambda_{y,l}\left(\hat{\lambda}_{n,l}\boldsymbol{L} + \lambda_{y,l}\boldsymbol{I}\right)^{-1}\boldsymbol{U}_y\boldsymbol{e}_l. \tag{29}$$

На основании определения $G \wedge$ аналитическое выражение для субоптимального линейного фильтра задаётся как:

$$\hat{\boldsymbol{H}} = \sum_{l=1}^{k} \boldsymbol{V}\lambda_{y,l}\left(\lambda_{y,l}\boldsymbol{I} + \boldsymbol{M}\hat{\lambda}_{n,l}\right)^{-1}\boldsymbol{V}^{\#}\boldsymbol{u}_{y,l}\boldsymbol{u}_{y,l}^{\#}. \tag{30}$$

Следует отметить, что (30) имеет единственное решение тогда и только тогда, когда матрица $\lambda_{y,l} I + M \lambda_{n,l}$ является несингулярной. Представленное приближенное решение не является оптимальным для цветного шума, но оптимально для белого шума. Таким образом, этот метод интересен как альтернатива с низкой вычислительной сложностью для подходов, основанных на процедуре отбеливания.

Например, если ограничения определены в KLT области, т.е. $V = U_y$, фильтр (30) упрощается до субоптимального SDC оператора [8]. В таком случае множители Лагранжа могут быть легко вычислены. Однако, если придать всем множителям фиксированные значения, например $\mu_l - \mu/k$, то получается субоптимальный TDC оператор [7]. В противном случае множители должны быть вычислены тем же способом, как в PCSS методе. В частности, если ограничения в (4) выполнены с равенством, уровни остаточного шума могут быть записаны следующим образом:

$$\alpha_i = \boldsymbol{v}_i^{\#}\hat{\boldsymbol{H}}\boldsymbol{R}_n\hat{\boldsymbol{H}}^{\#}\boldsymbol{v}_i. \tag{31}$$

Подставляя (31) в правую часть неравенства (4) и используя аппроксимацию (24), получим, что:

$$\alpha_i = \sum_{l=1}^{k}\left|\boldsymbol{v}_i^{\#}\boldsymbol{u}_{l,y}\frac{\lambda_{y,l}}{\hat{\lambda}_{n,l}k\mu_i + \lambda_{y,l}}\boldsymbol{u}_{l,y}\right|^2\hat{\lambda}_{n,l}. \tag{32}$$

Таким образом, в данном случае имеется k независимых одномерных уравнений, и множители Лагранжа могут быть найдены численно для определённого набора уровней остаточного шума аналогичным образом, как было показано выше.

4. Практическая реализация алгоритма PCSS

4.1. Схема обработки

Реализация обработки речевого сигнала в соответствии с методом PCSS осуществляется поблочно. Сигнал делится на блоки длиной N_f с перекрытием N_o отсчётов. Каждый блок разбивается на $m = N_f - N_o$ меньших перекрывающихся k-мерных векторов. Пусть t-й вектор внутри блока определяется следующим образом:

$$\boldsymbol{x}_t = \begin{bmatrix} x(\ell(N_f - N_0) + t + 1) \\ x(\ell(N_f - N_0) + t + 2) \\ \vdots \\ x(\ell(N_f - N_0) + t + k) \end{bmatrix}, \tag{33}$$

где ℓ — индекс блока и $x(\cdot)$ отсчёты зашумлённой речи. Последовательность этих векторов может рассматриваться как траектории в k-мерном евклидовом пространстве, которая организована в так называемую матрицу траекторий размера $k \times m$.

$$\boldsymbol{X}^{(\ell)} = [\boldsymbol{x}_1\ \boldsymbol{x}_2\ \ldots \boldsymbol{x}_m]. \tag{34}$$

Произведение матриц траекторий используется для вычисления ковариационной матрицы зашумленной речи:

$$\boldsymbol{C}_x^{(\ell)} = \frac{1}{m}\boldsymbol{X}^{(\ell)}\left(\boldsymbol{X}^{(\ell)}\right)^T. \tag{35}$$

Эта оценка является основой для расчёта сингулярных структур шума (только в речевых паузах) и KLT отбеленного сигнала, соответственно:

$$\begin{aligned} \boldsymbol{C}_n &\approx \boldsymbol{U}_n \boldsymbol{\Lambda}_n \boldsymbol{U}_n^{\#}, \\ \boldsymbol{C}_y &= \boldsymbol{C}_n^{-0.5}\boldsymbol{C}_x\boldsymbol{C}_n^{-0.5} - \boldsymbol{I} \approx \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\#}. \end{aligned} \tag{36}$$

Выше опущен индекс блока l для краткости. Чтобы избежать численных проблем, квадратные корни из матриц рассчитываются, используя сингулярные структуры $U_n \Lambda_n$ ковариационной матрицы шума. Упрощённая схема обработки приведена на рисунке 3. Первым вычисляется эффективный фильтр H, а затем все вектора в блоке обрабатываются с помощью той же матрицы. Результат сохраняется в матрице траекторий $\hat{Y}^{(l)}$ речевого сигнала, очищенного от шума. Обработанные векторы получаются из матрицы $\hat{Y}^{(l)}$, используя технику диагонального усреднения [12]. Наконец, блоки умножаются на окно Хеннинга и обрабатываются с помощью метода перекрытия с суммированием.

Как видно из схемы *(рис. 3)*, для вычисления эффективного фильтра необходимо множество неортогональных проекций, собственные значения отбеленной чистой речи и множители Лагранжа. В данной схеме множители рассчитываются итеративно по методу Ньютона. Известно, что этот метод может быть неустойчивым вблизи локального экстремума или горизонтальной асимптоты. Поскольку первая производная (23) отрицательна для $\mu_i \geq 0$, т.е.:

$$\frac{dg_i(\mu_i)}{d\mu_i} = -2k\sum_{l=1}^{k}\frac{1}{k\mu_i + \lambda_l}\left|\boldsymbol{v}_i^{\#}\frac{\lambda_l}{k\mu_i + \lambda_l}\bar{\boldsymbol{q}}_l\right|^2 < 0, \tag{37}$$

соотношение (23) является монотонно убывающей функцией в промежутке $(0;\infty)$.

Таким образом, может возникнуть только вторая проблема. Если $\min(\phi_t(\omega_i)\ \alpha_{i,max}) \approx 0$, то $g_i(\mu_i) = 0$ для $\mu_i \to \infty$. Такая ситуация влечёт к образованию пауз в отфильтрованном сигнале, если мощность зашумленного сигнала очень низкая. Поскольку матрица R_n считается положительно определённой, то максимальный уровень остаточного шума для $r > 0$ всегда больше нуля. Если это не так, матрица R_n может быть реализована путём добавления малой положительной константы к оцененным собственным значениям. В начале работы каждый множитель μ_i может быть обнулён. Число итераций можно уменьшить, установив $\mu_i = \mu_{i-1}$ для $i > 2$ в первой итерации. Ограничения определяются на сглаженном спектре, следовательно, функции имеют схожие формы и свойства. В эксперименте решение было найдено за приемлемое число итераций 5–20. Так как спектр $\{\alpha_i,\ i = 1,2,\ ...,\ k\}$, симметричен и k — чётное число, то только $k/2 + 1$ множителей Лагранжа должны быть вычислены.

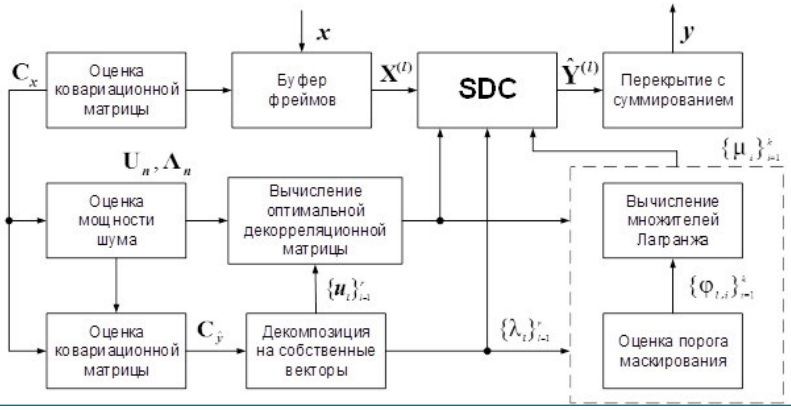

Рис. 3. Блок-схема вычислений по методу PCSS

4.2. Оценка порога маскирования

Прямого метода для оценки $\phi_t(\omega_i)$ по зашумлённому сигналу не существует. Обычно используемые методы работают на энергиях в критических частотных полосах, которые получены группированием соответствующих спектральных компонент мощности чистого речевого сигнала. Таким образом, здесь нужен спектр мощности чистой речи. Согласно определению, порог маскирования задаётся:

Ковариационная матрица чистой речи должна быть вычислена в первую очередь. Как правило, она может быть оценена из $R_y = R_x - R_n$, что эквивалентно технике спектрального вычитания. С другой стороны, может использоваться ковариационная матрица отбеленной речи. В целях ослабления музыкальных тонов предлагается восстановить спектр мощности чистой речи только из подпространства сигнала. Используя разложение (7), можно (38) переписать следующим образом:

$$\phi_y(\omega_i) = E\left[\left|v_i^{\#}\mathbf{y}\right|^2\right] = v_i^{\#}R_y v_i. \qquad (38)$$

$$\phi_y(\omega_i) = v_i^{\#}(Q^{-1})^{\#}\Lambda(Q^{-1})v_i = \sum_{l=1}^{r}\left|v_i^{\#}\bar{q}_l\right|^2\lambda_l, \qquad (39)$$

Полученные оценки при $i = 1,2, ..., k$ используются в качестве исходных данных для психоакустической модели Джонстона [12]. Параметры Q и Λ в (39) рассчитываются с использованием структур собственных значений из оценки ковариационной матрицы (36).

4.3. Следящий алгоритм оценки шума

В алгоритме PCSS преобразование Кархунена — Лоэва отбеленной речи можно вычислить, используя спектральное разложение ковариационной матрицы отбеленной зашумлённой речи. Другая возможность состоит в использовании следящих алгоритмов, работающих в подпространстве, для получения собственной структуры отбеленной речи напрямую из предварительно обработанного речевого сигнала. В обоих случаях отбеленная/неотбеленная матрицы необходимы. Они определяются, как обратная матрица из корня квадратного ковариационной матрицы шума и как корень квадратный из ковариационной матрицы шума, соответственно:

$$\boldsymbol{R}_{\boldsymbol{n}}^{\pm 0.5} = \boldsymbol{U}_{\boldsymbol{n}} \boldsymbol{\Lambda}_{\boldsymbol{n}}^{\pm 0.5} \boldsymbol{U}_{\boldsymbol{n}}^{\boldsymbol{T}}, \tag{40}$$

где U_n — матрица собственных векторов, Λ_n — диагональная матрица собственных значений. Следовательно, необходимы оценки собственных векторов и собственных значений ковариационной матрицы шума:

$$\boldsymbol{U}_{\boldsymbol{n}} = \left[\boldsymbol{u}_{\boldsymbol{n},1}, \boldsymbol{u}_{\boldsymbol{n},2}, \ldots, \boldsymbol{u}_{\boldsymbol{n},k}\right], \tag{41}$$

$$\boldsymbol{\Lambda}_{\boldsymbol{n}} = \mathrm{diag}\{\lambda_{n,1}, \lambda_{n,2}, \ldots, \lambda_{n,k}\}. \tag{42}$$

На практике ковариационная матрица шума неизвестна и должна быть оценена во время речевых пауз. К сожалению, такой подход требует устойчивого к ошибкам детектора речевой активности речи (voice activity detector, VAD). Предполагается, что KLT базис шума не изменяется быстро во время речевой активности, поэтому собственные вектора и собственные значения оцениваются раздельно. Оценка собственных значений регулируется вероятностями наличия речевой активности. Время от времени собственные вектора корректируются с использованием правила контроля по минимуму энергии. Таким образом, если текущее значение энергии шума, оцененное во временной области, опускается ниже порога (вычисленного по минимуму энергии на определённом интервале), то выполняется корректировка базиса KLT. Оценка вероятности присутствия речи основана на идее отслеживания минимума энергии, но реализована в KLT области. Блок-схема метода следящей оценки шума приведена на *рис. 4.*

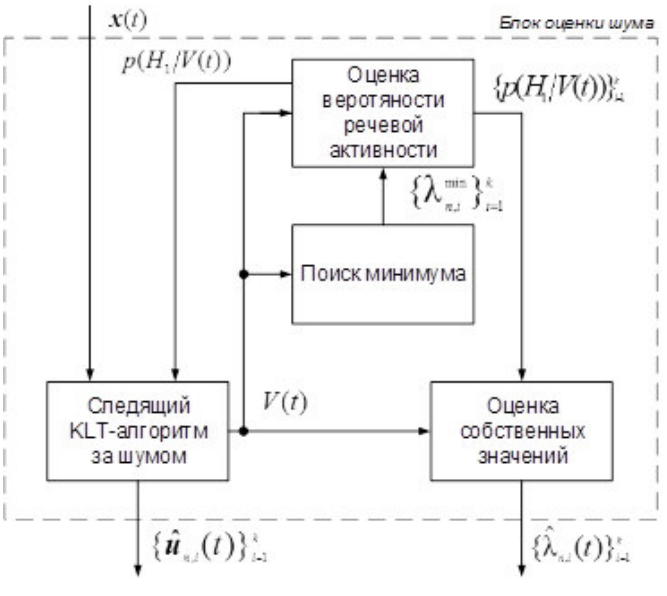

Рис.4. Блок-схема вычислений по методу PCSS

Собственные значения шума оцениваются следующим образом:

$$\lambda_{n,i}(t) = p_{post,i}(t)\lambda_{n,i}^{\min}(t) + \left(1 - p_{post,i}(t)\right)\min(B\lambda_{n,i}^{\min}(t), \hat{\lambda}_{x,i}^{w}(t)), \quad (43)$$

где B — смещающий компенсационный фактор, $\rho_{post,i}(t)$, $\lambda_{n,i}^{min}(t)$ и $\hat{\lambda}_{x,i}^{w}(t)$ обозначают апостериорную вероятность присутствия речи, уровень минимума зашумленной речи и усреднённую энергию зашумленной речи, соответственно, измеренные во время t для i-го собственного вектора.

Минимальный уровень энергии для i-го собственного вектора отслеживается на временном интервале в L блоков в соответствии со следующей процедурой:

IF $mod(t, L) = 0$

$\lambda_{n,i}^{\min}(t) = \min\{tmp_i, \hat{\lambda}_{x,i}^{w}(t)\}$

$tmp_i = \hat{\lambda}_{x,i}^{w}(t)$

ELSE

$\lambda_{n,i}^{\min}(t) = \min\{\lambda_{n,i}^{\min}(t-1), \hat{\lambda}_{x,i}^{w}(t)\}$

$tmp_i = \min\{tmp_i, \hat{\lambda}_{x,i}^{w}(t)\}$

Вероятность присутствия речи рассчитывается по Байесову правилу:

$$p_{post,i}(t) = \frac{\exp(LR_i(t))}{\exp(LR_i(t)) + (1.0 - p_{prio,i}(t))/p_{prio,i}(t)}, \quad (44)$$

где $\rho_{prio,i}(t)$, — априорная вероятность присутствия речи, $LR_i(t)$ — логарифмическая функция отношения правдоподобия. Априорной вероятности присутствия речи можно присвоить постоянное значение, однако, лучше использовать бинарную модель Маркова:

$$p_{prio,i}(t) = \Pi_{01} + (\Pi_{11} - \Pi_{01})p_{post,i}(t-1), \quad (45)$$

где Π_{ij} — вероятность перехода между состояниями H_i и H_j (т.е. присутствия i, $j = 1$ или отсутствия речи $i, j = 0$). В экспериментах принято, что $\Pi_{01} = 0{,}01$ и $\Pi_{11} = 0{,}9$.

В отличие от метода [13] отношение правдоподобия $LR_i(t)$ оценивается только с использованием смесей Гаусса. Это упрощение уменьшает вычислительную сложность и позволяет легко реализовать алгоритм в рамках обработки речевого сигнала поблочно, поскольку необходимыми данными являются только статистики второго порядка зашумлённого речевого сигнала и шума. В случае шума используется минимальный уровень энергии $\lambda_{n,i}^{min}(t)$ как её приблизительная оценка. Результаты можно несколько улучшить, если использовать (43) итеративно, т.е. для следующей итерации оценивается $\lambda_{n,i}$ вместо $\lambda_{n,i}^{min}(t)$ для вычисления $LR_i(t)$. На рисунке 5 показан пример слежения за шумом согласно данному методу оценки шума. *(См. рис. 5).*

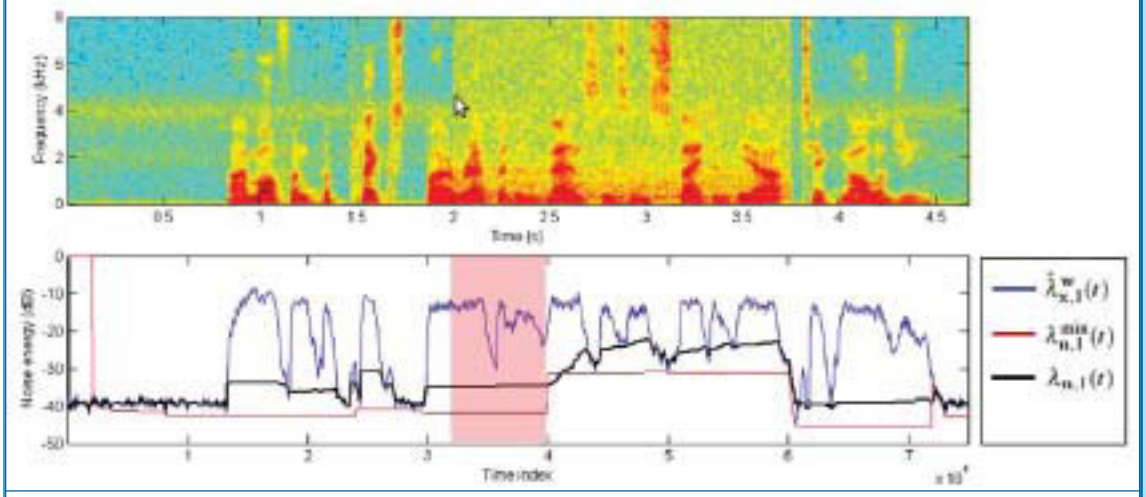

Рис. 5. Пример следящей оценки шума

Здесь показано быстрое изменение шумовых характеристик. Заметим, что структура спектра также изменяется. На нижнем графике приводится энергия флуктуаций в направлении главного собственного вектора шума. Синяя линия соответствует мощности зашумленной речи, красная линия — оцененная минимальная энергия, чёрная линия — оцененное собственное значение шума. На самом деле оценки собственных векторов шума на выбранном интервале некорректны (другими словами, слежение за KLT шума остановлено), также собственные значения шума недооценены. Однако после 0,5 секунд стабильность восстанавливается. Период адаптации зависит от длины интервала поиска минимума энергии (т.е. от параметра L). Тем не менее, этот параметр не может быть слишком мал, поскольку это ведёт к переоценке энергии шума на концах слов.

4.4. Вычислительная сложность реализации PCSS подхода

SS методы можно рассматривать, как обобщение методов спектрального взвешивания на KLT область. Известно, что KLT оптимально в смысле эффективности декорреляции. Поэтому обработка зашумленной речи в KLT области уменьшает артефакты музыкальных тонов и значительно превосходит методы, основанные на использовании ДПФ. Однако эти преимущества достигаются за счёт увеличения вычислительной сложности.

Сложность метода PCSS зависит от нескольких факторов: модели данных (высокий/низкий ранг), статистики сигнала и схемы обработки. Наиболее затратная с точки зрения времени выполнения операция PCSS алгоритмов — это вычисление KLT и реализация операторов 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) ковариационной матрицы. Вычислительная сложность ED равна $O(k^3)$, где k — размерность модели данных. Можно использовать любую процедуру слежения за подпространствами, т.е. алгоритмы аппроксимации проекций подпространств вместо ED. Такой метод зависит от данных и в худшем случае его сложность такая же, как у процедуры ED. К тому же метод не лишён ошибок оценивания и общая производительность системы, как правило, ухудшается. С другой стороны, в предложенном подходе KLT можно аппроксимировать дискретным косинусным преобразованием. Подобная структура, безусловно, является субоптимальной, зато требует меньше вычис-

лительных затрат. При практической реализации метода PCSS используется модель речевого сигнала с низким рангом и процедура ED из библиотеки LAPACK для симметричных матриц.

Как упоминалось выше, сложность метода PCSS зависит от схемы обработки, т.е. от эффективности реализации операторов. Существует два подхода, оба можно реализовать по схеме поблочной обработки. Первый подход — схема вычислений для коротких блоков. Построение обработки зашумленного сигнала по данной схеме основано на реализации PCSS подхода по структуре обработки, показанной на рисунке 2. Как можно заметить, она весьма подходит для параллельных вычислений. Её сложность равна

$$O((2k\log(k)(r+1) + 3kr)m + kr),$$

где m — число векторов на блок и r — размерность подпространства (которая зависит от данных). Второй подход обработки — схема для длинных блоков. В нём вначале вычисляется матрица операторов, которая затем умножается на все вектора блока. Сложность данного решения равна

$$O((k^2 + 4k\log(k)+2k)r + k^2m).$$

4.5. Экспериментальные исследования

Предложенный метод PCSS и его приближенная версия (PCSSa) были реализованы и протестированы в программной среде MATLAB. Для сравнения был выбран SDC оператор для белого шума [5]. Оценка шума осуществляется при следующих значениях параметров: частота дискретизации сигналов 16 кГц, $N_f = 400$, $N_0 = 200$ и $k = 40$. Набор из восьми предложений длительностью 5–8 с, произнесённых мужчиной и женщиной, был взят из базы данных TIMIT [14]. В качестве аддитивных помех были выбраны белый шум и два низкочастотных шума (шум двигателя автомобиля и шум в кабине самолёта F16). Эти шумы были программно добавлены к чистой речи, чтобы сегментное отношение сигнал/шум (SegSNR) было в пределах от 0 до 20 дБ. Для оценки эффективности реализованных алгоритмов использовались оценки SNR и перцептуальные измерения. Мера искажения речи (SD) определялась через SegSNR, где шум оценивался как разница между известным исходным и обработанным речевыми сигналами. Модифицированная оценка искажений спектра барков (MBSD) [11] была использована для оценки искажений речи. На ***рис. 6*** показаны результаты обработки для указанных шумов.

Для низкочастотных шумов оба предложенных PCSS метода работают заметно лучше по сравнению со стандартным SDC методом (SDCw). Даже при воздействии белого шума они обеспечивают чуть лучшие показатели. Это подтверждает наш тезис, что перцептуально мотивированные ограничения (7) являются более надёжными, чем винероподобное правило, использующееся в [5].

Лучшие результаты были получены для точного PCSS метода, использующего процедуру отбеливания, но за счёт повышения вычислительной сложности. Приближенная версия этого метода (PCSSa) гораздо проще и даёт похожие результаты для белого шума. В случае цветного шума здесь надо идти на компромисс между сложностью вычислений и качеством обработанной речи.

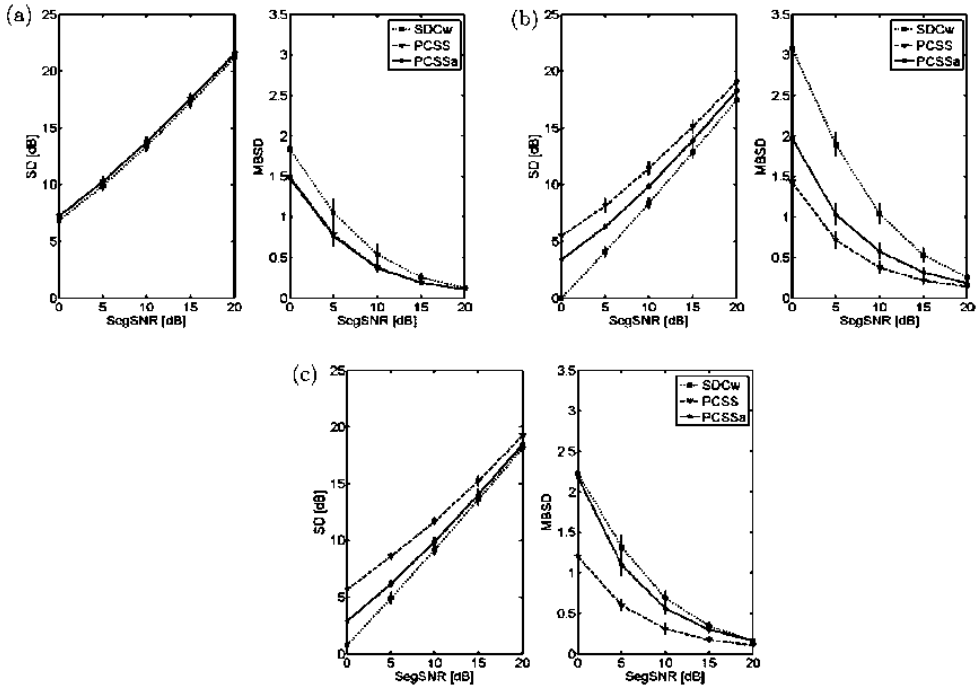

Рис. 6. Объективная оценка искажений речи на основе SD и перцептуальной оценки (MBSD): a — для белого шума, шума; b — для шума двигателя автомобиля; c — для шума кабины самолёта

Как и ожидалось, относительный прирост производительности приближенного метода PCSSa сильно зависит от типа шума. Лучшие результаты получены для автомобильного шума. Наши наблюдения показывают, что использованный в экспериментах оптимальный базис KLT для автомобильного шума похож на базис KLT на основе долгосрочной оценки ковариационной матрицы чистой речи. Таким образом, приближение (16) приводит к почти диагональной матрице. Однако это не относится к шуму в кабине самолёта F16, что говорит о том, что погрешность приближения может быть значительной в некоторых ситуациях. Оценки MSBD хорошо коррелируют с субъективными оценками искажений речи.

Как видно из спектрограмм (рисунок 7), стандартный SDC подход генерирует раздражительный низкочастотный остаточный шум. Обратите внимание, что он заметно отличается от музыкального шума, типичного для большинства ДПФ-методов. Приближенный PCSSa метод также генерирует аналогичный остаточный шум, но на более низком уровне. Эксперименты показали, что этот шум прослушивается в речевых паузах и практически не слышен во время речевой активности из-за явления маскировки. Исследования эффективности предложенных алгоритмов по критерию разборчивости речи выполнены в работе [16].

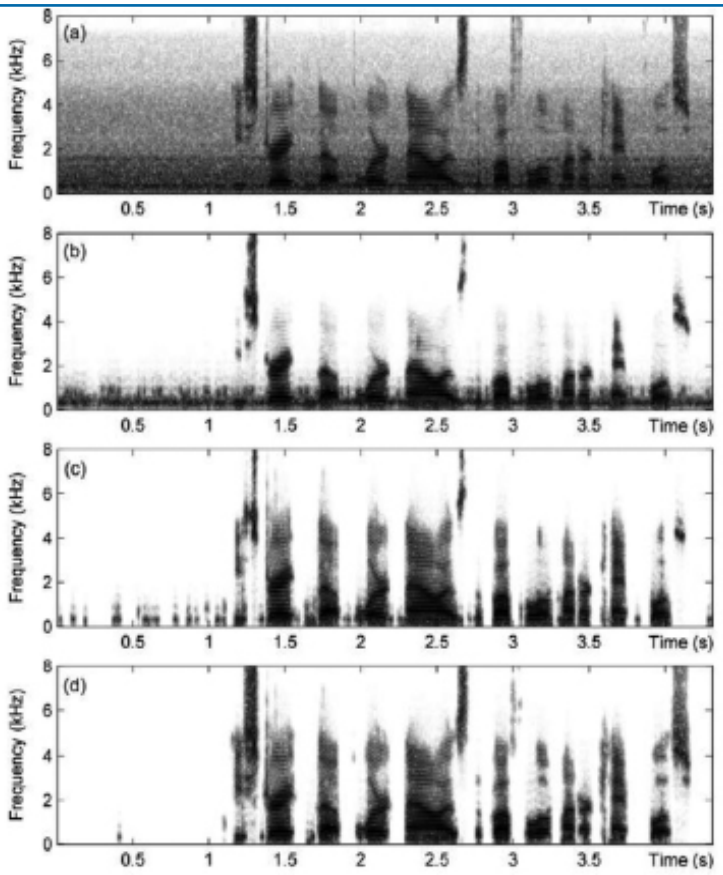

Рис. 7. Спектрограммы речевого сигнала:
a — аддитивная смесь речи и шум автомобиля (SNR_{seg}=5 дБ);
b — результат обработки алгоритмом SDCw;
c — результат обработки алгоритмом PCSSa;
d — результат обработки алгоритмом PCSS

Заключение

В данной работе предложен новый перцептуально мотивированный метод и алгоритм подавления шума окружающей среды на основе обработки речевого сигнала в подпространствах (PCSS), ядром которого является модифицированный оператор SDC [15]. Модифицированный SDC оператор получен в новой форме, которая делает реализацию подпространственного подхода более практичной. Ограничения остаточного шума определяются строго в частотной области с помощью векторного базиса на ДПФ основе и критериев восприятия акустической информации человеком. Эксперименты показали, что предложенный метод превосходит другие стандартные SS подходы, обеспечивая оптимальное восприятие остаточного шума и меньшие искажения речи.

Как упрощение метода PCSS, найдено приближенное решение с низкой вычислительной сложностью, которое не требует процедуры предварительного отбеливания. Эксперименты показали, что деградация обработанного речевого сигнала из-за приближения зависит от типа шума и ей можно пренебречь в случае шумов типа белых.

Литература

1. Loizou P.C. Speech enhancement: theory and practice. CRC Press, Taylor&Francis Group, NY. 2007.

2. Gustafson S., Jax P., Vary P. A novel psychoacoustically motivated audio enhancement algorithm preserving background noise characteristic. In: Proceedings of ICASSP, vol. 1, 1998. P. 397–400.

3. Petrovsky A.A., Parfieniuk M., Borowicz A. Warped DFT based perceptual noise reduction system. — AES, Convention Paper #6035, presented at the 116th Convention, 2004, May 8–11, Berlin, Germany.

4. Петровский А.А., Борович А., Парфенюк М. Дискретное преобразование Фурье с неравномерным частотным разрешением в перцептуальных системах редактирования шума в речи // Речевые технологии, 2008. № 3. С. 16–26.

5. Ephraim Y., Van Trees H. A signal subspace approach for speech enhancement. IEEE Trans. Speech, audio process, 1995, 3 (4). P. 251–266.

6. Mittal P., Phamdo N. Signal/noise KLT based approach for enhancing speech degraded by colored noise. IEEE Trans. Speech, audio process, 2000, 8 (2). P. 159–167.

7. Rezayee A., Gazor S. An adaptive KLT approach for speech enhancement. IEEE Trans. Speech, audio process, 2001, 9 (2). P. 87–95.

8. Jablom F., Champagne B. Incorporating the human hearing properties in the signal subspace approach for speech enhancement. IEEE Trans. Speech, audio process, 2003, 11 (6). P. 700–708.

9. Hu Y., Loizou P. A generalized subspace approach for enhancing speech corrupted by colored noise. IEEE Trans. Speech, audio process, 2003, 11 (4). P. 334–341.

10. Lev-Ari H., Ephraim Y. Extension of the signal subspace enhancement to colored noise. IEEE Sign. Process. Lett, 2003, 10 (4). P. 104–106.

11. Yang W., Benbouchta M., Yantorno R. Performance of a modified bark spectral distortion measure as an objective speech quality measure. In: Proc. ICASSP, Seattle, USA, 1998. P. 541–544.

12. Johnston J.D. Transform coding of audio signals using perceptuals noise criteria // IEEE Transactions on Selected Areas Communication. — February, 1988, vol. 6. P. 314–323.

13. Borowicz A., Petrovsky A. Minima controlled noise estimation for KLT-based speech enhancement // In Proc. 14th European Signal Processing Conference (EUSIPCO'2006), Florence, Italy, 4–8 Sep. 2006.

14. Garofolo J., Lamel L. and etc. DARPA TIMIT acoustic-phonetic continuous speech corpus. National institute of standards and technology (NIST), 1993.

15. Borowicz A., Petrovsky A. Signal subspace approach for psychoacoustically motivated speech enhancement. Speech communication, 2011, 53. P. 210–219.

16. Петровский А.А. и др. Шумоподавление на основе перцептуальных алгоритмов спектрального вычитания и обработки сигналов в подпространствах // Речевые технологии, 2012. № 4. С. 4–15.

Сведения об авторах

Борович Адам,
доктор-инженер факультета информатики Белостокского политехнического института, Польша. Область интересов: цифровая обработка речевых сигналов для целей редактирования шума, проектирование систем мультимедиа.

Петровский Александр Александрович,
доктор технических наук, профессор, Белорусский государственный университет информатики и радиоэлектроники (бывший Минский радиотехнический институт), кафедра электронных вычислительных средств. Главные научные интересы лежат в области цифровой обработки сигналов речи и звука для целей компрессии, распознавания, редактирования шума, а также проектирование проблемно-ориентированных средств вычислительной техники реального времени для систем мультимедиа. Член НТО РЭС им. А.С.Попова, IEEE, EURASIP, AES.