

Фильтрация речи на фоне полигармонических и стохастических помех

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

Азаров И.С., кандидат технических наук,

Лихачёв Д.С., кандидат технических наук, доцент,

Ромашкин Ю.Н., кандидат технических наук,

Жигулёвцев Ю.Н., кандидат технических наук, доцент,

Харламов А.А., доктор технических наук

В статье исследуются алгоритмы выделения речи на фоне аддитивных помех с полигармоническими и шумовыми составляющими. Сравниваются гибридный алгоритм (субполосная фильтрация — спектральное вычитание) и фильтрация в модуляционной области. По методу SII оценивается разборчивость речи, обеспечиваемая каждым из рассмотренных алгоритмов фильтрации. Даются выводы о наиболее эффективном алгоритме в зависимости от типа помехи.

• речевой сигнал • подавление шума • полигармоническая и стохастическая помехи.

The paper analyses different noise reduction algorithms for speech corrupted by additive noises with polyharmonic and stochastic components. A hybrid noise reduction algorithm based on subband filtering and spectral weighting, compares with speech filtering in modulation domain. The speech intelligibility is evaluated using SII measure for each considered technique. The paper discusses the choice of the most suitable processing algorithm depending on the noise properties.

• speech signal • noise reduction • polyharmonic and stochastic noise.

Введение

Широкий класс акустических помех может быть представлен в виде комбинации стохастической (шумовой) составляющей с гладким спектром и детерминированной (периодической) составляющей, спектр которой имеет ряд гармоник при медленно меняющейся основной частоте. Источником такого рода помех может быть электрическая сеть с переменным напряжением, трансформаторы, электродвигатели и т.д.

Большинство классических методов шумоподавления рассматривают шум как случайный сигнал с некоторыми априорно известными параметрами во временной или частотной области. Наиболее популярным является метод спектрального вычитания [1], где спектр шума оценивается в паузах речи и вычитается из спектра принятой аддитивной смеси. В случае же, когда помеха представляет собой сумму стохастических и полигармониче-



ских составляющих, оценка спектра шума часто не может быть выполнена достаточно точно, а речь — отфильтрована без слышимых артефактов. Данный эффект обусловлен следующими особенностями. С одной стороны, для подавления узкополосных гармоник оценка спектра должна проводиться с достаточно высоким частотным разрешением. С другой стороны, ослабление шумовой компоненты требует сглаженной оценки спектра, чтобы исключить появления музыкальных тонов в результате фильтрации. Необходимость выполнить одновременно оба этих требования сильно ограничивает подавление таких комбинированных помех алгоритмами на основе метода спектрального вычитания. Логичное решение данной проблемы — применение гибридной схемы с отдельной обработкой полигармонической и стохастической составляющих помехи.

Подавление квазипериодической компоненты помехи требует применения специальных методов, учитывающих её основные свойства:

- локализацию мощности в ряде узких частотных полос;
- наличие множества гармоник основной частоты;
- медленное изменение основной частоты во времени.

Учитывая эти свойства, можно предположить, что эффективными подходами для подавления полигармонической помехи могут быть адаптивная и узкополосная фильтрации. Эффективность первой объясняется высокой коррелированностью помехи, вследствие чего её легко предсказывать по предыдущим отсчётам. В гибридной системе шумоподавления оставшаяся стохастическая составляющая помехи может быть подавлена, например, при помощи традиционного метода спектрального вычитания либо при помощи обработки сигнала в подпространствах [2].

Ещё одним заслуживающим внимания методом подавления комбинированных стохастических/полигармонических помех является фильтрация в модуляционной области [3]. Общий алгоритм такой обработки сводится к субполосному разделению речевого сигнала, выделению огибающих в каждой полосе, фильтрации каждой огибающей полосовым (модуляционным) фильтром и обратному синтезу речевого сигнала. Фильтрация в модуляционной области может быть так же эффективна, поскольку отдельные составляющие помехи и речи занимают различные частотные полосы в модуляционном спектре.

В статье исследуются и сравниваются по разборчивости отфильтрованной речи различные методы подавления полигармонических/стохастических помех.

1. Подавление полигармонической помехи субполосным адаптивным фильтром

Для подавления узкополосных тональных составляющих полигармонической помехи могут быть использованы различные подходы. Простым в реализации и достаточно эффективным является способ, основанный на предсказании текущих отсчётов коррелированной помехи по предыдущим отсчётам с помощью стандартного адаптивного фильтра [4]. Для более полного подавления полигармонической помехи в схему должен быть включён банк полосовых фильтров (рис. 1).

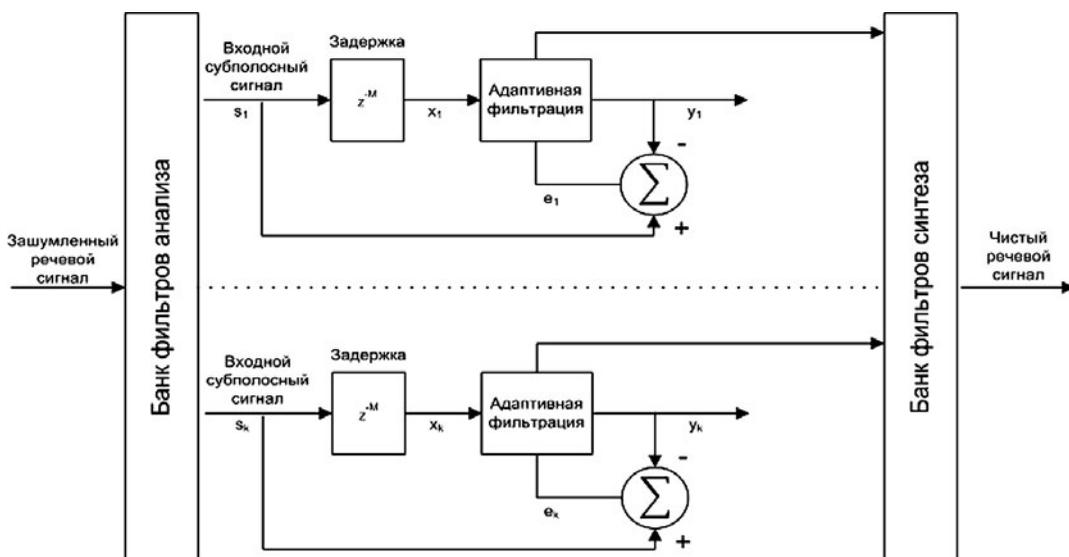


Рис. 1. Схема подавления полигармонической помехи с помощью алгоритма субполосной адаптивной фильтрации

Настройка коэффициентов адаптивного фильтра осуществляется в соответствии со следующими уравнениями [4]:

$$Y_k = \sum_{n=0}^{N-1} \omega_k(n) x_k(n), k = \overline{1, K}$$

$$e_k = s_k - y_k$$

$$\omega_{k+1}(n) = \omega_k(n) + 2\mu e_k x_k(n), n = \overline{0, N-1},$$

где s_k — зашумленный сигнал в k -й полосе частот;

$x_k(n)$ — задержанный отсчёт s_k ;

$\omega_k(n)$ — коэффициенты фильтра;

μ — коэффициент скорости сходимости;

y_k — выделенный узкополосный тональный сигнал;

K — количество полосовых фильтров.

Пример фильтрации речевого сигнала, принятого на фоне полигармонической сетевой помехи (с отношением сигнал/шум — 20 дБ), с помощью данного алгоритма представлен на рис. 2 (стр. 48). В алгоритме применялись банк узкополосных фильтров шириной 80 Гц, длительность задержки 100 мс и порядок адаптивного фильтра $N = 100$.

В отфильтрованном сигнале полигармоническая помеха ослаблена более чем на 20 дБ при слабых искажениях речи. Данный способ подавления полигармонической помехи является весьма удобным в реализации, если учесть, что при незначительной модификации схемы обработки субполосовые сигналы могут быть децимированы и, тем самым, существенно снижен требуемый порядок адаптивных фильтров.

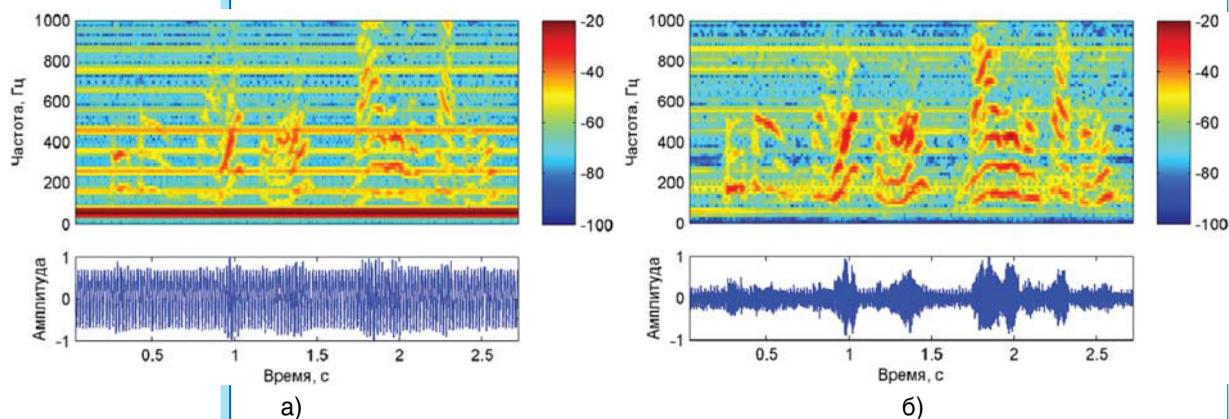


Рис. 2. Обработка речевого сигнала на фоне сетевой помехи (ОСШ = -20 дБ):
 а) исходный зашумлённый сигнал;
 б) выходной сигнал субполосного адаптивного фильтра

2. Подавление интенсивной полигармонической помехи

Для обеспечения большего подавления полигармонической помехи предлагается следующая схема обработки речевого сигнала (рис. 3).

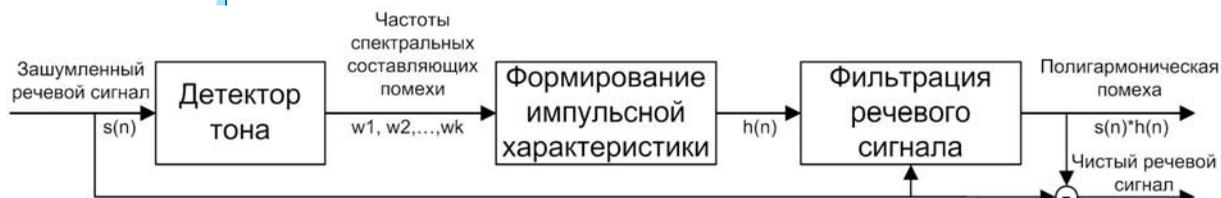


Рис. 3. Схема подавления интенсивной полигармонической помехи

Исходный сигнал обрабатывается детектором тона, выделяющим частоту каждой гармоники помехи. Затем в соответствии с найденными частотами формируется импульсная характеристика фильтра, пропускающего только полигармоническую помеху. Оценка речевого сигнала получается путём вычитания оценки помехи из принятого зашумленного сигнала. Импульсная характеристика фильтра обновляется в соответствии с изменениями параметров помехи.

Выделение тональных составляющих помехи в детекторе должно быть выполнено в соответствии со следующими требованиями:

- высокое частотное разрешение (до 1 Гц);
- нечувствительность детектора к гармоническим компонентам вокализованной речи;
- адаптация к медленным изменениям основной частоты помехи.

Для оценки текущего спектра в M заданных частотах $\omega_1, \omega_2, \dots, \omega_M$ отсчёт сигнала $s(n)$ умножается на вектор из M комплексных экспонент $(e^{-j\varphi_1}, e^{-j\varphi_1}, \dots, e^{-j\varphi_1})^T$, фазы которых $\varphi_1, \varphi_2, \dots, \varphi_M$ увеличиваются в соответствии с заданными частотами, где T — транспонирование. В результате формируется оценка кратковременного спектра в виде вектора отсчётов $(s_1, s_2, \dots, s_M)^T$,

по которому вычисляется среднее значение вектора спектральных амплитуд. Далее проводится оценка среднего спектра с использованием экспоненциального усреднения с коэффициентами $0 < \lambda_1 < 1$ (рис. 4).

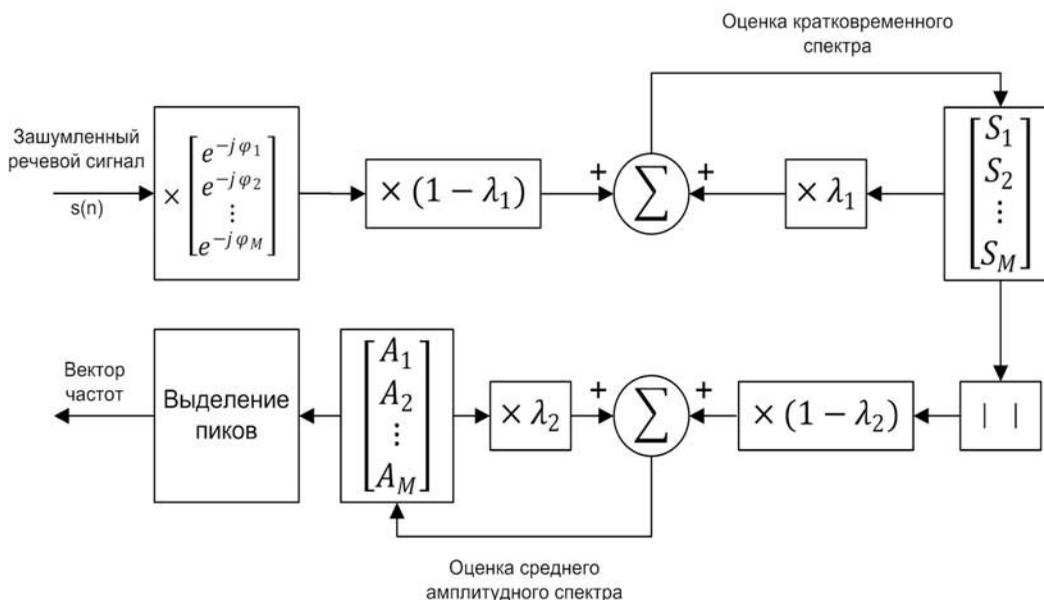


Рис. 4. Определение тональных компонент помехи

С помощью данной схемы можно получить оценку среднего спектра с требуемым частотным разрешением и длиной временного окна, регулируя значения коэффициентов. По среднему спектру вычисляются локальные максимумы, превышающие заданный порог (экспериментально определено, что в зависимости от типа помехи он составляет от 5 до 10 дБ).

В соответствии с предложенной схемой (рис. 3) блок фильтрации речевого сигнала, выполняющий выделение полигармонической помехи, должен пропускать лишь те частоты, которые выделены детектором тона, и задерживать все остальные. Удобным для этого является субполосный подход, поскольку общая импульсная характеристика фильтра может быть получена из соответствующих характеристик полосовых фильтров. Для увеличения скорости обработки полосовые фильтры предварительно синтезируются для каждой частоты $\omega_1, \omega_2, \dots, \omega_M$ и хранятся в оперативной памяти в процессе фильтрации.

Важным вопросом является выбор оптимальных значений для полосы пропускания (Δf) полосовых фильтров и длины импульсной характеристики. Расширение полосы пропускания способствует большему подавлению помехи, но в то же время увеличивает и подавление речевого сигнала. Увеличение длины импульсной характеристики позволяет уменьшить переходную полосу фильтра, но в то же время приводит к росту искажений сигнала.

Определение оптимальных значений выполнено опытным путём. К речевому сигналу (частота дискретизации 8 кГц) аддитивно добавлялась полигармоническая сетевая помеха с отношением сигнал/шум ОСШ = -20 дБ. Затем проводилась фильтрация сигнала предложенным алгоритмом (далее — алгоритм «TRem») и оценка качества обработанного сигнала по методу PESQ. Полученные результаты приведены в таблице 1 для различных значений Δf и длины импульсной характеристики фильтра.

Таблица 1

Зависимость оценки PESQ от ширины полосы пропускания и длины импульсной характеристики фильтра

| Δf , Гц | Длина импульсной характеристики, отсчёты | | | | | | | |
|-----------------|--|------|------|-------------|------|------|------|------|
| | 512 | 1024 | 1536 | 2048 | 2560 | 3072 | 3584 | 4096 |
| 4 | 1,34 | 1,41 | 1,48 | 1,57 | 1,67 | 1,78 | 1,88 | 1,97 |
| 7 | 1,37 | 1,52 | 1,70 | 1,93 | 2,14 | 2,27 | 2,32 | 2,32 |
| 10 | 1,40 | 1,66 | 2,00 | 2,28 | 2,37 | 2,36 | 2,34 | 2,33 |
| 13 | 1,44 | 1,84 | 2,26 | 2,37 | 2,35 | 2,33 | 2,32 | 2,3 |
| 16 | 1,48 | 2,03 | 2,37 | 2,34 | 2,32 | 2,3 | 2,28 | 2,27 |
| 19 | 1,53 | 2,19 | 2,34 | 2,31 | 2,29 | 2,27 | 2,26 | 2,25 |
| 22 | 1,58 | 2,29 | 2,31 | 2,28 | 2,26 | 2,24 | 2,23 | 2,21 |
| 25 | 1,64 | 2,30 | 2,28 | 2,25 | 2,23 | 2,21 | 2,2 | 2,18 |
| 28 | 1,70 | 2,27 | 2,24 | 2,21 | 2,19 | 2,17 | 2,16 | 2,14 |

Исходя из полученных результатов, оптимальная ширина полосовых фильтров составила 13 Гц, а длина импульсной характеристики фильтра — 2048.

На рис. 5 приведены спектрограммы и временные реализации исходного речевого сигнала на фоне сетевой помехи с ОСШ = -30 дБ и результата фильтрации алгоритмами «TRem» и спектрального вычитания. Анализ экспериментальных результатов показывает, что предложенный алгоритм «TRem» обеспечивает ослабление сетевой помехи более чем на 30 дБ и заметно меньшее искажение речевого сигнала.

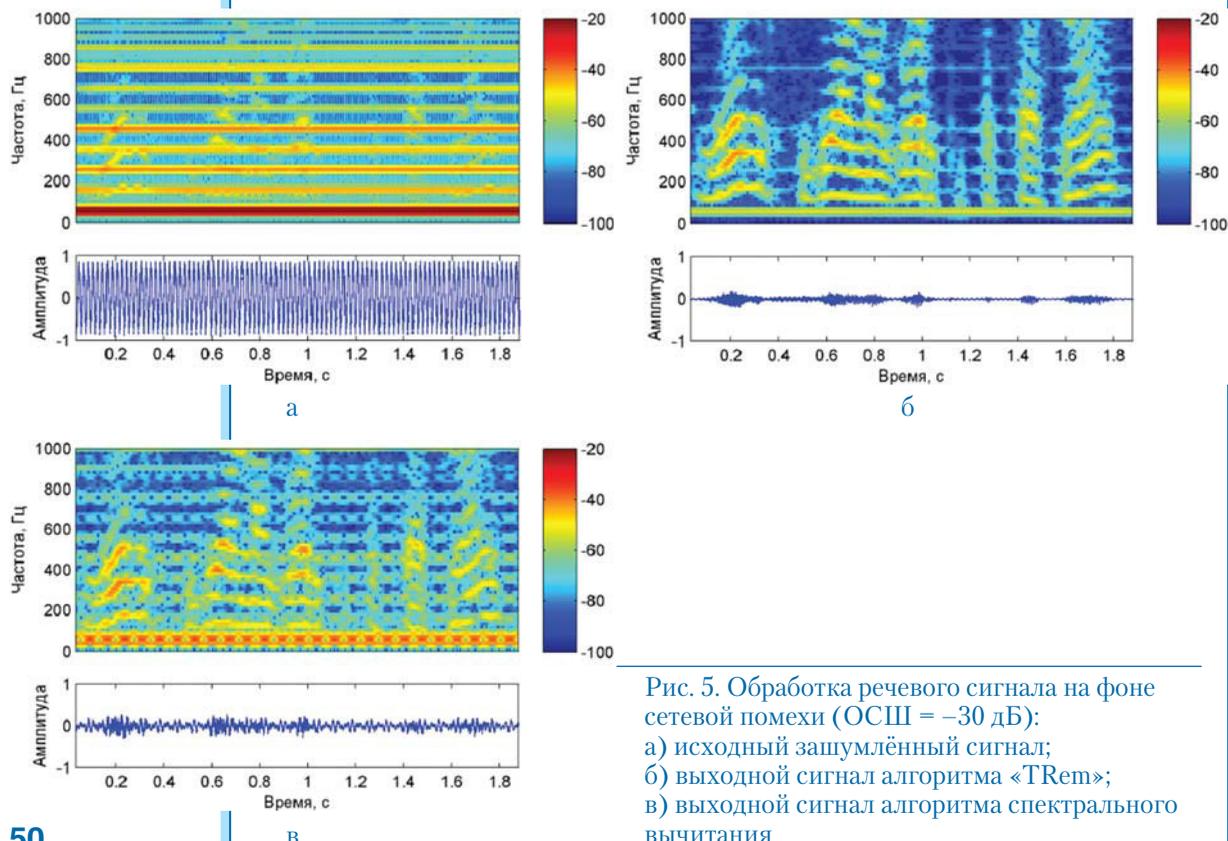


Рис. 5. Обработка речевого сигнала на фоне сетевой помехи (ОСШ = -30 дБ):
 а) исходный зашумлённый сигнал;
 б) выходной сигнал алгоритма «TRem»;
 в) выходной сигнал алгоритма спектрального вычитания

3. Гибридная фильтрация

Для подавления аддитивных помех, содержащих полигармонические и стохастические составляющие, может применяться последовательная схема обработки: сначала подавление полигармонических помех алгоритмом «TRem», а затем ослабление шумовой составляющей с помощью алгоритма спектрального вычитания. Результаты фильтрации речевого сигнала, принятого на фоне акустического шума пылесоса, показаны на рис. 6 а, б, в, г отдельно для метода спектрального вычитания, алгоритма «TRem» и гибридного метода соответственно. Из приведённых рисунков видно, что алгоритм спектрального вычитания не обеспечивает достаточного подавления периодических составляющих помехи. Гибридный метод позволяет получить более высокое подавление помехи при незначительном уровне искажений отфильтрованного речевого сигнала.

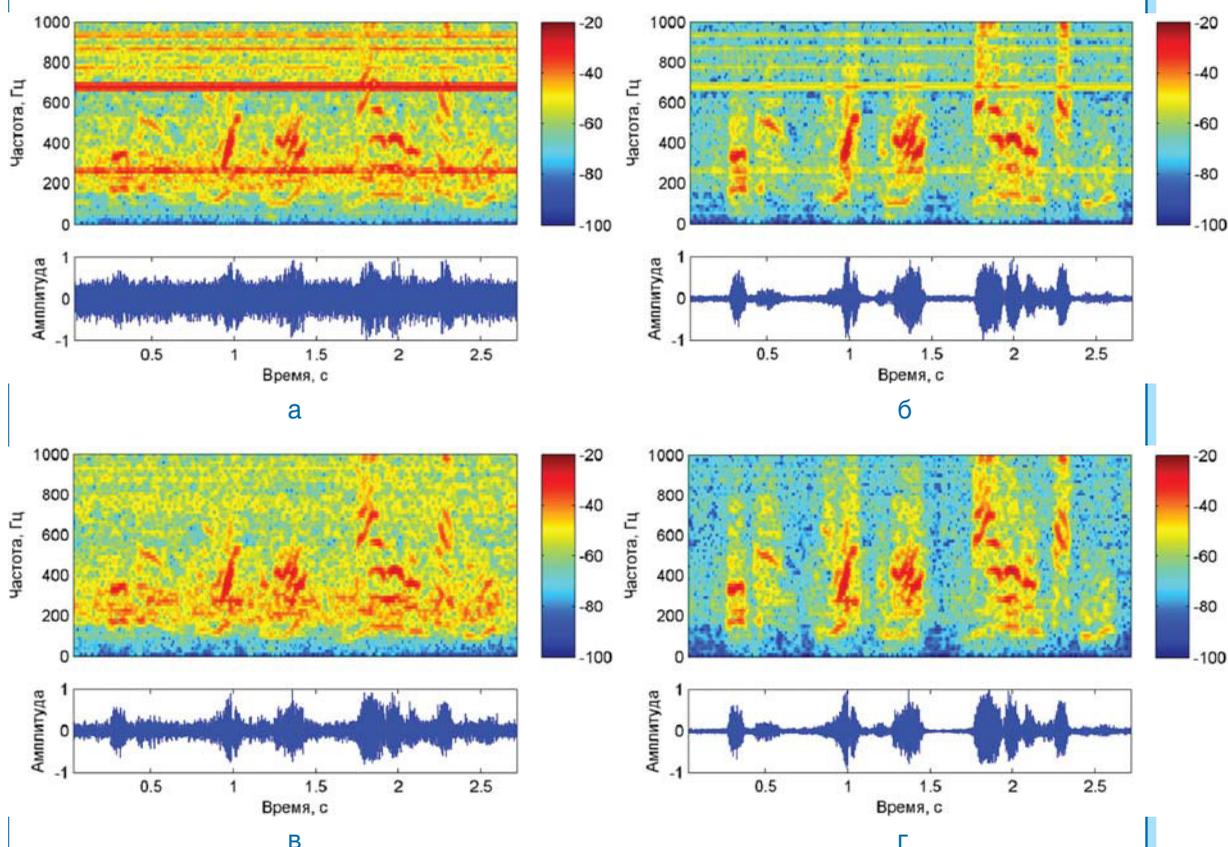


Рис. 6. Обработка речевого сигнала на фоне шума пылесоса (ОСШ = -10 дБ):

- а) исходный зашумлённый сигнал;
- б) выходной сигнал алгоритма спектрального вычитания;
- в) выходной сигнал алгоритма «TRem»;
- г) выходной сигнал гибридного алгоритма

4. Фильтрация в модуляционной области

Подход к обработке речи в модуляционной области появился сравнительно недавно и подтвердил информационную значимость амплитудных модуляций [5]. На основании ряда наблюдений было сделано предположение, что слуховой аппарат человека распознаёт речь при помощи частотно-временного анализа амплитудных модуляций

речи. В работе [6] показана тесная связь классических психоакустических принципов частотного и временного маскирования с аналогичными эффектами в модуляционном спектре.

Известны два метода обработки зашумленной речи: фильтрация модуляционного спектра — RASTA (RelAtive SpecTrA) и перцептуальное линейное предсказание — PLP (Perceptual linear prediction). В целом алгоритм на основе RASTA-PLP состоит из следующих шагов [6]:

- вычисление энергии сигнала в критических частотных полосах;
- компрессия амплитудного спектра путём нелинейной трансформации;
- фильтрация временной последовательности каждой из частотных компонент;
- растягивание амплитудного спектра путём обратной нелинейной трансформации
- умножение на кривую равной громкости для выравнивания громкости.

Ключевой идеей является подавление постоянных составляющих в кратковременном спектре принятого сигнала.

Для выполнения нелинейной трансформации обычно используются логарифмические функции. Они приводят к тому, что искажения, свёрнутые во временной области, проявляются как аддитивный шум в логарифмическом спектре. Трудности возникают, если помеха является некоррелированным аддитивным шумом. Такой шум аддитивен в линейном спектре, но в логарифмическом становится зависим от самого сигнала, что делает невозможным его удаление путём фильтрации в частотных полосах. Решение этой проблемы было предложено в [7] и заключается в использовании следующей функции преобразования:

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

где $X(k, n\Delta t)$ — амплитуда входного сигнала в k -й полосе частот, c — положительная константа, зависящая от сигнала.

Такое преобразование линейно для малых значений амплитуды, для больших — логарифмическое. Формула обратного преобразования имеет вид:

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

В качестве фильтра, используемого для обработки сигнала в модуляционной области, авторами RASTA был предложен БИХ фильтр с передаточной функцией [6]:

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

Нижняя граница полосы пропускания фильтра определяет максимальную скорость изменения логарифмического спектра, подавляемую фильтром. Верхняя граница определяет максимальную скорость изменения сохраняемого спектра. Начиная с частоты 12 Гц, амплитудно-частотная характеристика фильтра имеет крутизну спада 6 дБ/окт., достигая нуля при частоте 30 и 50 Гц.

Ниже приведены результаты обработки речи на фоне различных помех, полученные с помощью алгоритма фильтрации в модуляционной области. На рис. 7а, б приведён пример обработки аддитивной смеси речевого сигнала с белым шумом. Подавление шума составило в среднем 8 дБ.

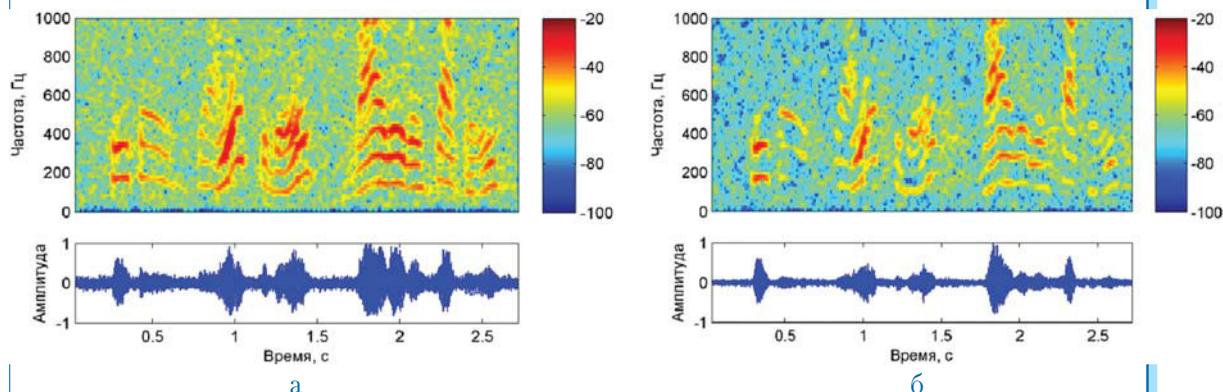


Рис. 7. Обработка речевого сигнала на фоне белого шума (ОСШ = -5 дБ):

- а) исходный зашумлённый сигнал;
- б) выходной сигнал модуляционного фильтра

Сетевая помеха, как отмечалось ранее, содержит множество гармоник с медленно изменяющимися амплитудами и потому может эффективно подавляться модуляционным фильтром. Пример обработки речевого сигнала на фоне сетевой помехи приведён на рис. 8а, б. На участках, где отсутствует голос диктора, снижение энергии сетевой помехи составляет, примерно, 33 дБ.

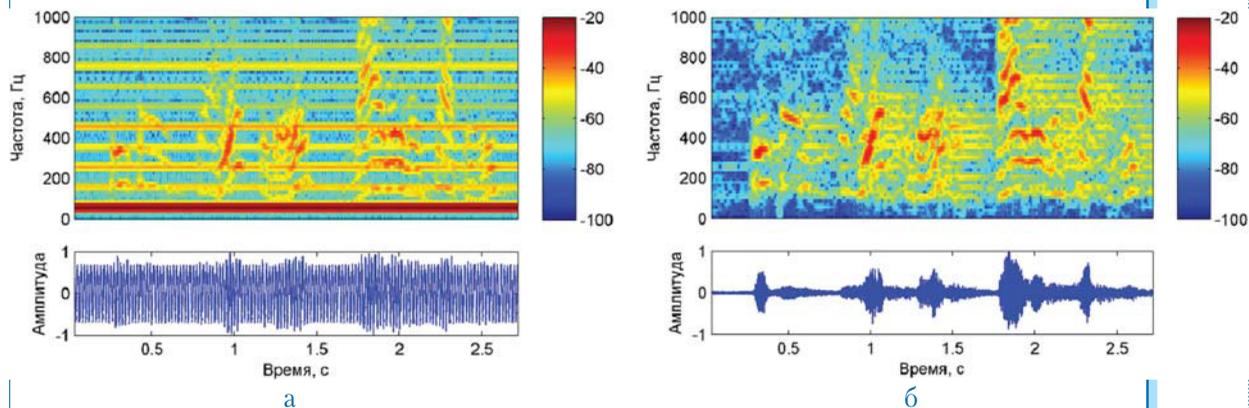


Рис. 8. Обработка речевого сигнала на фоне сетевой помехи (ОСШ = -20 дБ):

- а) исходный зашумлённый сигнал;
- б) выходной сигнал модуляционного фильтра

Акустический шум работающего пылесоса состоит из периодических и стохастических составляющих при его общей нестационарности. На рис. 9а, б (стр. 54) приведён пример обработки речевого сигнала на фоне шума пылесоса. В паузах речи подавление помехи составляет примерно 22 дБ.

Фильтрация речи на фоне музыкальной помехи одноканальными системами — крайне сложная задача. Тем не менее, применение модуляционного фильтра обеспечивает некоторое подавление тональных и переходных составляющих, характерных для музыкальных инструментов. Пример обработки речевого сигнала на фоне музыки приведён на рис. 10 (стр. 54). Снижение энергии музыки составило примерно 4 дБ.

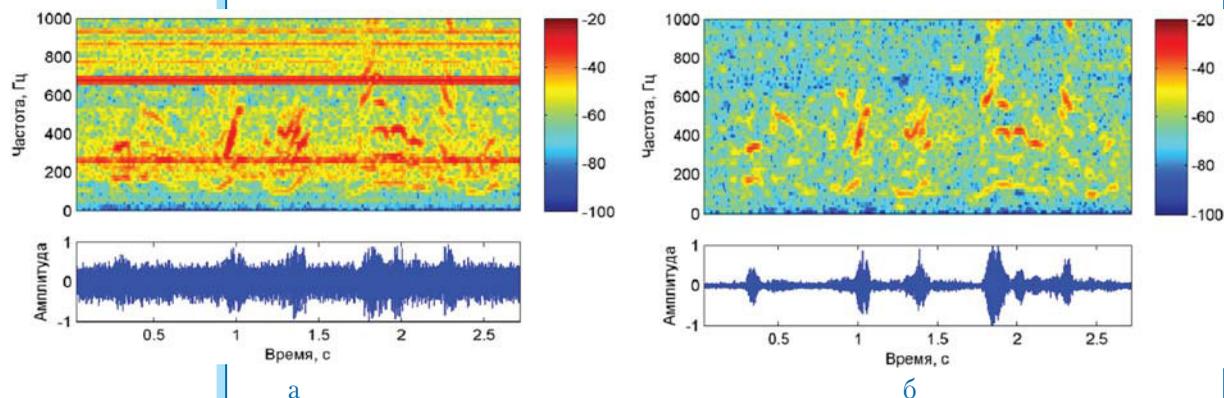


Рис. 9. Обработка речевого сигнала на фоне шума пылесоса (ОСШ = -10 дБ):
а) исходный зашумлённый сигнал;
б) выходной сигнал модуляционного фильтра

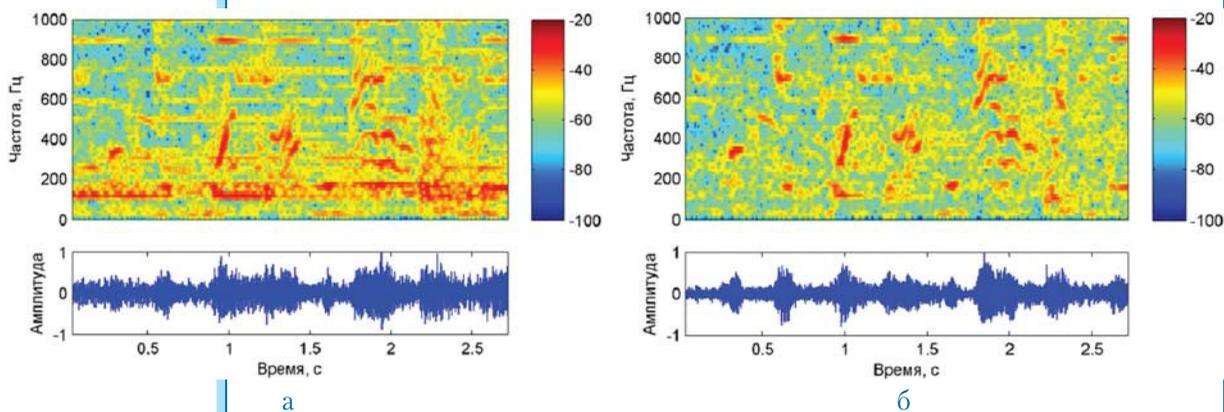


Рис. 10. Обработка речевого сигнала на фоне музыки (ОСШ = -5 дБ):
а) исходный зашумлённый сигнал;
б) выходной сигнал модуляционного фильтра

5. Оценка разборчивости речи

Оценка разборчивости по методу SII (speech intelligibility index) основана на вычислении т.н. индекса артикуляции, значения которого принято интерпретировать следующим образом [8]:

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

Для сравнения разборчивости отфильтрованной речи обрабатывались аддитивные смеси речевого сигнала (РС) на фоне трёх помех: белого шума, сетевой помехи и шума пылесоса с ОСШ от — 30 до 20 дБ. Эти смеси обрабатывались следующими алгоритмами: спектрального вычитания (NRS), алгоритмом «TRem», фильтром в модуляционной области («Modul»), а также алгоритмом обработки сигнала в подпространствах (perceptually constrained signal subspace — PCSS, [2]). Полученные результаты усреднялись по мужским и женским голосам длительностью около 2-х минут. На рис. 12–14 показаны экспериментальные зависимости индекса SII от ОСШ.

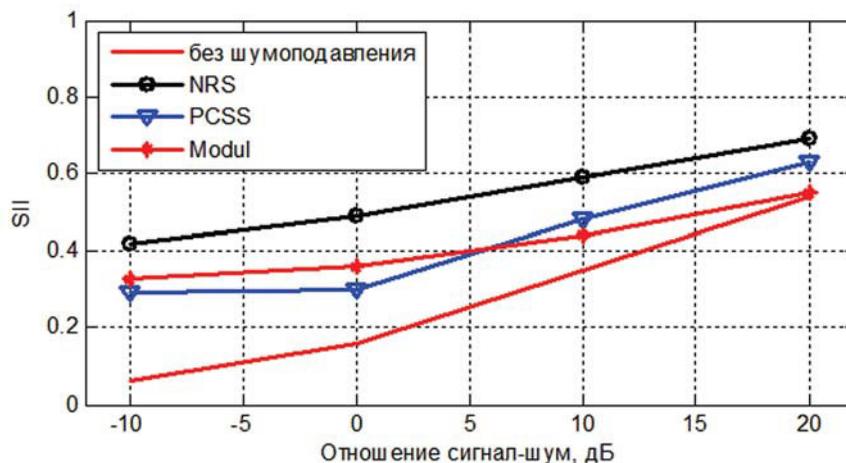


Рис. 11. Зависимость индекса SII от ОСШ для смеси PC с белым шумом

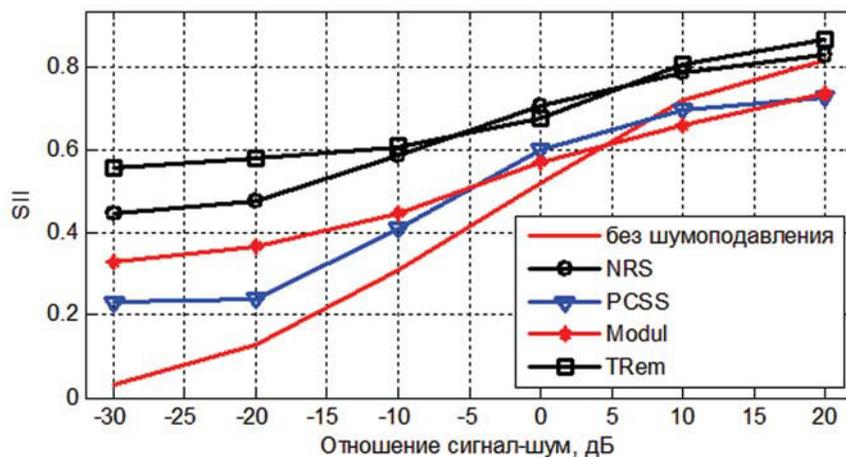


Рис. 12. Зависимость индекса SII от ОСШ для смеси PC с сетевой помехой

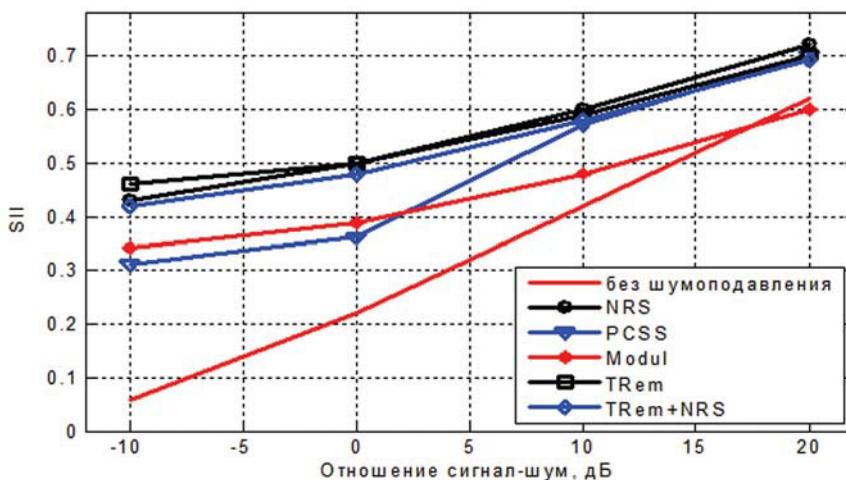


Рис. 13. Зависимость индекса SII от ОСШ для смеси PC с шумом пылесоса



Как видно из представленных результатов, при наличии белого шума алгоритмы спектрального вычитания, модуляционной фильтрации и PCSS позволяют заметно повысить разборчивость речи. Причём метод спектрального вычитания является предпочтительным при любом ОСШ.

Для случая воздействия полигармонической сетевой помехи алгоритм «TRem», предложенный в статье, позволяет достичь значительно большей разборчивости по сравнению с остальными алгоритмами.

При обработке речи на фоне акустического шума пылесоса, имеющего полигармонические и шумовые составляющие, эффективность метода PCSS значительно снижается при отрицательных ОСШ, в то время как алгоритмы «TRem» и спектрального вычитания заметно повышают разборчивость речи. Гибридный алгоритм «TRem+NRS» практически эквивалентен по разборчивости речи алгоритмам, его составляющим.

Алгоритм на основе модуляционной фильтрации в целом показывает средние результаты.

Выводы

Широко известные алгоритмы шумоподавления, такие как спектральное вычитание и субполосная адаптивная фильтрация, не обеспечивают наилучший результат в случае воздействия сложных полигармонических/стохастических помех.

Экспериментально доказано, что предложенный алгоритм подавления интенсивной полигармонической помехи обеспечивает хорошую разборчивость речи даже при ОСШ < -10 дБ.

Метод фильтрации в модуляционной области обеспечивает удовлетворительную разборчивость при различных видах помех. В частности, он подавляет белый шум на 8дБ, комбинированную полигармоническую/стохастическую помеху более чем на 20 дБ. Однако он уступает по разборчивости речи гибриднему алгоритму.

Список литературы

1. Loizou P. Speech Enhancement: Theory and Practice / P. Loizou. Boca Raton, FL: CRC Press, 2007.
2. Borowicz A., Petrovsky A. Signal subspace approach for psychoacoustically motivated speech enhancement. Speech communication 53 (2011), Elsevier. P.210–219.
3. Bashun J., Petrovsky A. Speech enhancement for cochlear implants based on the reducing slow temporal modulations // Proc. of The Acoustic Congress, Rom, Italy, 2000.
4. Haykin S. Adaptive filter theory, NJ.: Prentice — Hall, 1996.
5. Kowalski N., Depireux D. and Shamma S. Analysis of dynamic spectra in ferret primary auditory cortex: I. Characteristics of single unit responses to moving ripple spectra // Journal of Neurophysiology, 1996. Vol. 76, no. 5. P. 3503–3523.
6. Hermansky H. and Morgan N. RASTA processing of speech. IEEE Transaction on Speech and Audio Processing, 1994. Vol. 2, no. 4. P. 587–589.
7. Avendano Temporal processing of speech in a Time-Feature Space, Ph.D. thesis, Oregon Graduate Institute, April, 1997.
8. Основы психоакустики. Субъективные и объективные методы оценки разборчивости речи [Электронный ресурс] / И. Алдошина. Журнал «Звукорежиссёр», 2002. Режим доступа: www.kadva.ru/files/edu/AldoshinaPsychoacoustics.pdf — Дата доступа: 16.03.2011.

Сведения об авторах

Петровский Александр Александрович —

доктор технических наук, профессор. Работает в Учреждении образования «Белорусский государственный университет информатики и радиоэлектроники», кафедра «Электронные вычислительные средства». Закончил Учреждение образования «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. Учреждение образования «Белорусский государственный университет информатики и радиоэлектроники» (БГУИР), кафедра электронных вычислительных средств. Закончил Белорусский государственный университет, механико-математический факультет, отделение математической электроники. Область научных интересов: цифровая обработка речевых сигналов.

Лихачёв Денис Сергеевич —

кандидат технических наук, доцент. Учреждение образования «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.

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.