

Д.т.н., проф. Е.Г. Жилияков, А.В. Курлов,
д. ф.-м. н., проф. А.В. Глушак (Белгородский ГУ)

E.G. Zhilyakov, A.V. Kurlov, A.V. Glushak

**О КОМПЬЮТЕРНОЙ ОЧИСТКЕ РЕЧИ ОТ ШУМОВ
НА ОСНОВЕ ИЗБИРАТЕЛЬНОГО ВОЗДЕЙСТВИЯ
НА ЧАСТОТНЫЕ КОМПОНЕНТЫ РЕЧЕВЫХ СИГНАЛОВ**

**ABOUT COMPUTER CLEARING SPEECH FROM NOISE
ON THE BASIS SELECTIVE INFLUENCE ON FREQUENCY
COMPONENTS OF SPEECH SIGNALS**

В статье рассматриваются компьютерные процедуры очистки речи от шумов. В основе их используется математический аппарат субполосных матриц, позволяющий осуществлять избирательное воздействие на частотные компоненты речевых сигналов. Применение подобного математического аппарата является новым подходом к решению задачи очистки речи от шумов.

Keywords. speech, noise, frequency components, substrip matrixes, selective influence, weight factors

Введение

Речевые сообщения являются одной из наиболее часто используемых, естественных для человека, форм информационного обмена. В области цифровой обработки речевых сигналов особое внимание уделяется задачам, связанным с шумоочисткой и повышением разборчивости речевых сообщений, полученных в неблагоприятной помеховой обстановке. Примером могут быть диктофонные записи в условиях определенных помех или записи речевого сигнала, полученные по каналам связи низкого качества.

Об актуальности исследований в области речевых технологий свидетельствуют как многочисленные публикации результатов теоретических исследований [1,2,3,4,5], так и достаточно интенсивные разработки специализированных программных и аппаратных средств очистки речи от шумов, которые осуществляются такими фирмами как: ООО «Центр речевых технологий»; ООО "ГритТек"; Санкт-Петербургский институт информатики и автоматизации РАН; Центр цифровой обработки сигналов; компания СПИРИТ-Телеком [6,7,8,9].

В связи с этим, создание компьютерных технологий обработки речевых данных, позволяющих улучшить разборчивость и очистить от шумов речевые сообщения, является актуальной задачей.

Под речевыми сигналами принято понимать электрические колебания, регистрируемые на выходе микрофона под воздействием акустических волн, порождаемых речевым аппаратом человека. Для речевых сигналов спектр частот, достаточных для сохранения качества звучания, лежит в полосе от 0.3-3.5 кГц. Для сигналов с указанной полосой частот, согласно теореме Найквиста, необходимо использовать частоту дискретизации $f_0=8\text{кГц}$. Именно это значение частоты дискретизации является общепринятым стандартом при цифровой обработке речевых сигналов. [10].

Математические основы

В качестве основы для построения компьютерной модели используется математический аппарат субполосных матриц [11], который позволяет оценить точные значения долей энергий отрезков речевых сигналов, сосредоточенных в заданных частотных интервалах.

Известно, что энергия звуков русской речи сосредоточена в малом количестве достаточно узких частотных интервалов. Выбранный математический аппарат позволяет адекватно учесть данное свойство звуков русской речи, осуществить разделение обрабатываемого речевого сигнала на частотные компоненты, на которые затем оказывать избирательное воздействие.

В общем виде модель осуществления избирательного воздействия на компоненты речевого сигнала определяется следующими выражениями:

$$\bar{x}_N = (\hat{x}_1, \dots, \hat{x}_N)^T = \sum_{r=1}^R c_r \bar{u}_{rN}, \quad (1)$$

$$\bar{u}_{rN} = (u_{r1}, \dots, u_{rN})^T, \quad (2)$$

$$\bar{u}_{rN} = \sum_{k=1}^{J_r} \lambda_{kN}^r \alpha_{kN}^r \bar{q}_{kN}^r, \quad (3)$$

где \bar{x}_N - вектор значений выходного обработанного речевого сигнала c_r - весовые коэффициенты r -частотных компонент, \bar{u}_{rN} - векторы

частотных компонент речевого сигнала, соответствующие заданным частотным интервалам, $\lambda_{kN}^r, \bar{q}_{kN}^r$ - собственные числа и векторы субполосных матриц вида: $A_{rN} = \{\alpha_{ik}^r\}, i, k = 1, \dots, N; r = 1, \dots, R$.

$$a_{ik}^r = \begin{cases} \frac{\sin[V_r(i-k)] - \sin[V_{r-1}(i-k)]}{\pi(i-k)}, & i \neq k \\ \frac{V_r - V_{r-1}}{\pi}, & i = k \end{cases} \quad (4)$$

$$\Omega_r = [-V_r, -V_{r-1}) \cup [V_{r-1}, V_r), V_0 = 0; V_R = \pi \quad (5)$$

$$V_r - V_{r-1} = \pi / R. \quad (6)$$

$$\lambda_{kN}^r \bar{q}_{kN}^r = A_{rN} \bar{q}_{kN}^r, k = 1, \dots, N; r = 1, \dots, R \quad (7)$$

α_{kN}^r - скалярные произведения:

$$\alpha_{kN}^r = (\bar{q}_{kN}^r, \bar{x}_N), \quad (8)$$

$\bar{x}_N = (x_1, \dots, x_N)^T$ - вектор значений анализируемых отчетов отрезка обрабатываемого входного речевого сигнала, которые соответствуют значениям аргумента $i\Delta t$, т.е. $x_i = x(i\Delta t)$, $i = 1, \dots, N$, N - значение длительности обрабатываемого отрезка речевого сигнала.

Собственные векторы и числа субполосных матриц A_{rN} не зависят от анализируемого отрезка и вычисляются заранее для заданного значения длительности N для каждого $r = 1 - R$ частотного интервала. Для каждого отрезка речевого сигнала $\bar{x}_N = (x_1, \dots, x_N)^T$, вычисляются скалярные произведения α_{kN}^r анализируемого вектора \bar{x}_N и ортогонального собственного вектора \bar{q}_{kN}^r (8).

Избирательное воздействие на векторы частотных компонент \bar{u}_{rN} , предлагается осуществлять путем умножения их на весовые коэффициенты c_r (1). Данный подход предлагается использовать при разработке алгоритмов очистки речевых сигналов от шумов.

Оценить характеристику сосредоточенности энергии звуков русской речи в малом количестве достаточно узких частотных интервалах позволяет выражение вида

$$W_{NR}^i = f_{NR}^{im} / R \quad (8)$$

где f_{NR}^{im} - минимальное количество частотных интервалов (частотная концентрация), в которых сосредоточена заданная доля энергии звукового отрезка.

$$f_{NR}^{im} = \min d_{NR}^{im}. \quad (9)$$

Здесь для правых частей выполняется неравенство

$$\sum_{k=1}^{d_{NR}^{im}} P_{(k)N} \geq m \|\vec{x}_N\|^2 = m \sum_{i=1}^N x_i^2 \quad (10)$$

Верхний индекс i обозначает один из звуков русской речи, индекс в скобках у слагаемых суммы слева соотношения (10) означает, что доли энергий $P_{(k)N}$ упорядочиваются по убыванию:

$$P_{(k)N} \in \{P_{rN}, r = 1, \dots, R\}; P_{(k+1)N} \leq P_{(k)N}, k=1, \dots, R \quad (11)$$

m - задаваемая доля общей энергии, которая должна быть сосредоточена в указанном минимальном количестве частотных интервалов.

Частотная концентрация энергий звуков речи гораздо выше, чем у шумов. В частности для всех звуков речи даже при очень большой доле энергии $m=0.95$ для её частотной концентрации выполняется неравенство

$$W_{NR}' = f_{NR}^{im} / R \leq 0,5 \quad (12)$$

тогда, как в случае отрезков, порождённых только шумами, имеет место

$$W_{NR}' = f_{NR}^{im} / R > 0,7. \quad (13)$$

Данное свойство концентрации заданного количества энергии сигналов, порождаемых звуками русской речи, в малой доле частотной полосы, использовано для построения процедуры селекции информационных и не информационных частотных компонент.

Под информационными компонентами будем понимать компоненты, попадающие в частотные интервалы, сумма долей энергий которых составляет 0,92 от общей доли энергии, содержащейся во всех частотных интервалах. Остальные компоненты называются неинформационными.

Для определения информационных и неинформационных частотных компонент исходный сигнал разбивается на отрезки длительностью N отчетов и R частотных интервалов. Для ускорения вычислений долей энергии для каждого обрабатываемого отрезка речевого сигнала в заданных частотных интервалах используется выражение [12]

$$P_{rN} = \sum_{k=1}^{J_r} \lambda_{kN}^r (\alpha_{kN}^r)^2 \quad (14)$$

Количество элементов J_r определяется из условия

$$J_r = \left[\frac{N}{R} \right] + 2 \quad (15)$$

На основании выражения (14) вычисляются доли энергии для каждого частотного интервала анализируемого отрезка речевого сигнала (рис. 1).

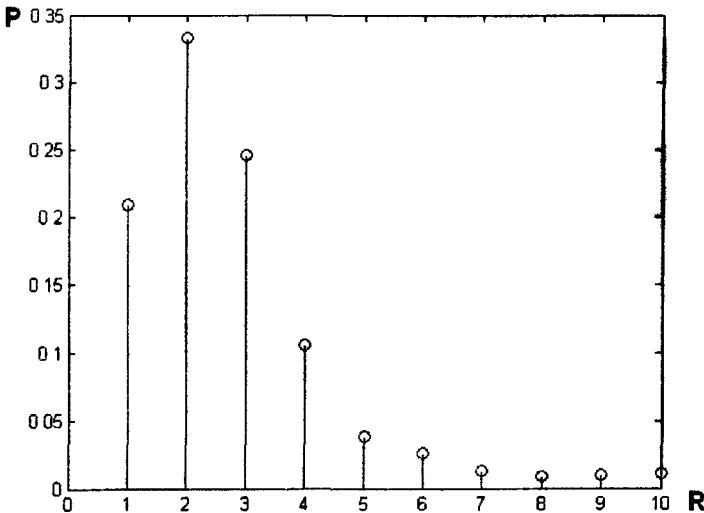


Рис.1

Распределение долей энергий отрезка анализируемого речевого сигнала для каждого частотного интервала

Далее, вычисленные доли энергии каждого частотного интервала ранжируются по убыванию (рис. 2). Те частотные интервалы, сумма долей энергий которых составляет значение 0.92 от общей доли энергии, содержащейся во всех частотных интервалах, принимаются за информационные. Частотные интервалы справа от границы энергии принимаются за неинформационные и путем избирательного воздействия (1) подавляются (рис.3).

Для информационных компонент использовались весовые коэффициенты $c_r = 1$, а для неинформационных $c_r = 0$.

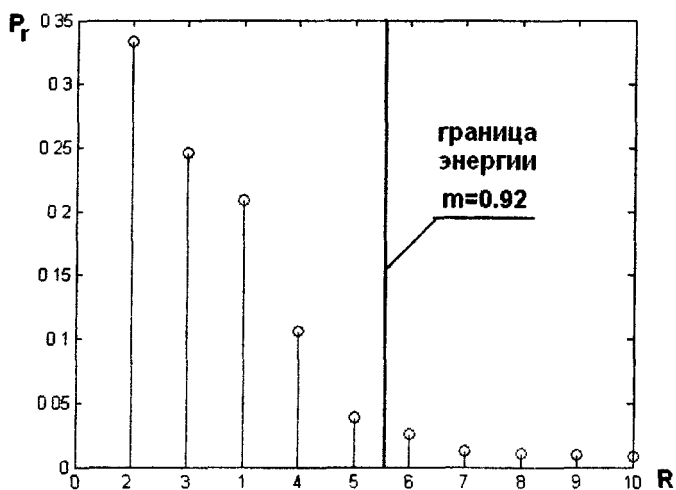


Рис.2

Упорядоченное по убыванию распределение долей энергии отрезка анализируемого речевого сигнала в выбранных частотных интервалах

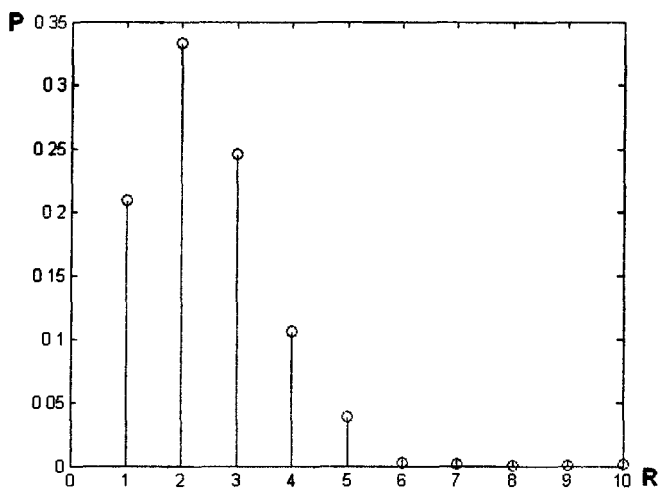


Рис.3

Распределение долей энергии отрезка анализируемого речевого сигнала в заданных частотных интервалах, после подавления неинформационных частотных интервалов

Вычислительные эксперименты

С целью проверки работоспособности предложенных алгоритмов было проведено компьютерное моделирование. Для анализа были взяты звук «а» и звук «ч», вырезанные из незашумленного фрагмента речевого сигнала с частотой дискретизации 8000 Гц и разрядностью 16 бит. Трансформанты Фурье исходных отрезков речевых сигналов, соответствующих звуку «а» и звуку «ч», показаны, соответственно, на рис. 4а и 4б.

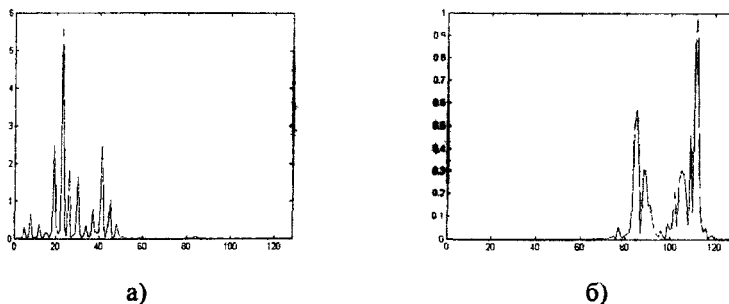


Рис.4

Трансформанты Фурье исходных отрезков речевых сигналов:
а) соответствующие звуку «а»; б) - звуку «ч»

Далее к фрагментам речевого сигнала был добавлен псевдослучайный белый шум с равномерным распределением. Трансформанты Фурье отрезков речевых сигналов, соответствующие зашумленному звуку «а» и звуку «ч», показаны на рис. 5а и 5б, соответственно.

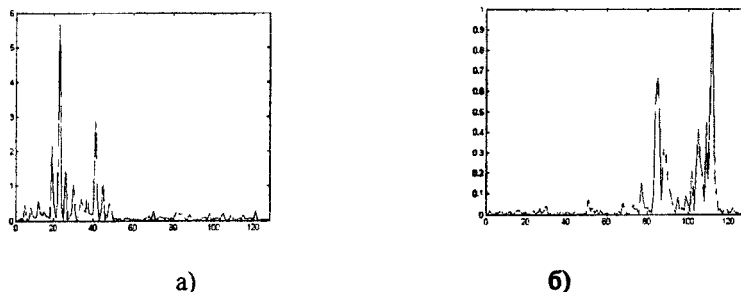


Рис. 5

Трансформанты Фурье отрезков речевых сигналов: а) соответствующие зашумленному звуку «а»; б) - зашумленному звуку «ч»

Затем к зашумленным фрагментам сигналов применялись процедуры, определяемые соотношением (1).

Трансформанты Фурье отрезков речевых сигналов, соответствующие звуку «а» и звуку «ч» после обработки, показаны на рис. 6а и 6б, соответственно.

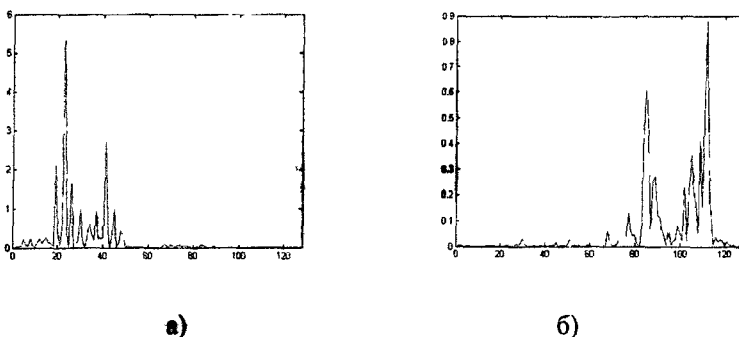


Рис.6

Трансформанты Фурье отрезков речевых сигналов: а) соответствующие обработанному звуку «а»; б) - обработанному звуку «ч»

Анализ результатов моделирования показывает, что после обработки зашумленного отрезка речевого сигнала с помощью предложенных процедур удается в значительной степени снизить уровень малоэнергетических составляющих частотных компонент в высокочастотной (рис 6а) и низкочастотной (рис 6б) областях спектра, соответствующих неинформационным компонентам сигнала. При этом частотные компоненты полезного сигнала, принимаемые как информационные, остаются без изменения.

Независимо от частотного диапазона, в котором сосредоточены информационные компоненты, достигается эффект шумоочистки, что позволяет говорить об адекватности использования рассмотренных компьютерных процедур в задаче очистки речевых сигналов от шумов.

Литература

1. Чучупал В.Я., Чичагов А.С. и Маковкин К.А. Цифровая обработка зашумленных речевых сигналов. М., 1998.52 с.
2. Анощенко А.Е. и Петровский А.А. Перцептуальный метод по-

вышения качества частотно-ограниченного речевого сигнала: одноканальный вариант, комбинированная система, частотный подход. – "Доклады БГУ информатики и радиоэлектроники", 2003, № 4, с. 71-81.

3. Сергеев С.В. и Червяков Н.И. Метод порогового подавления шума в сигнале на основе вейвлет преобразования. – "Инфокоммуникационные технологии", 2008, т. 6, № 2, с. 29-32.

4. Герасимов А.В. и Фидельман В.Р. Фильтрация речевого сигнала на основе анализа собственных чисел и его автокорреляционной матрицы. – В сб.: Труды II Всерос. науч. конф. Проектирование инженерных и научных приложений в среде «MATLAB».- М., 2004.

5. Будников Д.Н. и др. Цифровая фильтрация сигналов при высоком уровне коррелированных шумов в реальном времени. - Вестник НГУ им. Н.И. Лобачевского", 2000, № 1, с. 311-317.

6. Url://www.speechpro.ru

7. Url:// www.grittec.ru

8. Url:// www.dsp.sut.ru

9. Url:// www.spirit.rasu.ru

10. Рабинер Л.Р. Шафер Р.В. Цифровая обработка речевых сигналов. Пер. с англ. Под ред. М.В. Назарова и Ю.Н. Прохорова. М., Радио и связь, 1981. 495 с.

11. Жилияков Е.Г. Вариационные методы анализа и построения функций по эмпирическим данным. Белгород, БелГУ, 2007. 160 с.

12. Жилияков Е.Г., Белов С.П. и Черноорец А.А. Вариационные методы анализа сигналов на основе частотных представлений. - «Вопросы радиоэлектроники», сер. ЭВТ, 2010, вып. 1, с.10- 25.

Статья поступила 12.10.2010