



## О КОМПЬЮТЕРНОЙ ОЧИСТКЕ РЕЧИ ОТ ШУМОВ С ПРИМЕНЕНИЕМ ФИЛЬТРУЮЩЕЙ СУБПОЛОСНОЙ МАТРИЦЫ

**А.В. КУРЛОВ**  
**А.С. БЕЛОВ**  
**А.В. ЭСАУЛЕНКО**  
**С.Л. БАБАРИНОВ**

*Белгородский  
государственный  
университет*

*e-mail: Kurlov@bsu.edu.ru*

*e-mail: belov\_AS@bsu.edu.ru*

В статье рассматривается метод обработки речевых данных, позволяющий повысить разборчивость записанных речевых сигналов при их воспроизведении за счет очистки от шума. Проведено компьютерное моделирование предлагаемого метода с целью проверки его работоспособности.

Ключевые слова: речевые сообщения, субполосная матрица, фильтрация, частотные интервалы, шум.

Широкое использование речевых сообщений человеком в качестве формы обмена информацией является неотъемлемой частью современных тенденции развития информационного обмена. Для человека речевые сообщения являются одной из наиболее часто используемых естественных форм информационного обмена. Кроме непосредственного обмена речевыми сообщениями в настоящее время существенное значение приобрели технологии, обеспечивающие речевой информационный обмен на расстоянии, включая создание хранилищ речевых данных различного назначения: фонотеки, голосовая почта, аудиозаписи видеоконференций, записи лекционных курсов.

В процессе записи речевые данные подвержены влиянию внешних воздействий. Примером таких воздействий являются шумы, которые появляются вследствие неидеальности записывающей аппаратуры, собственных шумов микрофонов, внешней шумовой обстановки, различных электромагнитных наводок. Эти факторы вносят искажения в речевой сигнал, что приводит к ухудшению разборчивости человеком речевых сообщений при их воспроизведении.

В связи с этим, в области цифровой обработки речевых сигналов особое внимание уделяется задачам, связанным с шумоочисткой и повышением разборчивости речевых сообщений, что позволяет говорить об актуальности исследований в данной области [1].

В основе применяемого в статье метода очистки речи от шумов используется тот факт, что энергия звуков речи сосредоточена в малом количестве частотных интервалах. Поэтому для решения задачи определения информационных частотных интервалов, где сосредоточена подавляющая доля энергий, в работе, используется представление[2]:

$$P_{rN}(\vec{x}) = \vec{x}_N^T A_{rN} \vec{x}_N, \tag{1}$$

где  $A_{rN} = \{a_{ik}^r\}$  субполосная матрица с элементами вида:

$$a_{ik}^r = \begin{cases} \frac{\sin[V_r(i-k)] - \sin[V_{r-1}(i-k)]}{\pi(i-k)}, & i \neq k \\ \frac{V_r - V_{r-1}}{\pi}, & i = k \end{cases}, \tag{2}$$

границы частотных интервалов  $V_r = \pi, V_0 = 0; r = 1, \dots, R$ .



$\bar{x}_N = (x_1, \dots, x_N)^T$  – вектор значений анализируемых отчетов отрезка обрабатываемого входного речевого сигнала, которые соответствуют значениям аргумента  $i\Delta t$ , т.е.  $x_i = x(i\Delta t)$ ,  $i, k = 1, \dots, N$ ;  $N$  – значение длительности обрабатываемого отрезка речевого сигнала;  $V_{r-1}, V_r$  – границы частотного интервала. Ось частот в диапазоне от 0 до  $\pi$  разбивается на  $R$  частотных интервалов одинаковой ширины, равной  $\pi/R$ . В каждом частотном интервале определяются значения долей энергий. Для ускорения вычислений долей энергии для обрабатываемого отрезка речевого сигнала в заданных частотных интервалах используется выражение [2]

$$P_{rN} = \sum_{k=1}^{J_r} \lambda_{kN}^r (\alpha_{kN}^r)^2, \quad k = 1, \dots, N; r = 1, \dots, R, \quad (3)$$

где  $\lambda_{kN}^r$  – собственные числа субполосной матрицы  $A_{rN}$ ,  $\alpha_{kN}^r$  – скалярные произведения анализируемого вектора и соответствующего ортогонального собственного вектора  $\alpha_{kN}^r = (\bar{q}_{kN}^r, \bar{x}_N)$ ,  $\lambda_{kN}^r \bar{q}_{kN}^r = A_{rN} \bar{q}_{kN}^r, k = 1, \dots, N; r = 1, \dots, R$ . Количество элементов  $J_r$  в выражении (3) определяется из условия

$$J_r = \left\lceil \frac{V_r - V_{r-1}}{\pi} \cdot N \right\rceil + 2, \quad (4)$$

где,  $N$  – длительность анализируемого отрезка речевого сигнала.

Выражение вида [2]

$$W_{NR}^i = f_{NR}^m / R \quad (5)$$

позволяет оценить характеристику сосредоточенности энергии звуков русской речи в малом количестве достаточно узких частотных интервалах, где  $f_{NR}^m$  – минимальное количество частотных интервалов (частотная концентрация), в которых сосредоточена заданная доля энергии анализируемого отрезка речевого сигнала.

$$f_{NR}^m = \min d_{NR}^m. \quad (6)$$

Для правых частей выполняется неравенство

$$\sum_{r=1}^{d_{NR}^m} P_{(r)N} \geq m \|\bar{x}_N\|^2 = m \sum_{i=1}^N x_i^2. \quad (7)$$

Здесь  $m$  – задаваемая доля общей энергии, которая должна быть сосредоточена в указанном минимальном количестве частотных интервалов.  $P_{(r)N}$  – упорядоченные по убыванию доли энергий

$$P_{(r)N} \in \{P_{rN}, r = 1, \dots, R\}; P_{(r+1)N} \leq P_{(r)N}, r = 1, \dots, R, \quad (8)$$

Было установлено, что для отрезков сигнала, содержащего звуки речи выполняется неравенство

$$W_{NR}^i = f_{NR}^m / R \leq 0,5, \quad (9)$$

а для отрезков сигнала, порожденными только шумами

$$W_{NR}^i = f_{NR}^m / R > 0,7. \quad (10)$$

Данное свойство звуков русской речи, использовано для построения процедуры определения 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, не входящие в эту сумму будем считать неинформационными. Очевидно, что эти компоненты будут обладать малой энергией и вследствие чего могут быть отнесены к шумовым.



Далее, на основании выражения (3) вычисляются доли энергии для каждого частотного интервала, затем упорядочиваются по убыванию (8). При этом записываются номера частотных интервалов. Частотные интервалы, сумма долей энергий которых составляет значение  $m=0.92$ , принимаются за информационные, а частотные интервалы, не входящие в эту сумму долей энергий, принимаются за неинформационные.

Таким образом, получаем номера частотных интервалов  $r \in R_1$ , которые относятся к информационным.

Процедура фильтрации осуществляется в соответствии с принципом минимизации погрешности приближения к идеальному случаю:

$$Y(\omega) = \begin{cases} X(\omega), & \omega \in \Omega_r \\ 0, & \omega \notin \Omega_r \end{cases}, \quad (11)$$

$$\Omega_r = [-V_r, -V_{r-1}) \cup [V_{r-1}, V_r). \quad (12)$$

$$\frac{1}{2\pi} \int_{\omega \notin \Omega_r} |X(\omega) - Y(\omega)|^2 d\omega + \frac{1}{2\pi} \int_{\omega \in \Omega_r} |Y(\omega)|^2 d\omega = \min, \quad (13)$$

где  $X(\omega)$  – трансформанта Фурье исходного сигнала,  $Y(\omega)$  – трансформанта Фурье отфильтрованного сигнала.

Таким образом, фильтрация информационных частотных интервалов определяется выражением:

$$\tilde{x}_N = \sum_{r \in R_1} A_{rN} \cdot \bar{x}_N, \quad (14)$$

где  $\tilde{x}_N$  – вектор значений выходного отфильтрованного речевого сигнала,  $\bar{x}_N$  – вектор значений анализируемого речевого сигнала,  $r$  – информационные частотные интервалы,  $A_{rN}$  – фильтрующая субполосная матрица с элементами вида (2) с информационными частотными интервалами  $r \in R_1$ .

Для проверки работоспособности предложенного алгоритма было проведено компьютерное моделирование.

В качестве речевых сигналов были взяты отрезки, порождаемые звуками «а», «о», «ч», «ш», вырезанные из исходного анализируемого фрагмента сигнала с частотой дискретизации 8000 Гц и разрядностью 16 бит. Звуки выбирались исходя из того, что бы у одних (а, о) подавляющая доля энергии была сосредоточена в низкочастотной области, а у других (ч, ш) в высокочастотной.

Запумление фрагментов речевого сигнала проводилось по формуле:

$$U_k = X_k + \varepsilon_k, \quad (15)$$

где  $U_k$  – запумленный сигнал,  $X_k$  – исходный сигнал,  $\varepsilon_k = c \cdot \eta_k$ ,  $\eta_k$  – выборка из

нормального распределения,  $c = q \sqrt{\frac{\sum X_k^2}{\sum \eta_k^2}}$ , а  $q$  – определяет отношение шум/сигнал

задаваемое в разгах,  $k = 1, \dots, N$ . В данной работе были выбраны параметры моделирования  $q=0.3$ ,  $N=60$ .  $R=10$ .

Далее к запумленным отрезкам речевого сигнала применялась процедура, определяемая выражением (14). Используя представление (3), были получены распределения долей энергии по частотным интервалам для каждого звука. Результаты моделирования представлены на рис. 1-6.

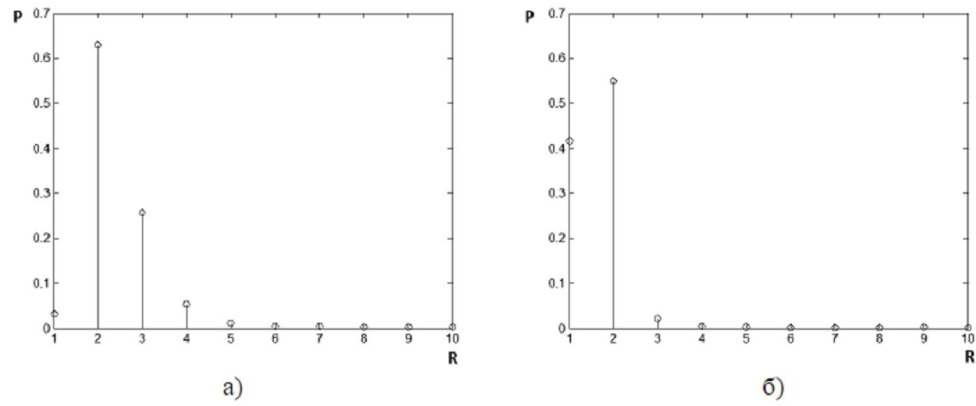


Рис. 1. Распределение долей энергии по частотным интервалам:  
а) соответствующие исходному звуку «а»; б) звуку «о»

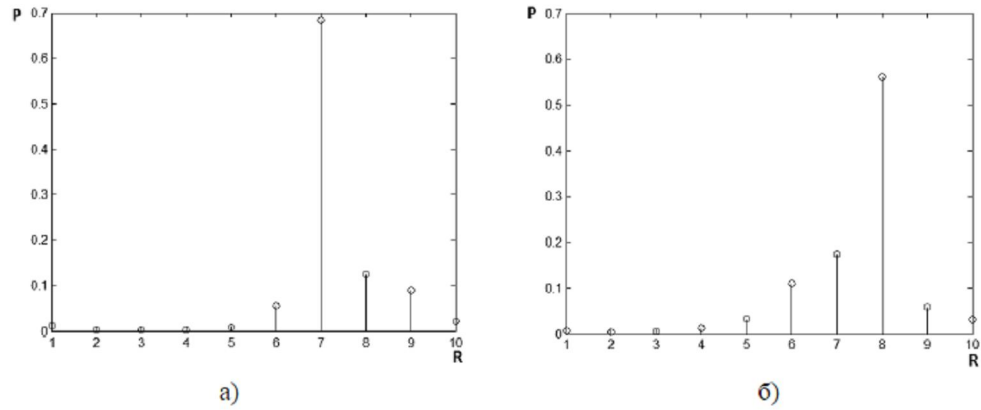


Рис. 2. Распределение долей энергии по частотным интервалам:  
а) соответствующие исходному звуку «ч»; б) звуку «ш»

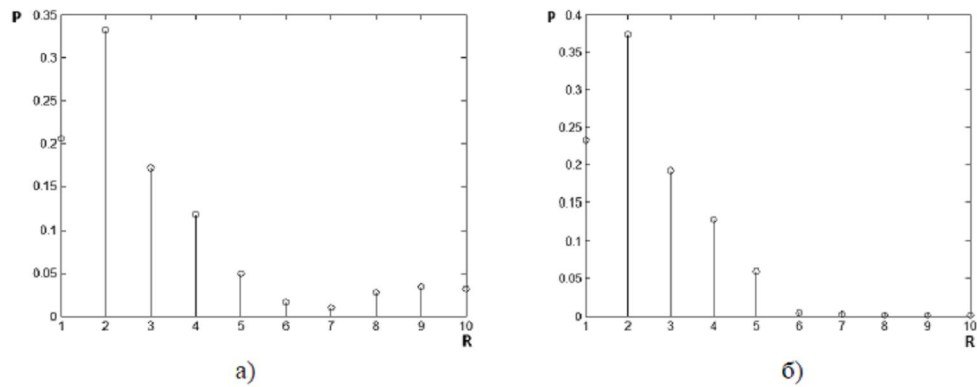
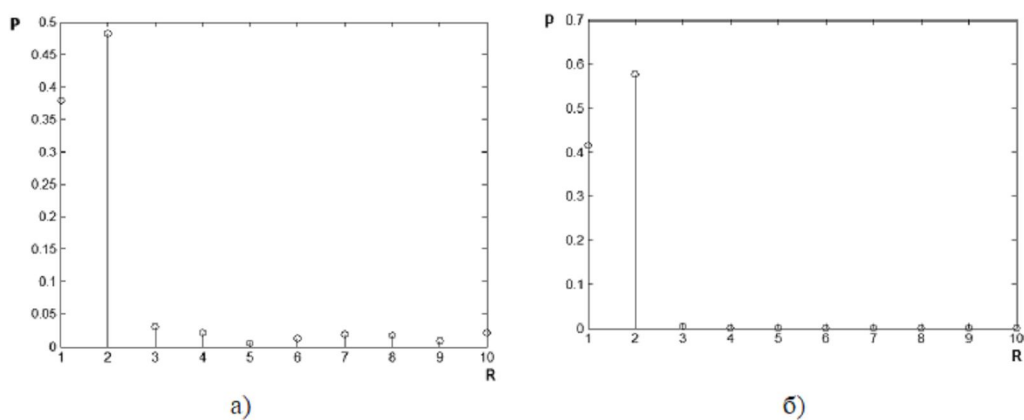
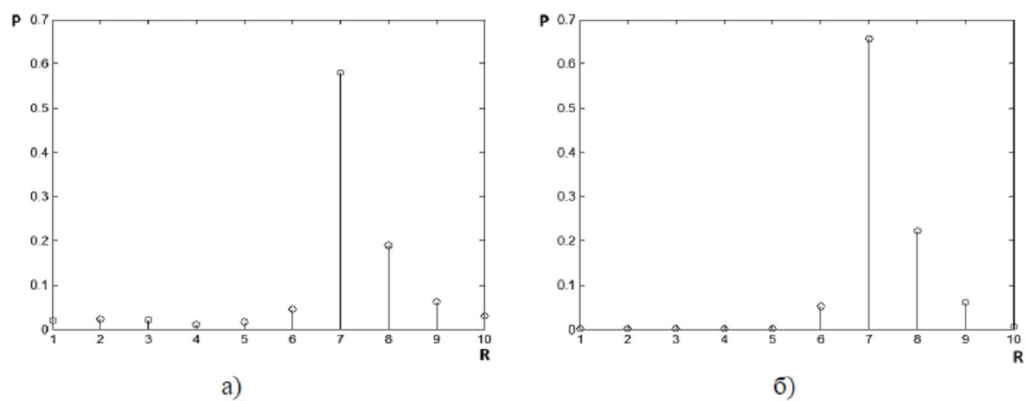


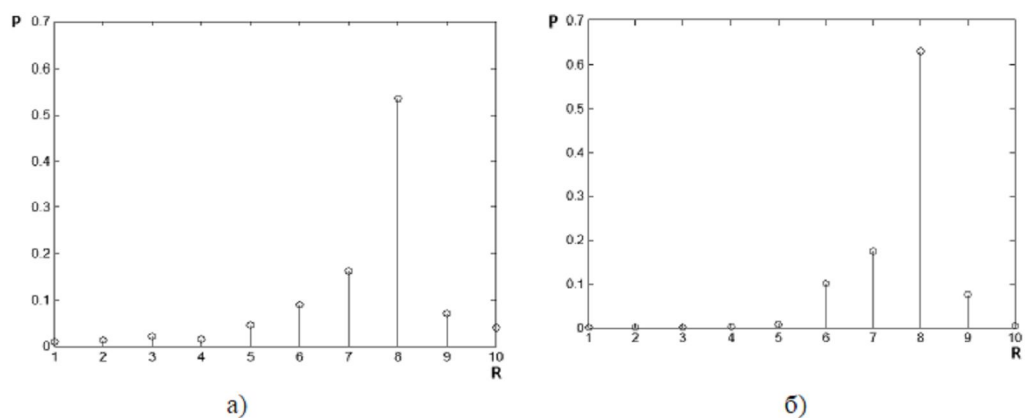
Рис. 3. Распределение долей энергии по частотным интервалам:  
а) соответствующие зашумленному звуку «а»; б) после процедуры фильтрации



**Рис. 4.** Распределение долей энергии по частотным интервалам:  
а) соответствующие зашумленному звуку «о»; б) после процедуры фильтрации



**Рис. 5.** Распределение долей энергии по частотным интервалам:  
а) соответствующие зашумленному звуку «ч»; б) после процедуры фильтрации



**Рис. 6.** Распределение долей энергии по частотным интервалам:  
а) соответствующие зашумленному звуку «ш»; б) после процедуры фильтрации





Анализ результатов моделирования показывает, что с помощью предложенных процедур фильтрации удастся в значительной степени снизить уровень частотных компонент, обладающих малой энергией. Причем эффект фильтрации достигается не зависимо от частотного диапазона, в котором сосредоточены информационные компоненты, будь то низкочастотная (рисунок 3 б, 4 б) или высокочастотная (рисунок 5 б, 6 б) области, что позволяет говорить об адекватности использования предлагаемого метода в задаче очистки речевых сигналов от шумов.

Необходимо отметить, что частотные компоненты полезного сигнала, определенные как информационные, остаются без изменения, в то время как неинформационные путем избирательного воздействия подавляются, благодаря чему удается в значительной степени снизить уровень шумов в речевом сигнале.

*Работа выполнена при финансовой поддержке РФФИ (проект № 10-07-00326-а).*

#### Литература

1. Чучупал В.Я. Цифровая обработка зашумленных речевых сигналов [Текст] / В.Я. Чучупал, А.С. Чичагов, К.А. Маковкин. – Вычислительный центр РАН, М.1998г. – 52 с.
2. Жилияков Е.Г. Вариационные методы анализа и построения функций по эмпирическим данным: монография. Белгород: Изд-во БелГУ, 2007. – 160 с.
3. Жилияков Е.Г. Вариационные методы анализа сигналов на основе частотных представлений [Текст]/ Е.Г. Жилияков, С.П. Белов, А.А. Черноморец. – «Вопросы радиоэлектроники», сер. ЭВТ, 2010, вып. 1. – С. 10-25.

## OF COMPUTER CLEANING OF SPEECH FROM NOISE USING FILTERS SUBSTRIP MATRIX

**A.V. KURLOV**  
**A.S. BELOV**  
**A.V. ESAULENKO**  
**S.I. BABARINOV**

*Belgorod State University*

*e-mail: Kurlov@bsu.edu.ru*  
*e-mail: belov\_AS@bsu.edu.ru*

The article discusses a method of processing voice data, which increases the intelligibility of recorded speech signals when they are playing through the removal of noise. A computer simulation of the proposed method to test its efficiency.

Key words voice messages, substrip matrix, filtering, frequency intervals, noise.