



ИССЛЕДОВАНИЕ РЕШАЮЩЕЙ ФУНКЦИИ МАКСИМАЛЬНОЙ ЧУВСТВИТЕЛЬНОСТИ К ИЗМЕНЕНИЮ ЧАСТЕЙ ЭНЕРГИИ В ЧАСТОТНЫХ ИНТЕРВАЛАХ

**С.П. БЕЛОВ
А.А. ФИРСОВА**

*Белгородский государственный
национальный исследовательский университет*

*e-mail:
Belov@bsu.edu.ru
Firsova_A@bsu.edu.ru*

В статье рассмотрена решающая функция максимальной чувствительности к изменению частей энергии в частотных интервалах. Проведена оценка чувствительности решающей функции на участке, соответствующем шуму в помещении, к увеличению энергии сигнала на всем диапазоне частот и оценка чувствительности решающей функции на участке, соответствующем звуку речи, к увеличению отношения шум/сигнал.

Ключевые слова: речевой сигнал, частотные представления, алгоритм обнаружение пауз в речевых сигналах, решающая функция.

Речевой сигнал состоит из участков, соответствующих шумам в паузах между звуками речи, а также участков, соответствующих сумме сигналов шумов и сигналов звуков речи. В различных алгоритмах обработки речевых сигналов на первом этапе используется алгоритм обнаружения пауз [1,2]. Алгоритмы обнаружения пауз используются при реализации алгоритмов сжатия, очистки от шумов, сегментации и распознавания речевых сигналов. При реализации алгоритма обнаружения пауз используются различия отрезков сигналов, соответствующих звукам речи, и сигналов, соответствующих шумам в паузах.

Исследования особенностей распределения энергии звуков русской речи и различных типов шумов показывает, что энергия сигналов, соответствующих звукам русской речи сосредоточена в достаточно узком частотном интервале, в то время как энергия сигналов, соответствующих шумам в паузах распределена более равномерно по всему частотному диапазону. Анализ распределения энергии отрезков сигнала, соответствующего шуму в паузе показывает, что различные отрезки имеют различное распределение энергии по частотным интервалам. При этом отклонение в распределении энергии каждого отрезка шума имеют меньшее отклонение от усредненного значения распределения энергии, чем отклонение суммы сигналов шума и звука речи.

На первом этапе формулируется нулевая гипотеза H_0 : анализируемый отрезок сигнала относится к шумам в паузе между звуками речи. В качестве решающей функции предлагается использовать выражение вида [3]:

$$S = \max_{r=1, \dots, R} \left(\frac{P_r}{P_{0,r}} \right), \tag{1}$$

где P_r – энергия анализируемого отрезка сигнала в r -ом частотном интервале,

$P_{0,r}$ – значение энергии в r -ом частотном интервале, полученное в результате усреднения по большому количеству отрезков сигнала, соответствующего шуму:

$$P_{0,r} = \frac{\sum_{ot=1}^{N_{ot}} P_{ot,r}}{N_{ot}}, \tag{2}$$

где $r=1, \dots, R$ – количество интервалов, на которые разбивается ось частот.

N_{ot} – количество окон анализа,

$P_{ot,r}$ – энергия в r -ом частотном интервале отрезка ot .

Для оценки распределения энергии сигнала по частотным интервалам предлагается использовать выражение вида [3]:

$$P_r = \|\vec{y}_r\|^2 = \sum_{k=1}^{J_{rr}} (y_{rk}^2), \tag{3}$$

где

$$y_{rk} = \sqrt{\lambda_{rk}} (\vec{q}_{rk}, \vec{x}), \tag{4}$$

где $r=1, \dots, R$ – частотных интервалов,



$k = 1, \dots, J$ – собственных чисел $\lambda \approx 1$,

\vec{q}_{rk} – собственный вектор субполосной матрицы A_r ,

\vec{x} – анализируемый отрезок сигнала.

Решение о принадлежности отрезка сигнала к участку шума в паузах или звуку речи осуществляется на основе неравенства:

$$S > h_\alpha, \quad (5)$$

где h_α – пороговое значение, обеспечивающее заданное значение вероятности ложной тревоги α .

Если неравенство (5) выполняется, то нулевая гипотеза отвергается, и анализируемый отрезок сигнала относится к участку смеси звука речи и шума. В противном случае, анализируемый отрезок сигнала относится к шуму в паузах.

Пороговое значение определяется на основе анализа достаточно большого количества отрезков сигнала, относящегося к шуму в паузе. Для оценки порогового значения h_α используются математическое ожидание и дисперсия решающей функции вида (1), рассчитанной для всех отрезков сигнала шума, используемого для обучения:

$$\bar{S}_\Pi = \sum_{k=1}^{N_{ot}} (S_k^\Pi) / N_{ot}, \quad (6)$$

$$D_\Pi^2 = \sum_{k=1}^{N_{ot}} (S_k^\Pi)^2 / N_{ot} - \bar{S}_\Pi^2, \quad (7)$$

где S_k^Π – значения решающей функции, рассчитанной для обучающей выборки сигнала, соответствующего шуму,

\bar{S}_Π – математическое ожидание решающей функции,

D_Π^2 – дисперсия решающей функции.

Оценка порогового значения при заданном уровне вероятности ложной тревоги осуществляется на основе неравенства Чебышева:

$$h_\alpha \leq \bar{S}_\Pi + D_\Pi / a\sqrt{\alpha}. \quad (8)$$

где \bar{S}_Π – математическое ожидание решающей функции,

D_Π – среднеквадратическое отклонение решающей функции,

a – значение, превышающее 2, уточняется в процессе обучения,

α – значение вероятности ложной тревоги.

Эффективность работы представленного алгоритма зависит от отношения шум/сигнал, а также от изменения уровня шумов.

Для исследования чувствительности алгоритма обнаружения пауз к увеличению отношения шум сигнал были оценены зависимости вероятности ошибочного принятия решения на участке, соответствующем шуму в помещении, к увеличению энергии сигнала на всем диапазоне частот и на участке, соответствующем звуку речи, к увеличению отношения шум/сигнал.

Для исследования зависимости вероятности ошибочного принятия решения на участке, соответствующем шуму в помещении, к увеличению энергии сигнала на всем диапазоне частот в качестве исходного сигнала был использован отрезок шума в комнате длительностью $Dl=25000$ отсчетов (3, 125 с). Оценка вероятностей ошибок первого рода оценивалась на основе анализа 24900 отрезков сигнала, соответствующего шуму в помещении. При этом производилась оценка пороговых значений для исходного сигнала, а затем энергия каждого отсчета шума увеличивалась в K раз. На рис. 1, 2 представлены графики зависимостей вероятности ошибки первого рода от коэффициента K при различных параметрах алгоритма.

Анализ рис.1 показывает, что для параметров $N=65$, $R=16$ с увеличением коэффициента K от 1,1 до 3 вероятность ошибочного принятия решения о наличии сигнала увеличивается от 0,0001 до 0,0061. Значение вероятности ошибки первого рода при значении коэффициента $K=3$ (0,0061) является приемлемым для алгоритма обнаружения пауз. На рисунке можно выделить три области: две области нечувствительности к изменению коэффициента K (от 1,1 до 1,2 и от 1,6 до 3) и одну область чувствительности к изменению коэффициента K (от 1,2 до 1,6). При этом важно отметить, что наибольшее изменение наблюдается при увеличении коэффициента от 1,2 до 1,3. В данном случае вероятность ошибки первого рода увеличивается на 52,32% (от 0,0006 до 0,0038). В свою

очередь в остальных случаях увеличение коэффициента на 0,1 приводит к увеличению вероятности ошибки первого рода в среднем на 2,51%.

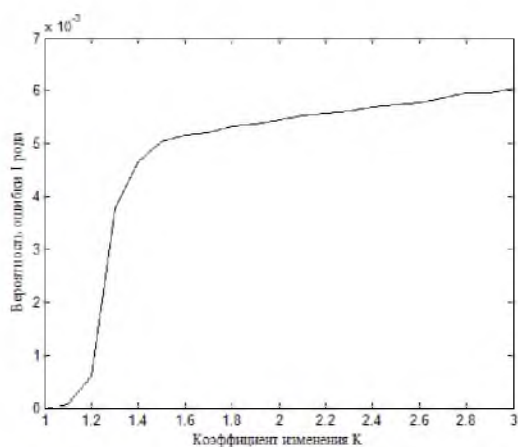


Рис. 1. График зависимости вероятности ошибки I рода от коэффициента K ($N=65, R=16$)

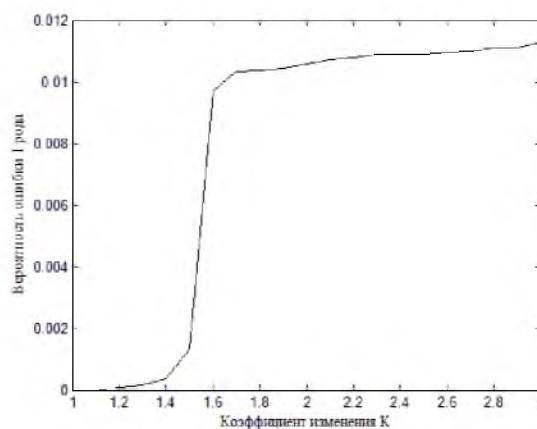


Рис. 2. График зависимости вероятности ошибки I рода от коэффициента K ($N=129, R=32$)

Анализ рис. 2 показывает, что при выборе параметров $N=129, R=32$ вероятность ошибки первого рода при увеличении коэффициента от 1,1 до 3 изменяется в диапазоне от 0,0001 до 0,0113. На графике можно выделить три области: две области нечувствительности к изменению коэффициента K (от 1,1 до 1,4 и от 1,7 до 3) и одну область чувствительности к изменению коэффициента K (от 1,4 до 1,7). Наибольшее изменение наблюдается при увеличении коэффициента от 1,5 до 1,6. В этом случае вероятность ошибки увеличивается на 74,02% (от 0,0013 до 0,0097). В остальных случаях увеличение коэффициента на 0,1 приводит к увеличению вероятности ошибки в среднем на 1,37%.

Для принятия решения о принадлежности анализируемого отрезка к речевому сигналу или к паузе предварительно было оценено пороговое значение при вероятности ошибки первого рода равной 0,002. Пороговое значение при $N=65$ составляет 397, а при $N=129$ – 330. На ри. 3 представлены графики зависимости порогового значения от коэффициента K при достижении вероятности ошибки первого рода 0,002 для различных параметров алгоритма.

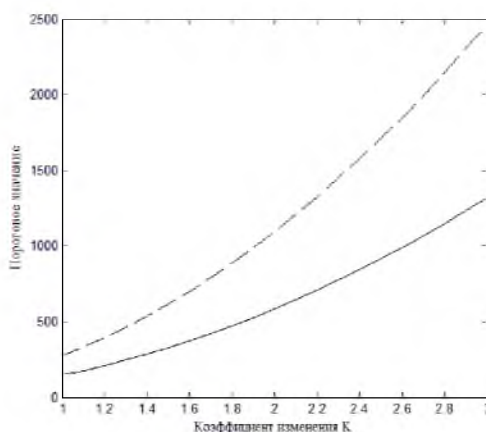


Рис. 3. Графики зависимости пороговых значений от коэффициента K при:
 – $N=129, R=32$;
 -- $N=65, R=16$

Анализ рисунков показывает, что графики зависимостей имеют одинаковый характер. При этом крутизна графика зависимости порогового значения от коэффициента изменения K при $N=65, R=16$ больше, чем при $N=129, R=32$.

На втором этапе исследований была оценена вероятность ошибочного принятия решения на участке, соответствующем звуку речи, к увеличению отношения шум/сигнал. Для исследования зависимости ошибки второго рода от отношения шум/сигнал использовались сигналы, соответст-



вующие звукам русской речи. На рис. 4, 5 представлены графики зависимостей вероятности ошибки второго рода от отношения шум/сигнал для звука «а» при различных параметрах.

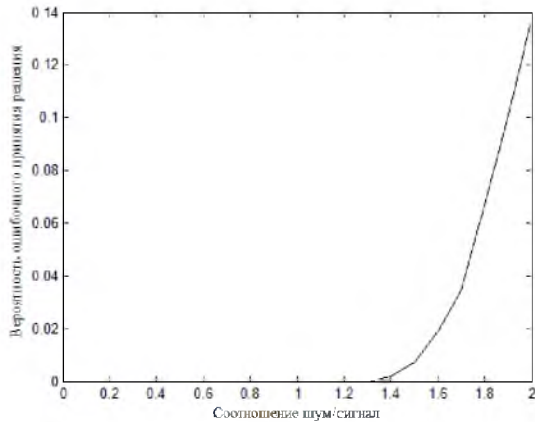


Рис. 4. График зависимости вероятности ошибки второго рода для звука «а» ($N=65, R=16, \alpha=0,002$)

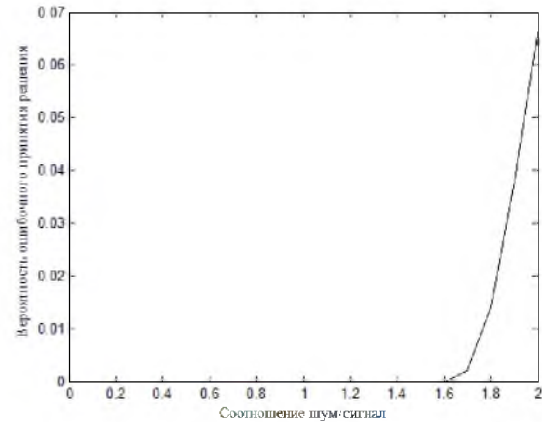


Рис. 5. График зависимости вероятности ошибки второго рода для звука «а» ($N=129, R=32, \alpha=0,002$)

Анализ результатов экспериментов показывает, что при выборе длины окна анализа $N=65$, вероятность ошибки второго рода резко возрастает при отношении шум/сигнал больше 1,2 для звуков: «з», «р», «п»; 1,3: «л»; 1,4: «д», «т», «ф», «э»; 1,5: «а», «г», «е», «й», «к», «о», «с», «ц»; 1,6: «б», «в», «ж», «и», «у», «х», «ч», «щ»; 1,7: «м», «ы», «ю», «я»; 1,8: «н», «ш»; 1,9: «ё». При выборе длины окна анализа $N=129$ вероятность ошибки второго рода меньше, чем для тех же звуков при выборе длины окна анализа $N=65$. Увеличение отношения шум/сигнал до 2 не приводит к увеличению вероятности ошибки второго рода для таких звуков как: «б», «в», «г», «ё», «м», «н», «у», «ш», «ы», «ю», «я». При значении отношения шум/сигнал больше, чем 1,4 вероятность ошибки второго рода резко возрастает для таких звуков, как: «р»; 1,5: «з», «п»; 1,6: «с», «т», «ц»; 1,7: «а», «ж», «л», «о», «ф», «х», «ч», «щ», «э»; 1,8: «д», «и», «й», «к»; 1,9: «е». Сравнение результатов показывает, что выбор длины окна анализа $N=129$ позволяет уменьшить вероятность ошибочного принятия решения второго рода.

На рис. 6 представлены результаты усреднения значений вероятностей ошибки второго рода по всем звукам русской речи при различных отношениях шум/сигнал.

Усреднение осуществлялось с использованием выражения:

$$\bar{P}_2(q) = \frac{\sum_{i=1}^M P_{2i}(q) \cdot p_i}{\sum_{i=1}^M p_i}, \quad (9)$$

где M – количество звуков русской речи;

$P_{2i}(q)$ – вероятность ошибки второго рода для i -го звука при отношении шум/сигнал q ;

p_i – вероятность появления i -го звука.

Вероятности появления звуков представлены в табл. 1 [4].

Таблица 1

Вероятность появления звуков русской речи							
звук	p	звук	p	звук	p	звук	p
а	0,121	з	0,016	п	0,023	ч	0,012
б	0,014	и	0,062	р	0,04	ш	0,006
в	0,038	й	0,01	с	0,045	щ	0,003
г	0,013	к	0,028	т	0,053	ы	0,016
д	0,025	л	0,035	у	0,021	э	0,086
е	0,072	м	0,026	ф	0,002	ю	0,006
ё	0,046	н	0,053	х	0,009	я	0,018
ж	0,007	о	0,09	ц	0,004		

На рис. 6, 7 представлены совмещенные графики зависимостей вероятности ошибки первого и второго рода при различных параметрах алгоритма.

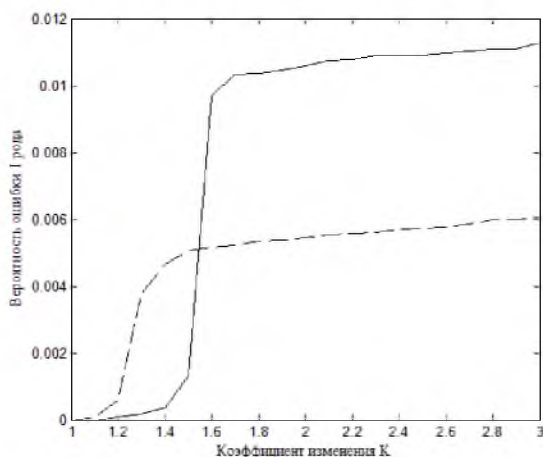


Рис. 6. Графики зависимости значений вероятностей ошибки первого рода от коэффициента K при:
 – $N=129, R=32$;
 -- $N=65, R=16$

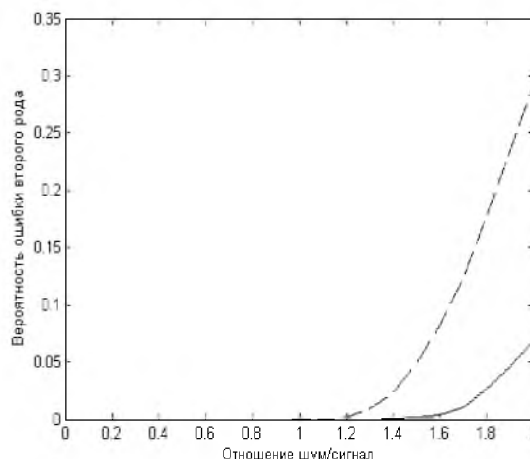


Рис. 7. Графики зависимости значений вероятностей ошибки второго рода от отношения шум/сигнал при:
 – $N=129, R=32$;
 -- $N=65, R=16$

Сравнение графиков, представленных на рис. 6, показывает, что при малых значениях коэффициента (1,1÷1,5) вероятность ошибки первого рода меньше при $N=129, R=32$. При увеличении значения коэффициента вероятность ошибки первого рода для $N=129, R=32$ значительно превышает значение вероятности ошибки первого рода для $N=65, R=16$. Величина ошибки первого рода при $N=129, R=32$ на этом участке является неприемлемой. Анализ рис. 7 показывает, что при любом значении отношения шум/сигнал вероятность ошибки второго рода больше при длине окна анализа $N=65$. При выборе длины окна анализа $N=65$ алгоритм может быть использован при соотношении шум/сигнал не более 1,2, а при выборе длины окна анализа $N=129$: не более 1,6.

Таким образом, исследование решающей функции максимальной чувствительности к изменению частей энергии в частотных интервалах показало, что при использовании представленного алгоритма наиболее эффективно при малом отношении шум/сигнал. Уменьшить вероятность ошибочного принятия решения алгоритма можно за счет изменения порогового значения в зависимости от уровня шумов в канале.

Исследования выполнены при поддержке ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009-2013 годы, ГК № П964 от 27 мая 2010г.

Список литературы

1. Рабинер, Л.Р. Цифровая обработка речевых сигналов / Л.Р. Рабинер, Р.В. Шафер – М.: Радио и связь, 1981. – 496с.
2. Шелухин, О.И. Цифровая обработка и передача речи [Текст] / О.И. Шелухин, Н.Ф.Лукьянцев; под ред. О.И. Шелухина. – М.: Радио и связь, 2000. – 456 с.: ил.
3. Жилияков Е.Г. Методы обработки речевых данных в информационно-телекоммуникационных системах на основе частотных представлений [Текст] / Е.Г. Жилияков, С.П. Белов, Е.И. Прохоренко – Белгород, 2007. – 136 с.
4. Яглом, А.М. Вероятность и информация / А.М. Яглом, И.М. Яглом – М.: КомКнига, 2007. – 512 с.

RESEARCH DECISION FUNCTION MAXIMUM SENSITIVITY TO CHANGES OF THE ENERGY IN THE FREQUENCY INTERVAL

**S.P. BELOV
 A.A. FIRSOVA**

*Belgorod National Research University
 e-mail:
 Belov@bsu.edu.ru
 Firsova_A@bsu.edu.ru*

In this article the crucial function of the maximum sensitivity to changes in the energy in the frequency domain. The estimation of the sensitivity of the decision function at the site corresponding to the noise in the room to increase the energy of the signal at the frequency range and sensitivity of the decision function at the site corresponding to the sound of speech, to the increase in the ratio noise / signal.

Keywords: speech signal, frequency representation, voice activity detector, the decision function.