

УДК 621.39

**Жиляков Е.Г. [Zhilyakov E.G.],
Белов С.П. [Belov S.P.],
Медведева А.А. [Medvedeva A.A.]****ИССЛЕДОВАНИЕ ЧУВСТВИТЕЛЬНОСТИ
СУБПОЛОСНОГО МЕТОДА
ОБНАРУЖЕНИЯ ПАУЗ
К ВОЗДЕЙСТВИЮ ШУМА*****Research of sensitivity of detection method
subband pause noise impact**

В статье представлен метод селекции пауз в речевых сообщениях на основе применения субполосного анализа. Исследование чувствительности проводилось при использовании аддитивного наложения нормально распределенного шума. Были определены вероятности ошибочного принятия решения при различных отношениях шум/сигнал. Результаты свидетельствуют об устойчивости предлагаемого субполосного метода к воздействию аддитивного шума.

Ключевые слова: шум, чувствительность, обнаружение пауз.

The article presents a selection method pauses in speech communications through the use of sub-band analysis. sensitivity study was conducted using an additive blending of normally distributed noise. the probability of wrong decision-making have been identified in various respects noise / signal. The results indicate that the stability of the proposed method of sub-band to the effects of the additive noise.

Key words: noise, sensibility, pause detection.

Одним из основных этапов обработки речевых сигналов при решении таких задач, как распознавание речи, сжатие речи, очистка речи от шума и т.п., является определение участков отсутствия речи – селекции пауз. Точность определения границ пауз влияет на эффективность дальнейших этапов анализа и обработки. В частности, на вероятность правильного распознавания, качество звучания речи и степень сжатия.

Известно, что различные звуки речи и шумы в паузах имеют разное распределение энергии в частотной области. Учет этих особенностей позволяет определить границы между звуками речи и паузами. Однако, как по-

* Работа подготовлена в рамках II Международной конференции «Параллельная компьютерная алгебра и ее приложения в новых инфокоммуникационных системах»

казывают исследования, при реализации методов селекции пауз необходимо учитывать несколько характеристик сравниваемых отрезков сигналов.

Основная проверяемая гипотеза формулируется следующим образом.

H_0 - анализируемый отрезок РС $\vec{x} = (x_1, \dots, x_N)^T$ порожден шумами в паузе речи

$$\vec{x} = \vec{u} = (u_1, \dots, u_N)^T. \quad (1)$$

Альтернатива H_1 заключается в том, что хотя бы часть компонент рассматриваемого вектора зафиксирована в присутствии звуков речи

$$\vec{x} = \vec{u} + \vec{s}, \quad \vec{s} = (s_1, \dots, s_N)^T \quad (2)$$

В качестве решающей функции при селекции пауз предлагается использовать статистику [1]:

$$F_u = W_u(x) \cdot \gamma_u(x), \quad (3)$$

где $W_u(x)$ – характеристика, учитывающая различия в значениях энергии,

а $\gamma_u(x)$ – мера различия в распределении долей энергии по частотной оси сравниваемых отрезков.

Мера различия в значениях энергии предлагается оценивать как отношение энергии анализируемого отрезка сигнала к средней энергии, определенной на основе обучения по фрагменту, соответствующему шуму в паузе:

$$W_u(x) = \|\vec{x}\|^2 / G_u, \quad (4)$$

где $|\vec{x}|^2$ – энергия (квадрат евклидовой нормы) анализируемого вектора;

G_u – математическое ожидание квадратов евклидовых норм отрезков шумов в паузах

$$G_u = M[|\vec{u}|^2]. \quad (5)$$

В свою очередь, меру различия распределения долей энергий по частотным интервалам предлагается определять в соответствии с выражением, в основе которого лежит аналог расстояния Питмена [2]:

$$\gamma_u(x) = \left(\sum_{n=0}^{N/2-1} ((Pd_n(\vec{x}))^{1/2} - D_n)^2 \right)^{1/2} = \left(2 \left(1 - \sum_{n=0}^{N/2-1} D_n (Pd_n(\vec{x}))^{1/2} \right) \right)^{1/2}, \quad (6)$$

где $Pd_n(\vec{x})$ – значение доли энергии, сосредоточенной в n -м частотном интервале:

$$Pd_n(\vec{x}) = P_n(\vec{x}) / \sum_{k=0}^{N/2-1} P_k(\vec{x}), \quad n = 0, \dots, N/2 - 1, \quad (7)$$

D_n^2 – математическое ожидание долей энергий отрезков шумов в паузах

$$D_n^2 = M[Pd_n(\vec{u})], \quad n = 0, \dots, N/2 - 1, \quad (8)$$

N – длительность отрезка анализа.

На рисунках 1-3 представлены фрагмент РС, порожденного словом «черепаша», и результат оценки характеристик (4) и (6). При этом значение математических ожиданий D_n^2 и G_u определялось на основе анализа фрагмента шума в паузе в начале фрагмента сигнала длительностью 0,19 сек.

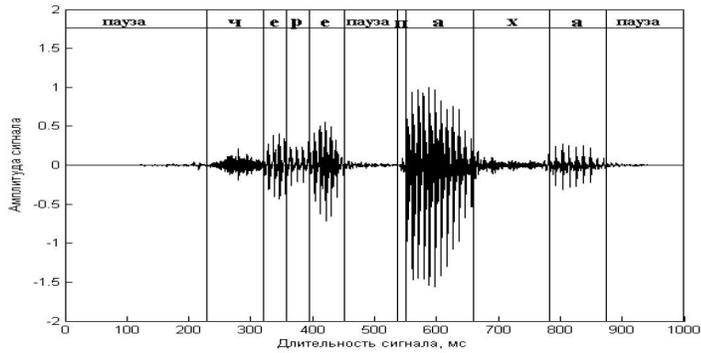


Рис. 1. Фрагмент РС, порожденного словом «черепаха» ($f_d = 16$ кГц).

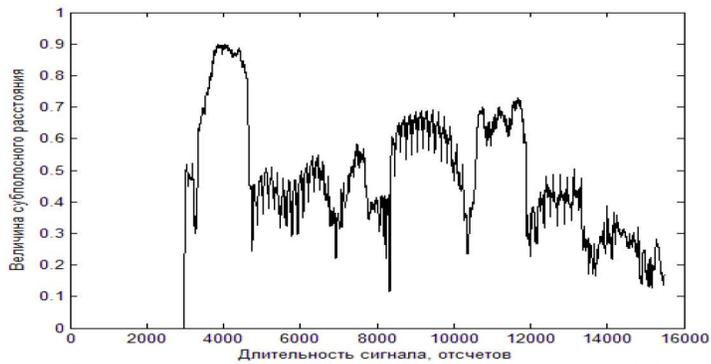


Рис. 2. Результат оценки субполосного расстояния $\gamma_u(x)$ фрагмента РС, порожденного словом «черепаха» ($f_d = 16$ кГц, $N = 256$).

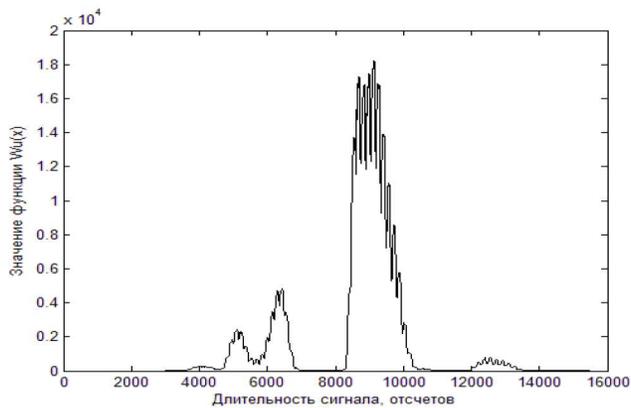


Рис. 3. Результат оценки функции $W_u(x)$ фрагмента РС, порожденного словом «черепаха» ($f_d = 16$ кГц, $N = 256$).

Функция $W_u(x)$ реагирует на изменение энергии по сравнению со средней, тогда как $\gamma_u(x)$ реагирует на изменение её распределения по частотным интервалам.

Гипотеза H_0 отвергается при выполнении неравенства

$$F_u > h_\alpha, \quad (9)$$

где $h_\alpha > 0$ – порог, удовлетворяющий условию

$$PR(F_u > h_\alpha / H_0) \leq \alpha. \quad (10)$$

Здесь PR – символ вероятности, а α – желаемый уровень вероятности ошибок первого рода.

Для оценки эффективности разработанного алгоритма были использованы оценки вероятностей ошибок первого и второго рода.

Оценка вероятности ошибки первого рода определена на основе анализа сигнала, соответствующего участку шума в паузах (185000 отсчетов). Величина вероятности определялась как:

$$P_{1ош} = N_{ошиб.речь} / N_{пауз}, \quad (11)$$

где $N_{ошиб.речь}$ – количество отрезков, ошибочно отнесенных к РС в присутствии звуков речи;

$N_{пауз}$ – количество отрезков РС, порожденных шумами, использованных для анализа (185 000 отрезков).

Для оценки вероятности ошибки второго рода был использован речевой материал с предварительно удаленными участками пауз (230000 отрезков). Вероятность ошибки определялась с использованием отношения вида:

$$P_{2ош} = N_{ошиб.пауз} / N_{речи} \quad (12)$$

ТАБЛ. 1. ЗНАЧЕНИЯ ВЕРОЯТНОСТЕЙ ОШИБОК ПЕРВОГО И ВТОРОГО РОДА ПРИ РАЗЛИЧНЫХ ПАРАМЕТРАХ ($f_d=16\text{кГц}$)

	N = 128		N = 256	
	$P_{1\text{ош}}$	$P_{2\text{ош}}$	$P_{1\text{ош}}$	$P_{2\text{ош}}$
без шума	0,0332	$< 10^{-4}$	0,0791	$< 10^{-4}$
$k = 0,1$	0,0374	0,0002	0,0861	$< 10^{-4}$
$k = 0,2$	0,0410	0,0015	0,0902	$< 10^{-4}$
$k = 0,3$	0,0462	0,0027	0,0963	0,0002
$k = 0,4$	0,0507	0,0092	0,1027	0,0006
$k = 0,5$	0,0537	0,0184	0,1081	0,0029
$k = 0,6$	0,0559	0,0298	0,1127	0,0057
$k = 0,7$	0,0573	0,0465	0,1164	0,0075
$k = 0,8$	0,0582	0,0621	0,1201	0,0109
$k = 0,9$	0,0584	0,0904	0,1235	0,0187
$k = 1$	0,0583	0,1161	0,1265	0,0256

где $N_{\text{ошиб.пауз}}$ – количество отрезков, ошибочно отнесенных к шуму в паузе; $N_{\text{речи}}$ – количество отрезков РС в присутствии звуков речи, использованных для анализа (230 000 отрезков).

В таблице 1 представлены результаты оценки вероятностей ошибок при различных значениях длительностей отрезков анализа и различных отношениях шум/сигнал k .

Полученные результаты показывают, что предлагаемый метод позволяет выделять участки пауз с малой вероятностью ошибочного принятия решений.

Работа выполнена при поддержке гранта РФФИ № 15-07-01463 "Разработка методов и алгоритмов автоматического распознавания устной речи с использованием субполосного анализа речевых сигналов".

Библиографический список

1. Жилияков Е.Г. Об одном алгоритме кодирования пауз в речевых данных / Е.Г. Жилияков, Е.И. Прохоренко, А.А. Фирсова, А.В. Болдышев // Вопросы радиоэлектроники, серия «Электронная вычислительная техника» (ЭВТ), вып. 1, 2013, с. 17-25.
2. Питмен, Э. Основы теории статистических выводов / Э. Питмен // Пер. с англ. М.: Мир, 1986. 104 с.