

УДК 621.391.037.372

Е.Г. Жилияков, С.П. Белов, А.А. Фирсова

firsova_A@bsu.edu.ru

ФГАОУ ВПО «Белгородский государственный национальный исследовательский университет»

СУБПОЛОСНАЯ РЕШАЮЩАЯ ФУНКЦИЯ СЕГМЕНТАЦИИ РЕЧЕВЫХ СИГНАЛОВ ПО ГРАНИЦАМ ЗВУКОВ РУССКОЙ РЕЧИ¹

Пусть $\vec{x} = (x_1, \dots, x_N)$ и $\vec{y} = (y_1, \dots, y_N)$ – непересекающиеся фрагменты речевого сигнала. Основная проверяемая гипотеза формулируется следующим образом: H_0 – фрагменты сигналов \vec{x} и \vec{y} соответствуют одному и тому же звуку русской речи.

Переход от отрезка речевого сигнала, порождаемого одним звуком русской речи, к отрезку, порождаемому другим звуком, сопровождается изменением характеристик этих сигналов, учет которых позволяет определять границы между звуками русской речи. Учитывая особенности воздействий, оказываемых различными звуками русской речи, по-видимому, единственной возможностью выявления этих изменений является использование в качестве признаков энергетических характеристик: энергия (квадрат евклидовой нормы) анализируемого вектора $\|\vec{x}\|^2$ и распределение её по частотным интервалам в виде наборов соответствующих долей:

$$Pd_n(\vec{x}) = P_n(\vec{x}) / 2 \|\vec{x}\|^2, n = 0, \dots, N/2, \quad (1)$$

где $P_n(\vec{x})$ – часть энергии, сосредоточенная в n -м частотном интервале:

$$P_n(\vec{x}) = \vec{x}^T A_n \vec{x}, \quad (2)$$

A_n – субполосная матрица с элементами вида [1]:

$$a_{ik}^0 = \sin(V_{20}(i-k) / \pi(i-k)), \quad (3)$$

$$a_{ik}^n = 2 * a_{ik}^0 \cos(n * 2\pi(i-k) / N), n = 1, \dots, N/2,$$

N – длительность окна анализа,

V_{20} – ширина первого частотного интервала.

При этом предполагается, что первый частотный интервал равен $2\pi/N$, а остальные $4\pi/N$, причем центры этих интервалов сдвинуты на $2\pi/N$ друг относительно друга.

Исходя из этого:

$$\sum_{n=0}^{N/2} Pd_n(\vec{x}) = 1. \quad (4)$$

Тогда в качестве одной из мер различий между векторами можно использовать аналог расстояния Питмена

$$\gamma(x, y) = \left(\sum_{n=0}^{N/2} ((Pd_n(\vec{x}))^{1/2} - (Pd_n(\vec{y}))^{1/2})^2 \right)^{1/2}, \quad (5)$$

которую естественно называть субполосным расстоянием.

С учетом свойства (4) соотношение (5) преобразуется к следующему виду

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

Для повышения чувствительности к изменени-

ям энергий отрезков предлагается использовать статистику:

$$W(x, y) = \max(\|\vec{x}\|^2, \|\vec{y}\|^2) / \min(\|\vec{x}\|^2, \|\vec{y}\|^2). \quad (7)$$

Данная характеристика всегда больше единицы, причем, чем больше ее значение, тем больше вероятность того, что сравниваемые фрагменты соответствуют различным звукам речи.

В качестве решающей функции предлагается использовать:

$$F(x, y) = W(x, y) * \gamma(x, y). \quad (8)$$

Ее естественно именовать субполосной РФ.

Гипотеза H_0 отвергается при выполнении неравенства:

$$F(x, y) > h_\alpha(K), \quad (9)$$

где α – вероятность ошибки первого рода, K – величина частотной концентрации при заданной доле энергии [2].

Если условие (9) выполняется, то в точке стыка сравниваемых фрагментов устанавливается граница, иначе принимается решение, что анализируемые отрезки принадлежат одному звуку речи. Использование представленного алгоритма при выборе длины окна анализа 16мс при частоте дискретизации 16кГц и уровне доли энергии 0,9 обеспечивает вероятности ошибок первого и второго рода 0,15÷0,20, что на 30% существующих алгоритмов [3].

Список литературы

1. Жилияков Е.Г. Вариационные методы анализа и построения функций по эмпирическим данным: моногр. / Е.Г. Жилияков. – Белгород: Изд-во, 2007. – БелГУ, 2007. – 160.
2. Фирсова А.А. О различии концентрации энергии по частотным диапазонам на отрезках сигналов, соответствующих шипящим звукам русской речи и шумам / Белов А.С., Курлов А.В., Фирсова А.А. / Научные ведомости Белгородского государственного университета №13(108) 2011 выпуск 19/1, серия Информатика, Белгород, 2011г. – стр. 186-191
3. Цыплихин А.И. Сегментация речи на кардинальные элементы / А.И. Цыплихин, В.Н. Сорокин // Информационные процессы, т.6, №3, с.177-207

¹ Работа выполнена в рамках гранта РФФИ № 12-07-00514-а