

## ОКОННЫЙ МЕТОД СЕГМЕНТАЦИИ РЕЧЕВОГО СИГНАЛА

А.В. Леонов

Решение задачи распознавания речи разделяется на несколько подзадач. Одна из таких подзадач – предварительная обработка сигнала, разбивающая его на сегменты для последующей их классификации. Известны различные варианты такого разбиения: сегменты фиксированного размера, скользящие окна фиксированного размера, скользящие окна плавающего размера. Основным недостатком окна фиксированного размера – это дополнительный объем вычислений, связанный с существенными индивидуальными частотными отличиями голосов различных дикторов. Развитие данного метода привело к реализации нескольких скользящих окон различного размера, учитывающих вероятный частотный диапазон для разных голосов. В этом случае для дальнейшей обработки сигнала в отдельном окне используются статистические методы, оконное преобразование Фурье, вейвлет преобразование [2]. При этом не удается получить однозначные оценки присутствия в отдельном окне конкретной частотной составляющей в силу неопределенности Гейзенберга для случая непрерывного речевого потока, но не для отдельно произнесенного звука. Для отдельного звука, время звучания которого укладывается в размер окна, вейвлет преобразование позволяет получить непрерывное разложение на некотором базисе ортогональных функций. Реализация окон изменяемого размера с адаптацией к особенностям одного голоса требует значительных вычислительных затрат на обучение и в дикторонезависимых системах неприменима. Для устранения перечисленных недостатков предлагается использовать скользящее окно достаточно большого размера с использованием в нем упрощенного преобразования Хаара [1] с последующей обработкой нейронной сетью.

Рассмотрим отсчеты речевого сигнала с частотой дискретизации 16 кГц. Учитывая частотный диапазон человеческого голоса, верхний предел в 8 кГц можно считать достаточным. Нижний предел частот для человека составляет около 30 Гц (по певческой классификации это нижняя граница для баса), который задается максимальным размером окна в 1024 дискретных отсчета. Такое значение выбрано для удобства вычисления преобразования Хаара. Коэффициенты преобразования вычисляются следующим образом:

$$S_i = (X_i + X_{i+1}) / 2, \quad (1)$$

$$C_i = (X_i - X_{i+1}) / 2, \quad (2)$$

где  $X_i$  -  $i$ -й отсчет сигнала,  $i = 0, 2, 4, 6 \dots 2n$ ,  $n$  – натуральное число.

Для сокращения вычислений можно отказаться от операции деления. Вычисление всех коэффициентов  $C_i$  для последовательности чисел, количеством равным целой степени 2 проводится для всех последующих последовательностей  $S$ .

Окно будем сдвигать на один отсчет в сторону увеличения времени. При этом мы будем получать повторы значений всех коэффициентов преобразования Хаара за исключением нулевых коэффициентов. Для дальнейшего анализа такая избыточность может оказаться лишней, и для передачи на обработку нейронной сети будем использовать только нормализованные значения нулевых коэффициентов. Для такой обработки требуется нейронная сеть с 11-ю входными нейронами для коэффициентов  $S_{9-0}, C_{0-0}, \dots, C_{9-0}$ , таким же количеством нейронов в скрытом слое и одним выходным нейроном. Значения весовых коэффициентов заданы случайным образом. Обучение нейронной сети не производится.

Основанием для отказа от первоначального обучения нейронной сети является ее биологический прототип.

Запись отдельных звуков речи русского языка (гласных и согласных) проводилась в лабораторных условиях на персональном компьютере без звукоизоляции и без подавления шума в исходном сигнале. Использовался микрофон NAIKO M22PC звуковая плата Creative PCI128. Графики исходных сигналов выборочных образцов звуков и выходной сигнал нейронной сети приведены на рисунках 1-4.

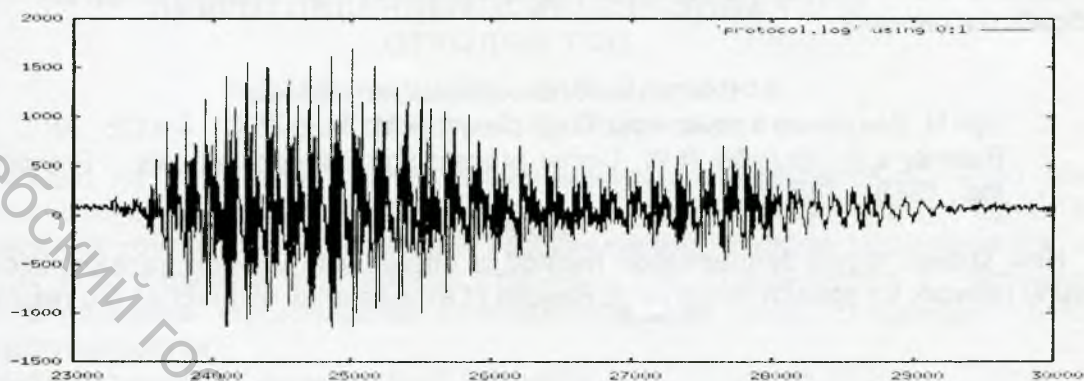


Рисунок 1 - Оцифрованный гласный звук «А»

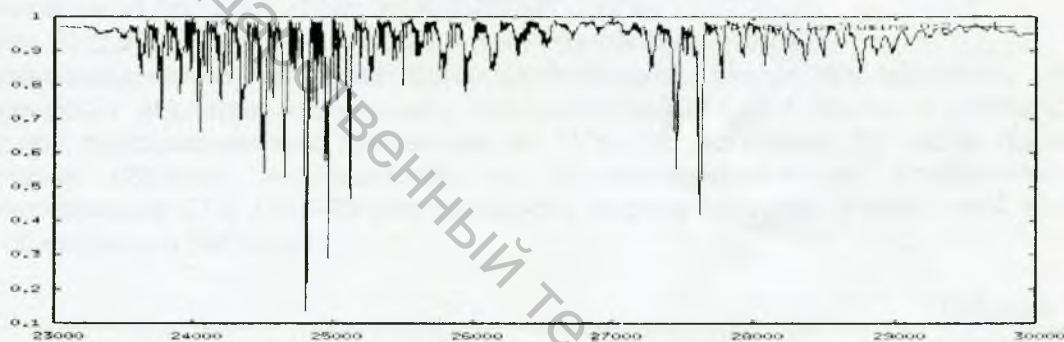


Рисунок 2 - Выходной сигнал нейронной сети для гласного звука «А»

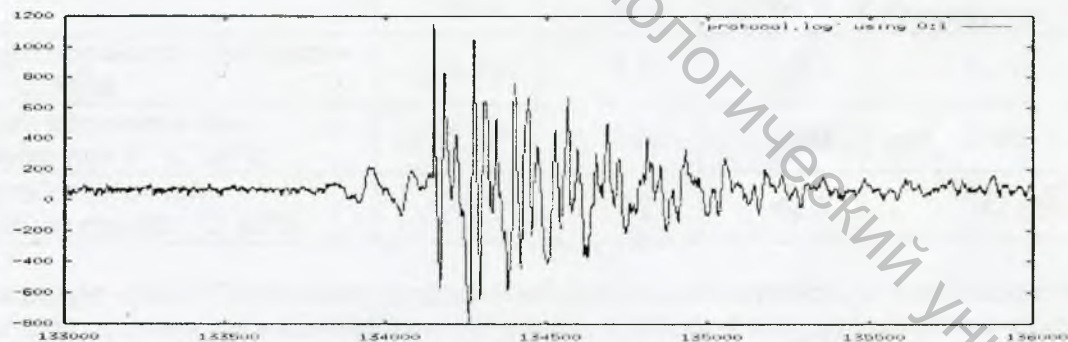


Рисунок 3 - Оцифрованный согласный звук «Б»

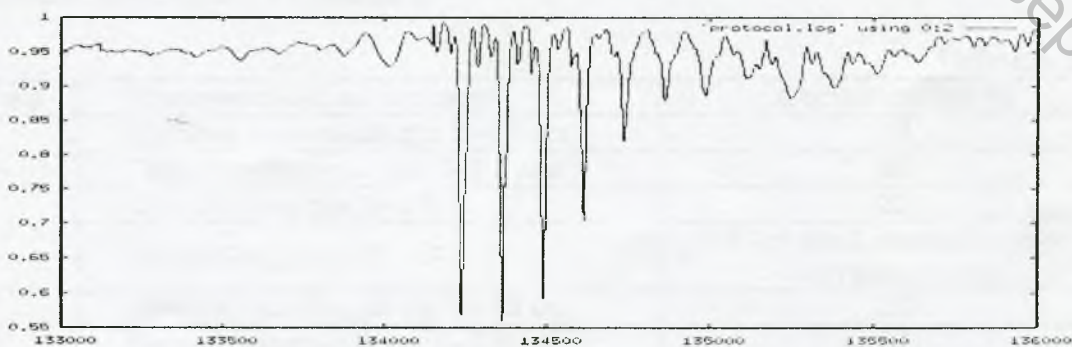


Рисунок 4 - Выходной сигнал нейронной сети для согласного звука «Б»

По выходному сигналу нейронной сети можно судить о присутствии наиболее значимого сочетания значений коэффициентов преобразования Хаара (в данном случае это минимумы) в конкретный момент времени, эти значения могут быть далее использованы для построения классификации конкретных звуков. По огибающей точек минимумов можно судить о присутствии речевого сигнала.

Полученные результаты позволяют сделать вывод о применимости предложенного метода в системах распознавания речи для предварительной обработки сигнала.

#### Список использованных источников

1. Чуи Ч. Введение в вэйвлеты: Пер. с англ. – М.: Мир, 2001. – 412с., ил.
2. Rabiner L.R., Schafer R.W. Digital processing of speech signals. - Prentice Hall Inc., 1978. - 265 p.

#### SUMMARY

New speech signal segmentation method is offered with wavelet transformation and neural network for speech recognition. Results of experimental researches are resulted.