

## ИССЛЕДОВАНИЕ РАСПОЛОЖЕНИЯ ФОНЕМ НА ОСНОВЕ ИСПОЛЬЗОВАНИЯ ЭНЕРГИИ СЕГМЕНТОВ ВЕЙВЛЕТ-СПЕКТРА

<sup>1</sup>Желтов П.В., <sup>1</sup>Семенов В.И., <sup>1</sup>Желтов В.П.

<sup>1</sup>ФГБОУ ВПО «ЧГУ им. И.Н. Ульянова», Чебоксары, Россия (428015, Россия, г. Чебоксары, Московский проспект, 15), e-mail: chnk@mail.ru

---

Разработана модель сегментации речи на основе анализа распределения локальных максимумов спектра в частотной области. Характеристики, вид, распределение локальных максимумов для речевых фрагментов и для фрагментов пауз существенно отличаются. Эти отличия проявляются в амплитудно-частотных соотношениях между локальными экстремумами (максимумами). Детальную картину расположения фонем в слове или предложении можно установить, исследуя зависимость энергии сегментов вейвлет-спектра от масштабного коэффициента  $a$ . Для исследования используется МНАТ-вейвлет. Объектом исследования является речевой сигнал на фоне слабого шума, длительность речевого сигнала составляет 4 с. Для вычисления вейвлет-спектра речевого сигнала используется формула непрерывного вейвлет-преобразования. Вейвлет-анализ речевого сигнала показывает, что гласные фонемы имеют максимальные энергии при средних значениях масштабного коэффициента. Такая закономерность наблюдается при многократном повторении и не зависит от случайных факторов.

Ключевые слова: вейвлет-преобразование, речь, энергия сегментов фонем, математическая модель, распознавание речи, энергия сегментов, двумерный объект

## RESEARCH LOCATION PHONEMES BASED ENERGY SEGMENT WAVELET SPECTRUM

<sup>1</sup>ZheltoV P.V., <sup>1</sup>Semenov V.I., <sup>1</sup>ZheltoV V.P.

<sup>1</sup>Chuvash State University, Cheboksary, Russia (428015, Cheboksary, Russia, Moskovskiy prospect 15), e-mail: chnk@mail.ru

---

A model for the segmentation of speech by analyzing the distribution of local maxima of the spectrum in the frequency domain is worked out. Specifications, appearance, distribution of local maxima for speech fragments and fragments pauses differ greatly. These differences manifest themselves in the amplitude-frequency relationships between local extremes (highs). A detailed picture of the location of phonemes in a word or sentence can be established by examining the dependence of the energy segments of the wavelet spectrum of the scale factor  $a$ . For research use MNAT-wavelet. The object of the study is the voice signal on a background of weak noise, the duration of the speech signal is four seconds. To calculate the wavelet spectrum of a speech signal uses the formula continuous wavelet transformation. Wavelet analysis of the speech signal indicates that the vowel phonemes have maximum energy at the average values of the scale factor. This pattern is observed when repeated many times, and does not depend on random factors.

Keywords: wavelet-transformation, speech, phonemes segments energy, mathematical models, speech recognition, segments energy, two dimensional object

Актуальной задачей является разработка модели сегментации речи на основе анализа распределения локальных максимумов спектра в частотной области. Основой данной модели является следующая физическая посылка.

Характеристики, вид, распределение локальных максимумов для речевых фрагментов и для фрагментов пауз существенно различаются. Эти отличия проявляются в амплитудно-частотных соотношениях между локальными экстремумами (максимумами).

Объектом исследования является речевой сигнал на фоне слабого шума, длительность речевого сигнала составляет 4 с. Для вычисления вейвлет-спектра речевого сигнала используется формула непрерывного вейвлет-преобразования.

$$W(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} S(t) \Psi\left(\frac{t-b}{a}\right) dt.$$

Для вычисления Фурье-спектра сегментов вейвлет-спектра используется преобразование Фурье:

$$F(\nu) = \int_{-\infty}^{\infty} f(t) e^{-i2\pi\nu t} dt.$$

Для вычисления энергии сегментов фонем используется формула Парсеваля:

$$\int_{-\infty}^{\infty} f^2(t) dt = \int_{-\infty}^{\infty} |F(\nu)|^2 d\nu.$$

Для исследования зависимости энергии сегментов вейвлет-спектра от масштабного коэффициента  $a$  вычисляется энергия сегментов функций  $W(a, b)$ . Математической моделью речевого сигнала в сегменте является:

$$d(n) = \frac{1}{M} \sum_{k=0}^{M-1} W(a, k) \cos\left(\frac{2\pi nk}{M}\right), \quad (1)$$

$$e(n) = \frac{1}{M} \sum_{k=0}^{M-1} W(a, k) \sin\left(\frac{2\pi nk}{M}\right). \quad (2)$$

Подставляя вместо вейвлет-спектра результат обратного Фурье-преобразования комплексно сопряженного спектра речевого сигнала и вейвлета, получаем:

$$d(n) = \frac{1}{M} \sum_{k=0}^{M-1} \left( \sum_{j=0}^{N-1} (c_1(j) + ic_2(j)) \exp\left(i \frac{2\pi kj}{N}\right) \right) \cos\left(\frac{2\pi nk}{M}\right),$$

$$e(n) = \frac{1}{M} \sum_{k=0}^{M-1} \left( \sum_{j=0}^{N-1} (c_1(j) + ic_2(j)) \exp\left(i \frac{2\pi kj}{N}\right) \right) \sin\left(\frac{2\pi nk}{M}\right).$$

По формуле:

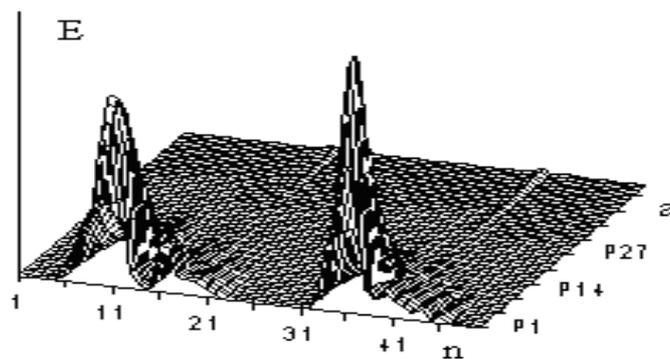
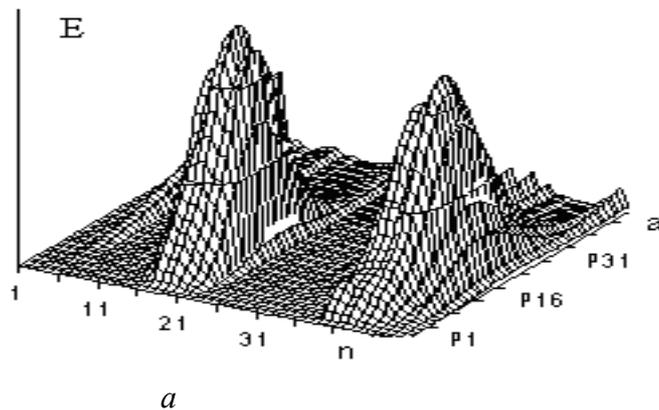
$$F(i) = d^2(i) + e^2(i) \quad (3)$$

вычисляется Фурье-спектр функций  $W(1, b)$ ,  $W(2, b)$ . Энергия сегментов для каждого масштабного коэффициента  $a$  вычисляется по формуле:

$$E = \sum_{i=1}^n F(i).$$

Таким образом, математической моделью речевого сигнала при выделении согласных фонем является энергия сегментов вейвлет-спектра речевого сигнала для разных масштабов. На рисунке 1 представлены графики зависимости энергии сегментов  $E$  от масштабного коэффициента  $a$  вейвлет-преобразования  $W(a, b)$  слова *часы*. На рисунке 1  $a$  масштабный коэффициент  $a$  меняется от 1 до 50 с шагом 1. При вычислении энергии просуммирована

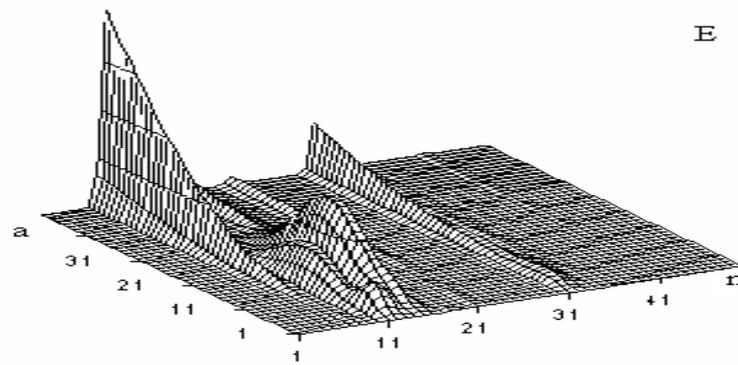
первая половина коэффициентов Фурье. На рисунке 1 б масштабный коэффициент  $a$  меняется от 0,4 до 2,9 с шагом 0,05. При вычислении энергии просуммирована вторая половина коэффициентов Фурье. На первом графике (рис. 1 а) видно, в каких сегментах выделяются гласные звуки а, ы. Фонемы ч, с не выделяются. На втором графике (рис. 1 б) фонемы ч и с (сегменты 4–13 и 32–38) хорошо видны и имеют энергию большую, чем гласные фонемы.



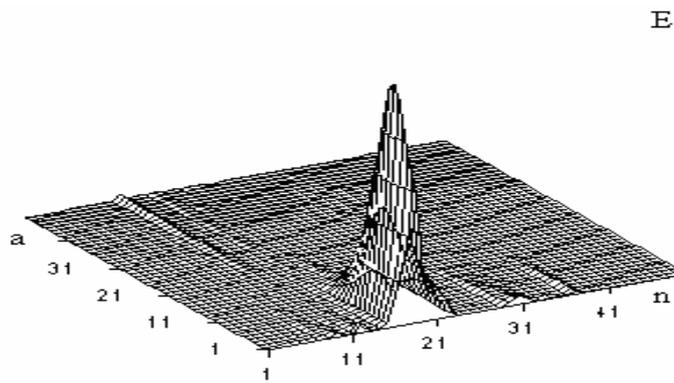
б

Рис. 1. Энергия сегментов ВП  $W(a,b)$  слова часы

На рисунке 2 представлены графики зависимости энергии сегментов  $E$  от масштабного коэффициента  $a$  ВП  $W(a,b)$  слова пуск. Параметры преобразования такие же, как для слова часы. Фонема  $n$  выделяется в сегментах 7–9. Фонема  $c$  выделяется (сегменты 15–21) так же, как выделяется фонема  $c$  в слове часы при небольших значениях масштабного коэффициента  $a$ . На графике хорошо видна пауза перед фонемой  $k$ .



*a*



*б*

Рис. 2. Энергия сегментов ВП  $W(a,b)$  слова пуск

Вейвлет-анализ речевого сигнала показывает, что гласные фонемы и фонемы н, м, л имеют максимальные энергии при средних значениях  $a$ . Энергия фонем н, м, л много меньше энергии гласных звуков речи, но значительно выше энергии шума. Фонемы к, т, п, д выделяются при больших значениях  $a$ . Перед фонемами к, т имеется пауза. Такая закономерность наблюдается при многократном повторении и не зависит от случайных факторов. Шипящие и свистящие фонемы при малых значениях масштабного коэффициента  $a$  имеют энергию  $W(a,b)$ , сравнимую с энергией гласных фонем. При средних значениях  $a$  они имеют энергию на уровне шума.

Можно заметить, что фонемы имеют отличающуюся друг от друга зависимость энергии  $W(a,b)$  от масштабного коэффициента  $a$  при использовании различного диапазона частот. На рисунках 3, 4 представлены графики зависимости энергии сегментов  $E$  от масштабного коэффициента  $a$  ВП  $W(a,b)$  слова стоп. На рисунке 3 масштабный коэффициент  $a$  меняется от 1 до 50 с шагом 1.

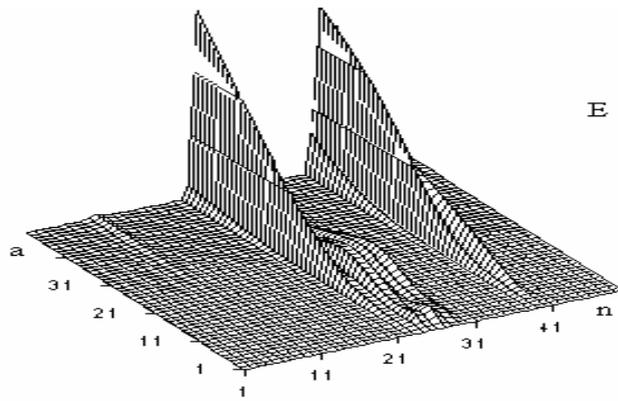


Рис. 3. Энергия сегментов ВП  $W(a,b)$  слова *стоп* для больших и средних значений масштабного коэффициента  $a$

На рисунке 4 масштабный коэффициент  $a$  меняется от 0,4 до 2,9 с шагом 0,05.

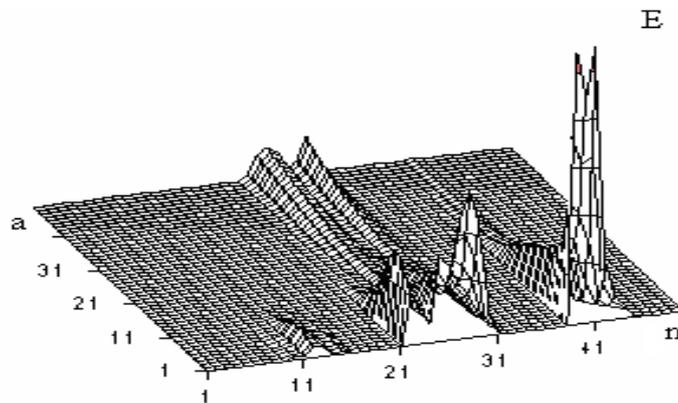


Рис. 4. Энергия сегментов ВП  $W(a,b)$  слова *стоп* для малых значений масштабного коэффициента  $a$

На рисунках 2, 3 и 4 местоположение фонем п, т, к при больших значениях масштабного коэффициента  $a$  для слов *пуск* и *стоп* различное, как и для гласных фонем при средних значениях масштабного коэффициента  $a$ .

Например, слова *выключить* и *отключить* имеют одинаковые функции  $W4(4,b)$ . Первое слово начинается с согласной, а второе – с гласной фонемы. Но отношение энергии сегментов ВП  $W(9,b)$  к энергии сегментов ВП  $W(2,b)$  первого слова больше, чем для второго в начале слова, так как для всех согласных фонем наблюдается такая закономерность. Для слова *выключить* пик в начале слова появляется, а для слова *отключить* – отсутствует.

На рисунке 5 представлена зависимость энергии сегментов ВП  $W(a,b)$  для масштабных коэффициентов  $a = 23$  и  $a = 47$ . На графике отчетливо выделяется фонема н при большом масштабном коэффициенте  $a = 47$ , а фонемы а, е имеют незначительную энергию. Также на графике видна пауза между слогами *лен* и *та*, сегменты которой имеют энергию,

незначительную для масштабных коэффициентов  $a = 23$  и  $a = 47$ . Энергия сегментов ВП  $W(a,b)$  меняется при переходе от фонемы л к фонеме е и от фонемы е к фонеме н.

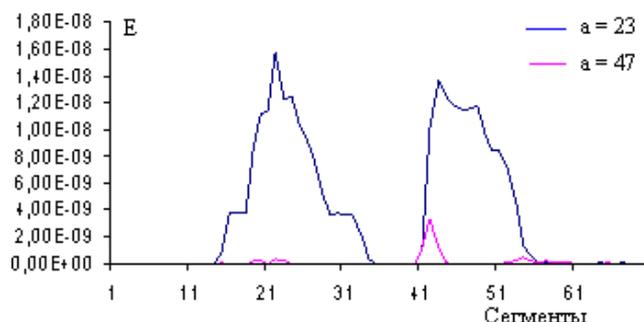


Рис. 5. Энергия сегментов слова лента для масштабных коэффициентов  $a = 23$  и  $a = 47$

Вейвлет-спектр  $W4(4,b)$  отчетливо выделяет границы между фонемами л, е, н и является основой для построения матрицы слова *лента*. В зависимости от количества максимумов этой функции прямоугольная поверхность (сегменты, масштабный множитель) разбивается на несколько областей. На рисунке б приведена проекция энергии сегментов ВП  $W(a,b)$  слова *умножить* на эту поверхность.

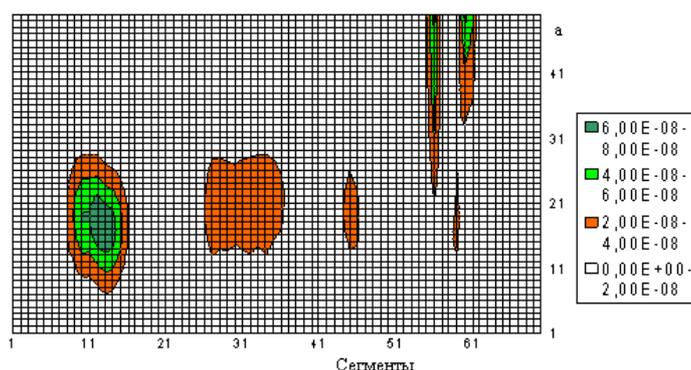


Рис. 6. Проекция энергии сегментов ВП  $W(a,b)$  слова *умножить* на прямоугольную поверхность (сегменты, масштабный множитель)

Для слова *умножить* функция  $W4(4,b)$  имеет 4 максимума, соответствующих фонемам у, о, и. Многомасштабный анализ, основанный на ВП, позволяет объединять слова в разные группы. В результате уменьшается время распознавания и увеличивается точность распознавания, так как базу данных слов можно разбить на подгруппы и представить в виде дерева поиска.

Таким образом, расположение фонем в слове или предложении можно установить, исследуя зависимость энергии сегментов вейвлет-спектра от масштабного коэффициента  $a$ . Для вычисления вейвлет-спектра речевого сигнала используется формула непрерывного вейвлет-преобразования. Вейвлет-анализ речевого сигнала показывает, что гласные фонемы имеют максимальные энергии при средних значениях масштабного коэффициента. Такая

закономерность наблюдается при многократном повторении и не зависит от случайных факторов.

*Работа выполнена при поддержке РФФИ, проект № 14-07-00143 а.*

### Список литературы

1. Вишнякова О.А., Лавров Д.Н. Алгоритм фонемной сегментации на основе анализа скорости изменения энергии дискретного вейвлет-преобразования // Вестник ОмГУ . 2011. № 4. С. 146–152.
2. Желтов П.В., Желтов В.П., Семенов В.И. Применение вейвлет-преобразования при сегментации речи // Современные проблемы науки и образования. – 2015. – № 1. URL: [www.science-education.ru/121-19205](http://www.science-education.ru/121-19205) (дата обращения: 03.06.2015)
3. Желтов П.В., Желтов В.П., Семенов В.И. Математическая модель распознавания слитной речи // Вестник ЧГУ. – 2012. – № 3. С. 210–212.
4. Желтов П.В., Семенов В.И., Трофимова А.И., Шурбин А.К. Алгоритмы идентификации фонем и формирования слова в системах распознавания речи на основе вейвлет-преобразования // Вестник ЧГУ. 2014. № 2. С. 98–102.
5. Семенов В.И. Методика определения границ между гласными и согласными звуками речи с применением быстрого непрерывного вейвлет-преобразования / В.И. Семенов, П.В. Желтов // Динамика научных исследований. 2011, Przemysl, 2011.

### Рецензенты:

Пряников В.С., д.т.н., профессор, профессор кафедры радиотехники и радиотехнических систем ФГБОУ ВПО «Чувашский государственный университет имени И.Н. Ульянова», г.Чебоксары;

Славутский Л.А., д.ф.-м.н., профессор кафедры автоматки и управления в технических системах ФГБОУ ВПО «ЧГУ им. И.Н. Ульянова», г. Чебоксары.