

Е.Г. ЖИЛЯКОВ, д.т.н., проф. НИУ "БелГУ", Белгород,
Е.И. ПРОХОРЕНКО, к.т.н., доц. НИУ "БелГУ", Белгород,
А.В. БОЛДЫШЕВ, аспирант НИУ "БелГУ", Белгород,
А.А. ФИРСОВА, аспирантка НИУ "БелГУ", Белгород,
М.В. ФАТОВА, магистр НИУ "БелГУ", Белгород

СЕГМЕНТАЦИЯ РЕЧЕВЫХ СИГНАЛОВ НА ОСНОВЕ АНАЛИЗА ОСОБЕННОСТЕЙ РАСПРЕДЕЛЕНИЯ ДОЛЕЙ ЭНЕРГИИ ПО ЧАСТОТНЫМ ИНТЕРВАЛАМ¹

В статье рассмотрены существующие подходы к сегментации речевых сигналов. Представлены результаты оценки особенностей распределения энергии речевых сигналов. Предлагается способ сегментации речи на основе учета особенностей распределения долей ее энергии по частотным интервалам. Ил.: 3. Табл.: 1. Библиогр.: 8 назв.

Ключевые слова: сегментация, речевые сигналы, распределение долей энергии.

Постановка проблемы и анализ литературы. Одной из проблем обработки речевых сигналов является сегментация сигнала на звуки. Точность алгоритмов сегментации определяет надежность и эффективность использования в дальнейшем таких алгоритмов, как распознавание речи, синтез, сжатие.

Сегментация речи – это процесс поиска границ между элементами речевого сообщения: фразами, словами, слогами, фонемами. Сегментация может осуществляться вручную или автоматически. Сегментация вручную является надежным, но трудоемким способом, особенно если это касается большого объема обрабатываемой информации. Также сегментация вручную невозможна при реализации обработки сигнала в режиме реального времени [1, 2, 3].

Наиболее интересной представляется автоматическая сегментация речевых сигналов. Эффективность алгоритма сегментации определяется точностью определения границы между различными звуками. Существует два подхода автоматической сегментации. Первый состоит в том, что при обработке речевого сигнала известна последовательность фонем, необходимо только определить границы между ними. Второй подход не использует априорную информацию о речевом сообщении, и сегментация осуществляется на основе изменения характера речевого сигнала. Можно выделить также третий подход, объединяющий два

¹ Исследования выполнены при поддержке гранта РФФИ № 0-07-00326-а.

перечисленных: использование априорной информации и анализ изменения характера сигнала [1, 4].

Все существующие алгоритмы сегментации речи основываются на статических или динамических характеристиках речи. Анализ статических характеристик не всегда приводит к точной сегментации. Оценка динамических характеристик сигнала позволяет увеличить точность сегментации [1, 4].

Одними из основных методов сегментации являются [1 – 4]:

- 1) сегментация по усредненному нормированному спектру;
- 2) сегментация по динамическим детекторам;
- 3) сегментация по корреляции между равноотстоящими спектрами;
- 4) сегментация с использованием дискретного вейвлет-преобразования.

Цель статьи – разработка нового метода сегментации речевых сигналов, основанного на учете особенностей распределения долей энергии по частотным интервалам для каждого звука речи [5].

Результаты исследований. Каждый звук имеет свое особенное распределение долей энергии по частотному диапазону. Звуки, соответствующие буквам русского алфавита, сосредоточены в узком частотном интервале, в то время как спектр шума распределен по всей области частот более равномерно. Эту особенность можно использовать для определения начала и конца слова или словосочетания.

На всей длительности любого звука можно выделить несколько участков, имеющих некоторые особенности. К таким участкам относятся начало звука, середина и конец. Это вызвано тем, что в слитной речи происходит переход одного звука в другой, для этого речевой аппарат человека некоторое время перестраивается. Для некоторых звуков, соответствующих таким буквам, как "е", "ё", "ю", "я", можно выделить большее число участков. Это связано с тем, что они состоят из двух звуков, плавно переходящих из одного в другой. Каждый из участков имеет свои особенности распределения энергии. Распределение энергии разных участков одного фрагмента звука отличается незначительно. Эти особенности могут быть использованы для сегментации речи.

Анализ распределения энергии отрезков сигналов по частотным интервалам предлагается проводить на основе точного метода [7]. В этом случае полный набор долей энергии отрезка сигнала определяется следующим образом:

$$P_r = \bar{x}^T A_r \bar{x}, \quad (1)$$

где $\vec{x}$ – анализируемый отрезок сигнала; r ($r = 1, \dots, R$) – номер частотного интервала; R – число частотных интервалов, на которые разбивается частотная ось; $A_r = \{a_{ik}^r\}$ – субполосная матрица, определяемая для каждого из R частотных интервалов, с элементами вида

$$a_{ik}^r = (\sin(v_{r+1}(i-k)) - \sin(v_r(i-k))) / (\pi(i-k)), \quad i, k = 1, \dots, N, \quad (2)$$

где v_r, v_{r+1} – границы r -го частотного интервала, причем:

$$0 \leq v_r < v_{r+1} \leq \pi, \quad v_{r+1} - v_r = \pi / R, \quad r = 1, \dots, R, \quad (3)$$

N – длительность анализируемого отрезка речевого сигнала.

Одной из характеристик, отражающей особенности звуков русской речи, является величина частотной концентрации, которая оценивается с использованием следующего выражения [8]:

$$W_{NR} = f_{NR}^m / R, \quad (4)$$

где f_{NR}^m – минимальное количество частотных интервалов (частотная концентрация), в которых сосредоточена заданная доля энергии m звукового отрезка, т.е.

$$f_{NR}^m = \min d_{NR}^m. \quad (5)$$

Здесь выполняется неравенство

$$\sum_{k=1}^{d_{NR}^m} P_{(k),N} \geq m \|\vec{x}_N\|^2 = m \sum_{i=1}^N x_i^2, \quad (6)$$

где $\vec{x}_N$ – анализируемый отрезок сигнала; m – заданное значение доли энергии сигнала; $P_{(k),N}$ – упорядоченные по убыванию доли энергий сигнала, попадающие в заданные частотные интервалы, т.е.

$$P_{(k),N} \in \{P_{rN}, r = 1, \dots, R\}, \quad P_{(k+1),N} \leq P_{(k),N}, \quad k = 1, \dots, R. \quad (7)$$

Для оценки возможности сегментации с использованием свойства частотной концентрации звуков русской речи было проведено большое количество экспериментов по оценке частотной концентрации различных фонем при различных значениях числа интервалов, на которые разбивается ось частот ($R = 4, 8, 16, 32, 64$), и значениях длины окна анализа ($N = 64, 128, 256$). В качестве исходного материала был использован фрагмент лекции, содержащий большое количество

различных фонем, записанный с частотой дискретизации $f_d = 8$ кГц с 16-битовым представлением в монорежиме.

Результаты экспериментов показали, что увеличение количества интервалов, на которые разбивается частотная ось, приводит к уточнению величины частотной концентрации отрезка сигнала.

В таблице представлены результаты оценки величины частотной концентрации для различных звуков русской речи.

Таблица. Распределение долей частотных интервалов, в которых сосредоточено 95% энергии при $N = 128$, $R = 32$ для различных звуков русской речи

гласные										
звук	а	е	ё	и	о	у	ы	э	ю	я
W_{NR}	0,31	0,13	0,09	0,09	0,19	0,09	0,16	0,31	0,13	0,09
сонорные согласные										
звук	й		л		м		н		р	
W_{NR}	0,19		0,19		0,19		0,16		0,34	
звонкие согласные										
звук	б		в		г		д		ж	
W_{NR}	0,22		0,28		0,19		0,16		0,25	
глухие согласные										
звук	к	п	с	т	ф	х	ц	ч	ш	щ
W_{NR}	0,22	0,16	0,25	0,28	0,16	0,25	0,19	0,44	0,47	0,34

Из таблицы видно, что для некоторых гласных и согласных звуков величины частотной концентрации совпадают. Особенно это проявляется для сонорных согласных.

На рис. 1 представлен фрагмент речевого сигнала, соответствующий звукосочетанию "шеч", выделенному из слова "шахматно-шашечный". Звук разбит на 16 равных окон анализа по 128 отсчетов.

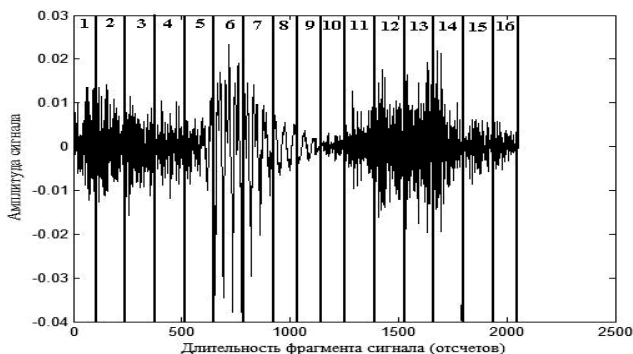


Рис. 1. Фрагмент речевого сигнала (слог "шеч" – безударный)

На рис. 2 представлен график распределения величины частотной концентрации для звуко сочетания "шеч" из слова "шахматно-шашечный", при длине окна анализа $N = 128$ и количестве частотных интервалов $R = 32$, для различных значений доли энергии.

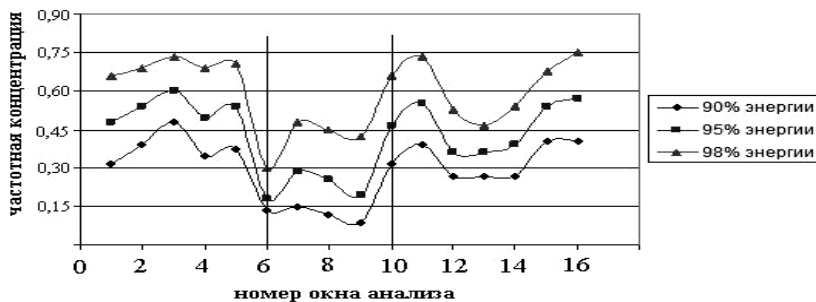


Рис. 2. Распределение величины частотной концентрации для звуко сочетания "шеч" ($N = 128$ и $R = 32$)

Анализ результатов экспериментов, представленных на рис. 2, показывает, что при выборе 95% энергии при переходе от 5 окна к 6-му, а также от 9-го к 10-му, наблюдается наибольшее изменение величины частотной концентрации. Эта особенность может быть использована для определения перехода между звуками. На рис. 1 видно, что окна 5 и 9 соответствуют переходу соответственно от звука "ш" к звуку "е", и от звука "е" к звуку "ч". Таким образом, увеличение разности частотной концентрации между соседними окнами может быть использовано для определения границы перехода между звуками.

На рис. 3 представлены результаты сегментации речевого сигнала с использованием различия величины частотной концентрации для различных звуков речи.

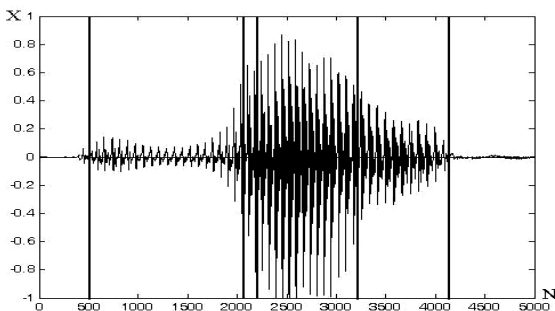


Рис. 3. Разбиение слова "заяц" на сегменты при $N = 128$, $R = 32$

Анализ рис. 3 показывает, что предлагаемый алгоритм позволил отделить звуки "з" и "а", "а" и "я", "я" и "ц". Проявилась дополнительная граница на фрагменте, соответствующем переходу между звуками "з" и "а".

Выводы. В результате проведенных экспериментов было выявлено, что длина фонем изменяется в пределах 1000 – 4000 отсчетов и зависит от типа звукосочетания: открытый слог, закрытый слог, ударный слог, безударный слог и т.д. Сравнение ударных и безударных слогов показало, что если гласный стоит под ударением, то длительность слога возрастает примерно в 1,25 раза. Спектры соответствующих звуков в ударном и безударном слогах отличаются незначительно. Использование представленного метода позволяет выявить место перестройки речевого аппарата с согласной на гласную и с гласной на согласную. Таким образом, данный метод может быть использован как один из элементов алгоритма сегментации речевого сигнала на отдельные звуки.

Список литературы: 1. *Сорокин В.Н.* Сегментация речи на кардинальные элементы / *В.Н. Сорокин, А.И. Цыплихин* // Информационные процессы. – 2006. – Т. 6. – № 3. – С. 177–207. 2. *Дремин И.М.* Вейвлеты и их использование / *И.М. Дремин, О.В. Иванов, В.А. Нечитайло* // Успехи физических наук. – 2001. – Т. 171. – №5. – С. 465–500. 3. *Ермоленко Т.Н.* Алгоритмы сегментации с применением быстрого вейвлет-преобразования / *Т.Н. Ермоленко, В.И. Шевчук* // Статьи, принятые к публикации на сайте международной конференции Диалог'2003. www.dialog-21.ru. 4. *Сорокин В.Н.* Сегментация и распознавание гласных / *В.Н. Сорокин, А.И. Цыплихин* // Информационные процессы. – 2004. – Т. 4. – № 2 – С. 202–220 5. *Жиляков Е.Г.* Методы обработки речевых данных в информационно-телекоммуникационных системах на основе частотных представлений: монография / *Е.Г. Жиляков, С.П. Белов, Е.И. Прохоренко*. – Белгород: Изд-во БелГУ, 2007. – 136 с. 6. *Шелухин О.И.* Цифровая обработка и передача речи / *О.И. Шелухин, Н.Ф. Лукьянцев*. Под ред. О.И. Шелухина. – М.: Радио и связь, 2000. – 456 с. 7. *Жиляков Е.Г.* Вариационные методы анализа и построения функций по эмпирическим данным: монография / *Е.Г. Жиляков*. – Белгород: Изд-во БелГУ, 2007. – 160 с. 8. *Фирсова А.А.* О различиях распределения энергии звуков русской речи и шума / *А.В. Болдышев, А.А. Фирсова* // Материалы 12-ой Международной конференции и выставки "Цифровая обработка сигналов и её применение. – "DSPA'2010". – Москва. – 2010. – С. 204–207.

УДК 621.391

Сегментація мовних сигналів на основі аналізу особливостей розподілу часток енергії за частотним інтервалам / *Є.Г. Жиляков, Є.І. Прохоренко, А.В. Болдышев, А.А. Фірсова, М.В. Фатова* // Вісник НТУ "ХПІ". Тематичний випуск: Інформатика і моделювання. – Харків: НТУ "ХПІ". – 2011. – № 17. – С. 44 – 50.

У статті розглянуті існуючі підходи до сегментації мовних сигналів. Представлені результати оцінки особливостей розподілу енергії мовних сигналів. Пропонується спосіб сегментації мови на основі врахування особливостей розподілу часток її енергії за частотними інтервалами. Іл.: 3. Табл.: 1. Бібліогр.: 8 назв.

Ключові слова: сегментація, мовні сигнали, розподіл часток енергії.

UDC 621.391

Segmentation of speech signals based on analysis of distribution of shares on energy frequency band/ Zhilyakov E.G., Prokhorenko E.I., Boldyshev A.V., Firsov A.A., Fatova M.V. // Herald of the National Technical University "KhPI". Subject issue: Information Science and Modelling. – Kharkov: NTU "KhPI". – 2011. – №. 17. – P. 44 – 50.

The paper considers existing approaches to segmentation of speech signals. Presents the results of evaluation of the features of the energy distribution of speech signals. Provides a method of speech segmentation on the basis of features of the distribution of shares of its energy on the frequency range. Figs.: 3. Tabl.: 1. Ref.: 8 titles.

Key words: segmentation, voice signals, the distribution of shares of energy.

Поступила в редакцию 01.02.2011