

Литература

1. Савченко Н.А. Аналитический обзор алгоритмов квантования как эффективного инструмента для внедрения скрытой информации. – В сб.: Инфоком-4 инфокоммуникационные технологии в науке, производстве и образовании: Тез. докл. IV междунар. НТК, Ставрополь, 2010, т. 2, с. 15-17.
2. Кошкина Н.В. Внедрение ЦВЗ в аудиосигналы на основе пакетной вейвлет-декомпозиции и частотного маскирования. – "Искусственный интеллект", 2010, вып. 4, с. 381-387.
3. Рублёв Д.П., Макаревич О.Б., Федоров В.М. Метод стеганографического встраивания сообщений в аудиоданные на основе вейвлет-преобразования. "Известия ЮФУ. Технические науки", 2010, с. 119-206.
4. Жилияков Е.Г., Белов С.П., Черноморец А.А. Вариационные методы синтеза сигналов на основе частотных представлений. – "Вопросы радиоэлектроники", сер. ЭВТ, 2010, вып.1.

Статья поступила 09.12.2011

Д.т.н., проф. Е.Г. Жилияков, к.т.н. Е.И. Прохоренко,
А.А. Фирсова, Д.К. Секисов (НИУ «БелГУ»)

E.G. Zhilyakov, E.I. Prokhorenko, A.A. Firsova, D.K. Sekisov

КОМПЬЮТЕРНАЯ СЕГМЕНТАЦИЯ РЕЧЕВЫХ СИГНАЛОВ

COMPUTER SEGMENTATION OF SPEECH SIGNALS

В статье рассмотрены решающие функции алгоритмов сегментации речевых сигналов, основанные на анализе распределения энергии по частотным интервалам. Проведен анализ возможности использования предложенных решающих функций для осуществления сегментации речевого сигнала.

The article deals with the crucial function of the segmentation algorithm of speech signals based on an analysis of the distribution of energy over the frequency intervals. The analysis of the possibility of using the proposed decision functions for the segmentation of the speech signal.

Ключевые слова: речевой сигнал, распределение энергии по частотным интервалам, сегментация речевых сигналов, решающие функции.

Key words: speech signal energy distribution over the frequency range segmentation of speech signals, the crucial function.

Сегментация речевого сигнала – это процесс определения границ участков, соответствующих одному звуку или сочетанию звуков речи.

В рамках статьи рассматриваются решающие функции, основанные на анализе распределения энергии по частотным интервалам. Для оценки энергии в заданном частотном интервале предлагается использовать выражение [2]:

$$P_{rN} = \bar{x}_N A_{rN} \bar{x}_N^T, \quad (1)$$

где A_{rN} – субполосная матрица:

$$A_{rN} = \{a_{ik}^r\}$$

Частотную ось предлагается разбивать на R частотных интервалов, причем граничные частотные интервалы должны быть вдвое меньше, чем остальные интервалы, и составлять π/R .

Пусть рассматриваются два отрезка сигнала. Тогда, если эти отрезки принадлежат одному и тому же звуку, то основная энергия этих отрезков сосредоточена в одних и тех же частотных интервалах. Для принятия решения о принадлежности сравниваемых отрезков одному звуку или разным предлагается использовать решающую функцию вида [3]:

$$S_2 = \begin{cases} \max \left\{ \frac{\sum_{r \in G_{12}} P_{r1}}{\sum_{r \in G_{12}} P_{r2}}, \frac{\sum_{r \in G_{12}} P_{r2}}{\sum_{r \in G_{12}} P_{r1}} \right\}, & \text{при } S_1 > 0 \\ \gg 1, & \text{при } S_1 = 0 \end{cases} \quad (2)$$

где

$$S_1 = 2 \frac{IntG_{12}}{IntR_{11} + IntR_{12}} \leq 1, \quad (3)$$

где $IntR_{11}, IntR_{12}$ – мощность множеств R_{11} и R_{12} соответственно, где:

$$\frac{\sum_{r \in R_{11}} P_{r1}}{\|\bar{x}_1\|^2} = \frac{\sum_{r \in R_{12}} P_{r2}}{\|\bar{x}_2\|^2} \approx m, \quad (4)$$

где: R_{11}, R_{12} – наименьшее количество частотных интервалов, в которых сосредоточена заданная доля энергии m , соответственно, для первого и второго отрезка анализа,

m – доля энергии, выбираемая порядка 0,9,

P_{r1}, P_{r2} – значение энергии в r -ом частотном интервале соответственно для первого и второго отрезка анализа,
 $IntG_{12}$ – мощность множества $G_{12}=R_{11} \cap R_{12}$ – пересечения множеств R_{11} и R_{12} .

Данная функция отражает во сколько раз отличается энергия одного отрезка сигнала от другого в тех частотных интервалах, которые попали в множество G_{12} . Чем больше значение решающей функции, тем больше вероятность того, что сравниваемые отрезки сигналов принадлежат различным звукам речи. Исследования показали, что значения решающей функции колеблются в диапазоне от 1 до 3.

В рамках данной работы использовались следующие параметры для анализа значений решающей функции: длина окна анализа $N=65$ отсчетов (при частоте дискретизации 8 кГц), количество интервалов, на которые разбивается ось частот, $R=16$, шаг смещения анализируемых отрезков $\Delta t=32$ отсчета относительно начала окна анализа и доля энергии $m=0,9$. Выбор такой длины окна анализа позволяет отразить периодическую структуру фрагмента сигнала. В то же время такая длина окна анализа достаточно мала и позволяет уменьшить вероятность захвата соседних звуков. Выбор такого шага смещения достаточно мал, что позволяет производить более точный анализ и, в то же время, достаточно велик, чтобы иметь возможность оценить изменение в распределении энергии между анализируемыми отрезками сигналов.

На рис. 1-и 2 представлен речевой сигнал, соответствующий слову «черепаха», и график изменения решающей функции S_2 для этого сигнала.

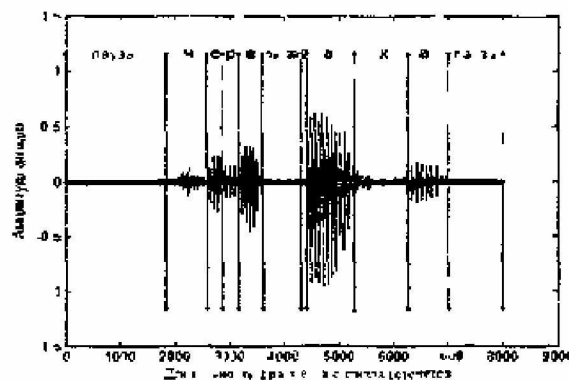


Рис 1

Фрагмент речевого сигнала, соответствующий слову «черепаха»

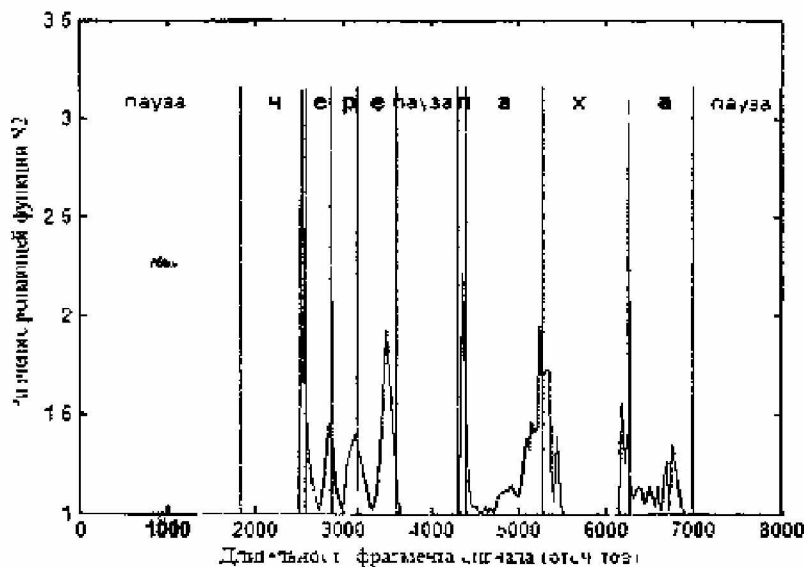


Рис 2

График изменения решающей функции S_2 речевого сигнала, соответствующего слову «черепаха» ($N=65$, $R=16$, $\Delta t=32$)

Анализ рисунков показывает, что значения решающей функции S_2 изменяются в диапазоне [1,3]. Выбор порогового значения 1,2 позволит обнаружить все границы. Важно также отметить, что данная решающая функция имеет локальные максимумы на всех границах между звуками. Наблюдаются также локальные максимумы на фрагментах, соответствующих одному звуку, но они расположены близко к границам звуков, что может быть использовано для устранения явления пересегментации.

На рис. 3-5 представлены результаты сегментации речевых сигналов, соответствующих различным словам и фразам с использованием решающей функции S_2 . Важно отметить, что при реализации алгоритмов на первом этапе осуществлялось выделение участков, относящихся к паузе [4]. Дальнейший анализ осуществлялся только на участках, соответствующих отрезкам звуков речи.

Анализ рисунков показывает, что при использовании решающей функции S_2 были установлены все границы между звуками. Установлены также четыре лишние границы, которые выделяют переходы от одного звука к другому. Погрешность между границами, установленными «на слух» и автоматически не превышает 125 отсчетов (15,6 мс).

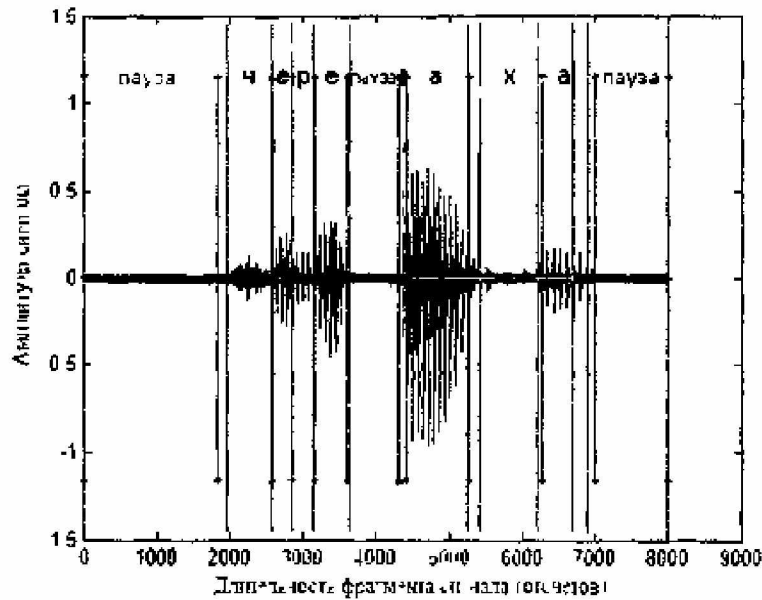


Рис 3

Результат сегментации речевого сигнала, соответствующего слову «черепаха» ($N=65$, $R=16$, $\Delta t=32$) (более короткие линии с маркерами – границы, установленные «на слух», более длинные линии – автоматически установленные границы)

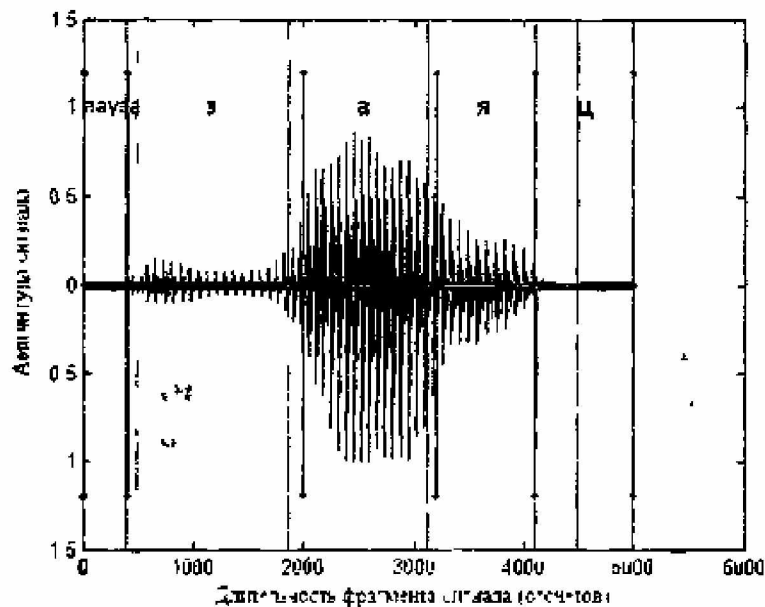


Рис 4

Результат сегментации речевого сигнала, соответствующего слову «заяц» ($N=65$, $R=16$, $\Delta t=32$) (более короткие линии с маркерами – границы, установленные «на слух», более длинные линии – автоматически установленные границы)

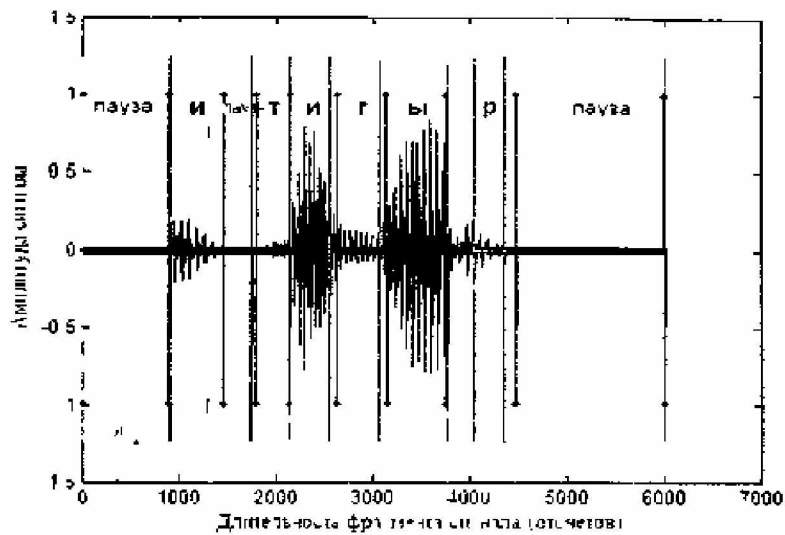


Рис 5

Результат сегментации речевого сигнала, соответствующего слову «и тигр» ($N=65$, $R=16$, $\Delta t=32$) (более короткие линии с маркерами – границы, установленные «на слух», более длинные линии – автоматически установленные границы)

Таким образом, в качестве решающей функции для обнаружения границ между звуками русской речи может быть использована решающая функция S_2 , основанная на учете изменения энергии в пересекающихся частотных интервалах, содержащих заданную долю энергии сигнала. Данная функция позволяет определять границы между звуками русской речи, при этом обеспечивается малый процент ложно установленных границ.

Литература

1. Сорокин В.Н., Цыплихин А.И. Сегментация речи на кардинальные элементы. – "Информационные процессы", 2006, т. 6, № 3, с. 177-207.
2. Жилияков Е.Г., Белов С.П., А.С., Фирсова А.А., Глушак А.В. Об эффективности различных подходов к сегментации речевых сигналов на основе обнаружения пауз. – "Научные ведомости Белгородского государственного университета", сер. Информатика, 2010, № 7(78), вып. 14/1, с. 187-193.
3. Жилияков Е.Г., Фирсова А.А., Секисов Д.К. Исследование решающих функций при сегментации речевых сигналов для различных звуков

русской речи. – В сб.: Труды Второй Междунар. НТК. Белгород, 2011, с. 579-584

4. Фирсова А.А., Чеканов А.Н. Компьютерное моделирование алгоритмов обнаружения пауз в IP-телефонии. - «Вопросы радиоэлектроники», сер. ЭВТ, 2011, вып. 1, с.72-82.

Статья поступила 09 12 2011

**К.т.н. И.И. Чижов, к.т.н. Т.Н. Балабанова, А.Р. Победа,
И.В. Деев (НИУ «БелГУ»)**

I.I. Chizhov, T.N. Balabanova, A.R. Pobeda, I.V. Deev

ОБ ОДНОМ АЛГОРИТМЕ СЖАТИЯ РЕЧЕВОГО СИГНАЛА

ABOUT THE ONE ALGORITHM OF SPEECH COMPRESSION

В статье предложен новый метод неэквидистантной дискретизации, позволяющий передать максимум энергии сигнала минимумом его значений. Приведены результаты вычислительных экспериментов, показывающие сохранение частотных компонент сигнала выше частоты Найквиста при применении данной передискретизации.

In the article the new method of nonequidistant sampling is given. The main idea of the method is to save the maximum part of signal energy by minimum number of digits. The results of experiments which proves the theoretical ideas are also given.

Ключевые слова: речевой сигнал, сжатие данных, прореживание, интерполяция, дискретизация

Key words: speech signal, data compression, decimation, interpolation, sampling

В настоящее время, объем аудиоданных в информационно-телекоммуникационных системах ежегодно значительно увеличивается. В первую очередь это обусловлено тем, что звук является наиболее естественной для человека формой информационного обмена, поэтому с ростом производительности информационно-телекоммуникационных систем тенденция к переходу на подобные формы обмена информацией будет только усиливаться.

Таким образом, для повышения эффективности функционирования информационно-телекоммуникационных систем необходимо существенно уменьшить объем речевых данных, передаваемых по каналам связи и хранимых на локальных накопителях, для чего