



О РАЗЛИЧИИ КОНЦЕНТРАЦИИ ЭНЕРГИИ ПО ЧАСТОТНЫМ ДИАПАЗОНАМ НА ОТРЕЗКАХ СИГНАЛОВ, СООТВЕТСТВУЮЩИХ ШИПЯЩИМ ЗВУКАМ РУССКОЙ РЕЧИ И ШУМАМ¹

**А.С. БЕЛОВ
А.В. КУРЛОВ
А.А. ФИРСОВА**

*Белгородский государственный
национальный исследовательский
университет*

e-mail: belov_as@bsu.edu.ru

В статье рассматривается различие концентрации энергии по частотным диапазонам на отрезках сигналов, соответствующих шипящим звукам русской речи и шумам. Полученные знания можно использовать в алгоритмах обнаружения пауз в звуковых сигналах.

Ключевые слова: речевой сигнал, анализ речевого сигнала, частотные представления, распределение энергии по частотным интервалам.

Введение.

Представление речевой информации в частотной области обладает рядом преимуществ. Во-первых, акустический анализ механизма речеобразования показывает, что распределение нулей и полюсов коэффициента передачи дает достаточно четкое описание звуков речи. Во-вторых, совершенно очевидно, что в начальной стадии процесса восприятия ухо производит некоторый грубый частотный анализ. Таким образом, характерные особенности, которые проявляются в результате анализа на основе частотных представлений, играют важную роль в процессах восприятия и воспроизведения.

Для непериодических сигналов конечной длительности $f(t)$, какими и являются речевые сигналы, используется форма разложения, при которой дискретность или шаг вычисления спектра стремится по величине к нулю и дискретный ряд Фурье переходит в интеграл Фурье или преобразование Фурье:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\nu) e^{j\nu t} d\nu \quad (1)$$

Это выражение представляет непериодическую функцию $f(t)$, как бесконечную сумму экспоненциальных функций $\exp(\nu t)$ с частотами на интервале $(-\infty < \nu < \infty)$ и весами, определяемыми для каждой частоты величиной $S(\nu)$.

При анализе речевых сигналов очень важным является иметь информацию о распределении их энергии. Для этого используется равенство Парсеваля, которое в математическом виде может быть записано следующим образом:

$$\|\bar{x}\|^2 = \sum_{k=1}^N x_k^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |X(\nu)|^2 d\nu \quad (2)$$

так что

$$P_r(\bar{x}) = \frac{1}{2\pi} \int_{\nu \in V_r} |X(\nu)|^2 d\nu \quad (3)$$

представляет собой долю энергии отрезка сигнала (евклидовой нормы вектора), соответствующую частотному интервалу.

¹ Исследования выполнены при поддержке ФЦП «Научные и научно-педагогические кадры для инновационной России» на 2009-2013 годы, гос.контракт № 16.740.11.0312.



Сосредоточенность энергии сигнала в узком частотном интервале позволяет человеку, который воспринимает звук, соотнести его с каким-либо объектом, такие типы сигналов будем называть информационными. В противном случае, если энергия сигнала распределена равномерно по частотному диапазону, сигнал неинформационный.

В качестве показателя сосредоточенности энергии предлагается использовать отношение

$$S = \frac{\sum_{k=1}^{r^* \alpha} P_{sort\ k}}{\sum_1^r P_r} > h_\alpha, \quad r = 1, \dots, R. \quad (4)$$

Здесь R – общее количество частотных интервалов;

P_r – значение доли энергии, сосредоточенной в r -ом частотном интервале;

$P_{sort\ k}$ – сумма долей энергии частотных интервалов, имеющая максимальное значение (значения энергии в частотных интервалах заранее отсортировываются по убыванию);

h_α – порог, определяемый на основе анализа эмпирических данных.

Математические основы метода.

В качестве основы для моделей анализа служит понятие частей энергий отрезков речевых сигналов, попадающих в соответствующий частотный интервал, определяемых интегралами вида:

$$P_r(\vec{x}) = \frac{1}{2\pi} \int_{v \in I^r} |X(v)|^2 dv. \quad (5)$$

В качестве инструмента проводимых исследований в статье использовалось полученное в [1] представление

$$P_r(\vec{x}) = \vec{x}^T A_r \vec{x}, \quad (6)$$

которое позволяет вычислять интегралы вида (5) непосредственно во временной области (не вычисляя трансформанты Фурье анализируемого отрезка), где

$$A_r = \{a_{ik}^r\}, \quad a_{ik}^r = \begin{cases} \frac{\sin[\nu_2^r(i-k)] - \sin[\nu_1^r(i-k)]}{\pi(i-k)}, & i \neq k \\ \frac{\nu_2^r - \nu_1^r}{\pi}, & i = k \end{cases}$$

Эти матрицы естественно называть субполосными.

Предлагаемый подход позволит оптимизировать процедуру принятия решений о наличии в анализируемом отрезке информационного сигнала или неинформационной помехи, в том смысле, что при заданной вероятности сосредоточенности энергии в определённом количестве частотных интервалов максимизируется вероятность правильного обнаружения информационных сигналов за счет адекватного учета сосредоточенности их энергий в нескольких узких частотных интервалах.

Результаты вычислительных экспериментов

Экспериментальные исследования проводятся с целью выявления закономерностей сосредоточенности энергий по частотным интервалам в сигналах принадлежащих шипящим звукам речи (информационным сигналам) и шумам (неинформационным помехам).



В основе вычислительных экспериментов по обработке отрезков речевых сигналов лежит разработанный алгоритм разделения информационных сигналов и неинформационных помех на основе частотных представлений [2,3].

В качестве эмпирических данных были использованы отрезки звуковых файлов, записанных на диктофон в стандартных останках (в аудитории на лекции, в лифте, на улице) с параметрами: частота дискретизации 8000 Гц, количество разрядов квантования 16.

Длина анализируемого отрезка выбрана равной $N=60$.

Область определения трансформант Фурье дискретных сигналов $[0, \pi]$ разбивается на R одинаковых частотных интервалов, таких что $\nu_{2r} - \nu_{1r} = \Delta\nu = \text{const}$, причем такие R , что M является целым числом ($N=MR$).

В ходе экспериментов для всех значений N используется различное разбиение оси частот на R частотных интервалов, а именно:

при $N=60$ $R=10, 15$.

Длина анализируемых сигналов варьирует от 200 до 1000 отсчетов.

На рис. 1-8 представлены фрагменты сигналов и распределения энергий по частотным интервалам некоторых отрезков этих сигналов.

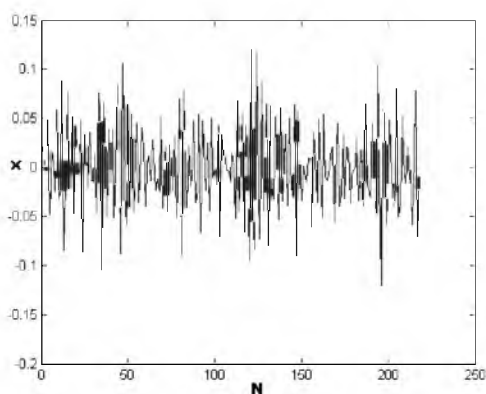


Рис. 1. Фрагмент отрезка речевого сигнала, соответствующего звуку «ш»

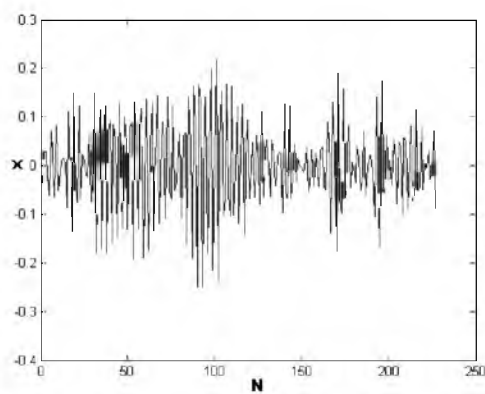


Рис. 2. Фрагмент отрезка речевого сигнала, соответствующего звуку «щ»

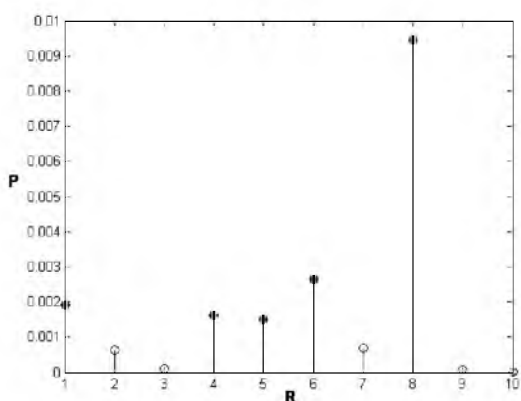


Рис. 3. Распределение значений долей энергии звука «ш» в выбранных частотных интервалах (черным маркером отмечены частотные интервалы, сумма долей энергий которых составляет значение $m=0.92$)

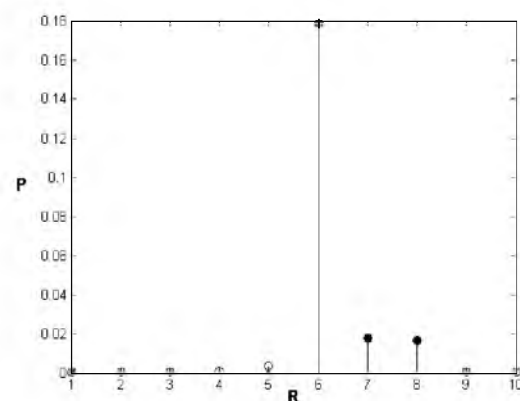


Рис. 4. Распределение значений долей энергии звука «щ» в выбранных частотных интервалах (черным маркером отмечены частотные интервалы, сумма долей энергий которых составляет значение $m=0.92$)

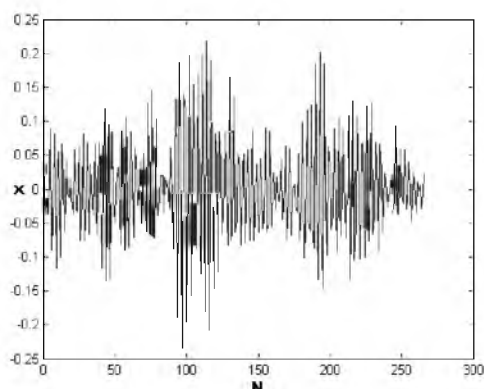


Рис. 5. Фрагмент отрезка речевого сигнала, соответствующего звуку «ч»

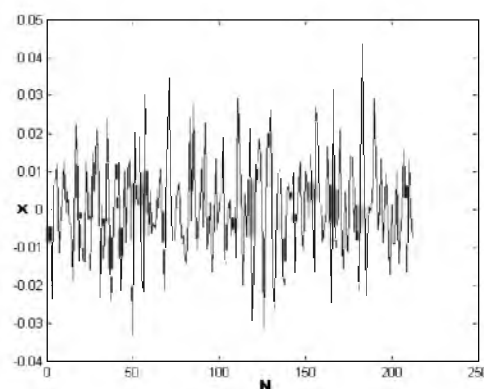


Рис. 6. Фрагмент отрезка речевого сигнала, соответствующего звуку «ц»

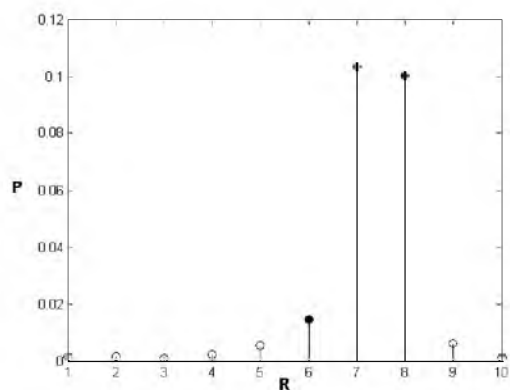


Рис. 7. Распределение значений долей энергии звука «ч» в выбранных частотных интервалах (черным маркером отмечены частотные интервалы, сумма долей энергий которых составляет значение $m=0.92$)

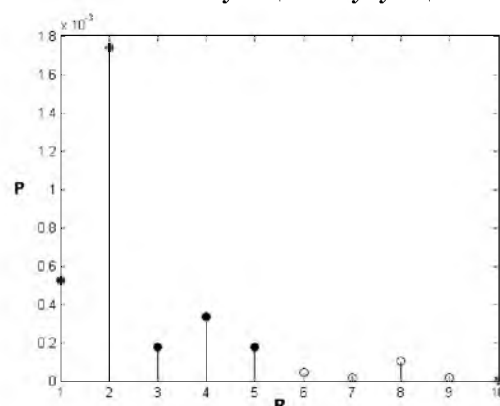


Рис. 8. Распределение значений долей энергии звука «ц» в выбранных частотных интервалах (черным маркером отмечены частотные интервалы, сумма долей энергий которых составляет значение $m=0.92$)

Анализ результатов экспериментов показывает, что основная доля энергии шипящих звуков русской речи сосредоточена не больше, чем в 50% частотных интервалах.

В табл. 1 представлены результаты оценки доли частотных интервалов, в которых сосредоточена заданная доля энергии анализируемого отрезка речевого сигнала.

Таблица 1

Сосредоточенность доли энергии в заданном количестве частотных интервалов

звук	N= 60, R=10				N= 60, R=15			
	80%	85%	90%	95%	80%	85%	90%	95%
ш1	0,3	0,3	0,4	0,5	0,2	0,3	0,3	0,5
ш2	0,3	0,4	0,5	0,7	0,3	0,4	0,5	0,7
ш3	0,3	0,4	0,4	0,6	0,3	0,3	0,4	0,6
ш4	0,5	0,6	0,6	0,7	0,3	0,4	0,5	0,7
ш5	0,5	0,6	0,6	0,6	0,5	0,5	0,6	0,7
щ1	0,2	0,2	0,3	0,4	0,1	0,2	0,3	0,3
щ2	0,1	0,1	0,3	0,4	0,1	0,1	0,2	0,3
щ3	0,2	0,2	0,3	0,4	0,2	0,2	0,3	0,4



Продолжение табл. 1

ш4	0,2	0,2	0,3	0,4	0,1	0,2	0,3	0,3
ш5	0,2	0,2	0,3	0,4	0,2	0,3	0,3	0,3
ч1	0,2	0,3	0,3	0,4	0,2	0,3	0,3	0,4
ч2	0,1	0,2	0,2	0,3	0,1	0,1	0,2	0,3
ч3	0,2	0,3	0,3	0,4	0,2	0,3	0,3	0,4
ч4	0,1	0,2	0,3	0,3	0,1	0,1	0,2	0,3
ч5	0,3	0,4	0,4	0,5	0,2	0,3	0,4	0,4
ц1	0,5	0,6	0,7	0,9	0,5	0,5	0,7	0,8
ц2	0,6	0,7	0,8	0,9	0,6	0,7	0,7	0,9
ц3	0,5	0,6	0,7	0,8	0,5	0,6	0,7	0,8
ц4	0,3	0,4	0,5	0,7	0,4	0,5	0,6	0,7
ц5	0,2	0,5	0,6	0,7	0,5	0,6	0,7	0,7
шум1	0,3	0,4	0,5	0,7	0,3	0,4	0,5	0,6
шум3	0,5	0,6	0,6	0,7	0,5	0,5	0,6	0,7
шум4	0,3	0,4	0,5	0,6	0,3	0,4	0,5	0,6
шум5	0,4	0,5	0,6	0,7	0,4	0,5	0,5	0,6

Заключение.

В ходе проведённых экспериментов было установлено, что шипящие звуки русской речи обладают большей концентрацией, чем участки сигнала, соответствующего шуму. Это позволяет сделать вывод о возможности использования характеристики концентрации энергии по частотным диапазонам для различения информационных и не информационных компонентов звуковых сигналов.

Список литературы

1. Жилияков, Е.Г. Вариационные методы частотного анализа звуковых сигналов [Текст] / Е.Г. Жилияков, С.П.Белов, Е.И. Прохоренко // Труды учебных заведений связи. – СПб, 2006. вып. 174. – с. 163-172
2. Жилияков, Е.Г. Уменьшение объема битового представления речевых данных на основе нового метода удаления пауз [Текст] / Е.Г. Жилияков, С.П.Белов, Е.И. Прохоренко // Вопросы радиоэлектроники. Серия «Электронная вычислительная техника (ЭВТ)». – М., 2007. – Вып. 2. – с. 82-92
3. Белов С. П., Белов А. С., Жилияков Е.Г., Прохоренко Е.И. Способ обнаружения пауз в речевых сигналах и устройство его реализующее.

ON THE DIFFERENCE BETWEEN CONCENTRATION OF ENERGY BANDWIDTHS ON THE SEGMENT SIGNALS CORRESPONDING TO A HISSING SOUND AND NOISE RUSSIAN SPEAKING

A.S. BELOV
A.V. KURLOV
A.A. FIRSOVA

*Belgorod National
Research University*

e-mail: Belov_AS@bsu.edu.ru

This article discusses the difference of concentration of energy in frequency bands on the segments of signals corresponding to the hissing sound of Russian speech and noise, the knowledge can be used in the algorithms for the detection of pauses in the audio signals.

Key words: speech signal, analysis of the speech signal, frequency representation, the distribution of energy over the frequency bands.