



ИССЛЕДОВАНИЕ АЛГОРИТМА СИНТЕЗА РЕЧЕВЫХ СИГНАЛОВ В ЦСА НА ОСНОВЕ ИЗБИРАТЕЛЬНОГО ВОЗДЕЙСТВИЯ НА ЧАСТОТНЫЕ ИНТЕРВАЛЫ

**А.С. БЕЛОВ
А.В. КУРЛОВ
А.А. ФИРSOVA**

*Белгородский государственный
национальный
исследовательский университет*

*e-mail:
Belov_AS@bsu.edu.ru
Kurlov@bsu.edu.ru
Firsova_A@bsu.edu.ru*

В статье рассматривается алгоритм синтеза речевых сигналов. Представлено описание алгоритма обнаружения пауз. Показаны результаты оценки информационных и неинформационных частотных интервалов и результаты оценки качества звучания результатов синтеза речи.

Ключевые слова: речевой сигнал, анализ речевого сигнала, частотные представления, информационные частотные интервалы, неинформационные частотные интервалы.

В настоящее время существует развитая индустрия по производству слуховых аппаратов (СА), предназначенных для протезирования людей с ослабленным слухом, которых по данным Всемирной организации здравоохранения в мире насчитывается свыше 10%. В большинстве случаев потери слуха проявляются в виде снижения или даже полной потери подвижности (эластичности) отдельных элементов механической системы человеческого уха, что приводит к ослаблению реакции на воздействие внешней акустической волны.

Для компенсации указанных потерь реакции на акустическое воздействие необходимо осуществлять довольно сложную обработку звуковых сигналов, что доступно только при использовании технических устройств на базе цифровых процессоров. Так как потери слуха индивидуальны для каждого пользователя, то предусматривается процедура адаптации к ним за счёт соответствующих настроек режимов функционирования СА на этапе их подбора (как правило, врачом – сурдологом).

В современных цифровых СА используется от трёх до 16 частотных диапазонов (каналов обработки). Применяемые при этом способы разделения на частотные компоненты (аппроксимации отрезков трансформант Фурье) и соответствующие процедуры синтеза являются ноу-хау и детально неизвестны.

Вместе с тем, несмотря на постоянное совершенствование СА, отзывы пользователей свидетельствуют о том, что в настоящее время отсутствуют эффективные алгоритмы разделения анализируемых отрезков сигналов на информационные частотные компоненты, то есть содержащие важную акустическую информацию, в том числе речевую, и неинформационные, обусловленные неизбежным присутствием, так называемых, помех окружающей среды и аппаратурных шумов. Ясно, что в последнем случае синтез выходных сигналов на основе аудиограммы за счёт усиления, в том числе и компонент, обусловленных помехами, приводит к некомфортности пользования СА.

Исследования особенностей распределения энергии по частотным интервалам сигналов, соответствующих звукам русской речи и шумам в паузах, показало, что звуки русской речи сосредоточены в достаточно узком частотном интервале, а энергия шумов распределена достаточно равномерно по всему частотному диапазону. Несмотря на различие длительностей различных звуков речи, оказалось возможным при различных сочетаниях длительности интервала анализа и количества частотных интервалов оценить характеристику сосредоточенности энергии вида:

$$W_{NR}^t = f_{NR}^{tm} / R, \quad (1)$$

где R – количество интервалов, на которые разбивается ось частот,

f_{NR}^{tm} – минимальное количество частотных интервалов (частотная концентрация), в которых сосредоточена заданная доля энергии звукового отрезка, так что имеет место:

$$f_{NR}^{tm} = \min d_{NR}^{tm}. \quad (2)$$

Здесь для правых частей выполняется неравенство



$$\sum_{k=1}^{d_{NR}^m} P_{(k)N} \geq m \|\vec{x}_N\|^2 = m \sum_{i=1}^N x_i^2, \tag{3}$$

где m – задаваемая доля общей энергии, которая должна быть сосредоточена в указанном минимальном количестве частотных интервалов,

\vec{x}_N – отрезок сигнала, длительностью N отсчетов,

$P_{(k)N}$ – значения энергий в заданных интервалах, после упорядочивания их по убыванию.

Верхний индекс t обозначает один из звуков русской речи, причём индекс в скобках у слагаемых суммы слева соотношения (3) означает, что части энергий P_{kN} упорядочиваются по убыванию, то есть имеет место

$$P_{(k)N} \in \{P_{rN}, r = 1, \dots, R\}; P_{(k+1)N} \leq P_{(k)N}, k = 1, \dots, R, \tag{4}$$

Проведенные исследования показали возможность использования данного метода для определения границ между звуками речи и паузами. Все исследования проводились на речевых сигналах, записанных с частотой дискретизации 8кГц и разрядностью кода 16 бит. В рамках данной работы был предложен алгоритм, в котором решение о принадлежности отрезка участку речи или паузы принимается на основе сравнения величины частотной концентрации вида (1) с пороговым значением. Если полученное значение величины частотной концентрации превышает заданный порог, принимается решение о том, что текущий фрагмент относится к шумам в паузе, в противном случае – к участку речи. Эксперименты показали, что наиболее эффективным при реализации алгоритма обнаружения пауз является использование следующих параметров: длительность окна анализа $N=256$, количество интервалов, на которые разбивается ось частот, $R=16$, доля энергии, при которой определяется величина частотной концентрации, $m=0,87$, пороговое значение величины частотной концентрации $h_{NR}=0,6$ ($P_1=0,08, P_2=0,00$).

На предыдущих этапах также было представлен алгоритм определения информационных и неинформационных частотных компонент отрезков сигналов, соответствующих участкам речи. К информационным частотным интервалам относятся те, которые попадают в минимальный набор интервалов, составляющих заданную долю энергии. На рис. 1-5 представлены фрагменты сигналов, соответствующих звукам русской речи и распределение их энергии по оси частот с указанием информационных и неинформационных частотных интервалов. На рисунках квадратным маркером отмечены частотные интервалы, которые были приняты как информационные.

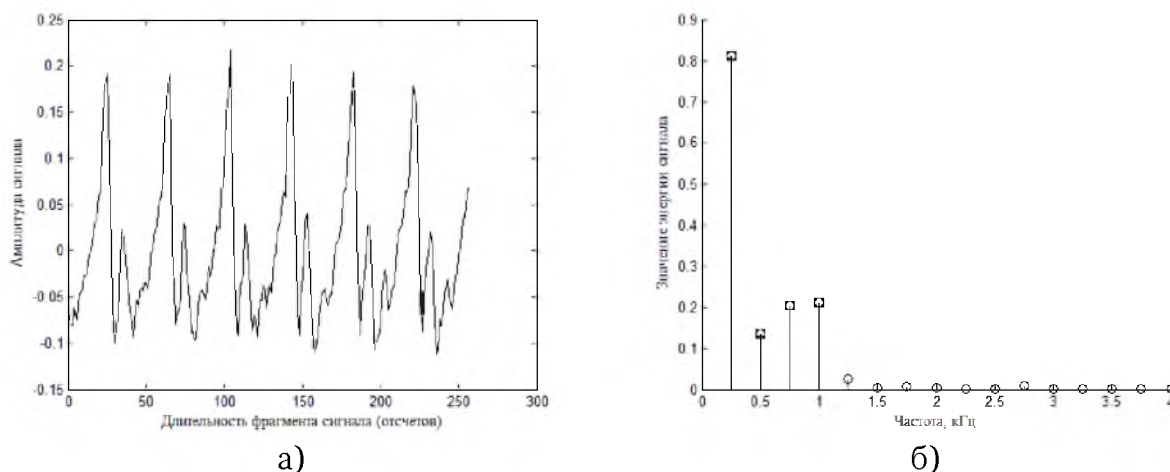
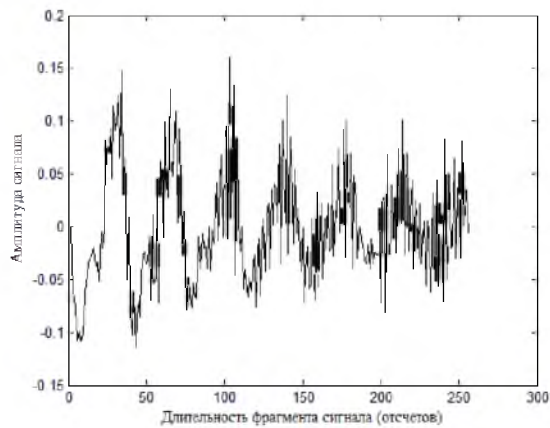
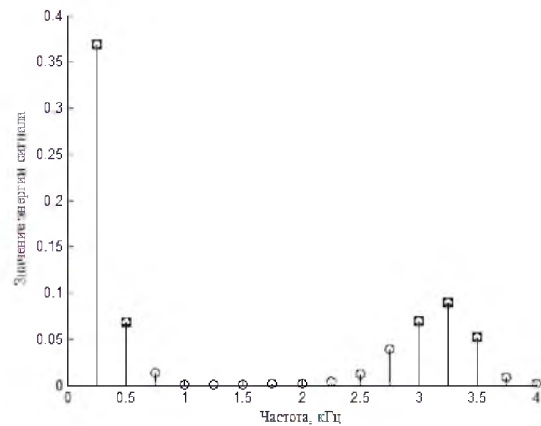


Рис. 1. Фрагмент сигнала, соответствующий звуку «а»:
а) изменение сигнала во времени ($f_d=8\text{кГц}$);
б) распределение энергии по частотным интервалам ($f_d=8\text{кГц}, N=256, R=16$)

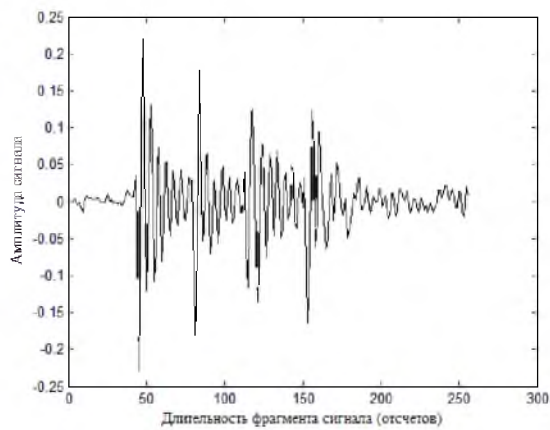


а)

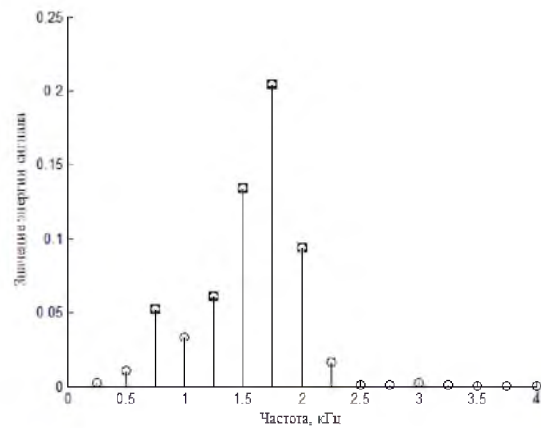


б)

Рис. 2. Фрагмент сигнала, соответствующий звуку «ж»:

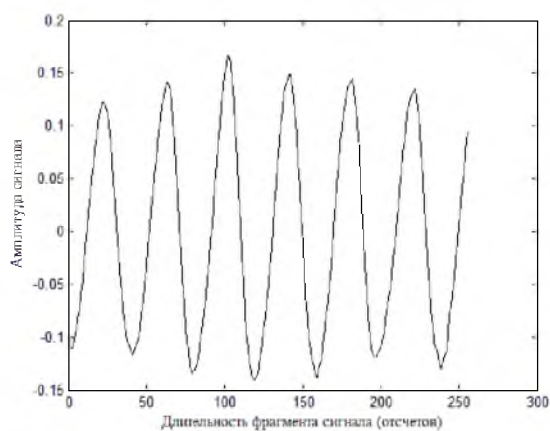
а) изменение сигнала во времени ($f_d=8\text{кГц}$);б) распределение энергии по частотным интервалам ($f_d=8\text{кГц}$, $N=256$, $R=16$)

а)

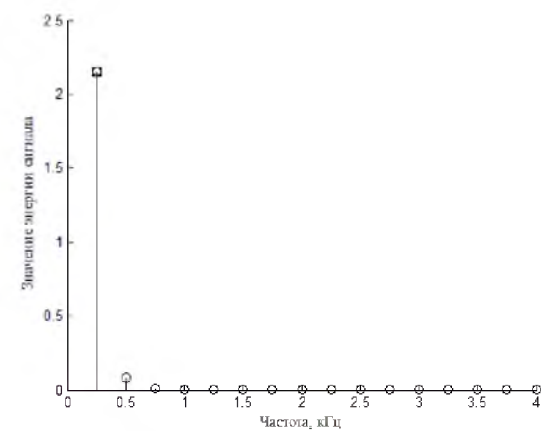


б)

Рис. 3. Фрагмент сигнала, соответствующий звуку «к»:

а) изменение сигнала во времени ($f_d=8\text{кГц}$);б) распределение энергии по частотным интервалам ($f_d=8\text{кГц}$, $N=256$, $R=16$)

а)



б)

Рис. 4. Фрагмент сигнала, соответствующий звуку «н»:

а) изменение сигнала во времени ($f_d=8\text{кГц}$);б) распределение энергии по частотным интервалам ($f_d=8\text{кГц}$, $N=256$, $R=16$)

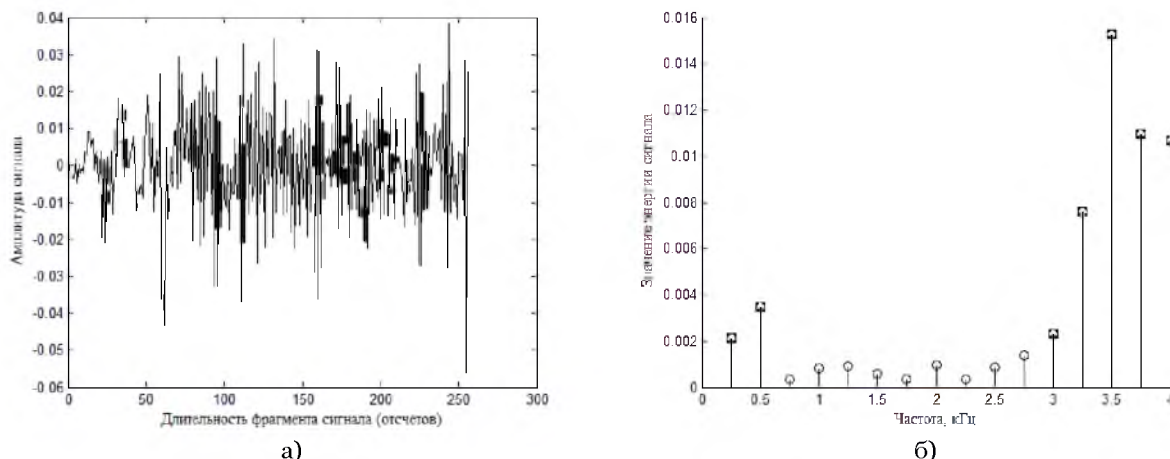


Рис. 5. Фрагмент сигнала, соответствующий звуку «ч»: а) изменение сигнала во времени ($f_d=8\text{кГц}$); б) распределение энергии по частотным интервалам ($f_d=8\text{кГц}$, $N=256$, $R=16$)

Анализ рисунков показывает, что количество информационных интервалов, а также их номера зависят от типа звука. Так для сигналов, соответствующих гласным и сонорным согласным, информационные частотные интервалы расположены в низкочастотной области. Для сигналов, соответствующих шумоподобным и взрывным согласным, информационные интервалы могут быть распределены более равномерно по частотной оси и преобладать в области средних и высоких частот.

В существующих слуховых аппаратах усиление сигнала осуществляется во всем диапазоне частот, что приводит также к усилению шумов. В рамках данной работы предлагается воздействовать только на информационные компоненты, а неинформационные компоненты при этом оставлять без изменений. Таким образом, именно коэффициенты при аппроксимациях $\hat{X}_{rN}(\omega)$ отрезков трансформант Фурье в частотных интервалах, где присутствует заметная часть энергии информационных компонент, должны компенсировать потери слуха.

$$\vec{w}_N = (w_1, \dots, w_N)^T = \sum_{r=1}^R c_r \vec{u}_{rN}, \tag{5}$$

где c_r – коэффициенты воздействия для каждого частотного интервала,

$\vec{u}_{rN} = (u_{r1}, \dots, u_{rN})^T$ – результата обратного преобразования Фурье:

$$u_{rk} = \int_{-\pi}^{\pi} X_{rN}(\omega) \exp(j\omega(k-1)) d\omega / 2\pi, k = 1, \dots, N,$$

где $X_{rN}(\omega)$ – результат оценки трансформанты Фурье,

N – длительность окна анализа.

Модель синтеза речевых сигналов представляет собой:

$$\vec{w}_N = (w_1, \dots, w_N)^T = \sum_{r=1}^R c_r \sum_{i=1}^J \lambda_{iN}^r \alpha_{iN}^r \vec{q}_{iN}^r, \tag{6}$$

где c_r – коэффициенты воздействия для каждого частотного интервала,

λ_{iN}^r – собственные числа субполосной матрицы r -го частотного интервала,

α_{iN}^r – скалярное произведение вида:

$$\alpha_{iN}^r = (\vec{q}_{iN}^r, \vec{x}_N), \tag{7}$$

\vec{q}_{iN}^r – собственные вектора субполосной матрицы r -го частотного интервала,

\vec{x}_N – анализируемый отрезок сигнала,



N – длина окна анализа фрагмента сигнала.

Анализ результатов экспериментов показывает, что общая средняя оценка качества звучания для проведенных экспериментов колеблется в диапазоне от 4,39 до 4,64 баллов по пятибалльной шкале оценок. Это свидетельствует о возможности выбора любых из перечисленных параметров. Наилучшее качество звучания наблюдается при выборе следующих параметров: $N=512$, $R=64$, $m=0,89$, $h_{NR}=0,5$. Выбор такой длины окна анализа может приводить к неточному определению границ между паузами и звуками речи. Поэтому рекомендуется выбирать меньшую длину окна анализа. Так при выборе параметров $N=256$, $R=64$, $m=0,81$, $h_{NR}=0,4$ средняя оценка качества звучания по всем проведенным экспериментам равна 4,6. Наименьший средний балл при следующих параметрах: $N=256$, $R=32$, $m=0,84$, $h_{NR}=0,5$. Важно отметить, что разные дикторы в качестве наилучших параметров отмечали различные сочетания длины окна анализа, количества интервалов, на которые разбивается ось частот, доли энергии, при которой оценивается величина частотной концентрации. Это связано с тем, что у каждого человека имеются свои особенности слухового аппарата.

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

Литература

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)». – Москва, 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.

RESEARCH SYNTHESIS ALGORITHM OF SPEECH SIGNALS IN DHA BASED SELECTIVE IMPACT ON THE FREQUENCY RANGES

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

*Belgorod National
Research University*

e-mail:

Belov_AS@bsu.edu.ru

Kurlov@bsu.edu.ru

Firsova_A@bsu.edu.ru

In the article the synthesis algorithm of speech signals. An algorithm for detecting pauses. Shows the results of evaluation of information and non-information on frequency domain and the evaluation results in sound quality speech synthesis.

Keywords: speech signal, analysis of the speech signal, the frequency representation, information frequency intervals, nondata frequency intervals.