

На правах рукописи



ФИРСОВА Александра Александровна

**РАЗРАБОТКА И ИССЛЕДОВАНИЕ СУБПОЛОСНЫХ МЕТОДОВ
И АЛГОРИТМОВ СЕГМЕНТАЦИИ РЕЧЕВЫХ СИГНАЛОВ**

Специальность 05.13.17 – Теоретические основы информатики

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук

Белгород –2013

Работа выполнена в федеральном государственном автономном образовательном учреждении высшего профессионального образования «Белгородский государственный национальный исследовательский университет» (НИУ «БелГУ»), на кафедре информационно-телекоммуникационных систем и технологий.

Научный руководитель **Жиляков Евгений Георгиевич**,
доктор технических наук, профессор

Официальные оппоненты: **Волчков Валерий Павлович**,
доктор технических наук, старший научный сотрудник,
Московский технический университет связи
и информатики,
профессор кафедры общей теории связи,
г. Москва

Черноморец Андрей Алексеевич,
кандидат технических наук, доцент,
Белгородский государственный национальный
исследовательский университет,
заведующий кафедрой прикладной
информатики, г. Белгород

Ведущая организация Федеральное государственное автономное
образовательное учреждение высшего
профессионального образования «Северо-
Кавказский федеральный университет»,
г. Ставрополь

Защита состоится 19 июня 2013 года в 15 часов 00 минут на заседании диссертационного совета Д 212.015.10 на базе ФГАОУ ВПО «Белгородский государственный национальный исследовательский университет» по адресу: 308015 г. Белгород, ул. Победы, д. 85, ауд. 3-8, к.15.

С диссертацией можно ознакомиться в научной библиотеке ФГАОУ ВПО «Белгородский государственный национальный исследовательский университет» (НИУ «БелГУ») по адресу: 308015 г. Белгород, ул. Победы, д. 85.

Автореферат разослан 17 мая 2013 г.

Ученый секретарь
диссертационного совета к.т.н., проф.,



С. П. Белов

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность работы.

Устная речь является одной из наиболее естественных для человека форм информационного обмена. В соответствии с этим существенное внимание уделяется разработке методов автоматического анализа устной речи на компьютерах, одним из важнейших направлений развития которых является совершенствование человеко-машинных интерфейсов. При этом обработке подвергаются речевые сигналы (РС), которые представляют собой результаты регистрации электромагнитных колебаний на выходе микрофонов, возбуждаемых акустическими воздействиями. В определенном смысле базовым этапом автоматического анализа устной речи служит сегментация РС на отрезки, которые порождены в разных условиях. Наибольший интерес представляет выделение участков, зарегистрированных при отсутствии звуков речи (селекция пауз, в том числе при сжатии речевых данных), определение границ участков, порождаемых разными звуками речи и определение среди последних отрезков, порождаемых вокализованными звуками (в задачах идентификации дикторов и оценки их психофизического состояния, распознавания речи и ее вокодерного синтеза с помощью липредеров и т.д.).

Проблема анализа речевых сигналов рассматривалась в достаточно большом количестве работ зарубежных и российских авторов, среди которых можно отметить Л. Рабинера, Б. Голда, М.А. Сапожкова, О.И. Шелухина, А.В. Аграновского, В.Н. Сорокина, Р.В. Гудонавичюса, Ю.Н. Прохорова, В.Г. Санникова, М.В. Хитрова, Е.Г. Жилякова и других. В основе многих из разработанных подходов используются частотные представления, так как порождаемые звуками речи отрезки РС обладают свойством концентрации энергии в достаточно «узких» полосах частотной оси. В связи с этим можно упомянуть рассматриваемое в литературных источниках разбиение частотной полосы на так называемые критические полосы слуха, которые опосредованно отражаются на частотных свойствах РС. До определенной степени эквивалентным (с учетом равенства Парсеваля) инструментом обработки РС представляется корреляционный анализ, особенно в задачах распознавания речевых конструкций из созданных заранее словарей.

Вместе с тем можно отметить, что предлагаемые в настоящее время методы сегментации РС на основе анализа их частотно-корреляционных свойств либо не отражают свойства концентрации энергии, либо слишком чувствительны к изменениям энергии в паузах.

Таким образом, адекватным подходом к обработке РС является субполосный анализ, когда их свойства соотносятся с некоторым разбиением оси частот на интервалы конечной ширины. Причем, в виду зависимости частотного распределения энергий от вида произносимого звука, анализу необходимо подвергать отрезки РС конечной длительности. Отметим необходимость проведения специальных исследований специфики частотного распределения энергий отрезков РС, порождаемых русской устной речью.

В рамках данной работы в качестве основных субполосных характеристик рассматриваются:

$$P_r(\bar{x}_N) = \int_{\omega \in V_r} |X(\omega)|^2 d\omega / 2\pi, \quad (1)$$

$$\Phi_N^r = \int_{\omega \in V_r} X(\omega) Y^*(\omega) d\omega / 2\pi, \quad (2)$$

где $X(\omega)$ и $Y(\omega)$ – трансформанты (спектры) Фурье анализируемых отрезков вида

$$Z(\omega) = \sum_{k=1}^N z_k \exp(-j\omega(k-1)), j = (-1)^{1/2}, \quad (3)$$

V_r – составной интервал оси нормированных частот

$$V_r = [-V_{1r}, -V_{2r}) \cup [V_{1r}, V_{2r}), 0 \leq V_{1r} < V_{2r} \leq \pi. \quad (4)$$

Соотношение (1) определяет часть энергии, попадающую в заданный частотный интервал V_r , а (2) – субполосную корреляцию спектров различных отрезков. В дальнейшем будет показано, что вычисление их точных значений можно осуществить непосредственно во временной области, не вычисляя спектров.

Вместе с тем следует отметить, что в настоящее время в качестве основного инструмента частотной обработки РС используется преобразование Фурье, преимущественно дискретное преобразование Фурье (ДПФ), что, по-видимому, обусловлено наличием алгоритма его быстрого вычисления (алгоритма БПФ). Важно иметь в виду, что ДПФ является вариантом разложения сигналов по ортонормированному базису специального вида и не позволяет в полной мере описать распределение энергий отрезков РС по частотным интервалам, то есть адекватно отразить их физическую сущность.

Использование ортогональных разложений по иным базисам, например, вейвлет-анализ, по тем же причинам нельзя признать адекватным подходом к описанию распределения энергий анализируемых отрезков по частотным интервалам.

Таким образом, разработка субполосных методов и алгоритмов сегментации речевых сигналов на основе адекватного анализа динамики изменений распределения энергий их отрезков по частотным интервалам является актуальной задачей.

Целью данной работы является совершенствование методов и алгоритмов компьютерного анализа русской устной речи на основе разработки субполосных методов сегментации РС на отрезки со значимыми различиями в распределениях энергий по частотным интервалам, которые обусловлены сменами звуков речи либо их отсутствием (наличием пауз).

Для достижения поставленной цели сформулированы и решены следующие задачи:

1. Анализ основных применяемых методов и алгоритмов сегментации РС.
2. Формирование моделей распределения энергий отрезков РС, порождаемых различными звуками русской речи, по частотным интервалам.
3. Разработка и исследование субполосного метода и алгоритма селекции пауз между звуками русской речи с предварительным обучением.

4. Разработка и исследование субполосного метода и алгоритма сегментации РС на отрезки, порождаемые различными звуками русской речи

5. Разработка и исследование субполосного метода и алгоритма селекции участков РС по вокализованным звукам речи и определения периода основного тона.

6. Разработка программных реализаций созданных алгоритмов.

Объект исследований: речевые сигналы.

Предмет исследования: методы сегментации речевых сигналов по границам звуков русской речи.

Методы исследований. В работе использованы методы субполосного анализа/синтеза, статистического анализа, цифровой обработки речевых данных, линейной алгебры и вычислительных экспериментов.

Научную новизну работы составляет следующее:

1. Модели частотного распределения энергий порождаемых звуками русской устной речи отрезков РС с учетом их длительности.

2. Решающие функции для сопоставления распределений энергий различных отрезков речевых сигналов по частотным интервалам в задачах селекции пауз и определения границ между звуками в русской устной речи.

3. Субполосный нормированный коэффициент корреляции, позволяющий повысить обоснованность решений о величине периода основного тона в задачах селекции вокализованных звуков русской устной речи.

Практическая значимость работы определяется разработанными алгоритмами сегментации РС по границам звуков русской устной речи, применение которых позволяет усовершенствовать процедуры решения задач сжатия речевых данных, распознавания речи, её синтеза, идентификации дикторов, проведения психофизических исследований и т.д.

Область исследования. Содержание диссертации соответствует паспорту специальности 05.13.17 «Теоретические основы информатики» (технические науки) по следующим областям исследований:

п. 5. Разработка и исследование моделей и алгоритмов анализа данных, обнаружения закономерностей в данных и их извлечениях, разработка и исследование методов и алгоритмов анализа текста, устной речи и изображений.

п. 6. Разработка методов, языков и моделей человекомашинного общения; разработка методов и моделей распознавания, понимания и синтеза речи, принципов и методов извлечения данных из текстов на естественном языке.

Связь с научными и инновационными программами.

Диссертационное исследование выполнялось в рамках следующих федеральных проектов: ГК № П964 от 27 мая 2010г., ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009-2013 годы; проект № 2.1.2/9382 АВЦП “Развитие научного потенциала высшей школы (2009-2010 годы)”;

ГК № 10-07-00326а, РФФИ; программа УМНИК-2010, проект № 13990; программа УМНИК-2012, проект № 16945.

Положения, выносимые на защиту:

1. Субполосные методы сегментации РС по границам звуков русской устной речи, позволяющие усовершенствовать процедуры селекции пауз, определения границ звуков русской речи, селекции вокализованных звуков русской речи, включая оценивание периода основного тона.

2. Алгоритмы субполосной обработки отрезков речевых сигналов при их сегментации по границам звуков русской устной речи.

3. Результаты сравнительных вычислительных экспериментов, иллюстрирующих эффективность разработанных алгоритмов субполосной сегментации РС.

Достоверность выводов и рекомендаций обусловлена корректностью применяемых математических преобразований, отсутствием противоречий с известными фактами теории и практики обработки речевых сигналов, а также иллюстрируется результатами вычислительных экспериментов с реальными речевыми данными.

Личный вклад соискателя. Все изложенные в диссертации результаты исследования получены либо соискателем лично, либо при его непосредственном участии.

Апробация результатов диссертационного исследования.

Результаты диссертационного исследования обсуждались на следующих научно-технических конференциях: Международная молодежная конференция «Прикладная математика, управление и информатика», (г. Белгород, 2012); 13-ая, 12-ая Международная конференция и выставка «Цифровая обработка сигналов и её применение - DSPA» (г. Москва, 2011, 2010 гг.); XIX Международная научно-практическая конференция «Информационные технологии: наука, техника, технология, образование, здоровье» (г. Харьков, 2011); II Международная научно-практическая конференция «Информационные технологии и компьютерная инженерия» (г. Харьков, 2011); Вторая Международная научно-техническая конференция, «Компьютерные науки и технологии» (г. Белгород, 2011); XXIII Международная научная конференция «Математические методы в технике и технологиях-ММТТ-23» (г. Саратов, 2010); Вторая Международная научно-практическая Интернет-конференция Белгородского филиала НАЧОУВПОСГА (г. Белгород, 2010); Всероссийская научная конференция с элементами научной школы для молодежи «Параллельная компьютерная алгебра» (г. Ставрополь, 2010); Всероссийская молодежная конференция «Теория и практика системного анализа» (г. Белгород, 2012).

Публикации. По теме диссертационного исследования опубликовано 24 печатные работы (из них 8 в журналах из списка ВАК РФ), в том числе 2 Свидетельства Роспатента РФ о государственной регистрации программ для ЭВМ.

Объем и структура работы. Диссертация состоит из Введения, четырех глав, Заключения и Приложений. Работа изложена на 165 страницах машинописного текста, включая 86 рисунков, 10 таблиц и список литературных источников из 110 наименований.

СОДЕРЖАНИЕ РАБОТЫ

Во введении обосновывается актуальность диссертационного исследования, формулируется его основная цель.

Глава 1. Современные методы и алгоритмы сегментации речевых сигналов. Анализируются основные характеристики существующих методов сегментации РС. Представлены классификации звуков русской речи с позиции особенностей их артикуляционных и акустических характеристик.

Показано, что в основе субполосных методов сегментации РС целесообразно использовать определение частей энергий (1) отрезка (вектора) дискретизованного РС $\vec{x}_N = (x_1, \dots, x_N)^T$ (T – символ транспонирования), попадающих в частотный интервал (4).

В рамках данной работы используется полученное в монографии¹ соотношение

$$P_r(\vec{x}_N) = \vec{x}_N^T A_r \vec{x}_N, \quad (5)$$

позволяющее, не переходя в частотную область, вычислять части энергии (1) с использованием субполосных матриц $A_r = \{a_{ik}^r\}$, $i, k = 1, \dots, N$, где

$$a_{ik}^r = (\sin(V_{2r}(i-k)) - \sin(V_{1r}(i-k))) / \pi(i-k); \quad a_{ii}^r = (V_{2r} - V_{1r}) / \pi. \quad (6)$$

Там же показано, что для удовлетворяющих уравнениям

$$\lambda_k^r \vec{q}_k^r = A_r \vec{q}_k^r, \quad k = 1, \dots, N, \quad (7)$$

собственных чисел λ_k^r и векторов \vec{q}_k^r субполосной матрицы с высокой точностью выполняются условия

$$\lambda_{J+k}^r = 0, \quad k = 1, \dots, N - J; \quad J = 2\lceil N(V_{2r} - V_{1r}) / 2\pi \rceil + 2 \quad \text{при} \quad \lambda_k^r \geq \lambda_{k+1}^r, \quad k = 1, \dots, N - 1. \quad (8)$$

Поэтому вычисления квадратичной формы (5) можно ускорить, если использовать представление вида

$$P_r(\vec{x}_N) = \sum_{k=1}^J \lambda_k^r \alpha_{kr}^2(\vec{x}_N), \quad (9)$$

где $\alpha_{kr}(\vec{x}_N)$ – скалярные произведения (проекции)

$$\alpha_{kr}(\vec{x}_N) = (\vec{x}_N, \vec{q}_k^r) = \sum_{i=1}^N x_i q_{ik}^r. \quad (10)$$

На основе проведенного анализа состояния вопроса и в соответствии с целью работы формулируются задачи исследования.

Глава 2. Субполосные модели отрезков речевых сигналов и селекция пауз между звуками речи.

Раздел 2.1. Вычислительные аспекты субполосного анализа отрезков РС. Пусть ось нормированных частот $[0, \pi]$ разбивается на ряд неперекрывающихся интервалов V_r , $r = 0, \dots, R$, границы которых определяются соотношениями:

$$V_{10} = 0; \quad V_{20} = 2\pi / N; \quad V_{1r} = V_{2, r-1}; \quad V_{2r} = V_{1r} + 4\pi / N, \quad r = 1, \dots, R, \quad (11)$$

$$R = N / 4. \quad (12)$$

¹ Жиликов, Е.Г. Вариационные методы анализа и построения функций по эмпирическим данным: моногр. / Е.Г. Жиликов. – Белгород: Изд-во БелГУ, 2007. – 160с.

Тогда с учетом определения (5) на основе равенства Парсеваля нетрудно установить справедливость следующего соотношения

$$0,5P_R(\bar{x}_N) + \sum_{r=0}^{R-1} P_r(\bar{x}_N) = \sum_{i=1}^N x_i^2 = \|\bar{x}_N\|^2. \quad (13)$$

Легко также показать, что при выборе (11) элементы субполосных матриц (6) могут быть представлены в виде:

$$a_{ik}^0 = \sin(V_{20}(i-k)) / \pi(i-k), \quad a_{ik}^r = 2 \cdot a_{ik}^0 \cos(\omega_r(i-k)), \quad \omega_r = r \cdot 4\pi / N, \quad r=1, \dots, R. \quad (14)$$

В соответствии с этим их собственные векторы можно аппроксимировать с помощью соотношений:

$$q_{i,2k-1}^r = q_{ik}^0 \cos(\omega_r(i-1)); \quad q_{i,2k}^r = q_{ik}^0 \sin(\omega_r(i-1)), \quad k=1, \dots, J/2, \quad i=1, \dots, N, \quad (15)$$

где q_{ik}^0 – компоненты соответствующих собственных векторов субполосной матрицы A_0 , которым отвечают собственные числа λ_k^0 , причем и для собственных чисел выполняются равенства

$$\lambda_{2k-1}^r = \lambda_{2k}^r = \lambda_k^0, \quad k=1, \dots, J/2. \quad (16)$$

Поэтому соотношения (10) можно преобразовать к виду

$$\alpha_{2k-1,r}(\bar{x}_N) = \sum_{i=1}^N z_{ki} \cos(\omega_r(i-1)), \quad \alpha_{2k,r}(\bar{x}_N) = \sum_{i=1}^N z_{ki} \sin(\omega_r(i-1)), \quad k=1, \dots, J/2, \quad (17)$$

где

$$z_{ki} = x_i q_{ik}^0, \quad k=1, \dots, J/2; \quad i=1, \dots, N. \quad (18)$$

Легко понять, что определяемые соотношениями (17) скалярные произведения являются выборками при $b=2 \cdot r$ из наборов вещественной и мнимой компонент (с обратным знаком) ДПФ вида

$$Z_k(b) = \sum_{i=1}^N z_{ki} \exp(-j2\pi / N \cdot b \cdot (i-1)), \quad b=0, \dots, N/2-1. \quad (19)$$

Таким образом, для их вычислений можно воспользоваться алгоритмом БПФ последовательностей с компонентами вида (18), выбирая соответствующие длительности анализируемых отрезков.

Использование всех компонент ДПФ вида (19) дает возможность более тонкого частотного анализа распределения энергий отрезков сигналов. На рисунке 1 представлена конфигурация частотных интервалов при использовании такого анализа.

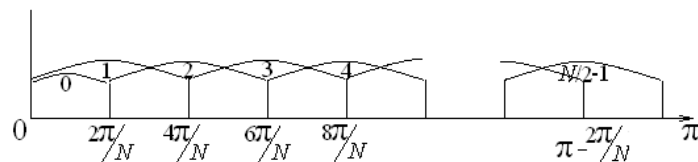


Рисунок 1 – Конфигурация частотных интервалов «с перекрытием».

Ниже в таблице 1 приведены значения собственных чисел субполосной матрицы A_0 (14).

Таблица 1 – Значения собственных чисел

Значение индекса k	1	2	3	4	5
Значение собственного числа λ_k^0	0,981	0,750	0,244	0,0246	0,001

Очевидно, что двумя последними значениями из приведенных собственных чисел можно пренебречь, так как их сумма по сравнению с суммой остальных почти в 80 раз меньше. Поэтому можно положить $J = 6$.

Отметим также, что определяемый соотношением (11) размер частотного интервала $\Delta V = 4\pi/N$ может служить единицей измерений при определении суммарной ширины распределения энергий подвергаемых анализу отрезков РС. В самом деле, пусть значения отсчетов анализируемого отрезка удовлетворяют условию $x_i = \cos(v(i - N/2))$. Тогда, подставляя это представление в определение (3), нетрудно получить

$$|X(z)|^2 = [SP(z - v) + SP(z + v)]^2$$

где

$$SP(y) = \sin(Ny/2) / \sin(y/2).$$

Очевидно, что слагаемые в предыдущем соотношении будут иметь максимумы в точках $z_{1,2} = \pm v$, а в точках $v \pm \Delta V/2$ будут равны нулю, что и позволяет указанную ширину частотного интервала использовать в качестве единицы измерений суммарной ширины распределения энергий анализируемых отрезков РС.

Раздел 2.2. Распределение энергий по частотным интервалам отрезков РС, порождаемых разными звуками русской речи. В разделе основное внимание уделено анализу распределения энергии с позиции выявления диапазонов, в которых сосредоточена основная доля энергии для РС, порождаемых различными звуками русской речи. Одной из характеристик частотного распределения энергии является величина частотной концентрации, характеризующая наименьшую ширину частотной области, в которой сосредоточена заданная доля энергии.

Как было отмечено ранее, в качестве единицы измерения целесообразно использовать величину $4\pi/N$.

Тогда частотная концентрация θ может быть оценена как

$$\theta = K \cdot (4\pi/N) / \pi = 4K/N, \quad (20)$$

где K – наименьшее количество частотных интервалов, для которых выполняется условие

$$\sum_{k=1}^K P_{(k)}(\bar{x}_N) \geq m \|\bar{x}_N\|^2, \quad (21)$$

m – задаваемая доля энергии; $P_{(k)}(\bar{x}_N)$ – упорядоченная по убыванию последовательность энергий (9) сигнала в частотных интервалах, т.е. выполняются следующие условия:

$$P_{(k)}(\bar{x}_N) \geq P_{(k+1)}(\bar{x}_N), \quad k = 1, 2, \dots, R. \quad (22)$$

$$P_{(k)}(\bar{x}_N) \in \{P_r(\bar{x}_N)\}, \quad r = 0, 1, \dots, R. \quad (23)$$

При этом частотная ось разбивается на R непересекающихся частотных интервалов (рис. 2).

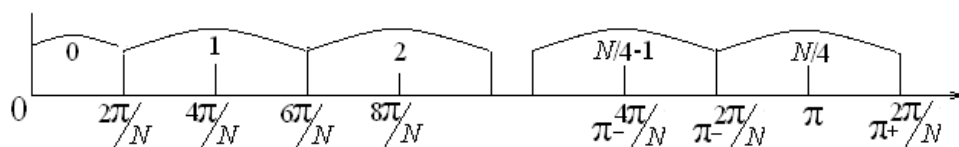


Рисунок 2 – Конфигурация частотных интервалов «без перекрытия».

В работе представлена методика и алгоритм проведения эксперимента, позволяющая осуществить оценку величины частотной концентрации отрезков РС.

В работе для каждого звука русской речи получены оценки величины частотной концентрации при различных значениях длительности анализируемого отрезка РС и задаваемой доли энергии. При этом учитывалось, что длительность РС, порождаемых такими звуками, как: «к», «п», «т», «г», «д», не превышает 18мс (при $f_d=16\text{кГц}$ $N=288$). Поэтому для анализа РС использована длительность сигнала 16мс (при $f_d=16\text{кГц}$ $N=256$).

В таблице 2 представлены результаты оценки средней величины частотной концентрации для РС, порождаемых разными звуками русской речи и шумами при различных значениях задаваемой доли энергии m . В таблице шум1 – означает последовательность псевдослучайных гауссовых чисел; шум2 – фрагмент сигнала, соответствующего шуму в помещении.

Таблица 2 – Значения величины θ для различных звуков русской речи ($f_d=16\text{кГц}$, $N=256$)

тип звука	$m=0.80$	$m=0.86$	$m=0.90$	$m=0.96$	$m=0.98$
гласные					
а	0,08	0,09	0,11	0,16	0,20
и	0,05	0,06	0,08	0,13	0,19
высокие шумные согласные					
ж	0,14	0,19	0,23	0,38	0,48
ч	0,22	0,28	0,34	0,50	0,50
шумы в паузах					
шум1	0,59	0,69	0,75	0,88	0,92
шум2	0,39	0,47	0,55	0,72	0,81

Анализ результатов исследования всех звуков русской речи показал, что энергия РС, порождаемых гласными, сонорными и низкими звонкими согласными звуками, сосредоточена в меньшем количестве частотных интервалов, чем энергия РС, порождаемых низкими глухими и высокими согласными звуками. Причем, величина частотной концентрации θ при $m=0,9$ не превосходит 0,47.

Анализ результатов оценки частотного распределения энергии отрезков РС, порождаемых различными звуками русской речи показывает, что основная энергия ($m=0,9$) РС, порождаемых гласными, сонорными и низкими звонкими согласными, сосредоточена в диапазоне от 0 до 2кГц (исключение составляют такие звуки как «й», «р», для которых энергия распределена до 4кГц). Энергия РС, порождаемых низкими глухими и высокими шумными

согласными, распределена вдоль оси частот до 7кГц. Таким образом, для более точного анализа особенностей частотного распределения энергии РС, порождаемых различными звуками русской речи, необходимо использовать частоту дискретизации не меньше 16кГц.

Раздел 2.3. Селекция пауз между звуками речи. Выше было показано, что энергии порождаемых звуками речи отрезков РС распределены в совокупности достаточно узких частотных интервалов, верхняя граница которых не превышает 7 килогерц, а суммарная ширина не превышает половины этого частотного диапазона. В свою очередь, посторонние шумы при отсутствии речи имеют иное частотное распределение энергии, что позволяет на этой основе построить решающие процедуры селекции пауз.

Основная проверяемая гипотеза формулируется следующим образом.

H_0 – анализируемый отрезок РС $\vec{x} = (x_1, \dots, x_N)^T$ порожден шумами в паузе речи

$$\vec{x} = \vec{u} = (u_1, \dots, u_N)^T. \quad (24)$$

Альтернатива H_1 заключается в том, что хотя бы часть компонент рассматриваемого вектора зафиксирована в присутствии звуков речи

$$\vec{x} = \vec{u} + \vec{s}, \quad \vec{s} = (s_1, \dots, s_N)^T. \quad (25)$$

Здесь и в дальнейшем, где это не вызывает затруднений, индекс у векторов опущен.

В виду неодинаковости и изменчивости во времени воздействий, оказываемыми на микрофон различными звуками речи, по-видимому, единственной возможностью является использование в качестве признаков энергетических характеристик: энергия (квадрат евклидовой нормы) анализируемого вектора $\|\vec{x}\|^2$ и распределение её по частотным интервалам в виде наборов соответствующих долей

$$Pd_n(\vec{x}) = P_n(\vec{x}) / \sum_{k=0}^{N/2-1} P_k(\vec{x}), \quad n = 0, \dots, N/2 - 1. \quad (26)$$

Непосредственно из этого определения следует равенство

$$\sum_{n=0}^{N/2-1} Pd_n(\vec{x}) = 1. \quad (27)$$

В качестве меры различий между сравниваемыми векторами предлагается использовать соотношение

$$\gamma(x, y) = \left(\sum_{n=0}^{N/2-1} ((Pd_n(\vec{x}))^{1/2} - (Pd_n(\vec{y}))^{1/2})^2 \right)^{1/2}, \quad (28)$$

которое естественно называть субполосным расстоянием.

Отметим, что использование квадратных корней позволяет до определенной степени выровнять вклады разностей долей энергий из различных частотных интервалов.

С учетом свойства (27) соотношение (28) преобразуется к следующему виду

$$\gamma(x, y) = \left(2 \left(1 - \sum_{n=0}^{N/2-1} (Pd_n(\vec{x}) \cdot Pd_n(\vec{y}))^{1/2} \right) \right)^{1/2}. \quad (29)$$

Пусть теперь D_n^2 и G_u – математические ожидания долей энергий отрезков шумов в паузах

$$D_n^2 = M[Pd_n(\bar{u})] \quad (30)$$

и их квадратов евклидовых норм

$$G_u = M[\|\bar{u}\|^2]. \quad (31)$$

Легко показать, что в виду (27) будет иметь место равенство

$$\sum_{n=0}^{N/2-1} D_n^2 = M\left[\sum_{n=0}^{N/2-1} Pd_n(\bar{u})\right] = 1. \quad (32)$$

В качестве решающей функции (РФ) при селекции пауз предлагается использовать статистику

$$F_u(x) = W_u(x) \cdot \gamma_u(x), \quad (33)$$

где

$$\gamma_u(x) = \left(\sum_{n=0}^{N/2-1} ((Pd_n(\bar{x}))^{1/2} - D_n)^2 \right)^{1/2} = \left(2 \left(1 - \sum_{n=0}^{N/2-1} D_n \cdot (Pd_n(\bar{x}))^{1/2} \right) \right)^{1/2}, \quad (34)$$

$$W_u(x) = \|\bar{x}\|^2 / G_u. \quad (35)$$

Отметим, что первый сомножитель в (33) реагирует на изменение энергии по сравнению со средней, тогда как второй сомножитель реагирует на изменение её распределения по частотным интервалам.

Гипотеза H_0 отвергается при выполнении неравенства

$$F_u(x) > h_\alpha, \quad (36)$$

где $h_\alpha > 0$ – порог, удовлетворяющий условию

$$PR(F_u(x) > h_\alpha / H_0) \leq \alpha. \quad (37)$$

Здесь PR – символ вероятности, а α – желаемый уровень вероятности ошибок первого рода.

Оценивание значений математических ожиданий (30) и (31), а также порогов для правила (36) осуществляется в режиме обучения по совокупности реализаций шумов в паузах речи.

Разработаны алгоритмы обучения и селекции пауз, описание которых представлено в диссертации.

Для оценки эффективности разработанного алгоритма была использованы оценки вероятностей ошибок первого и второго рода:

$$P_{1ош} = N_{ошиб.речь} / N_{пауз}, \quad (38)$$

$$P_{2ош} = N_{ошиб.пауз} / N_{речи}, \quad (39)$$

где $N_{ошиб.речь}$ – количество отрезков, ошибочно отнесенных к РС в присутствии звуков речи; $N_{ошиб.пауз}$ – количество отрезков, ошибочно отнесенных к шуму в паузе; $N_{пауз}$ – количество отрезков РС, порожденных шумами, использованных для анализа (185000 отрезков); $N_{речи}$ – количество отрезков РС в присутствии звуков речи, использованных для анализа (230000 отрезков).

Важно отметить, что вероятность ошибки первого рода оценивалась по фрагменту сигнала в паузе большему, чем длительность фрагмента участка пауз, использованного для обучения.

Разработаны методики проведения экспериментов, включающие концептуальные основы, цели, план и порядок выполнения.

В работе представлены результаты сравнения эффективности предлагаемого алгоритма с существующими при различных параметрах. В таблице 3 представлены результаты сравнений с методом, основанным на применении РФ максимальной чувствительности².

Таблица 3 – Значения вероятностей ошибок второго рода при различных параметрах ($f_d=16\text{кГц}$, $\alpha=0,05$)

	предлагаемая РФ		РФ максимальной чувствительности	
	$N=128$	$N=256$	$N=128$	$N=256$
$P_{1ош}$	0,0332	0,0791	0,0659	0,0984
$P_{2ош}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$

Эти и приведенные в диссертации результаты сравнений показывают преимущества предлагаемого метода.

Раздел 2.4. Результаты и выводы главы.

Глава 3. Сегментация речевых сигналов по границам звуков русской речи.

Раздел 3.1. Метод и алгоритм определения границ звуков русской речи. Пусть $\vec{x} = (x_1, \dots, x_N)$ и $\vec{y} = (y_1, \dots, y_N)$ – непересекающиеся фрагменты РС. Необходимо определить соответствуют ли фрагменты \vec{x} и \vec{y} одному и тому же звуку русской речи.

Основная проверяемая гипотеза формулируется следующим образом: H_0 – фрагменты сигналов \vec{x} и \vec{y} соответствуют одному и тому же звуку русской речи.

В основе решающей процедуры используется введенное и описанное в разделе 2.3 субполосное расстояние (29).

Для повышения чувствительности к изменениям энергий отрезков также предлагается использовать статистику

$$W(x, y) = \max(\|\vec{x}\|^2, \|\vec{y}\|^2) / \min(\|\vec{x}\|^2, \|\vec{y}\|^2). \quad (40)$$

Данная характеристика всегда больше единицы, причем, чем больше ее значение, тем больше вероятность того, что сравниваемые фрагменты соответствуют различным звукам речи.

В качестве РФ предлагается использовать

$$F(x, y) = W(x, y) \cdot \gamma(x, y). \quad (41)$$

Ее естественно именовать субполосной РФ.

Гипотеза H_0 отвергается при выполнении неравенства

$$F(x, y) > h_\alpha(\theta), \quad (42)$$

где α – вероятность ошибки первого рода, определенная ранее (37), θ – величина частотной концентрации отрезка \vec{x} , определенная в разделе 2.2.

² Белов, С.П. Исследование решающей функции максимальной чувствительности к изменению частот энергии в частотных интервалах / С.П. Белов, А.А. Фирсова // Научные ведомости Белгородского государственного университета. Сер. История. Политология. Экономика. Информатика – 2012. – №13(132) выпуск 23/1. – С. 227-231

Для установления пороговых значений проведены вычислительные эксперименты, которые показали, что для всех звуков русской речи величина порога зависит от значения частотной концентрации отрезка \bar{x} : чем больше значение θ , тем выше требуемая величина порога. В таблице 4 представлены значения порогов при длительности отрезка анализа 16мс ($N=256$ при $f_d=16$ кГц), полученные при анализе не менее 500 отрезков для каждого значения θ . Важно отметить, что значения порогов для величин частотной концентрации, попадающих в интервалы между значениями, приведенными в таблице 4, могут быть определены при использовании линейной интерполяции.

Таблица 4 – Значения порогов $h_\alpha(\theta)$ при различных значениях частотной концентрации θ ($f_d=16$ кГц, $N=256$, $m=0.9$, $\alpha=0,05$)

θ	0,1	0,2	0,3	0,4	0,5
$h_\alpha(\theta)$	0,88	0,98	1,02	1,03	1,09

Если условие (42) выполняется, то в точке стыка сравниваемых фрагментов устанавливается граница, иначе принимается решение, что анализируемые отрезки принадлежат одному и тому же звуку речи.

Разработан алгоритм сегментации РС по границам звуков речи, описание которого представлено в диссертации.

Для оценивания вероятностных характеристик принятия решений используются вычислительные эксперименты, методики проведения, включая концептуальные основы, цель, план и порядок проведения, которых описаны в диссертации.

В работе представлены результаты сегментации РС с использованием субполосной РФ при различных параметрах. Для оценки эффективности были использованы оценки вероятностей ошибки первого и второго рода вида:

$$P_{1ош} = (N_{авт} - N_{совнав}) / N_{актов}, \quad (43)$$

$$P_{2ош} = 1 - N_{совнав} / N_{руч}, \quad (44)$$

где $N_{совнав}$ – количество границ, установленных в автоматическом режиме, совпавших с количеством границ, установленных в ручную; $N_{руч}$ – количество границ, установленных в ручную (6000 границ); $N_{авт}$ – количество границ, установленных автоматически; $N_{актов}$ – количество актов проверки гипотезы H_0 (3000000 проверок).

В работе представлены методики проведения экспериментов и результаты сравнения разработанного алгоритма сегментации с существующими. В таблице 5 представлены некоторые из них.

Таблица 5 – Значения вероятностей ошибок первого и второго рода ($f_d=16$ кГц, $N=256$, $m=0.9$)

	метод субполосной РФ	по корреляции между равноотстоящими спектрами ³
$P_{1ош}$	0,0004	0,0007
$P_{2ош}$	0,25	0,39

³ Цыплихин, А.И. Сегментация речи на кардинальные элементы / А.И. Цыплихин, В.Н. Сорокин // Информационные процессы. – 2006. – Т. 6, № 3. – С. 177-207.

Полученные результаты показывают, что алгоритм сегментации РС по границам звуков, основанный на субполосной РФ, обеспечивает меньшие значения вероятностей ошибок первого и второго рода.

Раздел 3.2. Селекция вокализованных звуков русской речи и оценка периода основного тона. Вокализованные звуки являются периодическими, что может быть положено в основу их селекции. При этом необходимо оценивать период основного тона, если отвергается основная гипотеза:

H_0 – анализируемый отрезок $\vec{x}_1 = (x_1, x_2, \dots, x_N)^T$ порожден невокализованным звуком речи.

В настоящее время существует два основных подхода к определению периода основного тона (ОТ): на основе анализа спектров и корреляционном анализе.

Спектральный метод заключается в определении в диапазоне априори допустимых значений частоты ОТ такой из них, в которой наблюдается максимум модуля спектра Фурье. Такая частота при выполнении некоторых условий принимается за частоту основного тона.

Основной недостаток спектрального оценивания заключается в следующем. Пусть последовательность отсчетов сигнала (x_1, x_2, \dots) имеет периодический характер так что

$$x_{i+kM} = x_i, \quad k=0, 1, \dots \quad (45)$$

Тогда соответствующая трансформанта Фурье (спектр)

$$X(\omega) = \sum_{i=1}^{CM} x_i e^{-j\omega(i-1)}$$

может быть представлена в виде

$$X(\omega) = \sum_{k=1}^C e^{-j\omega M(k-1)} \sum_{i=1}^M x_i e^{-j\omega(i-1)}. \quad (46)$$

Таким образом

$$|X(\omega)|^2 = |X_p(\omega)|^2 \cdot \sin^2(CM\omega/2) / \sin^2(M\omega/2), \quad (47)$$

где $X_p(\omega)$ – внутрипериодный спектр сигнала (вторая сумма в (46)).

Легко понять, что первый множитель будет достигать максимального значения в следующих точках оси частот:

$$\omega_m = m2\pi/M, \quad m=1, 2, \dots, \quad (48)$$

причем именно значение $2\pi/M$ ($m=1$) соответствует частоте основного тона.

Однако влияние $|X_p(\omega)|^2$ может проявляться в том, что максимум правой части будет соответствовать другому значению m . Именно это не позволяет методически надежно определять период основного тона по спектру анализируемого отрезка сигнала.

В основе корреляционного метода определения периода основного тона используется характеристика

$$\rho_{\tau, N} = \sum_{i=1}^N x_i x_{i+\tau} / \sqrt{\sum_{i=1}^N x_i^2 \sum_{k=1}^N x_{k+\tau}^2}, \quad (49)$$

которая является оценкой нормированного коэффициента корреляции.

Гипотеза H_0 отвергается при выполнении неравенства

$$\max_{0 \leq \tau \leq L} \rho_{\tau, N} \geq h \in (0.7 \div 1), \quad (50)$$

где L определяется отношением частоты дискретизации к минимально возможной частоте основного тона.

Тогда в качестве оценки периода основного тона принимается

$$M = \arg \max_{\tau, N} \rho_{\tau, N}, \quad 0 \leq \tau \leq L. \quad (51)$$

Иными словами, максимальное значение характеристики (49) должно превышать некоторый порог, что принимается за признак почти периодического поведения отрезков сигнала.

Одним из недостатков такого подхода является присутствие искажающих шумов, что маскирует наличие периодичности в сигнале.

Кроме того, концентрация спектра $|X(\omega)|^2$ вблизи частоты, не совпадающей с $2\pi/M$, приводит к тому, что максимальное значение (49) будет достигаться при меньшем, чем длина интервала между возбуждающими гортань воздействиями.

На рисунке 3 представлены фрагмент РС, порожденного звуком «и», и значения нормированных коэффициентов корреляции при различных значения смещения.

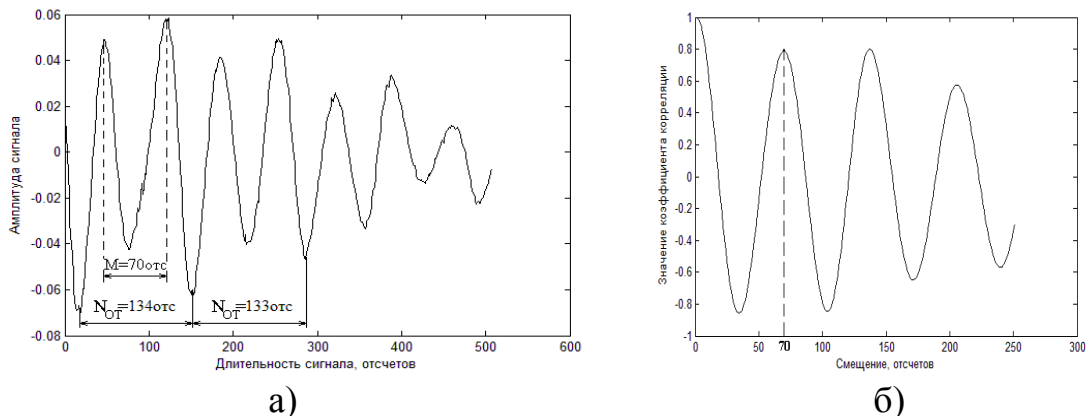


Рисунок 3 – а) Фрагмент РС, порожденного звуком «и» ($f_d=16\text{кГц}$); б) значения нормированных коэффициентов корреляции ($N=256, f_d=16\text{кГц}$).

На рисунке 3а N_{OT} – «истинный» период основного тона РС, а M – его оценка при использовании корреляционного метода. На рисунке 3б пунктиром отмечен максимум характеристики (49). Видно, что максимальное значение характеристики (49) достигается при значении в 2 раза меньшем, чем величина периода основного тона.

Таким образом, необходимо использовать иные методы определения периода основного тона, устойчивые как к воздействию шумов, так и к влиянию периодичности сигнала между двумя последовательными возбуждающими гортань воздействиями.

Представляется естественным ориентироваться на поиск наименьшей частоты из набора (48).

Для этого в работе введено понятие нормированной субполосной корреляции (НСК)

$$\varphi_{\tau}^n = \Phi_{\tau}^n / \sqrt{P_n(\bar{x}_1)P_n(\bar{x}_{\tau})}, \quad (52)$$

где

$$\Phi_{\tau}^n = \int_{\omega \in V_n} X_1(\omega) X_{\tau}^*(\omega) d\omega / 2\pi, \quad (53)$$

$$X_1(\omega) = \sum_{i=1}^N x_i e^{-j\omega(i-1)}, \quad X_{\tau}(\omega) = \sum_{i=1}^N x_{i+\tau} e^{-j\omega(i-1)}, \quad (54)$$

$$\vec{x}_1 = (x_1, x_2, \dots, x_N)^T, \quad \vec{x}_{\tau} = (x_{1+\tau}, x_{2+\tau}, \dots, x_{N+\tau})^T, \quad (55)$$

$$V_n = [-V_{2n}, -V_{1n}) \cup [V_{1n}, V_{2n}), \quad (56)$$

$$P_n(\vec{y}) = \int_{\omega \in V_n} |Y(\omega)|^2 d\omega / 2\pi, \quad (57)$$

$$Y(\omega) = \sum_{i=1}^N y_i e^{-j\omega(i-1)}. \quad (58)$$

Наличие в (52) знаменателя позволяет обеспечить чувствительность к частотным полосам с малым уровнем энергии.

Подставляя в соотношение (52) и (53) определения (57) и (58), нетрудно получить представления для НСК непосредственно во временной области

$$\varphi_{\tau}^n = \vec{x}_1^T A_n \vec{x}_{\tau} / \sqrt{\vec{x}_1^T A_n \vec{x}_1 \vec{x}_{\tau}^T A_n \vec{x}_{\tau}}, \quad (59)$$

где $A_n = \{a_{ik}^n\}$ – субполосные матрицы (14).

Очевидно, что

$$\varphi_{\tau}^n \leq 1, \quad (60)$$

причем правая часть достигается только при выполнении условия пропорциональности:

$$\vec{x}_{\tau} = c \vec{x}_1. \quad (61)$$

Таким образом, определение периода основного тона сводится к вычислению характеристик (52) при разных значениях τ и ω_n .

Если условие

$$\max_{0 \leq \tau \leq L} \varphi_{\tau}^n \geq h_{\varphi} \quad (62)$$

для некоторого частотного интервала выполняется, то принимается решение, что анализируемый отрезок РС относится к вокализованному звуку. Рекомендуется использовать частотный интервал $n=1$.

Для установления порога h_{φ} проведены вычислительные эксперименты, которые показали, что наименьшая вероятность ошибок в селекции вокализованных звуков русской речи обеспечивается при величине порогового значения

$$h_{\varphi} = 0,8. \quad (63)$$

Использование z-преобразования Фишера вида

$$z_{\tau}^n = 1 / 2 \ln \left((1 + \varphi_{\tau}^n) / (1 - \varphi_{\tau}^n) \right), \quad (64)$$

позволяет повысить чувствительность НСК в области значений φ_{τ}^n близких к максимальному, что способствует более точному оцениванию периода ОТ. При этом рекомендуется использовать соотношение

$$z_{\tau} = \sum_{n=0}^{N/2-1} k_n \cdot z_{\tau}^n, \quad (65)$$

где $k_n = Pd_n(\bar{x}_1)$ – весовой коэффициент, представляющий собой долю энергии (26).

Тогда в качестве оценки периода основного тона принимается значение

$$M = \arg \max z_\tau, \quad 0 \leq \tau \leq L, \quad (66)$$

если выполнено условие

$$z_M \geq h_z = 1,1. \quad (67)$$

Указанное значение h_z получено при подстановки (63) в (64) с учетом свойства

$$\sum_{n=0}^{N/2-1} k_n = 1.$$

Вычисление значения числителя НСК вида (59) ускоряется за счет использования представления вида

$$\Phi_\tau^n(\bar{x}_1, \bar{x}_\tau) = \sum_{k=1}^{J/2} \lambda_k^r (\alpha_{2kr}(\bar{x}_1) \alpha_{2kr}(\bar{x}_\tau) + \alpha_{(2k-1)r}(\bar{x}_1) \alpha_{(2k-1)r}(\bar{x}_\tau)),$$

где $\alpha_{kr}(\bar{x}_1)$, $\alpha_{kr}(\bar{x}_\tau)$ - скалярные произведения (проекции) вида (17).

Как было отмечено ранее, выбор соответствующих длительностей анализируемых отрезков и конфигурации частотных интервалов «с перекрытием» позволяет воспользоваться алгоритмом БПФ с компонентами вида (18).

Разработаны алгоритмы селекции вокализованных звуков русской речи и определения периода основного тона, описание которых представлено в диссертации.

На рисунке 4 представлен фрагмент РС, порожденного звуком «и» и результат оценки на основе соотношений (65), (66) для этого фрагмента.

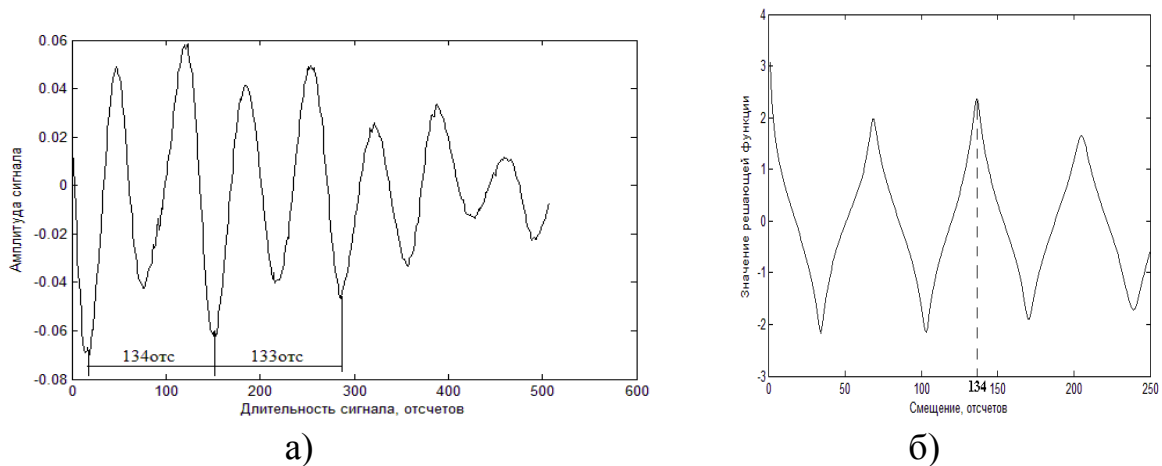


Рисунок 4 – а) Фрагмент РС, порожденного звуком «и» ($f_d=16\text{кГц}$); б) значения РФ НСК ($N=256, f_d=16\text{кГц}$).

Из рисунка 4б видно, что максимум РФ достигается при сдвиге равном периоду основного тона, в отличие от корреляционного метода оценивания периода основного тона (рис.3б).

В работе представлены результаты исследования поведения НСК для различных звуков русской речи. Для оценки эффективности предлагаемой

решающей процедуры были оценены вероятности ошибок первого и второго рода:

$$P_{1ош} = 1 - N_{авт.невокал} / N_{невокал}, \quad (68)$$

$$P_{2ош} = 1 - N_{авт.вокал} / N_{вокал}, \quad (69)$$

а также погрешность оценки периода основного тона:

$$\sigma^{zv} = \sqrt{\sum_{i=1}^{N_{zv}} (M_i - N_{OT})^2 / \sum_{i=1}^{N_{zv}} N_{OT}^2}. \quad (70)$$

Здесь $N_{авт.невокал}$ – количество правильно определенных невокализованных отрезков; $N_{авт.вокал}$ – количество правильно определенных вокализованных отрезков; $N_{вокал}$ – количество отрезков, соответствующих вокализованным звукам речи, использованных для анализа (66000 отрезков); $N_{невокал}$ – количество отрезков, соответствующих невокализованным звукам речи, использованных для анализа (45000 отрезков); N_{zv} – количество отрезков анализа одного звука; M_i – значение периода основного тона, определенное автоматически; N_{OT} – значение периода основного тона, определенное в ручную.

В работе представлены методики проведения экспериментов и результаты сравнения эффективности разработанного алгоритма с методом, основанным на корреляционном анализе.

В таблицах 6-7 представлены некоторые результаты этих сравнений для реальных РС при длительности окна анализа 16мс (256 отсчетов при $f_d=16кГц$).

Таблица 6 – Значения вероятностей ошибок первого и второго рода при селекции вокализованных участков сигналов ($f_d=16кГц$, $N=256$ (16мс))

	НСК	на основе корреляции между отрезками
$P_{1ош}$	0,107	0,118
$P_{2ош}$	0,130	0,132

Таблица 7 – Значения погрешности определения периода основного тона ($f_d=16кГц$, $N=256$ (16мс))

тип звука	НСК	на основе корреляции между отрезками
«ё»	0,026	0,041
«и»	0,018	0,025
«у»	0,018	0,081
«ю»	0,076	0,088
«я»	0,001	0,002
по всем звукам	0,024	0,032

Результаты проведенных экспериментов показывают, что использование НСК позволяет осуществлять селекцию вокализованных звуков русской речи с вероятностью ошибки не больше, чем существующие алгоритмы, при этом погрешность оценки периода основного тона для вокализованных звуков речи меньше, чем у существующих, иногда в разы.

Раздел 3.3. Результаты и выводы главы.

Глава 4. Программная реализации алгоритмов сегментации речевых сигналов. В работе представлено описание структуры и интерфейса разработанной программной поддержки алгоритмов сегментации РС. Также приведены рекомендации по выбору параметров: частоты дискретизации (f_d), длительности анализируемого отрезка (N), шага смещения на различных этапах алгоритмов (Δt , τ), значение доли энергии (m), желаемые значения вероятностей ошибок первого рода (α), пороговые значения ($h_\alpha(\theta)$, h_z).

Функциональная схема алгоритма сегментации РС и оценки периода основного тона представлена на рисунке 5.

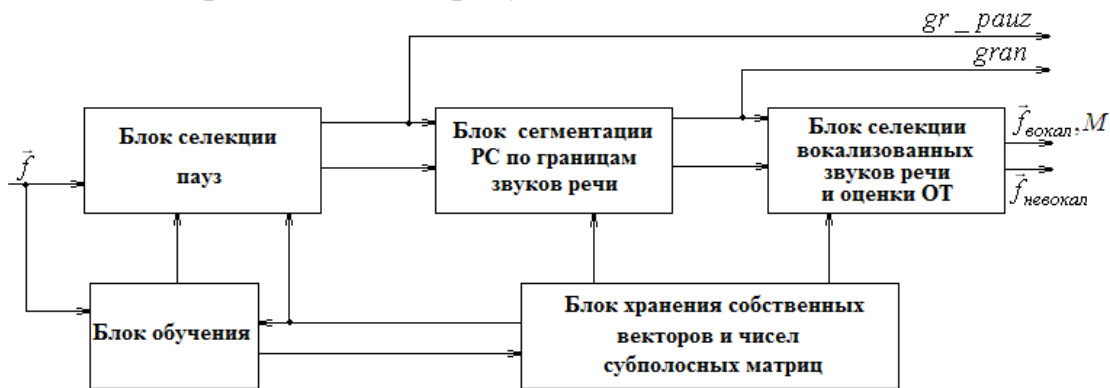


Рисунок 5 – Функциональная схема алгоритма сегментации РС.

В блоке обучения оцениваются средние значения распределений долей энергий, средние значения энергий отрезков сигналов, соответствующих шуму в паузе, а также определяется пороговое значение для решающего правила алгоритма селекции пауз. В блоке селекции пауз оцениваются значения РФ и определяются границы участков пауз. В блоке определения границ звуков русской речи рассчитываются субполосные РФ и определяются границы звуков. На основе полученных данных в блоке селекции вокализованных отрезков сигналов принимается решение о принадлежности отрезков участкам вокализованных или невокализованных звуков речи, а также оцениваются мгновенные значения периода основного тона вокализованных звуков речи. Выходными данными этого блока являются отрезки невокализованных участков РС, отрезки вокализованных участков РС с указанием мгновенных значений периода основного тона, а также вектор с указанием границ участков звуков русской речи.

Анализ одного отрезка РС длительностью 16мс (при $f_d=16кГц$) с целью селекции пауз и определения границ в программной среде Matlab занимает 2мс. Оценка периода основного тона одного отрезка анализа занимает 60мс. Таким образом, анализ одного отрезка РС длительностью 16мс без учета предварительного обучения занимает порядка 62мс.

В заключении сформулированы основные результаты и выводы диссертационной работы.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЫВОДЫ РАБОТЫ

1. Построены модели частотного распределения энергий порождаемых звуками русской устной речи отрезков РС с учетом их длительности, которые используются при создании методов и алгоритмов сегментации РС.

2. Разработаны субполосные методы и алгоритмы сегментации РС на отрезки со значимыми различиями в распределениях энергий по частотным интервалам, позволяющие усовершенствовать методы и алгоритмы компьютерного анализа русской устной речи, включая:

2.1. метод и алгоритм селекции пауз с обучением на основе введенной субполосной решающей функции;

2.2. метод и алгоритм определения границ отрезков РС, порождаемых разными звуками русской речи, на основе субполосной решающей функции;

2.3. метод и алгоритм селекции вокализованных участков русской речи на основе введенного субполосного нормированного коэффициента корреляции.

3. Разработаны метод и алгоритм оценивания периода основного тона РС, порождаемых вокализованными звуками русской речи, на основе использования введенного субполосного нормированного коэффициента корреляции.

4. Разработан прототип программной поддержки созданных алгоритмов сегментации РС по границам звуков русской речи.

5. Результаты проведенных вычислительных экспериментов иллюстрируют обоснованность основных выводов и преимущества созданных методов и алгоритмов.

6. Вычислительные эксперименты показали, что алгоритм селекции пауз с обучением на основе введенной субполосной решающей функции обеспечивает вероятности ошибочного принятия решения не более 0,08; алгоритм определения границ отрезков РС, порождаемых разными звуками русской речи, на основе субполосной решающей функции обеспечивает вероятности ошибочного принятия решения в границах $0,15 \div 0,2$; алгоритм селекции вокализованных участков русской речи, на основе введенного субполосного нормированного коэффициента корреляции обеспечивает вероятности ошибочного принятия решения не больше 0,12 при обеспечении погрешности оценивания периода основного тона не более 0,02.

7. Из полученных результатов исследования следует вывод о том, что разработанные методы и алгоритмы позволяют усовершенствовать процедуры автоматического компьютерного анализа русской устной речи.

В шести Приложениях приведены результаты некоторых вычислительных экспериментов и свидетельства о регистрации Программ для ЭВМ.

ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи в журналах из списка ВАК

1. Жилияков, Е.Г. Об эффективности различных подходов к сегментации речевых сигналов на основе обнаружения пауз [Текст] / Е.Г. Жилияков, С.П. Белов, А.С. Белов, А.А. Фирсова, А.В. Глушак // Научные ведомости Белгородского государственного университета. Сер. История. Политология. Экономика. Информатика. – 2010. – №7(78) выпуск 14/1. – С. 187-193.

2. Фирсова, А.А. Компьютерное моделирование алгоритмов обнаружения пауз в IP-телефонии [Текст] / Н.А. Чеканов, А.А. Фирсова // Вопросы радиоэлектроники. Сер. Электронная вычислительная техника. – 2011. – С. 72-82.

3. Жилияков, Е.Г. Сегментация речевых сигналов на основе анализа распределения энергии по частотным интервалам [Текст] / Е.Г. Жилияков, Е.И. Прохоренко, А.В. Болдышев, А.А. Фирсова, М.В. Фатова // Научные ведомости Белгородского государственного университета. Сер. История. Политология. Экономика. Информатика. – 2011. – № 7 (102) выпуск 18/1. – С. 187-196.

4. Фирсова, А.А. О различии концентрации энергии по частотным диапазонам на отрезках сигналов, соответствующих шипящим звукам русской речи и шумам [Текст] / А.С. Белов, А.В. Курлов, А.А. Фирсова // Научные ведомости Белгородского государственного университета. Сер. История. Политология. Экономика. Информатика. – 2011. – № 13 (108) выпуск 19/1. – С. 186-191.

5. Жилияков, Е.Г. Компьютерная сегментация речевых сигналов [Текст] / Е.Г. Жилияков, Е.И. Прохоренко, А.А. Фирсова, Д.К. Секисов // Вопросы радиоэлектроники. Сер. Электронная вычислительная техника. – 2012. – выпуск 1. – С. 48-54.

6. Жилияков, Е.Г. Алгоритмы обнаружения основного тона речевых сигналов [Текст] / Е.Г. Жилияков, А.А. Фирсова, Н.А. Чеканов // Научные ведомости Белгородского государственного университета. Сер. История. Политология. Экономика. Информатика. – 2012. – № 1 (120) выпуск 21/1. – С. 135-143.

7. Фирсова, А.А. Исследование решающей функции максимальной чувствительности к изменению частот энергии в частотных интервалах [Текст] / С.П. Белов, А.А. Фирсова // Научные ведомости Белгородского государственного университета. Сер. История. Политология. Экономика. Информатика. – 2012. – № 13 (132) выпуск 23/1. – С. 227-231.

8. Жилияков, Е.Г. Об одном алгоритме кодирования пауз в речевых данных [Текст] / Е.Г. Жилияков, Е.И. Прохоренко, А.В. Болдышев, А.А. Фирсова // Вопросы радиоэлектроники. Сер. Электронная вычислительная техника. – 2013. – выпуск 1. – С. 17-25.

Публикации в научных журналах и сборниках трудов научных конференций

9. Жилияков, Е.Г. Сегментация речевых сигналов на основе анализа особенностей распределения долей энергии по частотным интервалам [Текст] / Е.Г. Жилияков, Е.И. Прохоренко, А.В. Болдышев, А.А. Фирсова, М.В. Фатова // Вестник национального технического университета «ХПИ». – 2011. – 17'2011, Тематический выпуск «Информатика и моделирование», №16. – С. 44-50.

10. Фирсова, А.А. О различиях распределения энергии звуков русской речи и шума [Текст] / А.В. Болдышев, А.А. Фирсова // Материалы 12-ой Международной конференции и выставке «Цифровая обработка сигналов и её применение - DSPA'2010». – Москва. – 2010. – С. 204-207.

11. Жилияков, Е.Г. Модели распределения энергии звуков русской речи на основе частотных представлений [Текст] / Е.Г. Жилияков, А.В. Болдышев, А.А. Фирсова // Материалы XXIII Международной научной конференции «Математические методы в технике и технологиях – ММТТ-23». – Саратов. – 2010. – С. 236-239.

12. Фирсова, А.А. О возможности использования особенностей распределения энергии по частотным диапазонам в задачах сегментации речевых сигналов [Текст] / А.А. Фирсова, М.В. Фатова // Материалы 13-ой Международной конференции и выставки «Цифровая обработка сигналов и её применение - DSPA'2011». Москва. – 2011. – С. 255-258.

13. Жилияков, Е.Г. Обзор существующих методов сегментации речевых сигналов [Текст] / Е.Г. Жилияков, С.П. Белов, А.А. Фирсова // Тезисы II Международной научно-практической конференции «Информационные технологии и компьютерная инженерия». – Харьков. – 2011. – С. 129-130.

14. Жилияков, Е.Г. Сегментация речевых сигналов на основе анализа особенностей распределения долей энергии по частотным интервалам [Текст] / Е.Г. Жилияков, Е.И. Прохоренко, А.В. Болдышев, А.А. Фирсова, М.В. Фатова // Тезисы XIX Международной научно-практической конференции «Информационные технологии: наука, техника, технология, образование, здоровье». – Харьков. – 2011. – С. 28.

15. Жилияков, Е.Г. Обзор алгоритмов обнаружения основного тона речевых сигналов [Текст] /Е.Г. Жилияков, С.П. Белов, А.А. Фирсова // Системы обработки информации. 2012. – Выпуск 4(102), Т.1 Информационные технологии и защита информации. – С. 193-194.

16. Жилияков, Е.Г. Исследование алгоритмов определения частоты основного тона речевых сигналов [Текст] / Е.Г. Жилияков, А.А. Фирсова // Сб. трудов Международной молодежной конференции «Прикладная математика, управление и информатика». – Белгород. – 2012. – Т.1. – С. 371-375.

Программы для ЭВМ

17. Свидетельство о государственной регистрации для программы для ЭВМ №2012660160 «Программная система сегментации речевых сигналов на основе анализа распределения энергии по частотным интервалам», авторы: Жилияков Е.Г., Прохоренко Е.И., Болдышев А.В., Фирсова А.А., Секисов Д.К. 12 ноября 2012 г.

18. Свидетельство о государственной регистрации для программы для ЭВМ №2012660685 «Программная система определения частоты основного тона речевых сигналов», авторы: Жилияков Е.Г., Прохоренко Е.И., Болдышев А.В., Фирсова А.А., Курлов А.В. 12 ноября 2012 г.

Подписано в печать 13.05.2013. Гарнитура Times New Roman.

Формат 60×84/16. Усл. п. л. 1,0. Тираж 100 экз. Заказ 195.

Оригинал-макет подготовлен и тиражирован в ИД «Белгород» НИУ «БелГУ»
308015, г. Белгород, ул. Победы, 85

