

На правах рукописи



**Трубицын Владимир Геннадьевич**

**МОДЕЛИ И АЛГОРИТМЫ В СИСТЕМАХ АНАЛИЗА  
РЕЧЕВЫХ СИГНАЛОВ**

по специальности 05.13.17 – Теоретические основы информатики

**АВТОРЕФЕРАТ**

диссертации на соискание ученой степени  
кандидата технических наук

Белгород – 2013

Работа выполнена в Федеральном государственном бюджетном образовательном учреждении высшего профессионального образования «Государственный университет – учебно-научно-производственный комплекс», на кафедре «Электроника, вычислительная техника и информационная безопасность».

Научный руководитель:

**Ерёменко Владимир Тарасович,**  
доктор технических наук, профессор

Официальные оппоненты:

**Кузичкин Олег Рудольфович,**  
доктор технических наук, профессор,  
Муромский институт (филиал) Владимирского государственного университета им. А.Г и Н.Г Столетовых, профессор кафедры «Управление и контроль в технических системах», г. Муром

**Прохоренко Екатерина Ивановна,**  
кандидат технических наук, доцент,  
ФГАОУ ВПО «Белгородский государственный национальный исследовательский университет» доцент кафедры «Информационно-телекоммуникационных систем и технологий» г. Белгород

Ведущая организация:

Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Брянский государственный технический университет», кафедра «Компьютерные технологии и системы»

Защита состоится «22» мая 2013 года в 16 часов 30 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 «12» апреля 2013 г.

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



С.П. Белов

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность работы:** В настоящее время наблюдается значительный рост исследований и разработок в области анализа, кодирования и синтеза речи. Одной из основных задач, стоящих перед речевыми технологиями является задача компактного цифрового представления устной речи для её хранения, обработки и передачи.

Интерес к системам анализа речевых данных при кодировании обусловлен повышением требований к качеству речевого сигнала в системах различного назначения:

1. В криптографических системах государственного, военного и корпоративного назначения.
2. В авиационных системах связи для реализации возможности определения психофизиологического состояния летчика по голосу.
3. В системах защиты информации по каналам диапазонов сверх длинных и длинных волн при организации связи типа "электронной речевой почты".
4. В системах хранения речи для автоматизированных, электронных, речевых справочных служб.
5. В мультимедиа ориентированных системах коммуникации.
6. В системах речевого управления.

Существующие модели речеобразования не в полной мере соответствуют их адекватному представлению в речевых данных, что ограничивает применение речевых технологий. Это связано с возникающим противоречием: с одной стороны длительность временного участка речевых отсчётов для анализа должна быть малой, чтобы последовательность таких участков точно отражала динамику изменений в речевом сигнале, в частности появление коротких пауз и участков переходных звуков. С другой стороны длительность анализируемого участка должна быть достаточно большой, чтобы обеспечить точность оценки характеристик речевого сигнала.

Данное противоречие приводит к проблемам человеко-машинного взаимодействия и искусственного интеллекта возникающим при автоматическом распознавании и понимании речи, синтезе и воспроизведении речи из данных, записи и хранении больших объёмов речевых данных отражающих человеческие знания в различных областях, верификации и идентификация диктора.

Разрешение противоречия возможно за счёт принятия соотношения локально-стационарного характера модели речеобразования с квазистационарным характером участков речевого сигнала, соответствующих режиму установившихся звуков.

Стратегия определения роли анализа субъективных факторов в деле создания речевых интерактивных систем имеет три аспекта:

1. Разработка методов для определения областей применения технологии автоматического распознавания, кодирования и генерации речи.
2. Выбор подходящих алгоритмов распознавания, кодирования, генерации речи и характеристик соответствующих систем.
3. Включение речевых подсистем в контекст задач пользователя.

Исходя из анализа литературы и изделий, представленных на тематическом рынке программного обеспечения или компьютерного оборудования, следует, что ни одна из подобных систем не лишена принципиальных недостатков. Иными словами, необходимого уровня эффективного представления речевых данных с точки зрения занимаемого объёма и качества воспроизведения существующим изделиям пока достичь не удалось. Такое положение вещей, по-видимому, связано не с частными недостатками конкретных систем, а с необходимостью изменения представлений о закономерностях речевого процесса.

В основе настоящего исследования лежат результаты работ в области: **теории и методов параметрического кодирования речи** (Д. Фланаган, Г. Фант, М.А. Сапож-

ков, А.А. Пирогов); **теории кодирования на основе линейного предсказания** ( Л.Р. Рабинера, Р.В. Шафера, Дж.Д. Маркела, Б.С. Атола, Д. Раймда, Ф. Итакура, М.Р. Шредера), **методов сегментации речевого сигнала** (П.Мермельштайн, Дж. Ван Хемерт, Б. Пеллом, В.Н. Сорокин, Е.Г. Жилияков). Развитие и использование цифровых процессоров обработки сигналов и труды таких ученых как Б.С. Атол, Д. Серено, Л. Селарио, А. Гершо, Е. Пайксоу, Д.Г. Рове, Т. Ванг **привели к усложнению алгоритмов, появлению новых моделей, развитию методов анализа речевых сигналов через синтез.**

В указанных трудах имеются фундаментальные научные результаты для дальнейшего развития речевых технологий. Между тем, существующие подходы к решению задач исследования моделей и алгоритмов для анализа речевых данных и обнаружение в них закономерностей, а так же разработка и исследование методов и алгоритмов представления устной речи носят локальный по областям применений и разрозненный по методам характер. Это делает актуальной тему диссертационных исследований и обуславливает выбор объекта, предмета и цели исследования.

**Объект исследования:** Процесс анализа речевых сигналов для систем представления данных устной речи.

**Предмет исследования:** Алгоритмы, модели, методы анализа речевых сигналов для систем представления данных устной речи.

**Целью работы является** Совершенствование методов анализа устной речи за счёт установления закономерностей в речевых сигналах, при сегментации вокализованных участков для эффективного представления речевых данных.

Для достижения цели были сформулированы и решены частные задачи:

1. Анализ статистических, временных, спектральных, корреляционных характеристик и параметров речевых сигналов фонетических групп звуков русской речи.
2. Исследование существующих методов, моделей и алгоритмов цифрового преобразования речи в системах кодирования речевых сигналов.
3. Разработка методов и алгоритмов формирования участков квазистационарности в реализациях речевого сигнала, соответствующих вокализованным звукам.
4. Разработка метода уменьшения объёма речевых данных за счёт изменения частоты выборки речевых отсчётов в сформированных границах участков квазистационарности речевого сигнала.
5. Оценка эффективности предложенных методов в системах анализа речевых сигналов для представления устной речи.

**Методы исследований** базируются на использовании математического аппарата теории вероятностей и математической статистики, теории корреляционного и спектрального анализа, цифровой обработки сигналов, методов регрессионного анализа и методов моделирования систем, а так же исследование речевых процессов с использованием средств и методов имитационного моделирования.

**Научная новизна:**

1. Математическая модель анализа речевых сигналов, базирующаяся на методе корреляционного анализа, позволяющая представлять речевые данные в виде временных рядов интервалов автокорреляции смежных сегментов речевых отсчётов фиксированной длины.
2. Методика формирования участков речевых отсчётов, порождаемых вокализованными звуками, по принятым решениям на основе аппроксимации временных рядов интервалов автокорреляции смежных сегментов речевых отсчётов линейными моделями тренда, позволяющая определять границы таких участков для эффективного представления речевых отсчётов в данных.
3. Алгоритм передискретизации за счёт изменения частоты выборки речевых отсчётов на выделяемых участках с помощью разработанной методики, позволяющий

сформировать необходимое и достаточное количество речевых отсчётов для их эффективного представления в речевых данных.

**Практическая значимость:**

Разработанные теоретические положения доведены до патентов на изобретения:

1. Патент на изобретение № 2248619 от 20.03.2005 г “Способ и устройство преобразования речевого сигнала методом линейного предсказания с адаптивным распределением информационных ресурсов”.
2. Патент на изобретение № 2400831 от 21.09.2010 “Способ выделения сегментов квазистационарности при анализе речевого сигнала в вокодерах с линейным предсказанием”.
3. Патент на изобретение № 2432624 от 11.05.2011 “Способ уменьшения объема данных при широкополосном кодировании речевого сигнала”.
4. Патент на изобретение № 2445718 от 20.03.2012 “Способ выделения сегмента обработки речи на основе анализа корреляционных зависимостей в речевом сигнале”.

**Реализация результатов исследования осуществлена:** в ФГБОУ ВПО «Государственный университет – УНПК» на кафедре «Электроника, вычислительная техника и информационная безопасность» при подготовке специалистов по направлениям 210201 «Проектирование и технология радиоэлектронных средств», 210202 «Проектирование и технология электронно-вычислительных средств», и 090900 «Информационная безопасность».

Внедрения подтверждаются соответствующими документами.

**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, обнаружение закономерностей в данных и их извлечения, разработка и исследование методов и алгоритмов анализа текста, устной речи и изображений”;

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

**Положения, выносимые на защиту:**

1. Математическая модель анализа речевых данных в системах представления устной речи, базирующаяся на методе корреляционного анализа.
2. Методика формирования участков речевых отсчётов, порождаемых вокализованными звуками, базирующаяся на оценке соотношения между снижением остаточной дисперсии и потерей числа степеней свободы при переходе от единой линейной регрессионной модели тренда временного ряда к кусочно-линейной регрессионной модели.
3. Алгоритм передискретизации речевых отсчётов за счёт изменения частоты выборки речевых отсчётов на выделяемых участках с помощью разработанной методики, базирующийся на использовании процедур децимации и интерполяции.

**Достоверность выводов и рекомендаций** обусловлена корректностью применяемых преобразований с использованием теории вероятностей, математической статистики, непротиворечивостью и воспроизводимостью результатов, полученных теоретическим путем и в ходе выполнения экспериментов, проведением оценки адекватности разработанных моделей и устойчивости алгоритмов.

**Апробация результатов диссертационного исследования.**

Результаты диссертационного исследования обсуждались на следующих научно-технических конференциях: IX Международная научная конференция. Сборник трудов. – М.: Академия управления МВД России, 2000. – С. 372 –377; 5-я международная конференция “Цифровая обработка сигналов и ее применение”. Доклады. Том

2. – М: Радиотехника, 2003. – С. 405 – 407; Proceedings of International Workshop "Speech and computer" SPECOM'2003. – М.: MSLU, 2003. – P. 168 – 169;

**Публикации.** По теме диссертационного исследования опубликовано 14 печатных работ (из них 6 в журналах из списка ВАК РФ), в том числе четыре патента на изобретение зарегистрированных в Государственном реестре изобретений РФ.

**Объём и структура работы.** Диссертация состоит из Введения, пяти глав, Заключения и Приложения. Работа изложена на 134 страницах машинного текста, включая 19 рисунков, 10 таблиц и списка литературных источников из 109 наименований.

### СОДЕРЖАНИЕ РАБОТЫ

**Во Введении** обосновывается актуальность диссертационного исследования и формулируется его основная цель.

**Глава 1. Анализ процесса речеобразования, характеристик, свойств и моделей речевого сигнала. Обзор существующих методов и алгоритмов сегментации речевого сигнала для его представления. Постановка задачи исследования.**

В главе представлен анализ общепринятой модели речеобразования, временных, спектральных и корреляционных характеристик, свойств и параметров речевых сигналов. Дается обзор особенностей формантного состава и длительностей фонем русской речи и существующих методов сегментации речи.

Проведён анализ недостатков спектральных и энергетических методов сегментации речевых сигналов. Выявлено, что использование динамических и статических свойств речевого сигнала позволяет эффективно производить анализ речевого сигнала для его сегментации. Установлено, что дальнейшее развитие методов сегментации речевого сигнала возможно за счёт классифицирования финитных участков речи на квазистационарные и переходные. В работах современных исследователей отмечен тот факт, что основой такого классифицирования является учёт автокорреляции и распознавания кардинальных изменений в речевых отсчётах.

Сделан вывод о необходимости проведения дальнейшего исследования проблем поиска границ сегментов на речевом сигнале за счёт распознавания типов сегментов в интересах решения практических задач.

Поставлена научная задача совершенствования существующих методов сегментации речевых сигналов за счёт выявления квазистационарных свойств речевых сигналов в системах анализа для эффективного представления речевых данных.

**Глава 2. Математическая модель анализа речевых данных в системах представления устной речи.**

Установлено, что речь при её представлении в виде речевых сигналов обладает значительной статистической и психофизической избыточностью. Сокращение статистической избыточности базируется на учёте свойств и закономерностей речевого сигнала и определяется наличием корреляционной связи между соседними речевыми отсчётами временных реализаций при их дискретизации.

Математическая модель анализа речевых данных представлена системой отображений 1. Отличительной особенностью модели является представление устной речи речевыми отсчётами на выделяемых участках переменной длительности и отвечающие требованиям квазистационарности, для которых сохраняются параметры модели голосового тракта и модели сигнала возбуждения.

$\{S(t)\} \xrightarrow{P_0} \{\vec{S}(m)\}$  - отображение пространства реализаций непрерывного речевого сигнала  $S(t)$  в пространство анализируемых речевых отсчетов  $\vec{S}(m)$ . Для отображения  $P_0$  реализуется АЦП непрерывного речевого сигнала и описывается рядом Котельникова В.А., уравнениями линейного квантования и неэффективного кодирования.

$$\left\{ \begin{array}{l}
\{\bar{S}(t)\} \xrightarrow{\Pi_0} \{\bar{S}(m)\}, \text{ где } m = 0,1,2,\dots,M, M < \infty; \\
\{\bar{S}(m)\} \xrightarrow{\Pi_1} \{\bar{S}^l(i)\}, \text{ где } i = 0,1,2,\dots,I, I < \infty; l = 0,1,2,\dots,L = \frac{M}{I}; \\
\{\bar{S}^l(i)\} \xrightarrow{\Pi_2} \{\bar{S}_k^l(i)\}, \text{ где } k = 0,1,2,\dots,N_{\text{сезм}} = \frac{2/3 I}{R}, \text{ где } R = 24; \\
\{\bar{S}_k^l(i)\} \xrightarrow{\Pi_3} \{\bar{B}_k^l(j)\}, \text{ где } j = 0,1,2,\dots,\frac{I}{3}-1; \\
\{\bar{B}_k^l(j)\} \xrightarrow{\Pi_4} \{\tau_{k\text{кор}}^l\}; \\
\{\tau_{k\text{кор}}^l\} \xrightarrow{\Pi_5} \{\bar{\tau}_{\text{кор}}^l(k)\}; \\
\{\bar{\tau}_{\text{кор}}^l(k)\} \xrightarrow{\Pi_6} \{\hat{k}^l\}; \\
\{\hat{k}^l\} \xrightarrow{\Pi_7} \{\hat{k}_{\text{кд}}^l\}; \\
\{\hat{k}_{\text{кд}}^l\} \xrightarrow{\Pi_8} \{S_{\text{гр}}^l\}; \\
\{S_{\text{гр}}^l\} \xrightarrow{\Pi_9} \{\bar{S}_{\text{выд уч}}^\gamma\}, \text{ где } \gamma = 0,1,2,\dots,\gamma', \gamma' < \infty
\end{array} \right. \quad (1)$$

$\{\bar{S}(m)\} \xrightarrow{\Pi_1} \{\bar{S}^l(i)\}$  - отображение пространства речевых отсчётов на финитном интервале, соответствующем длительности  $t_\phi = IT_\phi$  мс с частотой дискретизации  $f_\phi$  из пространства речевых отсчётов активного участка речи. Отображение  $\Pi_1$  представляет собой формирование из  $M$  речевых отсчётов всего активного участка  $L$  совокупностей по  $I$  речевых отсчёта, соответствующих одинаковым финитным интервалам. Количество речевых отсчётов  $I$  выбираются исходя представленных в литературе результатов исследований продолжительности участков вокализованной и невокализованной речи. Рядом исследователей установлено, что для вокализованных участков распределение количества сегментов речевых отсчётов по 20 мс носит экспоненциальный характер. При этом, с вероятностью 0,9 вокализованный участок будет состоять из трёх следующих друг за другом сегментов. Исходя из этих соображений длительность финитного участка для анализа предлагается в 72 мс, что при  $f_\phi = 8 \text{ кГц}$  соответствует 576 речевым отсчётам.

$\{\bar{S}^l(i)\} \xrightarrow{\Pi_2} \{\bar{S}_k^l(i)\}$  - отображение пространства отсчетов речевого сигнала  $\{\bar{S}^l(i)\}$  в пространство речевых отсчётов для смежных сегментов фиксированной длины, сдвинутых относительно друг друга на  $R$  отсчётов  $\{\bar{S}_k^l(i)\}$ . Отображение  $\Pi_2$  реализуется сдвигом последовательности в  $\frac{I}{3}$  речевых отсчёта из последовательности финитного участка на кратные  $R$  отсчета по мере поступления данных. Количество сегментов  $k = \frac{2/3 I}{R}$ . Количество речевых отсчётов  $R$  зависит от средней скорости изменения временной огибающей речевых отсчётов на вокализованных участках. Исходя из этих соображений длительность сдвига предлагается использовать в 3 мс, что при  $f_\phi = 8 \text{ кГц}$  соответствует 24-м речевым отсчётам.

$\{\bar{S}_k^l(i)\} \xrightarrow{\Pi_3} \{\bar{B}_k^l(j)\}$  - отображение пространства речевых отсчётов смежных сегментов на финитном участке речевого сигнала  $\{\bar{S}_k^l(i)\}$  в пространство автокорреляционных функций  $\{\bar{B}_k^l(j)\}$ .

$$\Pi_3 : \bar{B}_k^l(j) = \sum_{i=0}^{\frac{l}{3}-1} S_k^l(i+k \cdot R) \cdot S_k^l(i+k \cdot R+j), \quad (2)$$

$\{\bar{B}_k^l(j)\} \xrightarrow{\Pi_4} \{\tau_{k \text{ кор}}^l\}$  - отображение пространства автокорреляционных функций смежных сегментов речевых отсчётов в множество значений интервалов автокорреляции этих сегментов  $\{\tau_{k \text{ кор}}^l\}$ . Отображение  $\Pi_4$  реализуется определением интервалов автокорреляции для каждого сегмента подсчётом количества дискретных отсчетов автокорреляционной функции, для которых значение автокорреляционной функции будет меньше или равно  $0,05$  её максимального значения.

$$\Pi_4 : B_k^l(\mathcal{F}_k) \geq 0,05 B_k^l(j=0); \quad \tau_{k \text{ кор}}^l = (\mathcal{F}_k), \quad j=0,1,2,\dots,J = \frac{l}{3}-1; \quad (3)$$

$\{\tau_{k \text{ кор}}^l\} \xrightarrow{\Pi_5} \{\bar{\tau}_{k \text{ кор}}^l(k)\}$  - отображение множества значений интервалов автокорреляции смежных сегментов  $\{\tau_{k \text{ кор}}^l\}$  во временной ряд  $\{\bar{\tau}_{k \text{ кор}}^l(k)\}$  для финитного участка.

$\{\bar{\tau}_{k \text{ кор}}^l(k)\} \xrightarrow{\Pi_6} \{\mathcal{K}\}$  - отображение пространства временных рядов значений интервалов автокорреляции  $\{\bar{\tau}_{k \text{ кор}}^l(k)\}$  в множество индексов сегментов со структурными изменениями во временных рядах  $\{\mathcal{K}\}$  финитных участков. Для отображения  $\Pi_6$  используется формальный статистический тест (тест Чоу), предполагающий расчёт параметров уравнений трендов (парной линейной регрессии), описывающих динамику поведения речевых отсчётов в реализациях финитных участков.

$$\Pi_6 : F_{\text{набл}} > \text{или} < F_{\text{кр}}; \quad F_{\text{набл}} = \frac{D_{\Delta C_{\text{ост}}}}{D_{\text{кл}}} = \frac{\Delta C_{\text{ост}} : (p_1 + p_2 - p_3)}{C_{\text{ост}}^{\text{кл}} : (n - p_1 - p_2)}. \quad (4)$$

где  $C_{\text{ост}}^{\text{кл}} = C_{\text{ост}}^1 + C_{\text{ост}}^2$  - остаточная сумма квадратов кусочно-линейной модели,  $F_{\text{набл}}$  - наблюдаемое значение критической статистики,  $p_1 = p_2 = p_3$  - число параметров первого, второго и третьего уравнения регрессии,  $n$  - число наблюдений.

$$C_{\text{ост}}^{\text{кл}} = C_{\text{ост}}^1 + C_{\text{ост}}^2; \quad (5)$$

$$C_{\text{общ}}^1 = \sum_{i=1}^{n1} (\tau_{\text{кор}}^i - \bar{\tau}_{\text{кор}}^1)^2, \quad C_{\text{ост}}^1 = (1-r) \cdot C_{\text{общ}}^1; \quad (6)$$

$$C_{\text{общ}}^2 = \sum_{i=1}^{n2} (\tau_{\text{кор}}^i - \bar{\tau}_{\text{кор}}^2)^2, \quad C_{\text{ост}}^2 = (1-r) \cdot C_{\text{общ}}^2. \quad (7)$$

где  $C_{\text{общ}}^1$  - общая сумма квадратов разности первого уравнения регрессии,  $C_{\text{общ}}^2$  - общая сумма квадратов разности второго уравнения регрессии,  $\tau_{\text{кор}}^i$  -  $i$ -е значение интервала автокорреляции временного ряда,  $\bar{\tau}_{\text{кор}}^1$  - среднее значение интервалов автокорреляции первого уравнения регрессии,  $\bar{\tau}_{\text{кор}}^2$  - среднее значение интервалов автокорреляции второго уравнения регрессии,  $n1$  - количество значений интервалов автокорреляции описываемых первым уравнением регрессии,  $n2$  - количество значений интервалов автокорреляции описываемых вторым уравнением регрессии,  $r$  - коэффициент детерминации изменяется от 0 до 1, поскольку рассчитывается, как отношение суммы квадратов объяснённой регрессией к общей сумме квадратов.

$\Delta C_{\text{ост}} = C_{\text{ост}}^3 - C_{\text{ост}}^{\text{кл}}$  - сокращение остаточной дисперсии при замене единого уравнения регрессии кусочно-линейной моделью.

$$\Delta C_{ост} = C_{ост}^3 - C_{ост}^{кл} . \quad (8)$$

где  $C_{ост}^3$  – остаточная сумма квадратов разности единого уравнения регрессии.

$$C_{общ}^3 = \sum_{i=1}^n (\tau_{кор}^i - \bar{\tau}_{кор})^2 , C_{ост}^3 = (1-r) \cdot C_{общ}^3 . \quad (9)$$

где  $C_{общ}^3$  – общая сумма квадратов разности единого уравнения регрессии,  $\tau_{кор}^i$  –  $i$ -е значение интервала автокорреляции единого уравнения временного ряда,  $\bar{\tau}_{кор}$  – среднее значение интервалов автокорреляции единого уравнения регрессии,  $n$  – количество значений интервалов автокорреляции единого уравнения регрессии,  $r$  – коэффициент детерминации.

$\{\mathcal{K}\} \xrightarrow{\Pi_7} \{\mathcal{K}_{код}^l\}$  – отображение множества индексов сегментов со структурными изменениями во временных рядах  $\{\mathcal{K}\}$  во множество граничных речевых отсчётов, соответствующих структурным изменениям на финитном участке  $\{\mathcal{K}_{код}^l\}$ .

Отображение  $\Pi_7$  определяет номер отсчёта в сегменте, для которого сделан вывод о структурных изменениях. Номер отсчёта формируется суммой числа речевых отсчётов начального сегмента финитного участка и отсчётов, определяемых произведением индекса сегмента со структурными изменениями на число отсчётов сдвига.

$$\Pi_7 : \mathcal{K}_{код} = I/3 + \mathcal{K} \cdot R . \quad (10)$$

$\{\mathcal{K}_{код}^l\} \xrightarrow{\Pi_8} \{S_{зр}^l\}$  – отображение множества граничных речевых отсчётов, соответствующих структурным изменениям в пределах финитного участка  $\{\mathcal{K}_{код}^l\}$  в множество граничных отсчётов на активных участках речи  $\{S_{зр}^l\}$ .

$$\Pi_8 : S_{зр}^l = S^l(0) + \mathcal{K}_{код} . \quad (11)$$

где  $S^l(0)$  – начальный речевой отсчёт на  $l$  - ом финитном участке.

$\{S_{зр}^l\} \xrightarrow{\Pi_9} \{S_{выд\ уч}^\gamma\}$  – отображение множества граничных отсчётов на активном участке речи  $\{S_{зр}^l\}$  в пространство выделяемых участков квазистационарности  $\{S_{выд\ уч}^\gamma\}$ .

$$\Pi_9 : S_{выд\ уч}^\gamma = S_{зр}^{l+1} - S_{зр}^l . \quad (12)$$

где  $\gamma = 0,1,2,\dots,\gamma'$ ;  $\gamma' < \infty$ .

Предлагаемая математическая модель позволяет выделять речевые отсчёты квазистационарных участков переменной длины большей продолжительности с целью повышения эффективности их представления.

### Глава 3. Методика формирования участков речевых отсчётов, порождаемых вокализованными звуками.

В диссертации временной ряд интервалов автокорреляции аппроксимируется парной линейной регрессионной моделью, описывающей отношение между двумя группами переменных. В качестве входных (независимых) переменных, определяющих условия функционирования системы кодирования, принимаются индексы интервалов автокорреляции  $k$ .

Выходными (зависимыми) переменными, характеризующие результат корреляционного анализа, являются значения интервалов автокорреляции  $\tau_{кор}(k)$ . Для определения аналитического вида зависимости между  $\tau_{кор}(k)$  и  $k$  использовался парный линейный регрессионный анализ.

$$\tau_{кор}^{мод}(k) = b_0 + b_1 \cdot k. \quad (13)$$

Для формирования уравнения регрессии применялся метод наименьших квадратов, позволяющий найти координаты  $(b_0, b_1)$  минимизирующие целевую функцию.

$$\min F(b_0, b_1) = \sum_{i=1}^{\frac{2/3 I}{R}} \varepsilon_i^2 = \sum_{i=1}^{\frac{2/3 I}{R}} (\tau_{кор_i}^{мод} - \tau_{кор_i}^{набл})^2 \quad (14)$$

$$b_0 = \frac{\sum_{i=1}^{\frac{2/3 I}{R}} k_i^2 \cdot \sum_{i=1}^{\frac{2/3 I}{R}} \tau_{кор_i}^{набл} - \sum_{i=1}^{\frac{2/3 I}{R}} k_i \cdot \sum_{i=1}^{\frac{2/3 I}{R}} k_i \cdot \tau_{кор_i}^{набл}}{\frac{2/3 I}{R} \cdot (\sum_{i=1}^{\frac{2/3 I}{R}} k_i^2) - (\sum_{i=1}^{\frac{2/3 I}{R}} k_i)^2} \quad (15)$$

$$b_1 = \frac{\frac{2/3 I}{R} \cdot (\sum_{i=1}^{\frac{2/3 I}{R}} k_i \cdot \tau_{кор_i}^{набл}) - \sum_{i=1}^{\frac{2/3 I}{R}} k_i \cdot \sum_{i=1}^{\frac{2/3 I}{R}} \tau_{кор_i}^{набл}}{\frac{2/3 I}{R} \cdot (\sum_{i=1}^{\frac{2/3 I}{R}} k_i^2) - (\sum_{i=1}^{\frac{2/3 I}{R}} k_i)^2} \quad (16)$$

где  $\tau_{кор_i}^{мод}$  - модельные значения интервалов автокорреляции,  $\tau_{кор_i}^{набл}$  - наблюдаемые значения интервалов автокорреляции формируемых временных рядов,  $k_i$  - индексы интервалов автокорреляции временных рядов.

Для оценки точности полученной регрессионной модели использовался коэффициент детерминации, т.е. рассчитывалась выборочная дисперсия, характеризующая разброс значений зависимой переменной  $\tau_{кор}(k_i)$  относительно общего среднего.

$$r(\tau_{кор}, k) = \frac{\frac{\frac{2/3 I}{24}}{\sum_{i=1}^{\frac{2/3 I}{24}} (\tau_{кор_i}^{мод} - \bar{\tau}_{кор})^2}}{\frac{\frac{2/3 I}{R} - 1}{\sum_{i=1}^{\frac{2/3 I}{R}} (\tau_{кор_i}^{набл} - \bar{\tau}_{кор})^2}} = \frac{S_{объяс}^2(\tau_{кор}^{мод})}{S_{общ}^2(\tau_{кор})} = \frac{\frac{\frac{2/3 I}{R}}{\sum_{i=1}^{\frac{2/3 I}{R}} (\tau_{кор_i}^{мод} - \bar{\tau}_{кор})^2}}{\frac{\frac{2/3 I}{R}}{\sum_{i=1}^{\frac{2/3 I}{R}} (\tau_{кор_i}^{набл} - \bar{\tau}_{кор})^2}} \quad (17)$$

где  $\tau_{кор_i}^{мод}$  - модельные значения интервалов автокорреляции,  $\tau_{кор_i}^{набл}$  - наблюдаемые значения интервалов автокорреляции формируемых временных рядов,  $\bar{\tau}_{кор}$  - общее среднее значение интервалов автокорреляции временных рядов.

Известно, что чем большую долю в общей дисперсии составляет объясненная дисперсия  $S_{объяс}^2(\tau_{кор}^{мод})$ , тем точнее полученная функция регрессии соответствует эмпирическим данным. Начиная с некоторого момента времени  $t^*$ , которому соответст-

вует  $k$ -ый из  $k \in \left\{ 0, 1, 2, \dots, \frac{2/3 I}{R} \right\}$  смежных сегментов, в котором происходят изменения,

обуславливает изменение параметров тренда временного ряда.

В диссертации предложено для определения структурных изменений временного ряда использовать кусочно-линейные регрессионные модели, предполагающие разделение исходной последовательности значений интервалов автокорреляции на две части – до момента времени  $t^*$  и после момента времени  $t^*$ . Для каждой части формируется уравнение регрессии. Если изменения незначительно повлияли на характер тренда временного ряда, то ее описывают единым для всей совокупности уравнением.

В результате исследований установлено, что построение единого для всей последовательности уравнения тренда позволяет сохранить исходное число наблюдений, но остаточная сумма квадратов увеличивается по сравнению с кусочно-линейной моделью. В тоже время, выбор между двумя моделями будет зависеть от соотношения между снижением остаточной дисперсии и потерей числа степеней свободы при переходе от единого уравнения к кусочно-линейной модели.

Для оценки этого соотношения использовался статистический тест (тест Чоу), основанный на расчете параметров уравнений регрессии. При этом, выдвигалась гипотеза  $H_0: C_{ост}^{кл} = C_{ост}^3$  о структурной стабильности тенденции временного ряда (совпадение параметров регрессии единой модели и кусочно-линейной модели). Соответственно гипотеза структурной нестабильности  $H_1: C_{ост}^{кл} < C_{ост}^3$ .

Остаточная сумма квадратов разности по кусочно-линейной модели  $C_{ост}^{кл}$  определяется выражением (5) с соответствующим ей числом степеней свободы  $v_2 = n - p_1 - p_2$ , а остаточная сумма квадратов разности единого уравнения регрессии  $C_{ост}^3$  определяется выражением (9). Снижение остаточной дисперсии  $\Delta C_{ост}$  при замене единого уравнения регрессии кусочно-линейной моделью определяется выражением (8) с соответствующим ей числом степеней свободы  $v_1 = p_1 + p_2 - p_3$ .

Для проверки гипотезы  $H_0$  в качестве критерия предложено использовать наблюдаемое значение  $F$ -критерия по дисперсиям  $\Delta C_{ост}$  и  $C_{ост}^{кл}$  на одну степень свободы вариации, представленное в выражении (4) математической модели.

Найденное значение критерия  $F_{набл}$  сравнивалось с критическим  $F_{кр}$ , полученным по справочным таблицам распределения Фишера для заданного уровня значимости  $\alpha$  (ошибка первого рода) и числа степеней свободы  $v_1 = p_1 + p_2 - p_3$  и  $v_2 = n - p_1 - p_2$ , при этом использовалась правосторонняя критическая область. Если  $F_{набл} > F_{кр}$ , то нулевая гипотеза отклонялась, а влияние структурных изменений на динамику стабильности временного ряда признавалось значимым. Если  $F_{набл} < F_{кр}$ , то нет оснований отвергать нулевую гипотезу.

Применительно к анализу процесса речеобразования в диссертации решалась обратная задача – определения значения интервала автокорреляции, которое соответствует структурным изменениям речи.

Для решения поставленной задачи предложена методика определения индекса структурных изменений значений временного ряда интервалов автокорреляции для совокупности смежных сегментов речевых отсчетов фиксированной длительности:

1. Выбор интервала автокорреляции  $\tau_{кор}^*$ , соответствующий предполагаемым структурным изменениям в речи, на основе разности предыдущего и последующего значений интервалов автокорреляции временного ряда. В случае отсутствия равенства пер-

вых пяти элементов временного ряда, временной ряд финитного участка формировался с учётом пяти последних элементов временного ряда предыдущего участка.

2. Формирование структуры трех моделей регрессии: двух моделей регрессии для кусочно-линейной модели, разбивающих всю последовательность на две части и единой модели тренда для всей совокупности интервалов автокорреляции.

3. Оценивание параметров моделей с использованием МНК.

4. Оценка значимости и адекватности полученных моделей. Для оценки значимости использовался  $F$ -критерий (для оценки значимости всего уравнения регрессии) и  $t$ -критерий (для оценки значимости каждого параметра модели). Оценка адекватности моделей осуществлялась на основании анализа среднего квадрата ошибки и исправленного коэффициента детерминации.

5. Расчет остаточной суммы квадратов разности для кусочно-линейной модели регрессии  $C_{ост}^{кл}$  в соответствии с выражением (5):

6. Расчет значения сокращения остаточной дисперсии при замене единого уравнения регрессии кусочно-линейной моделью (8).

7. Определение значения статистического критерия  $F_{набл}$  по выражению (4).

8. Определение значения статистического критерия  $F_{кр}$  по таблицам распределения Фишера в соответствии с величинами  $\nu_1 = p_1 + p_2 - p_3$ ,  $\nu_2 = n - p_1 - p_2$  и  $\alpha$ .

9. Сравнение полученных значений  $F_{набл}$  и  $F_{кр}$ . Если  $F_{набл} > F_{кр}$ , то в точке  $\tau_{кор}^*$  произошли структурные изменения в речевых отсчётах и определяется граница участка квазистационарности. С зафиксированной границы, характеризуемой значением  $\tau_{кор}^*$ , описанные шаги методики выполняются для нового временного участка речи. Если  $F_{набл} < F_{кр}$ , то структурные изменения отсутствуют, и необходимо анализировать следующее значение интервала корреляции, т. е. вернуться к первому этапу.

#### **Глава 4. Алгоритм передискретизации речевых отсчётов за счёт изменения частоты выборки речевых отсчётов на выделяемых участках.**

Оценки спектральных характеристик речевого сигнала, проведенные Д. Фланганом и Сапожковым А.А. показывают, что для гласных звуков основную роль в восприятии речи играют две первые форманты. Первая форманта определяет интенсивность звука, а вторая важна для слухового ощущения. При этом, соотношение между формантными частотами, а так же форма спектральной огибающей представляет одну из основных характеристик распознавания звуков. Для распознавания гласных звуков “а, о, у” достаточно первой форманты, а для звуков “э, ы, и” требуется две форманты. Исходя из анализа формантного состава, первая и вторая форманты для гласных “а, о, у” сосредоточены до 1200 Гц, для “и” в полосе до 2300 Гц, а для “э, ы” до 2000 Гц. Это позволяет утверждать, что имеется возможность уменьшения количества речевых отсчётов при их представлении, что приводит к уменьшению объёма речевых данных.

В соответствии с математической моделью, известными процедурами многоскоростной цифровой обработки сигналов и рекомендациями исследователей в области речевого кодирования разработан алгоритм передискретизации посредством изменения частоты выборки речевых отсчётов на выделяемых участках.

Алгоритм включает этапы представленные на рисунке 2.

1. Формирование финитного участка речевых отсчётов  $t_\phi = IT_\phi$  сводится к последовательной выборке из всей совокупности речевых отсчётов последовательностей по  $I = 576$  отсчётов, что при  $f_{дискр} = 8 \text{ кГц}$  соответствует временному интервалу в 72 мсек.

2. Осуществляется последовательное выделение из субпоследовательности финитного участка смежных сегментов по  $I/3 = 192$  отсчёта, сдвинутых на  $R = 24$  отсчёта от

носителем друг друга. Речевые отсчёты сегмента в 192 отсчёта соответствует временному отрезку в 24 мсек. Сдвиг в 24 отсчёта смежных сегментов относительно друг друга соответствует временному интервалу в 3 мсек.

3. Производится расчет автокорреляционных функций для смежных сегментов речевых отсчётов фиксированной длительности в пределах финитного участка.

4. Осуществляется классифицирование по признаку активность / пауза смежных сегментов в пределах финитного участка осуществляется путем сравнения значений автокорреляционной функции сегмента с нулевым значением аргумента со значением автокорреляционной функции сегмента паузы, так же с нулевым значением.

5. Реализуется расчёт интервалов автокорреляции для каждого сегмента путем подсчёта количества интервалов дискретизации дискретной автокорреляционной функции до индекса интервала дискретизации, когда значение самой автокорреляционной функции будет больше или равно 0,05 от её максимального значения. На основе проведённых расчётов формируется временной ряд значений интервалов автокорреляции, характеризующий изменения в речевых отсчётах всего финитного участка.

6. Формируется временной ряд интервалов автокорреляции финитного участка с учётом поведения речевых отсчётов на временном промежутке, которому соответствуют  $r$  последних интервалов автокорреляции предыдущего финитного участка.

7. Осуществляется классифицирование финитного участка с учётом сегментов предыдущего финитного участка по признаку активность / пауза проверкой равенства с нулём, суммы всех значений временного ряда интервалов автокорреляции. Если сумма не равна нулю, то финитный участок является активным и анализируется далее. Если сумма равна нулю, участок является паузой и алгоритм возвращается к первому этапу.

8. Фиксируется наличие изменений значений интервалов автокорреляции во временном ряде посредством итерационного расчёта суммы разности между предыдущим и последующим значением интервалов автокорреляции всего временного ряда. Равенство данной суммы нулю показывает отсутствие изменений во временном ряде.

9. Определяется индекс значения интервала автокорреляции, в котором произошли изменения в пределах финитного участка по значению разности между значениями интервалов автокорреляции предыдущего и последующего в пределах этого участка.

10. Проводится анализ изменений в значениях интервалов автокорреляции с найденными индексами (предыдущего и последующего) с точки зрения структурных изменений. Этот анализ осуществляется в соответствии с предложенной методикой в третьей главе. Если в зафиксированных интервалах не произошло структурных изменений, тогда осуществляется возвращение к этапу №9 по определению следующих индексов интервалов автокорреляции, для которых имеет место разность значений предыдущего и последующего сегментов в пределах анализируемого участка.

11. Определяется индекс речевого отсчёта в пределах финитного участка подсчётом суммы  $I/3 = 192$  речевых отсчёта начального сегмента и произведения индекса интервала автокорреляции, в котором зафиксированы структурные изменения на количество речевых отсчётов сдвига  $R = 24$ .

12. Осуществляется определение граничных индексов речевых отсчётов, соответствующих структурным изменениям на активных участках посредством следующего структурного изменения или регистрации паузы.

13. Формируются участки речевых отсчётов переменной длины для различных состояний речевого процесса: вокализованный, невокализованный, переходной с вокализованного и на вокализованный, переходной на невокализованный и с него. Участки формируются управлением выборкой индексов речевых отсчётов участков переменной длины от одного регистрируемого структурного изменения до другого.

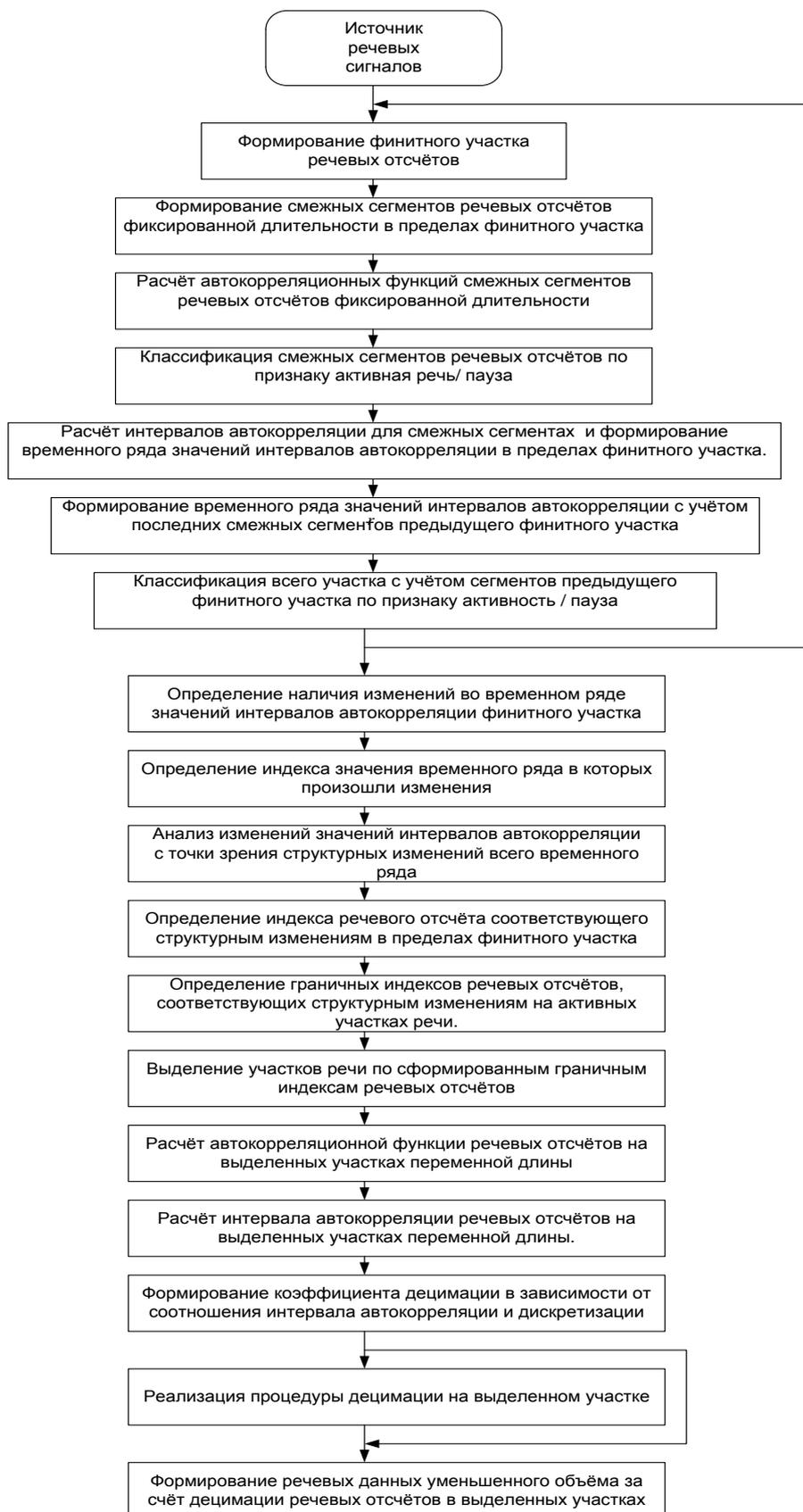


Рисунок 2. Алгоритм передискретизации речевых отсчётов на участках квазистационарности

14. Производится расчёт автокорреляционной функции, определение интервала автокорреляции для выделенных участков переменной длины. Осуществляется вычисление соотношения интервала автокорреляции и дискретизации, что позволяет с помощью устанавливаемых отношений, классифицировать участки, соответствующие во-

кализованным, невокализованным и переходным звукам. Если соотношение интервалов равно единице, тогда участок воспринимается как невокализованный, если больше 10, то это переходной. В случае, когда соотношение интервалов находится в интервале от 2 до 9, участок принимается вокализованным.

15. Осуществляется расчёт коэффициента децимации по результатам предыдущего этапа. Для случаев, когда соотношение интервала автокорреляции и дискретизации равно 1 или больше 10 коэффициент равен 1. В случае, когда участок воспринимается, как вокализованный, коэффициент децимации выбирается равный 2.

16. Процедура децимации реализуется только для тех участков, у которых соотношение интервала автокорреляции и дискретизации изменяется от 2 до 9.

17. Формирование речевых отсчётов осуществляется последовательным объединением участков переменной длины в сформированных границах между существенными изменениями с использованием процедуры децимации и без неё.

В результате передискретизации на выделяемых участках нарушается структура речевых отсчётов, что приводит к пропаданию звука при его воспроизведении. Для восстановления структуры речевых отсчетов необходимо применить процедуру интерполяции к речевым отсчётам на участках с децимацией. Процедуры децимации и интерполяции описываются линейно-разностными уравнениями вида:

$$\begin{cases} S_{\text{выд}}^{\gamma}(m_M) = \sum_{n=0}^{\infty} h(n) S_{\text{выд}}^{\gamma}(Mm_M - n); \\ m = Mm_M \end{cases} \quad (18)$$

где  $m = 0, M, 2M, 3M, \dots$ ,  $m_M = 0, 1, 2, 3, \dots$ ,  $n = 0, 1, 2, 3, \dots$

$$\begin{cases} S_{\text{выд}}^{\gamma}(m_L) = \sum_{n=0, L, 2L, L, \dots}^{\infty} h(m_L - n) S_{\text{выд}}^{\gamma}\left(\frac{n}{L}\right); \\ m_L = Lm_M \end{cases} \quad (19)$$

где  $m_L = 0, L, 2L, 3L, \dots$ ,  $m_M = 0, 1, 2, 3, \dots$ ,  $n = 0, L, 2L, 3L, \dots$

Для синтеза КИХ фильтров с линейной фазо-частотной характеристикой в работе применялся метод с использованием окон. Анализ ряда окон показал, что уменьшение коэффициента пульсаций сопровождается увеличением ширины главного лепестка, которое регулируется изменением длины окна. Проблемой использования таких окон является либо чрезмерная ширина переходной полосы, что снижает избирательность, либо слишком большая длина  $N$ , что приводит к трудностям реализации и возникновению задержки реакции из-за высокого порядка фильтра. Описанное противоречие в значительной степени может быть преодолено компромиссом, который достигается применением окна Кайзера.

Окно Кайзера обладает одним замечательным свойством. Для фиксированного отклонения  $\delta$  амплитудно-частотной характеристики в полосе задерживания произведение порядка фильтра  $N-1$  на нормированную ширину переходной полосы  $\Delta F$  в АЧХ фильтра является практически постоянной величиной. Данное свойство позволяет синтезировать цифровой фильтр, добиваясь компромиссного значения длины фильтра за счёт снижения коэффициента пульсаций и расширения переходной полосы.

В ходе проводимых экспериментов при восстановлении речевых отсчётов после передискретизации для различных значений минимального затухания в полосе задерживания  $a_{\min}$  и нормированных граничной частоты полосы задерживания и частоты среза полосы пропускания  $f_k$ ,  $f_x$  относительно новой частоты дискретизации получены соответствующие значения порядков фильтра. Порядок фильтра возрастает при увели-

чении значений коэффициентов децимации/интерполяции и при уменьшении переходной полосы частот  $\Delta F = f_k - f_z$  в АЧХ фильтра. При этом в многоскоростных системах с увеличением коэффициента децимации/интерполяции на единицу приводит к увеличению порядка фильтра в полтора раза.

В ходе проведённых экспериментов с речевым материалом установлено, что наилучший эффект от использования алгоритма передискретизации речевых отсчётов с точки зрения среднего квадрата разности речевых отсчётов до и после передискретизации достигается при:

$$\Delta F = f_k - f_z = 0.125 ; a_{\min} = 40 \div 60 [dB]; M(L) = 2 \quad (20)$$

При этом порядок фильтра  $N$  изменяется от 19 до 29 соответственно для 40 и 60 дБ. Увеличение коэффициента децимации/интерполяции приводит к увеличению временной задержки реакции фильтра, что отрицательно сказывается на стыках объединения участков подверженных многоскоростной обработке и не подверженных ею.

Таким образом, применение процедур многоскоростной обработки ограничивается не только формантным составом звуков речи, но и возможностями процедур многоскоростной обработки с точки зрения адекватного слухового восприятия после восстановления.

### **Глава 5. Оценка эффективности предложенных: методики формирования участков речевых отсчётов, порождаемых вокализованными звуками и алгоритма передискретизации речевых отсчётов на выделяемых участках.**

В исследовании подтверждено, что повышение эффективности систем представления речевых данных на активных участках зависит от частоты появления вокализованных звуков и их местоположения в словах, а так же от слогового состава русской речи. Вокализованные звуки составляют 68 % общего числа звуков.

В качестве показателя эффективности разработанной методики формирования участков речевых отсчётов предложена разность между границами речевых отсчётов формируемыми оператором и системой, реализующей процедуры методики.

Экспериментальные исследования показали, что величина разности распределена по нормальному закону распределения для гласных а, о, у, ы, э, и, при их расположении в начале, середине и конце слова и взаимодействии с различными классами согласных. Результаты интервальной оценки математического ожидания разности доверительной вероятностью  $\gamma = 0,95$  представлены в таблице 5.1.

Таблица 5.1

	Сонорные, носовые	Сонорные, ротовые	Звонкие, взрывные
средняя разность по выборке (мсек)	$\Delta \bar{t} = 2,96 \text{ мсек}$	$\Delta \bar{t} = 2,63 \text{ мсек}$	$\Delta \bar{t} = 2,35 \text{ мсек}$
доверительный интервал (мсек)	$2,45 < \Delta \bar{t} < 3,47$	$2,08 < \Delta \bar{t} < 3,18$	$1,95 < \Delta \bar{t} < 2,75$
	Звонкие, щелевые	Глухие, взрывные	Глухие, щелевые
средняя разность по выборке (мсек)	$\Delta \bar{t} = 2,36 \text{ мсек}$	$\Delta \bar{t} = 2,31 \text{ мсек}$	$\Delta \bar{t} = 1,67 \text{ мсек}$
доверительный интервал (мсек)	$1,96 < \Delta \bar{t} < 2,75$	$1,75 < \Delta \bar{t} < 2,85$	$1,35 < \Delta \bar{t} < 1,99$

В качестве показателя эффективности алгоритма передискретизации в работе использовался коэффициент уменьшения речевых отсчетов, который определяется отношением числа речевых отсчетов с равномерной дискретизацией к количеству отсчетов полученных с использованием децимации.

$$\Phi = \frac{M_{рав.дискр.}}{M_{передискр.}} \quad (21)$$

При оценке эффективности передискретизации в работе учитывались требования к качеству восстанавливаемой речи. Объективным показателем качества в работе принят квадрат разности между отсчетами речевого сигнала до и после обработки в пределах активного речевого участка, выраженный в процентном отношении к мощности исходной реализации.

$$P_{ucx} = \frac{1}{M} \sum_{m=0}^{M-1} S_{ucx}^2(m) \quad (22)$$

$$P_{обр} = \frac{1}{M} \sum_{m=0}^{M-1} S_{обр}^2(m) \quad (23)$$

$$P_{ном} = \frac{1}{M} \sum_{m=0}^{M-1} [S_{ucx}(m) - S_{обр}(m)]^2 \quad (24)$$

$$P_{ном}(\%) = \frac{P_{ном}}{\frac{P_{ucx}}{100}} \quad (25)$$

Проведенные экспериментальные исследования алгоритма передискретизации, показали, что величина потерь при восстановлении участков, для которых была реализована процедура децимации, распределена по нормальному закону. Для оценки математического ожидания потерь использовались интервальные оценки с доверительной вероятностью  $\gamma = 0,9$  для шести основных гласных звуков: а, о, у, ы, э, и, при их расположении в начале, середине и конце слов во взаимодействии со звонкими и глухими согласными. Результаты оценок представлены в таблице 5.2.

Таблица 5.2

	А	О	У
среднее потерь по выборке (%)	$\bar{P}_{ном} = 5,11 \%$	$\bar{P}_{ном} = 2,84 \%$	$\bar{P}_{ном} = 1,22 \%$
доверительный интервал	$4,09 \% < P_{ном} < 6,13 \%$	$2,25 \% < P_{ном} < 3,4 \%$	$4,09 \% < P_{ном} < 6,13 \%$
	Ы	Э	И
среднее потерь по выборке (%)	$\bar{P}_{ном} = 5,2 \%$	$\bar{P}_{ном} = 10,51 \%$	$\bar{P}_{ном} = 5,7 \%$
доверительный интервал	$4,12 \% < P_{ном} < 6,15 \%$	$8,45 \% < P_{ном} < 12,5 \%$	$4,22\% < P_{ном} < 7,17\%$
среднее потерь по всей выборке (%)	$\bar{P}_{ном}^{\Sigma} = 4,93 \%$	доверительный интервал для всей выборки	$4,28\% < P_{ном}^{\Sigma} < 5,57\%$

Результаты интервальной оценки коэффициента уменьшения речевых отсчетов:

$$\bar{\Phi}_{выборки} = 1,15; \quad 1,12 < \Phi_{выборки} < 1,175.$$

**В Заключение** кратко сформулированы основные результаты, полученные в диссертационной работе.

#### Основные результаты и выводы работы:

1. В исследовании решена новая научная задача обнаружения существенных изменений в речевом сигнале и сегментирования речевых отсчетов во временной облас-

ти, имеющая существенное значение для области разработки моделей и алгоритмов при анализе устной речи.

2. В результате анализа существующих методов представления речи и рекомендаций известных исследователей доказано, что повышение эффективности методов и алгоритмов преобразования речи может быть достигнуто за счет учета корреляционных свойств речевого сигнала в рамках локально-стационарной модели речеобразования. Это позволяет представить задачу эффективного представления, как задачу формирования необходимого объема речевых данных на участках переменной длины, соответствующих вокализованным звукам.

3. Определено, что автокорреляционные функции смежных сегментов речевых отсчетов на финитном участке речевого сигнала соответствуют состояниям вокализованного, переходного с вокализованного, переходного к невокализованному и от невокализованного звука. Установлено, что динамику изменения речевых отсчетов соответствующих состояний можно проследить по изменениям значений временного ряда интервалов автокорреляции смежных сегментов финитного участка.

4. Установлено, что для состояний речевого сигнала соотношение значения интервала автокорреляции к интервалу дискретизации имеет существенные различия. Для состояния вокализованного звука соотношение изменяется от 2 до 9, для невокализованного равно 1, а для финитных участков, классифицируемых как переходные, принимает значения десятков и сотен единиц.

5. Выявлено, что качественная оценка динамики изменений речевых отсчетов имеющая линейное нарастание значений интервалов автокорреляции во временном ряде соответствует переходу реализации речевого сигнала в низкочастотную часть спектра, а уменьшение в высокочастотную область.

6. Установлено, что существенные изменения во временных рядах, соответствующих финитным участкам речевых отсчетов, соответствуют изменениям в динамике речевого сигнала. Для определения границ участка квазистационарности необходима оценка структурной стабильности временного ряда интервалов автокорреляции.

7. Предложена методика формирования участков речевых отсчетов, порождаемых вокализованными звуками, основанная на изменениях в регрессионной модели тренда временного ряда, вызванных структурными изменениями в процессе речеобразования.

8. Выявлено, что изменение значений интервалов автокорреляции влечет за собой изменение в структуре временного ряда. Для определения структурных изменений временного ряда предложено оценивать снижение остаточной суммы квадратов при переходе от общего уравнения тренда для всей последовательности к кусочно-линейной модели временного ряда.

9. В результате анализа проведенных экспериментов установлено, что математическое ожидание погрешности положения границ участков квазистационарности для вокализованных гласных изменяется в пределах от 1,67 миллисекунд при их взаимодействии с глухими щелевыми согласными до 2,96 миллисекунд с сонорными согласными. Доверительные интервалы погрешности положения границ составляют 0,6 миллисекунды для глухих щелевых и 1,1 миллисекунды для сонорных.

10. Проведенные эксперименты, с помощью разработанного алгоритма передискретизации речевых отсчетов позволяют утверждать, что определение временных и спектральных границ временного отрезка речевых данных устраняет несоответствие между двумя противоречивыми требованиями к речевому сигналу. С одной стороны сегменты речевого сигнала должны быть короткими, чтобы их последовательность точно отражала динамику изменений речевого сигнала (в частности появление коротких пауз и переходных звуков). С другой стороны отрезки речевого сигнала для обра-

ботки должны быть продолжительными, что позволяет на выявленных участках квазистационарности представлять речевой сигнал меньшим объемом данных.

11. Доказано, что при соответствующей частоте дискретизации уменьшение объёма речевых данных может быть реализовано процедурой децимации, посредством адаптивного изменения частоты выборки речевых отсчётов для участков, соответствующих условиям квазистационарности. Адаптивное изменение частоты выборки осуществляется выбором коэффициента децимации в соответствии с рассчитанными соотношениями между интервалами автокорреляции и дискретизации.

12. Выявлено, что ограничением на значение коэффициента децимации/интерполяции при передискретизации на выделяемых участках квазистационарности является временная задержка формирования реакции фильтра при восстановлении. Для систем, в которых общая последовательность речевых отсчётов формируется из совокупностей выделяемых участков временные задержки, возникающие в результате линейных преобразований таких участков должны быть незначительными. Порядок цифрового фильтра, используемого при осуществлении процедур многоскоростной обработки сигналов должен оставаться низким. Установлено, что обеспечение основного требования при использовании процедур децимации/интерполяции достигается при значениях порядка фильтра равного 19 и 29 при обеспечении затухания в полосе задерживания 40 и 60 [дБ] соответственно.

13. Результаты определения границ участков квазистационарности вокализованных звуков являются развитием методов сегментации речи за счёт применения нового способа определения существенных изменений в речевом сигнале без применения спектральных преобразований с погрешностями, вызванными усечением ряда Фурье.

14. Анализ результатов применения алгоритма передискретизации речевых отсчётов на формируемых участках вокализованных гласных позволяет утверждать, что только при значениях коэффициентов децимации/интерполяции  $M(L) = 2$  обеспечиваются требования по качеству воспроизведения.

15. Повышение эффективности представления речевых отсчётов лежит в области параметрического представления речи в выделяемых участках квазистационарности переменной длительности.

## **2. ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ**

### *Статьи в журналах из списка ВАК*

1. Трубицын. В.Г. Стартовый набор TMS320VC5416 DSK. Состав, основные характеристики и возможности. Организация памяти [Текст]. / А.А.Рыболовлев, В.Г. Трубицын // Научно–технический журнал // Цифровая обработка сигналов. – 2003. – № 2. – С. 44 – 49.

2. Трубицын. В.Г. Технология eXpressDSP проектирования и отладки систем цифровой обработки сигналов при работе с TMS320VC5416 DSK. Интегрированная среда разработки Code Composer Studio v.2 для DSK [Текст]. / А.А. Рыболовлев, В.Г. Трубицын // Научно–технический журнал // Цифровая обработка сигналов. – 2003. – № 4 – С. 24 – 32.

3. Трубицын. В.Г. Цифровое кодирование речевого сигнала методом линейного предсказания с адаптивным распределением информационных ресурсов [Текст]. / А.А. Рыболовлев, Г.В. Богачёв, В.Г. Трубицын // ООО “Наука и Технологии”, журнал “Телекоммуникации”. – 2005. - № 9. – С. 15 – 19.

4. Трубицын. В.Г. Выделение сегментов квазистационарности при анализе речевого сигнала в вокодерах с линейным предсказанием [Текст]. / Научно – технический журнал // Фундаментальные и прикладные проблемы техники и технологии. – март – апрель 2010 г. С. 97 – 104.

5. Трубицын. В.Г. Уменьшение объема данных при широкополосном кодировании речевого сигнала [Текст]. / А.А. Афанасьев, В.Г. Трубицын. // Научно – технический журнал // Цифровая обработка сигналов. – 2011. – № 1 – С. 53 – 56.

6. Трубицын. В.Г. Способ выделения участков квазистационарности для кодирования речи на основе анализа автокорреляционных зависимостей в речевом сигнале [Текст]. / ООО “Наука и Технологии” // Научно – технический журнал “Телекоммуникации”. – 2012. - № 4 – С. 20 – 27

#### ***Статьи в научных журналах и сборниках трудов***

7. Трубицын. В.Г. Оценка потенциально достижимого коэффициента сжатия речевого сигнала липредером с переменной скоростью кодирования [Текст]. / А.А. Рыболовлев, В.Г. Трубицын // Защита информации в сетях и системах связи. Тезисы докладов. – Пенза: ПНИЭИ, 2000. – С. 24;

8. Трубицын. В.Г. Перспективы разработки и применения в современных информационно-телекоммуникационных системах липредеров с адаптивным распределением информационных ресурсов [Текст]. / А.А. Рыболовлев, В.Г. Трубицын // Информатизация правоохранительных систем. IX Международная научная конференция. Сборник трудов. – М.: Академия управления МВД России, 2000. – С. 372 –377;

9. Трубицын. В.Г. Математическая модель структурно адаптивного липредера [Текст]. / А.А. Рыболовлев, В.Г. Трубицын // 5-я международная конференция "Цифровая обработка сигналов и ее применение". Доклады. Том 2. – М: Радиотехника, 2003. – С. 405 – 407;

10. Трубицын. В.Г. The statistical characteristics of a speech signal parameters used in codecs, based on the linear prediction method [Текст]. / А.А. Рыболовлев, В.Г. Трубицын // Proceedings of International Workshop "Speech and computer" SPECOM'2003. – М.: MSLU, 2003. – P. 168 – 169;

#### ***Патенты на изобретения***

11. Трубицын. В.Г. Патент на изобретение "Способ и устройство преобразования речевого сигнала методом линейного предсказания с адаптивным распределением информационных ресурсов" / Рыболовлев А.А., Богачев Г. В., Азаров И. А. // № 2248619. Зарегистрировано в Государственном реестре изобретений Российской Федерации 20 марта 2005 г;

12. Трубицын. В.Г. Патент на изобретение "Способ выделения сегментов квазистационарности при анализе речевого сигнала в вокодерах с линейным предсказанием" / А.А. Афанасьев // №2400831 Зарегистрировано в Государственном реестре изобретений Российской Федерации 27 сентября 2010 г;

13. Трубицын. В.Г. Патент на изобретение "Способ уменьшения объема данных при широкополосном кодировании речевого сигнала / А.А. Афанасьев // №2432624 Зарегистрировано в Государственном реестре изобретений Российской Федерации 27 октября 2011 г;

14. Трубицын. В.Г. Патент на изобретение "Способ выделения сегментов обработки речи на основе анализа корреляционных зависимостей в речевом сигнале / Е.И. Новиков // №2445718 Зарегистрировано в Государственном реестре изобретений Российской Федерации 20 марта 2012 г;

ЛР ИД № 00670 от 05.01.2000 г.

Подписано к печати « 15 » марта 2013г.

Усл. печ. л.1,5 Тираж 100 экз.

Заказ № 15.

---

Полиграфический отдел ГУ – УНПК  
302005, г. Орел, ул. Московская, 65