

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ  
ДОНЕЦКОЙ НАРОДНОЙ РЕСПУБЛИКИ  
ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ  
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ  
«ДОНЕЦКИЙ НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ»**

*На правах рукописи*

**Третьяков Игорь Александрович**

**АВТОМАТИЗАЦИЯ ПРОЦЕДУРЫ СТРУКТУРНОГО АНАЛИЗА  
МАССИВОВ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ НАУЧНЫХ  
ИССЛЕДОВАНИЙ**

Специальность 05.13.06 – Автоматизация и управление технологическими процессами и производствами (по отраслям) (технические науки)

**АВТОРЕФЕРАТ**  
диссертации на соискание ученой степени  
кандидата технических наук

Донецк – 2020

Работа выполнена в ГОУ ВПО «ДОННУ» Министерства образования и науки Донецкой Народной Республики (г. Донецк)

Научный руководитель: доктор технических наук, профессор  
**Данилов Владимир Васильевич**  
ГОУ ВПО «ДОННУ» (г. Донецк), заведующий кафедрой радиофизики и инфокоммуникационных технологий

Официальные оппоненты: доктор технических наук, доцент  
**Лобков Николай Иванович**  
Республиканский академический научно-исследовательский и проектно-конструкторский институт горной геологии, геомеханики, геофизики и маркшейдерского дела (РАНИМИ) (г. Донецк), заведующий отделом управления геомеханическими и технологическими процессами

кандидат технических наук  
**Волуева Ольга Сергеевна**  
ГОУВПО «ДОННТУ» (г. Донецк), доцент кафедры автоматизации и телекоммуникаций

Ведущая организация: **Государственная образовательная организация высшего профессионального образования «Донецкий институт железнодорожного транспорта»** (г. Донецк)

Защита состоится «22» декабря 2020 г. в 12:00 часов на заседании диссертационного совета Д 01.024.04 при ГОУВПО «ДОННТУ» и ГОУ ВПО «ДОННУ» по адресу: 283001, г. Донецк, ул. Артема, 58, корп. 1, ауд.203. Тел./факс: +38(062) 304 30 55, e-mail: uchensovets@donntu.org

С диссертацией можно ознакомиться в библиотеке ГОУВПО «ДОННТУ» по адресу: 283001, г. Донецк, ул. Артема, 58, корпус 2. Адрес сайта университета: <http://donntu.org>

Автореферат разослан «    » ноября 2020 г.

Ученый секретарь  
диссертационного совета Д 01.024.04  
кандидат технических наук, доцент



Т.В. Завадская

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность темы исследования.** На современном этапе развития всех сторон деятельности человека решающую роль играют процессы получения, хранения, обработки и представления информации. В частности, в области развития научных исследований первостепенное значение приобретают методы и вычислительные алгоритмы анализа, фильтрации, преобразования и классификации экспериментальных данных как средства обоснования принимаемых решений и выводов.

Значительные объемы научной информации представляются в виде экспериментальных кривых. Данные этого типа широко используются в автоматизации управления технологическими процессами в промышленности, так как являются одним из способов представления результатов в автоматизированных системах научных исследований. Таким образом представляют, например, хроматограммы в анализе физико-химических свойств веществ, электро-фонокардиограммы и электроэнцефалограммы в медицине, спектры колебаний молекул в спектроскопии.

В связи с тем, что в современном мире постоянно возрастает сложность технологических процессов и сложность новых научных теорий, результаты научных исследований в виде массивов экспериментальных данных содержат десятки и сотни тысяч компонентов. Такие массивы экспериментальных данных не содержат в явном виде информации о свойствах исследуемого процесса, а наиболее существенные свойства и характеристики исследуемого процесса оказываются недоступными для непосредственного измерения. Поэтому возникает необходимость в разработке специальных вычислительных алгоритмов и эффективных методов анализа, аппроксимации, классификации и построения точного сжатого описания экспериментальных данных. Учитывая вышесказанное, автоматизация процессов анализа массивов экспериментальных данных, их классификации и представления в виде сжатого описания является актуальной научно-технической задачей, имеющей отраслевое значение.

**Степень разработанности темы исследования.** Анализ литературных источников и результатов исследований, полученных рядом авторов, обеспечил наличие достаточно обширного статистического материала, что, в свою очередь, обеспечило высокую аргументированность полученных научных результатов проведенного исследования и подтвердило актуальность выбранной темы. Из анализа литературных источников следует, что в настоящее время четко определены два крупных класса методов анализа экспериментальных кривых: интегральный и структурный. В интегральном подходе сжатое описание кривой строится без предварительного выделения какого-либо её сегмента либо разделения ее на однородные фрагменты. Недостатками такого подхода являются: необходимость задания класса экспериментальных кривых, что на практике зачастую невозможно, особенно для слабо изученных процессов; необходимость достаточного уровня знаний об анализируемом процессе;

сильная корреляция переменных при построении моделей, возникающая из-за плохой обусловленности системы нормальных уравнений. Методы структурного подхода представляют сжатое описание экспериментальной кривой в два основных этапа, а именно: разделение кривой на однородные участки и построение сжатого описания кривой в целом. Применение методов структурного класса в действительности является приемлемым в большинстве случаев. Таким образом для экспериментальных кривых, являющихся неоднородными на всей области определения, возможно получить разделение на такие интервалы, на каждом из которых кривая оказывается более простой, что позволяет использовать для представления сжатого описания кривой достаточно типичные локальные модели. Полученные описания участков таких кривых несут важные данные о функционировании анализируемого процесса, например, о режимах работы исследуемого объекта. В рамках структурного подхода к анализу экспериментальных кривых разработано много различных методов, однако их практическое использование часто является неэффективным, а в некоторых случаях и невозможным, что определяет необходимость дальнейших разработок.

**Цели и задачи исследования.** Целью работы является обоснование модернизированных методов и алгоритмов автоматизации процессов анализа массивов экспериментальных данных, их классификации и представления в виде компактных структур. Для достижения данной цели сформулированы и решены следующие задачи:

1. Разработка комбинированных вычислительных методов сегментации экспериментальных кривых.
2. Алгоритмическая реализация процедуры структурного анализа экспериментальных данных.
3. Разработка методов лингвистического описания экспериментальных кривых.
4. Исследование применимости разработанной системы структурного анализа в процессах электромагнитной совместимости радиоэлектронных средств.

**Объект исследования.** Объектом исследования являются массивы данных экспериментальных исследований.

**Предмет исследования.** Предметом исследования являются вычислительные методы автоматизации структурного анализа массивов данных.

**Научная новизна полученных результатов** состоит в следующем:

1. Дальнейшее развитие получил метод структурного анализа данных, в рамках лингвистического подхода к анализу экспериментальных кривых, в котором анализируемая кривая описывается в виде сжатого описания из последовательного ряда символов либо целых слов из определенного алфавита. Каждый элемент такого ряда представляет соответствующий участок либо группу участков, определенный сегментацией анализируемой экспериментальной кривой.

2. Впервые предложен вычислительный метод сегментации массивов экспериментальных данных с использованием функций сложности, отличающийся способностью осуществлять бинарную классификацию помимо сегментации, а также наличием дополнительных условий во избежание выделения ложных экстремумов.

3. Впервые предложен метод лингвистического описания участков экспериментальных кривых на основе сравнения с эталонами, в котором через конечное число циклов достигается устойчивая классификация и ни один вектор не переносится из одного класса в другой, и отличающийся способностью классификации по признаку минимума расстояния до эталона.

4. Впервые применены вычислительные методы сегментации и лингвистического описания для автоматизации процедуры структурного анализа экспериментальных данных к исследованию спектрограмм радиосигналов. При этом обоснована процедура составления более расширенного лингвистического описания экспериментальных кривых, позволяющая составлять это описание с учетом местоположения участков кривой на оси абсцисс.

**Теоретическая значимость** диссертационной работы состоит в установлении закономерностей и теоретическом обосновании структурных методов анализа данных и принимаемых решений в области автоматизации процессов анализа массивов экспериментальных данных, их классификации и представления.

**Практическая значимость** результатов диссертационной работы заключается в классификации экспериментальных данных и представлении их в виде сжатого описания с использованием вычислительных алгоритмов сегментации экспериментальных кривых на основе функции сложности, а также вычислительных алгоритмов лингвистического описания участков экспериментальных кривых для автоматизации процедуры структурного анализа массивов экспериментальных данных научных исследований при решении практических задач. Практическая ценность работы подтверждается:

а) внедрением вычислительного алгоритма сегментации массивов экспериментальных данных с использованием функций сложности и вычислительного алгоритма лингвистического описания участков экспериментальных кривых на основе сравнения с эталонами в научно-исследовательский процесс Государственного учреждения «Донецкий физико-технический институт им. А.А. Галкина»;

б) использованием вычислительных алгоритмов сегментации экспериментальных кривых на основе функции сложности и вычислительных алгоритмов лингвистического описания участков экспериментальных кривых для автоматизации процедуры структурного анализа массивов экспериментальных данных научных исследований при выполнении научно-исследовательских работ кафедры радиофизики и инфокоммуникационных технологий ГОУ ВПО «Донецкий национальный университет» в 2018-2019 гг. (справка №5516/01-27/6.20 от 12.12.2019 г.);

в) внедрением методики компьютерной обработки для выделения и анализа экспериментальных данных, методики сегментации и анализа экспериментальных данных и вычислительных алгоритмы сегментации на основе функции сложности для выделения информативных участков экспериментальных кривых в учебный процесс ГОУ ВПО «Донецкий национальный университет» путем использования при чтении лекций и проведении лабораторных занятий по дисциплине «Цифровая обработка сигналов» для подготовки бакалавров по направлениям подготовки 03.03.03 Радиофизика и 10.03.01 Информационная безопасность (справка №5515/01-27/6.20 от 12.12.2019 г.).

**Методология и методы исследования.** Для решения поставленных задач и получения результатов применялись современные научные методы исследования: систематизация, обобщение и статистический анализ экспериментальных данных, методы анализа данных, методы сравнения с эталонами, методы лингвистического описания, а также практическая апробация результатов исследований на экспериментальных данных, полученных на лабораторных макетах.

**Положения, выносимые на защиту:**

1. Показано, что применение структурного подхода целесообразно для анализа массивов экспериментальных данных, а применение разработанных вычислительных методов сегментации с использованием функций сложности и лингвистического описания участков экспериментальных кривых на основе сравнения с эталонами позволяют осуществлять бинарную классификацию во избежание выделения ложных экстремумов и устойчивую классификацию по признаку минимума расстояния до эталона.

2. Установлено, что разработанные вычислительные методы и алгоритмы позволяют автоматизировать процессы структурного анализа экспериментальных данных научных исследований, осуществлять их классификацию и представление в виде сжатого описания. При этом обоснована процедура составления расширенного лингвистического описания экспериментальных кривых, позволяющая составлять это описание с учетом местоположения участков кривой на оси абсцисс.

**Степень достоверности и апробация результатов.** Обоснованность и достоверность научных результатов исследования подтверждается значительным количеством наблюдений, современными методами исследования, которые соответствуют поставленным в работе целям и задачам. Научные положения и выводы, сформулированные в диссертации, подтверждаются убедительными фактическими данными, наглядно представленными в приведенных таблицах и графиках.

Полученные результаты диссертации соответствуют областям исследования паспорта научной специальности 05.13.06 – Автоматизация и управление технологическими процессами и производствами (по отраслям) (технические науки), а именно: п. 8 «Формализованные методы анализа, синтеза,

исследования и оптимизация модульных структур систем сбора и обработки данных в АСУТП, АСУП, АСТПП и др.», п. 18 «Средства и методы проектирования технического, математического, лингвистического и других видов обеспечения АСУ», п. 20 «Разработка автоматизированных систем научных исследований».

Основные положения диссертационного исследования в достаточной степени апробированы на следующих научных и научно-практических конференциях:

II Международная научная конференция студентов и молодых ученых «Донецкие чтения 2017: Русский мир как цивилизационная основа научно-образовательного и культурного развития Донбасса», г. Донецк, 17-20 октября 2017 г.;

Международная научно-практическая конференция «Социально-гуманитарные и естественно-технические науки и вызовы современности», г. Ставрополь, 22 декабря 2017 г.;

Международная научно-практическая конференция "Открытые физические чтения – 2018", г. Луганск, 18-19 мая 2018 г.;

III Международная научная конференция «Донецкие чтения 2018: образование, наука, инновации, культура и вызовы современности», г. Донецк, 25 октября 2018 г.;

IV Международная научная конференция «Донецкие чтения 2019: образование, наука, инновации, культура и вызовы современности» г. Донецк, 31 октября 2019 г.;

Международная научно-практическая конференция молодых исследователей им. Д.И. Менделеева, г. Тюмень, 15 ноября 2019 г.

**Личный вклад соискателя.** Основные научные результаты диссертационного исследования получены соискателем самостоятельно. Личный вклад соискателя заключается в обосновании идеи исследования и ее реализации, постановке целей и задач исследования, выборе методологии и методов исследования, проведении теоретических и экспериментальных исследований, а также во внедрении результатов диссертации.

Вклад соискателя в работы, опубликованные в соавторстве, конкретизирован в списке работ, опубликованных по теме диссертации.

**Публикации.** Основные научные результаты диссертационного исследования опубликованы в 13 научных изданиях, из них: 5 - в научных изданиях, включенных в перечень ВАК ДНР, 3 – в других изданиях, 5 – в материалах научных конференций.

**Структура и объем диссертации.** Диссертация состоит из введения, четырех разделов с выводами по каждому из них, заключения, списка литературы из 153 наименований и 1 приложения. Полный объем диссертации составляет 150 страниц, включая 22 рисунка и 2 таблицы.

## ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

**В первом разделе** проведен сравнительный анализ методов анализа экспериментальных кривых. Показано, что применение интегральных методов нецелесообразно при анализе малоизученных процессов. Проанализированы существующие методы сегментации, из которых выделены две группы параллельные и последовательные. При сегментации параллельными методами экспериментальные кривые обрабатываются полностью, и вводится некоторый параметр оценки качества сегментации. При сегментации последовательными методами на данный момент анализируется только определенная часть экспериментальной кривой, на которой в дальнейшем выделяются границы соответствующих участков. Обе группы методов сегментации имеют как преимущества, так и недостатки, учитывая которые определено, что оптимальным решением для этапа сегментации экспериментальных кривых является комбинирование последовательных и параллельных методов. Проанализированы две крупные группы методов описания экспериментальных кривых, а именно: дискриминантные и лингвистические. Лингвистические методы являются более применимыми для анализа малоизученных процессов. Таким образом, для анализа экспериментальных кривых целесообразно использовать структурный метод с использованием комбинированных методов сегментации и лингвистического метода описания.

**Во втором разделе** решалась задача разработки комбинированных вычислительных методов сегментации экспериментальных кривых и алгоритмической реализации процедуры структурного анализа экспериментальных кривых. Представлена методика сегментации, в которой переходы от одного события к другому в исследуемом процессе рассматриваются как быстрые изменения его формы. Благодаря этому, приведенное описание таких экспериментальных кривых сохраняет данные о структуре и может быть получено путем выделения участков, в которых форма существенно отличается от граничащих участков и рассматривается как пики, рывки, резкие изменения и т.д. Эти участки называются переходными или сложными, в отличие от однородных простых, у которых форма не изменяется и которые разделены сложными участками.

Введение понятия «сложности» участков при сегментации кривых двух указанных выше типов не требует предварительной информации о том, к какому типу относится кривая. Таким образом, сложные участки могут быть интерпретированы как изменения состояния процесса, как фоновые возмущения некоторого постоянного состояния, или, наконец, как кратковременные возмущения, которые переводят процесс из одного состояния в другое.

Пусть  $\omega = (t_1, t_2)$  некоторый участок области определения экспериментальной кривой  $f(t)$  и задана реальная функция  $\phi(f, \omega)$ , которая зависит от формы кривой на участке  $\omega$  и является представлением о ее изменчивости в течение этого интервала. Таким образом, величина  $\phi(f, \omega)$



постоянна, если кривая  $f(t)$  однородна, и изменяется, если ее форма быстро изменяется в течение этого интервала. Значение этой функции может быть интерпретировано как степень изменчивости формы кривой (назовем ее, для определенности, функцией сложности).

Пусть в интервале  $\omega^*$  находится точка соединения двух однородных участков экспериментальной кривой, которые отличаются по форме или отображают аномалию на однородном фоне, тогда, с небольшими сдвигами интервала  $\omega$  фиксированной длины справа и слева от  $\omega^*$ , значение функции сложности уменьшается или увеличивается в зависимости от ее конкретной формы. То есть функция  $\phi(f, \omega)$  имеет локальный экстремум при  $\omega = \omega^*$ . Таким образом, в зависимости от выбора функции сложности, сложные участки могут быть локальными экстремумами или участками, которые являются "более сложными", чем граничащие с ними. Учитывая сказанное, методика сегментации для обработки экспериментальной кривой, должна содержать следующее:

- экспериментальная кривая  $f(t)$  делится на ряд элементарных участков  $\omega_j, j=1, \dots, N$  одинаковой длины  $l$ , которые следуют с определенным шагом  $\Delta$  вдоль оси изменения аргумента (шаг  $\Delta$  может быть равен или меньше, чем  $l$ );

- выбирается определенная функция сложности  $\phi(f, \omega)$ , при этом каждый элементарный участок  $\omega_j$  связан с реальной величиной в виде  $\phi_j = \phi(\omega_j)$ ;

- выделяются сложные участки, т.е. участки с локально экстремальными значениями  $\phi_j$  (максимальным или минимальным, в зависимости от выбранной функции сложности).

**В третьем разделе** решалась задача алгоритмической реализации процедуры структурного анализа экспериментальных кривых и задача разработки алгоритмов лингвистического описания экспериментальных кривых. Полученный в результате классификации алфавит символов будет являться классификатором необычных явлений в ходе анализируемого процесса. При применении алгоритма построения опорных участков через конечное число циклов работы будет достигнута устойчивая классификация, а именно для любого вектора  $f^j$ , принадлежащего классу  $A^i$ , расстояние  $r$  до эталона своего класса не будет превышать расстояние до других эталонов, что свидетельствует о том, что ни один вектор не будет перенесен из одного класса в другой. Разбиение полученного текста на непересекающиеся отрезки задает некоторый словарь, формирование которого можно рассматривать как выделение набора макрособытий в развитии исследуемого процесса. Эти макрособытия отражены в символическом представлении кривых в виде стабильных цепочек символов. Экспериментальные кривые можно описать правильными текстами вполне строго. Обработка кривых нацелена на автоматизацию анализа данных научных экспериментов, в которых смена состояний исследуемого процесса не имеет жесткой логической закономерности. Морфологическую грамматику следует

рассматривать лишь как модель, более или менее точно аппроксимирующую тексты, порождаемые экспериментальными кривыми. Для этого задается некоторая мера сходства двух произвольных цепочек символов алфавита, которая отражает степень искажения при переходе от одной из цепочек к другой, а роль искажающего механизма необходимо возложить на трансформационную грамматику, содержащую некоторое множество элементарных трансформаций, т.е. единичных искажений. Тогда в качестве меры сходства между двумя цепочками принимается минимальное число элементарных трансформаций, необходимых для перехода от одной цепочки к другой.

**В четвертом разделе** решалась задача исследования применимости разработанной системы структурного анализа в процессах электромагнитной совместимости радиоэлектронных средств. Для ее реализации проведен ряд экспериментов. Эксперименты проводились как отдельно по каждому из этапов обработки кривой (выделения и распознавания характерных участков, присвоения выделенным участкам символов некоторого алфавита, анализа полученных последовательностей символов), так и автоматизированной системы лингвистического анализа в целом.

В качестве экспериментальных данных для проведения сравнительного исследования работы двух алгоритмов сегментации была использована кривая акустических колебаний в металлическом цилиндре, возбуждаемых ударным методом, полученная в ГОУ ВПО «Донецкий национальный университет» на лабораторном макете. Удар и съем осуществлялся по оси вращения цилиндра. Исследуемая экспериментальная кривая (Рисунок 1) содержала 5000 точек отсчета.

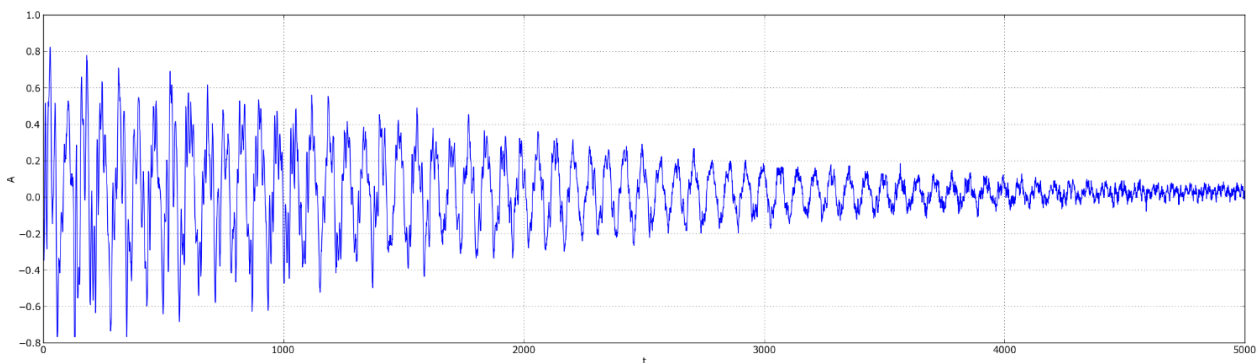


Рисунок 1 – Экспериментальная кривая акустических колебаний в металлическом цилиндре

Два алгоритма сегментации были реализованы на высокоуровневом языке программирования общего назначения Python и применены к анализу выделения переходных участков кривой акустических колебаний. Исследование содержало два эксперимента.

Первый эксперимент. Исследуемая кривая разбивается на 100 непересекающихся участков, каждый в 50 точек отсчета. Последовательность действий первого алгоритма сегментации, основанного на оценке подобия граничащих участков, описана далее. Сперва предварительно разбитый массив участков нормировался. Нормировка заключалась в делении каждого вектора значений на величину  $\|g^j\| = \sqrt{\sum_{i=1}^l |f_i^j|^2}$ , вычисленную для данного участка. Величина  $\|g^j\|$  также имеет название метрика L2, норма или евклидова норма и является геометрическим расстоянием между двумя точками в многомерном пространстве. Затем, для каждого участка вычислялось значение функции сложности  $\phi(\omega_j) = \frac{1}{2}(g^j, g^{j-1} + g^{j+1})$ , являющейся средним значением скалярного произведения вектора  $g^j$  с граничащими предварительно нормированными векторами этой же кривой  $g^{j-1}$  и  $g^{j+1}$ . Значения функции сложности для первого и последнего участка, граничащих всего с одним участком, вычислялись как скалярные произведения:  $\phi(\omega_1) = (g^1, g^{j+1})$  и  $\phi(\omega_{end}) = (g^{end-1}, g^{end})$  соответственно. Далее алгоритмом выделялись участки, соответствующие локальным минимумам, а именно участки, значения функции сложности  $\phi(\omega_j)$  которых принимали значения менее -0.4, полученного экспериментальным путем.

Второй алгоритм сегментации, основанный на исключении переходных участков, аналогично предыдущему, сперва нормировал предварительно разбитый массив участков. Нормировка так же заключалась в делении каждого вектора значений на евклидову норму данного участка. Затем для каждого участка вычислялся набор коэффициентов Фурье. Далее для каждого участка вычислялось значение функции сложности

$$\phi(\omega_j) = \sqrt{\sum_{i=1}^n \left[ g_i^j - \left( \frac{C_0^j}{2} + \sum_{s=1}^m C_{s1}^j \cos(2\pi s \frac{i}{n}) + \sum_{s=1}^m C_{s2}^j \sin(2\pi s \frac{i}{n}) \right) \right]^2},$$

являющейся характеристикой качества аппроксимации кривой на участке рядом Фурье, где  $n$ -количество точек на участке,  $m$  принимает значения от 1 до  $n$ , а  $s$  от 1 до  $m$ . После этого алгоритм выделял участки, значения функции сложности  $\phi(\omega_j)$  которых отличались от граничащих на заданную величину.

Графики функций сложности, полученные в результате работы двух алгоритмов представлены на рисунках 2 и 3.

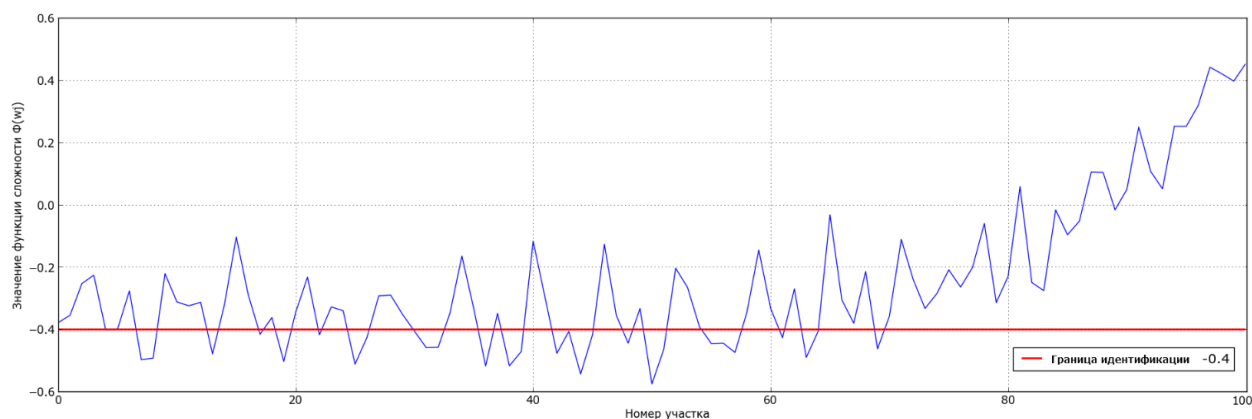


Рисунок 2 – График функций сложности при применении первого алгоритма

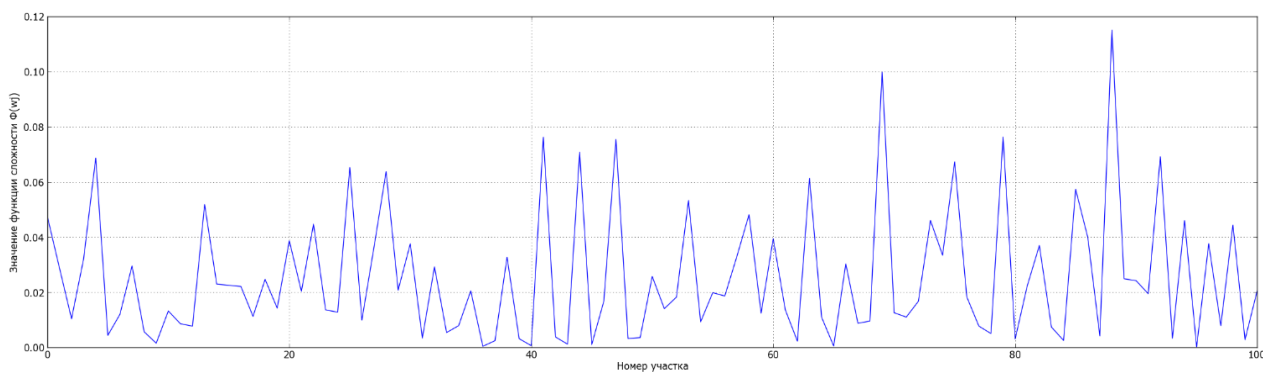


Рисунок 3 – График функций сложности при применении второго алгоритма

Первым алгоритмом выделено 29 переходных участков, вторым - 31. Следовательно, данный эксперимент показывает сходства в результате применения предложенных алгоритмов, несмотря явные отличия способов вычисления функции сложности. Примеры выделенных участков показаны на рисунке 4.

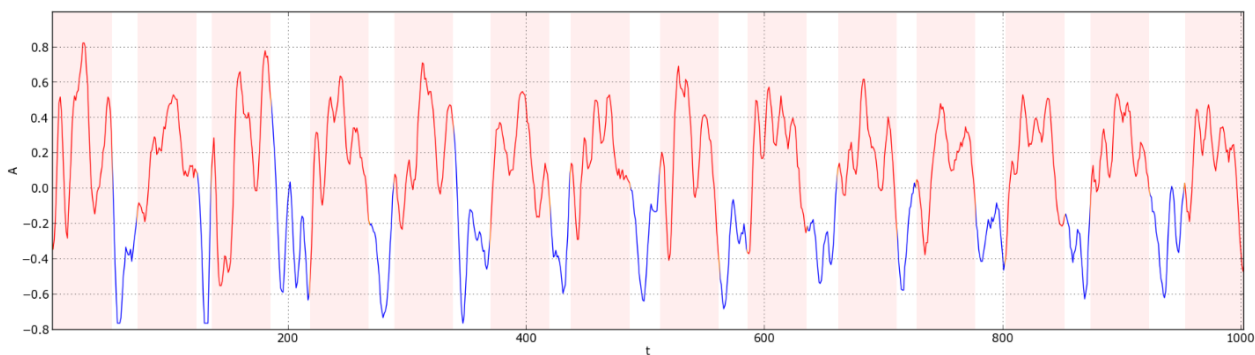


Рисунок 4 – Примеры выделенных участков в первом эксперименте

Второй эксперимент. Исследуемая кривая разбивается на 500 участков, каждый в 10 точек отсчета, с шагом в 5 точек. Таким образом граничащие участки имели перекрытие размером в половину длины участка.

Первым алгоритмом выделено 114, вторым - 119 переходных участков. Данный эксперимент так же показал наличие сходства в результате применения предложенных алгоритмов. Примеры выделенных участков показаны на рисунке 5.

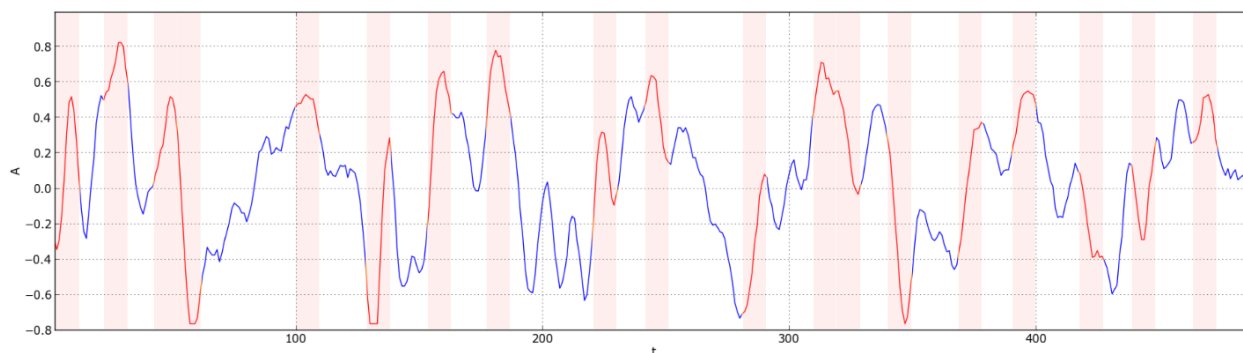


Рисунок 5 – Примеры выделенных участков во втором эксперименте

Проведенные эксперименты показали, что выделенные участки непосредственно и на вид являются более сложными, чем остальные. Исходя из этого, можно заключить что предложенные алгоритмы с применением функции сложности целесообразно применять к задачам сегментации переходных участков экспериментальных кривых акустических колебаний. Первый эксперимент показал сходства в результаты работы обеих алгоритмов. Второй эксперимент так же подтвердил наличие сходства, а также показал, что разбиение на большее количество участков и применение перекрытия приводит к увеличению точности выделения.

Предлагаемые алгоритмы лингвистического описания участков экспериментальных кривых были реализованы на высокоуровневом языке программирования общего назначения Python и применены к задаче присвоения символов анализируемым участкам экспериментальных кривых. В качестве экспериментальных данных были выбраны записи ЭКГ, с использованием одной из схем усиления биопотенциала из лабораторного сеанса длительностью 16 минут на частоте дискретизации 100 Гц. Экспериментальные данные ЭКГ содержали более 95000 точек отсчета. Сегмент экспериментальной кривой, построенной по этим данным изображен на рисунке 6.

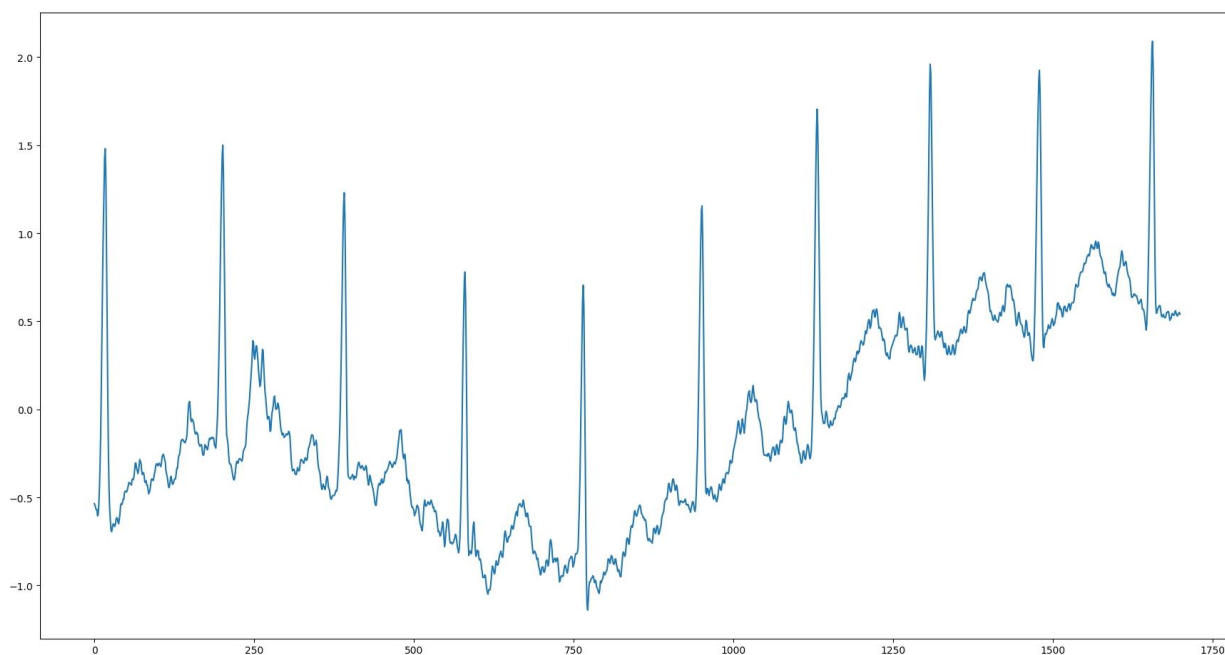


Рисунок 6 – Сегмент кривой ЭКГ

Разработанным алгоритмом экспериментальная кривая была классифицирована на 5 классов. Классам были присвоены соответствующие символы P, Q, R, S, T. После анализа всех точек отсчета экспериментальной кривой в первый класс было определено 378 участков, во второй - 420, в третий - 480, в четвертый - 452, и наконец в пятый - 370. Сегменту экспериментальной кривой, построенной на участке от 0 до 500 точек, изображенному на рисунке 7, соответствует цепочка символов: PQRSTPQRST.



Рисунок 7 – Присвоение символов алфавита сегментам экспериментальных кривых

В результате применения системы лингвистического анализа к анализу данных электрокардиограмм можно сделать вывод, что полученная классификация находит значительную схожесть с реальной расшифрованной электрокардиограммой (рисунок 8).

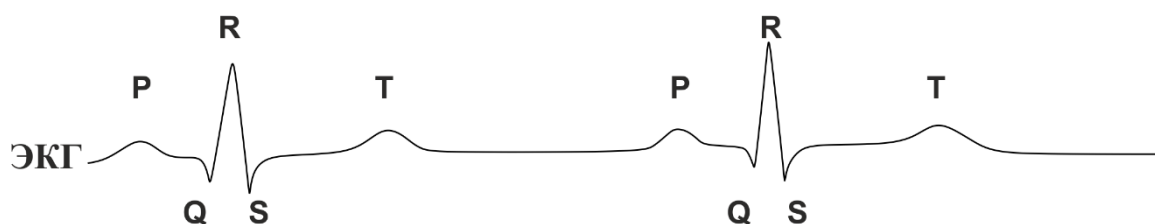


Рисунок 8 – Электрокардиограмма человека

Для исследования разработанной системы лингвистического анализа был проведен эксперимент. Исходной информацией для эксперимента послужили экспериментальные кривые спектрограмм радиочастот FM-диапазона, полученные в ГОУ ВПО «Донецкий национальный университет» на лабораторном макете. Экспериментальные данные были сняты с помощью SDR приемника на базе RTL2832 и R820T. На рисунке 9 представлен характер поведения экспериментальных кривых при различных состояниях.

Исследуемый диапазон ограничен шириной полосы приёмника от 0,25 МГц до 3,2 МГц. Для удобства выбрана частота дискретизации RTL приемника 2432 MSPS (~2.400 МГц). Регистрировалась зависимость частоты от относительной мощности сигнала dBFS в диапазоне частот 95.6-98 МГц. Данный диапазон выбран в связи с большим количеством источников сигналов (радиостанций). Каждая кривая представлялась набором 2400 значений ординат. Эти числа соответствовали точкам отсчета, шагом в 1 КГц. В исследовании применялось частотное подавление сигналов. Подавление осуществлялось лабораторным широкополосным генератором качающейся частоты. Кривые регистрировались при 4 состояниях: без подавления, с подавлением 1-го одного сигнала, 2-х сигналов и 3-х сигналов соответственно. Всего было зарегистрировано и проанализировано 60 таких кривых (по 15 кривых для каждого состояния).

Сегментация анализируемых кривых осуществлялась алгоритмами на основе функции сложности. Каждая кривая разбивалась на равные участки длиной в 200 КГц, взятые с шагом следования вдоль кривой в 100 КГц. Таким образом имелось перекрытие в половину длины участка. Алгоритмом определялись участки, для которых функция сложности принимала локально минимальные значения. Количество выделенных участков, соответствующих сложному поведению кривой, для различных кривых составляло от 4 до 12. Итого на всех 60 кривых было выделено 452 таких участка, которые затем были разбиты алгоритмом классификации на 3 класса. На рисунке 10 представлены образцы эталонов каждого класса.

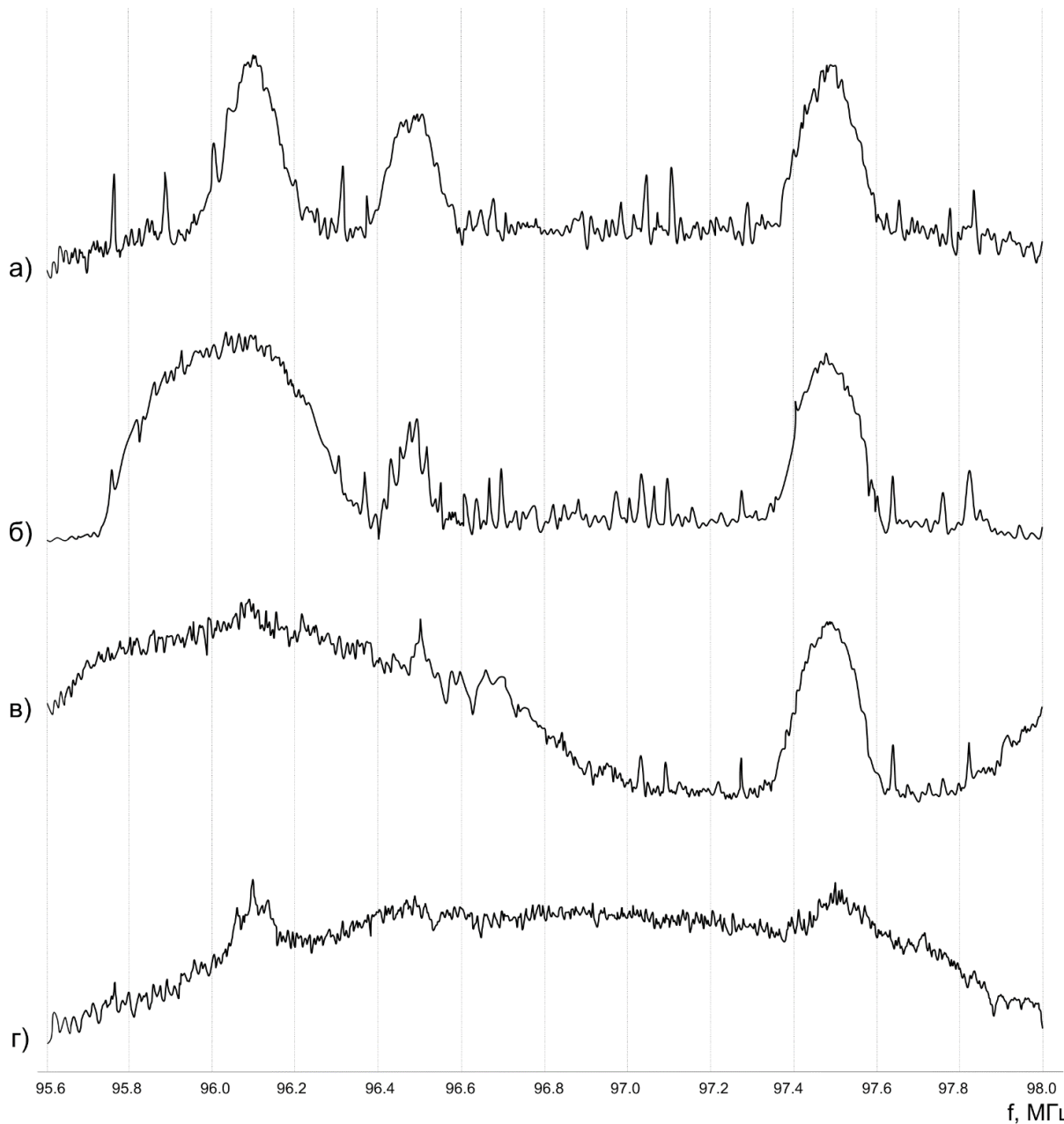


Рисунок 9 – Экспериментальные кривые спектрограмм радиочастот:  
 а) без подавления, б) с подавлением 1-го сигнала,  
 в) с подавлением 2-х сигналов, г) с подавлением 3-х сигналов

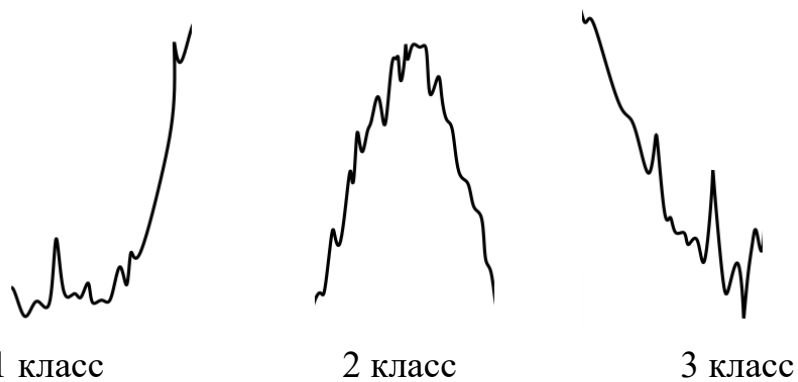


Рисунок 10 – Образцы эталонов классов



Каждый из этих эталонов представляет собой участок кривой, ординаты которого являются среднеарифметическим значением ординат всех участков соответствующего класса. Из рисунка 10 можно отметить, что участки 1-го класса характеризуют левый фронт сигнала, участки 2-го класса – центральную часть сигнала и участки 3-го класса – правый фронт сигнала. Обозначим участки следующим образом: 1 класс – L (left), 2 класс – C (central), 3 класс – R (right). Таким образом сформирован алфавит языка описания исследуемых экспериментальных кривых, состоящий из 3х символов  $A = \{L, C, R\}$ . Лингвистические описания исследуемых кривых представлены в таблице 1.

Таблица 1 – Лингвистические описания исследуемых кривых

1 группа	2 группа	3 группа	4 группа
LCCLC	LLCRCC	CCCCCRC	LLCCCCCCCCCR
LCCLC	LLCRCC	CCCCCRC	LLLCCCCCCCCCR
CCLC	LLCRCC	CCCCCRC	LCCCCCCCCCR
LCCLC	LCRCC	CCCCCRR	LCCCCCCCCCR
LCCLC	LCRCC	CCCCRLC	LCCCCCCCCCR
LCRCLC	LLCRCC	CCCCRC	LCCCCCCCCRR
LCRCLC	LLCRCLC	CCCCCRR	LLCCCCCCCCRR
LCCLC	LLCRCLC	CCCCCRC	LCCCCCCCCRR
LCCLC	LCRCLC	CCCCCRC	LCCCCCCCCCR
LCCC	LLCRCC	CCCCRLC	LCCCCCCCCCR
LCCLC	LCCRCC	CCCCCRC	LCCCCCCCCCR
LCCLC	LLCRCC	CCCCCRC	LCCCCCCCCCR
LCCLCR	LLCRCCR	CCCCRCR	LCCCCCCCCCR
LCCLC	LLCRCC	CCCCCRC	LCCCCCCCCCR
LCCLC	LLCRCC	CCCCCRC	LCCCCCCCCCR

Для удобства анализа, полученные описания кривых сортированы на группы, соответствующие различным состояниям, а именно: 1 группа - без подавления, 2 группа - с подавлением 1-го сигнала, 3 группа - с подавлением 2-х сигналов, 4 группа - с подавлением 3-х сигналов. Символы в описаниях расположены в соответствии с последовательностью их естественного следования вдоль кривых.

Из анализа таблицы 1 следует, что лингвистические описания, соответствующие одной группе очень близки, а описания, взятые из разных групп, заметно отличаются. Можно отметить, что все лингвистические описания третьей группы начинаются с символа C, а описания четвертой группы заканчиваются на R. В описаниях третьей группы практически не встречается символ L, в описаниях третьей и четвертой групп большую часть символов

составляет  $S$ . Так же описания первой группы не превышают пяти символов, а описания четвертой группы состоят из не менее чем десяти символов.

Таким образом, можно заключить что приведенные описания экспериментальных кривых, полученные разработанной автоматизированной системой лингвистического анализа, представляют короткие и надежные правила для анализа спектрограмм радиочастот.

Для более расширенного лингвистического описания экспериментальных кривых предлагается составлять это описание с учетом местоположения участков кривой на оси абсцисс.

Для составления такого расширенного лингвистического описания, учитывающего положение участков на исследуемой кривой, на примере спектрограмм радиочастот FM диапазона воспользуемся следующей процедурой. Область определения кривой разбивается на  $l$  зон (в нашем случае  $l=12$ ), одинаковых по длине (200 КГц). Алфавит языка описания кривых дополняется символом  $b$  – background, обозначающим не информативные фоновые участки. Таким образом алфавит состоит из четырех символов  $A = \{L, C, R, b\}$ . Исходя из этого, каждому участку будет присвоен символ, либо характеризующий его поведение на кривой, либо характеризующий фоновое и не информативное возмущение.

Полученное расширенное лингвистическое описание кривой будет отражать не только классы выделенных на ней элементарных событий, но и их фазы. Таким образом, цепочка символов  $\langle L C C L C \rangle$  будет представлена в виде  $\langle b L C b C b b b L C b b \rangle$ . При этом расширенные описания разных экспериментальных кривых с одинаковой длиной будут иметь одинаковое количество символов.

Исходной информацией для эксперимента послужили те же экспериментальные кривые спектрограмм радиочастот, что и в предыдущем эксперименте. В таблице 2 представлены расширенные лингвистические описания этих спектрограмм.

Положение символов описаний кривых в таблице 2 соответствуют естественным положениям участков на этих кривых. Каждой группе соответствует усредненная расширенная цепочка символов (эталон группы)

полученные по  $\phi(T) = \sum_{j=1}^n r(T, T_j) = \min$ , а справа для каждой кривой указаны расстояния  $r(\hat{T}, T_j)$  от соответствующей ей расширенной цепочки до эталонов всех четырех групп, отражающие минимальное число элементарных трансформаций, переводящих эталон в эту цепочку.

Таблица 2 – Расширенные лингвистические описания исследуемых кривых

Группы	Описания кривых	Эталонные группы и расстояние $r(\hat{T}, T_j)$			
		1	2	3	4
		bLCLCbCbbb bb	LLCRCbbbbC bb	CCCCCRbbC bb	bLCCCCCCCC CR
1	2	3	4	5	6
1 группа	bLCbCbbbLCbb	0	6	10	14
	bLCbCbbbLCbb	0	6	10	14
	bbCbCbbbLCbb	2	10	10	18
	bLCbCbbbLCbb	0	6	10	14
	bLCbCbbbLCbb	0	6	10	14
	bLCRCbbbLCbb	4	8	10	14
	bLCRCbbbLCbb	4	8	10	14
	bLCbCbbbLCbb	0	6	10	14
	bLCbCbbbLCbb	0	6	10	14
	bLCbCbbbbCbb	2	6	10	14
	bLCbCbbbLCbb	0	6	10	14
	bLCbCbbbLCbb	0	6	10	14
	bLCbCbbbLCRb	2	6	10	12
	bLCbCbbbLCbb	0	6	10	14
	bLCbCbbbLCbb	0	6	10	14
2 группа	LLCRCbbbbCbb	6	0	10	14
	LLCRCbbbbCbb	6	0	10	14
	LLCRCbbbbCbb	6	0	10	14
	bLCRCbbbbCbb	4	2	10	14
	bLCRCbbbbCbb	4	2	10	14
	LLCRCbbbbCbb	6	0	10	14
	LLCRCbbbLCbb	8	4	12	14
	LLCRCbbbLCbb	8	4	12	14
	bLCRCbbbLCbb	4	2	10	14
	LLCRCbbbbCbb	6	0	10	14
	LLCRCbbbbCbb	6	0	10	14
	LLCRCbbbbCbb	6	0	10	14
	LLCRCbbbbCRb	8	2	8	14
	LLCRCbbbbCbb	6	0	10	14
	LLCRCbbbbCbb	6	0	10	14

Продолжение Таблицы 2

1	2	3	4	5	6
3 группа	CCCCC <b>R</b> bbCb <b>b</b>	10	8	0	10
	CCCCC <b>R</b> bbCb <b>b</b>	10	8	0	10
	CCCCC <b>R</b> bbCb <b>b</b>	10	8	0	10
	CCCCC <b>R</b> bb <b>R</b> bb	10	10	2	12
	bCCCC <b>R</b> bLCb <b>b</b>	10	12	2	12
	bCCCC <b>R</b> bbCb <b>b</b>	10	10	2	12
	CCCCC <b>R</b> bb <b>R</b> bb	10	10	2	12
	CCCCC <b>R</b> bbCb <b>b</b>	10	8	0	10
	CCCCC <b>R</b> bbCb <b>b</b>	10	8	0	10
	bCCCC <b>R</b> bLCb <b>b</b>	10	12	2	12
	CCCCC <b>R</b> bbCb <b>b</b>	10	8	0	10
	CCCCC <b>R</b> bbCb <b>b</b>	10	8	0	10
	bCCCC <b>R</b> bb <b>C</b> Rb	10	12	4	12
	CCCCC <b>R</b> bbCb <b>b</b>	10	8	2	10
	CCCCC <b>R</b> bbCb <b>b</b>	10	8	0	10
4 группа	LLCCCCCCCC <b>R</b>	18	14	12	2
	LLLCCCCCCCC <b>R</b>	20	16	14	4
	bLCCCCCCCC <b>R</b>	14	14	10	0
	bLCCCCCCCC <b>R</b>	14	14	10	0
	bLCCCCCCCC <b>R</b>	14	14	10	0
	bLCCCCCCCC <b>R</b> R	14	14	10	2
	bLLCCCCCCCC <b>R</b> R	18	14	12	4
	bLCCCCCCCC <b>R</b> R	14	14	10	2
	bLCCCCCCCC <b>R</b>	14	14	10	0
	bLCCCCCCCC <b>R</b>	14	14	10	0
	bLCCCCCCCC <b>R</b>	14	14	10	0
	bLCCCCCCCC <b>R</b>	14	14	10	0
	bLCCCCCCCC <b>R</b>	14	14	10	0
	bLCCCCCCCC <b>R</b> b	12	12	8	2
	bLCCCCCCCC <b>R</b>	14	14	10	0

Применение к каждой группе цепочек из таблицы 2 алгоритма поиска разбора с минимальной неплотностью позволяет получить следующие эталоны групп: 1 группа – bLCbCb**b**bLCb**b**, 2 группа – LLCRCb**b**b**b**Cb**b**, 3 группа – CCCCC**R**bbCb**b**, 4 группа – bLCCCCCCCC**R**, что подтверждает правильность вычисления эталонов. Для удобства восприятия расширенных лингвистических описаний и наглядности, на рисунке 11 на кривых из разных групп отмечены символы, присвоенные каждому анализируемому участку.

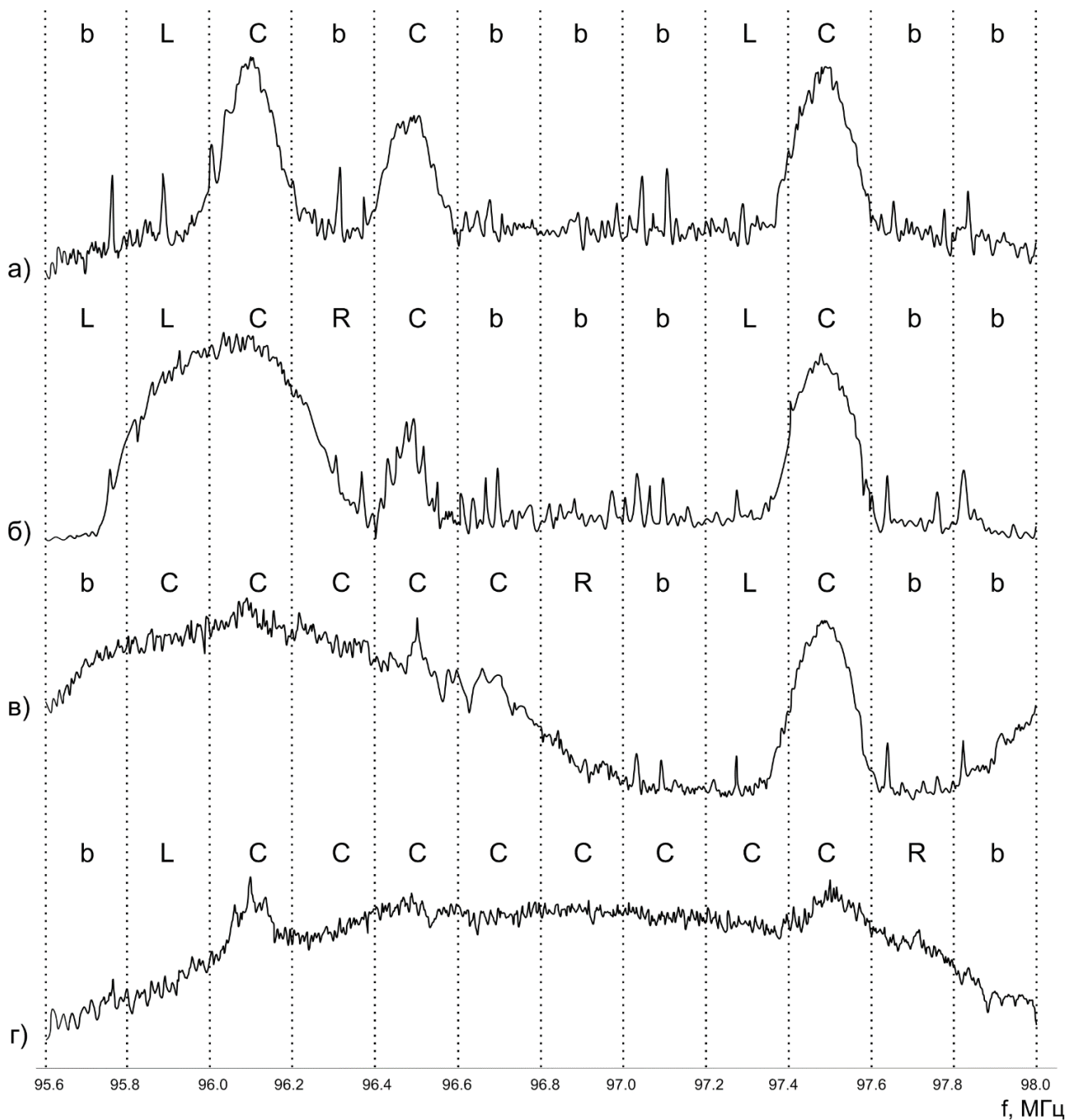


Рисунок 11 – Экспериментальные кривые спектрограмм радиочастот и их расширенные описания

Применение эталонов позволяет отнести каждую цепочку символов к своей группе по признаку минимума расстояния до эталона. Более детальное описание исследуемых кривых возможно получить путем изменения длины участков и шага их следования на этапе сегментации, что в случае с анализом приведенных спектрограмм не является необходимым. Полученные расширенные описания достаточно точно описывают характер поведения исследуемых кривых.

## ЗАКЛЮЧЕНИЕ

Диссертационная работа является завершенным научным исследованием, в котором получено решение важной научно-технической задачи, состоящей в обосновании модернизированных методов и алгоритмов автоматизации процессов анализа массивов экспериментальных данных, их классификации и представления в виде компактных структур, что позволяет использовать фактические результаты работы в практической, научно-исследовательской и преподавательской деятельности. Основные научные результаты и выводы, полученные при выполнении работы, состоят в следующем:

1. Дальнейшее развитие получил метод структурного анализа данных, в рамках лингвистического подхода к анализу экспериментальных кривых, в котором анализируемая кривая описывается в виде сжатого описания из последовательного ряда символов либо целых слов из определенного алфавита.

2. Разработаны вычислительные методы сегментации экспериментальных кривых с использованием функций сложности, способные осуществлять бинарную классификацию помимо сегментации, и содержащие дополнительные условия во избежание выделения ложных экстремумов.

3. Осуществлена алгоритмическая реализация процедуры структурного анализа данных, точнее лингвистического подхода к анализу экспериментальных кривых, в котором анализируемая кривая описывается последовательным рядом символов либо целых слов из определенного алфавита.

4. Разработаны методы лингвистического описания участков экспериментальных кривых на основе сравнения с эталонами, в котором через конечное число циклов достигается устойчивая классификация и ни один вектор не переносится из одного класса в другой, благодаря классификации по признаку минимума расстояния до эталона.

5. Обоснована применимость разработанной системы структурного анализа к выделению информативных участков кривой акустических колебаний, к анализу экспериментальных данных электрокардиограмм, к анализу спектрограмм радиочастот.

6. Обоснована процедура составления более расширенного лингвистического описания экспериментальных кривых, позволяющая составлять это описание с учетом местоположения участков кривой на оси абсцисс.

## СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ АВТОРОМ ПО ТЕМЕ ДИССЕРТАЦИИ

**- в рецензируемых научных изданиях, рекомендуемых ВАК:**

1. **Третьяков, И. А.** Функции сложности для выделения и распознавания характерных участков экспериментальных кривых / **И. А. Третьяков, В. В. Данилов** // Вестник ДонНУ. Сер. А: Естественные науки. – 2017. – № 2. – С. 101-107.

2. Данилов, В. В. Алгоритмы идентификации переходных участков экспериментальных кривых с применением аппроксимации / В. В. Данилов, **И. А. Третьяков**, А. В. Шалаев, Я. И. Рушечников // Сборник научных трудов ДОНИЖТ. – 2018. – № 48. – С. 19-23.

3. Данилов, В. В. Идентификация переходных участков кривой акустических колебаний / В. В. Данилов, **И. А. Третьяков**, А. В. Шалаев, Я. И. Рушечников // Сборник научных трудов ДОНИЖТ. – 2018. – № 49. – С. 10-16.

4. Данилов, В. В. Алгоритмы экстраполяции участков экспериментальных кривых / В. В. Данилов, **И. А. Третьяков**, Я. И. Рушечников, А. В. Шалаев // Сборник научных трудов ДОНИЖТ. – 2018. – № 50. – С. 10-15.

5. Данилов, В. В. Алгоритмизация присвоения символов анализируемым участкам экспериментальных кривых / В. В. Данилов, **И. А. Третьяков**, Я. И. Рушечников // Сборник научных трудов ДОНИЖТ. – 2018. – № 51. – С. 15-22.

**- в других научных изданиях:**

6. **Третьяков, И. А.** Исследование применимости функций сложности к задачам идентификации переходных участков экспериментальных кривых акустических колебаний / **И. А. Третьяков**, А. В. Шалаев, Я. И. Рушечников, В. В. Данилов // Вестник Луганского национального университета имени Владимира Даля. – 2018. – № 5(11). – С. 332-335.

7. **Третьяков, И. А.** Методы параллельной сегментации экспериментальных кривых / **И. А. Третьяков** // Вестник ДонНУ. Сер. Г: Технические науки. – 2018. – № 4. – С. 36-41.

8. **Третьяков, И. А.** Исследование алгоритмов лингвистического описания участков экспериментальных кривых / **И. А. Третьяков** // Вестник ДонНУ. Сер. Г: Технические науки. – 2019. – № 3. – С. 26-30.

**- в материалах научных конференций:**

9. **Третьяков, И. А.** Функции сложности для выделения и распознавания характерных участков экспериментальных кривых // **И. А. Третьяков**, А. В. Шалаев, В. В. Данилов // Донецкие чтения 2017: Русский мир как цивилизационная основа научно-образовательного и культурного развития Донбасса: Материалы Международной научной конференции студентов и молодых ученых (Донецк, 17-20 октября 2017 г.). – Том 1: Физико-математические и технические науки / под общей редакцией проф. С.В. Беспаловой. – Донецк: Изд-во ДонНУ, 2017. – С. 206-207.

10. **Третьяков, И. А.** Алгоритмы идентификации переходных участков экспериментальных кривых / **И. А. Третьяков** // Социально-гуманитарные и естественно-технические науки и вызовы современности: материалы международной научно-практической конференции (Ставрополь, 21 декабря 2017 г.) / С.Е. Шиянов, А.П. Федоровский (отв.ред.) – Ставрополь: АНО ВО СКСИ, 2017. – С. 824-828.

11. **Третьяков, И. А.** Алгоритмы прогнозного и формального типов экстраполяции участков экспериментальных кривых / **И. А. Третьяков** // Донецкие чтения 2018: образование, наука, инновации, культура и вызовы современности: Материалы III Международной научной конференции (Донецк, 25 октября 2018 г.). – Том 1: Физико-математические и технические науки / под общей редакцией проф. С.В. Беспаловой. – Донецк: Изд-во ДонНУ, 2018. – С. 162-164.

12. **Третьяков, И. А.** Сравнительный анализ методов параллельной сегментации экспериментальных кривых / **И. А. Третьяков** // Донецкие чтения 2019: образование, наука, инновации, культура и вызовы современности: Материалы IV Международной научной конференции (Донецк, 31 октября 2019 г.). – Том 1: Физико-математические и технические науки. Часть 2 / под общей редакцией проф. С.В. Беспаловой. – Донецк: Изд-во ДонНУ, 2019. – С. 176-178.

13. **Третьяков, И. А.** Реализация алгоритмов формирования алфавита символов для анализа экспериментальных кривых / **И. А. Третьяков** // Материалы международной научно-практической конференции молодых исследователей им. Д. И. Менделеева (г. Тюмень, 15 ноября 2019 г.): сборник статей / отв. ред. А. Н. Халин. – Тюмень: ТИУ, 2020. – С. 28-31.

**Личный вклад соискателя в публикациях:** [1, 9] - предложена методика для выделения и распознавания характерных участков экспериментальных кривых, предложены дополнительные условия для уменьшения ложных экстремумов при вычислениях, представлены несколько видов функции сложности; [2, 10] - рассмотрена формальная постановка задачи сегментации экспериментальных кривых, основанная на выделении переходных участков, разработаны и представлены два алгоритма сегментации переходных участков с применением аппроксимации; [3, 6] - проведено сравнительное исследование двух разработанных алгоритмов сегментации переходных участков кривой, представлены результаты экспериментальных исследований программы, реализующей эти алгоритмы, на примере кривой акустических колебаний; [4, 11] - представлена общая идея алгоритма, осуществляющего прогнозирование поведения кривой за границами определенных участков, основываясь на использовании функции сложности в качестве математико-статистической модели кривой, разработаны два алгоритма, формального и прогнозного типа, экстраполяции экспериментальных кривых; [5, 8, 13] - представлена реализация этапа присвоения выделенным участкам символов, соответствующих различным характерам поведения кривой, лингвистического анализа экспериментальных кривых, произведена реализация и исследование вычислительных алгоритмов лингвистического описания экспериментальных кривых; [7, 12] - произведен сравнительный анализ методов параллельной сегментации экспериментальных кривых, выявлены соответствующие недостатки, предложены решения и рекомендации для их устранения.



Подписано в печать 06.10.2020 г.  
Формат 60x84x1/16. Бумага офсетная.  
Печать цифровая. Усл. печ. л. 1,0.  
Тираж 100 экз. Заказ № 20окт144

Отпечатано в Издательстве ГОУ ВПО «Донецкий национальный университет»  
ДНР 83001, г. Донецк, ул. Университетская, д. 24.  
Тел.: +38 (062) 302 92 27.

Свидетельство о внесении субъекта издательской деятельности  
в Государственный реестр серия ДК № 1854 от 24.06.2004 г.