

ИССЛЕДОВАНИЕ СПЕКТРОГРАММ РАДИОЧАСТОТ МЕТОДАМИ ЛИНГВИСТИЧЕСКОГО АНАЛИЗА

И. А. Третьяков, В. В. Данилов

*Донецкий национальный университет,
Донецк, Донецкая Народная Республика*

Исходной информацией для эксперимента послужили экспериментальные кривые спектрограмм радиочастот FM-диапазона, полученные на лабораторном макете. Алгоритмом определялись участки, для которых функция сложности принимала локально минимальные значения. Для каждого класса был определен эталон, который представляет собой участок кривой, ординаты которого являются среднеарифметическим значением ординат всех участков соответствующего класса. Для более расширенного лингвистического описания экспериментальных кривых предлагается составлять это описание с учетом местоположения участков кривой на оси абсцисс. Полученное расширенное лингвистическое описание кривой будет отражать не только классы выделенных на ней элементарных событий, но и их фазы. В результате применения системы лингвистического анализа к анализу спектрограмм радиочастот можно заключить, что приведенные описания экспериментальных кривых представляют короткие и надежные правила для анализа спектрограмм радиочастот. Применение эталонов позволяет безошибочно отнести каждую цепочку символов к своей группе по признаку минимума расстояния до эталона. Полученные расширенные описания довольно точно описывают характер поведения исследуемых кривых.

Ключевые слова: анализ данных, экспериментальные кривые, лингвистическое описание, эталон, цепочка символов, спектрограммы.

Для цитирования: Третьяков И. А., Данилов В. В. Исследование спектрограмм радиочастот методами лингвистического анализа // Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика. 2020. № 3. С. 45–51. DOI: 10.24143/2072-9502-2020-3-45-51.

Введение

В настоящее время анализ данных [1, 2] стремительно развивается как научное направление, ориентированное на формирование и исследование математических методов и вычислительных алгоритмов фильтрации, преобразования и извлечения полезной информации из экспериментальных данных и принятия решений. Любые методы обработки данных применяются для структурирования и анализа существующей информации. Анализ данных имеет множество аспектов и подходов, охватывает разные методы в различных областях науки и деятельности. Значительная часть этого направления уделяется анализу данных, представленных в виде экспериментальных кривых. Данные этого типа широко используются в автоматизации управления технологическими процессами в промышленности. Также экспериментальные кривые 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.

Состояние проблемы

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. Также нет представления о том, какие именно данные являются полезной информацией в огромном массиве данных. Поэтому возни-

кает необходимость в разработке специальных вычислительных алгоритмов и эффективных методов анализа, аппроксимации и построения точного сжатого описания экспериментальных кривых с целью выделения полезной информации, а целью данной работы является исследование экспериментальных данных научных исследований методами лингвистического анализа, разработанными авторами статьи.

Процедура эксперимента лингвистического описания спектрограмм радиочастот

Разработанная авторами система лингвистического анализа реализована на высокоуровневом языке программирования общего назначения Python на основе алгоритмов [3, 4]. Для оценки эффективности разработанной системы лингвистического анализа экспериментальных данных научных исследований 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. Исследуемый диапазон ограничен шириной полосы приемника от 0,25 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 2 400 значений ординат. Эти числа соответствовали точкам отсчета, шагом в 1 КГц. В исследовании применялось частотное подавление сигналов. Подавление осуществлялось лабораторным широкополосным генератором качающейся частоты. На рис. 1 представлен характер поведения экспериментальных кривых при различных состояниях.

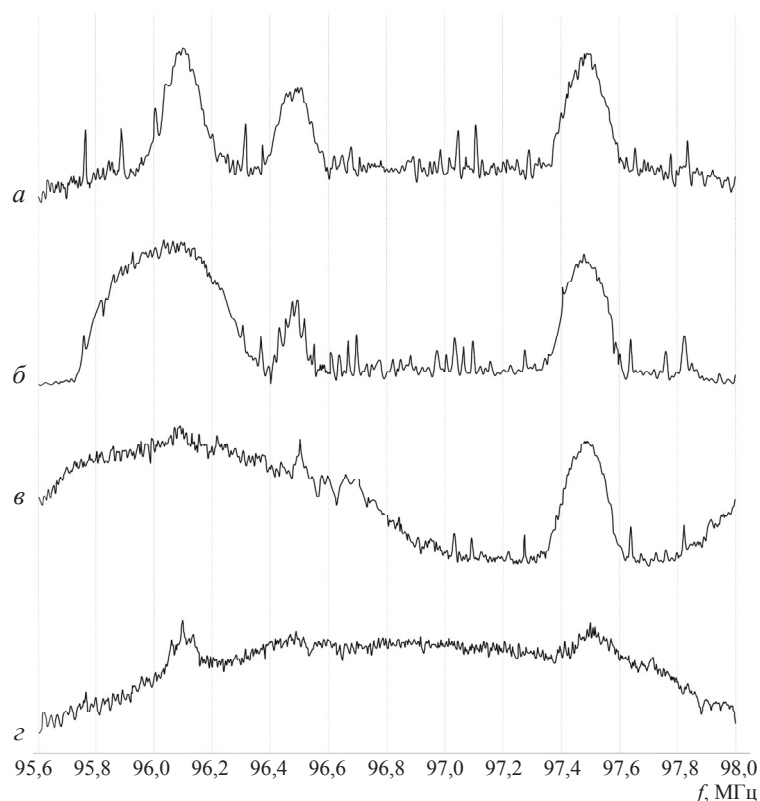


Рис. 1. Экспериментальные кривые спектрограмм радиочастот FM-диапазона: без подавления (а); с подавлением одного сигнала (б); 2-х сигналов (в); 3-х сигналов (з)

Кривые регистрировались при 4-х состояниях. Всего было зарегистрировано и проанализировано 60 таких кривых (по 15 кривых для каждого состояния).

Сегментация анализируемых кривых осуществлялась алгоритмами на основе функции сложности [5]. 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 [4] на 3 класса.

Для каждого класса был определен эталон, который представляет собой участок кривой, ординаты которого являются среднеарифметическим значением ординат всех участков соответствующего класса. На рис. 2 представлены образцы эталонов каждого класса.

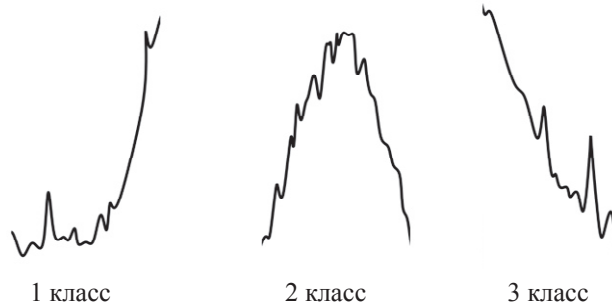


Рис. 2. Образцы эталонов классов

По виду эталонов определено, что участки 1-го класса характеризуют левый фронт сигнала, участки 2-го класса – центральную часть сигнала, 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 3-х символов $A = \{L, C, R\}$.

Полученные лингвистические описания исследуемых кривых представлены на рис. 3.

1 группа	2 группа	3 группа	4 группа
LCCLC	LLCRCC	CCCCCRC	LLCCCCCCCCCR
LCCLC	LLCRCC	CCCCCRC	LLLCCCCCCCCR
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

Рис. 3. Лингвистические описания исследуемых кривых

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 рис. 3 следует, что лингвистические описания, соответствующие одной группе, очень близки, а описания, взятые из разных групп, заметно отличаются. Можно отметить, что все лингвистические описания 3-й группы начинаются с символа *C*, а описания 4-й группы заканчиваются на *R*. В описаниях 3-й группы практически не встречается символ *L*, в описаниях 3-й и 4-й групп большую часть символов составляет *C*. Также описания 1-й группы не превышают 5 символов, а описания 4-й группы состоят из не менее чем 10 символов.

Процедура расширенного лингвистического описания спектрограмм радиочастот

Для более расширенного лингвистического описания экспериментальных кривых предлагается составлять это описание с учетом местоположения участков кривой на оси абсцисс. Такая информация не использовалась при описании в предыдущем эксперименте, однако эта информация может оказаться необходимой для более глубокого анализа представленной кривой. Например, при анализе спектрограмм важно учитывать, на какой из частот находится пик амплитуды. Кроме этого имеется большой класс экспериментальных кривых, для анализа которых необходима информация о характере поведения как на отдельных информативных участках, так и информация о положении этих участков на оси аргумента.

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 неинформативные 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 неинформативное возмущение. Полученное расширенное лингвистическое описание кривой будет отражать не только классы выделенных на ней элементарных событий, но и их фазы. Таким образом, цепочка символов $\langle L C C L C \rangle$ представляется 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.

При определении автоматизированной системой лингвистического анализа конечного подмножества \hat{L} в множестве всех таких цепочек L трансформационная грамматика сможет состоять из одной элементарной трансформации, а именно в замене одного символа другим. Минимальное количество элементарных трансформаций, переводящих цепочку символов T_1 в T_2 , является мерой отличия этих цепочек друг от друга. Рассмотрим метрику на множестве всех цепочек длины l , приводящую к числу несовпадающих символов на одинаковых местоположениях:

$$r(T_1, T_2) = \sum_{i=1}^l |a_i^1, a_i^2|, \quad (1)$$

$$\text{где } |a_i^1, a_i^2| = \begin{cases} 0, & a_i^1 = a_i^2 \\ 1, & a_i^1 \neq a_i^2 \end{cases}.$$

Так образуются символьные цепочки фиксированной длины, которые можно сравнивать посимвольно, например, как сравнивают числовые векторы одинаковой размерности. Для ядра языка, состоящего из одной символьной цепочки \hat{T} , степень принадлежности произвольной символьной цепочки T к этому языку определяется монотонной убывающей функцией расстояния в метрическом пространстве L от этой цепочки T до эталонной цепочки \hat{T} . Пусть имеется выборка экспериментальных кривых одной группы, представленная расширенными описаниями $T_j = \langle a_1, \dots, a_l \rangle$, $j = 1, \dots, n$, где n – количество кривых в данной выборке. Чтобы найти

символьную цепочку $\hat{T} = \langle \hat{a}_1, \dots, \hat{a}_l \rangle$, для которой $\varphi(T) = \sum_{j=1}^n r(T, T_j) = \min$, необходимо под-

ставить критерий $\varphi(T)$ в (1) и изменить порядок суммирования: $\varphi(T) = \sum_{i=1}^l \sum_{j=1}^n |\hat{a}_i, a_i^j|$. Минимум критерия будет обеспечен, когда \hat{a}_i будет символом из расширенного алфавита A , который чаще других оказывается на i -м месте в символьных цепочках T_j , $j = 1, \dots, n$.

Исходной информацией для эксперимента послужили те же экспериментальные кривые спектрограмм радиочастот, что и в предыдущем эксперименте. На рис. 4 и 5 представлены расширенные лингвистические описания этих спектрограмм.

Группы	Описания кривых	Эталон группы и расстояние $r(\hat{T}, T_j)$				Группы	Описания кривых	Эталон группы и расстояние $r(\hat{T}, T_j)$			
		1	2	3	4			1	2	3	4
		bLcCbCbbLcCb	LLCRcCbbCbb	CCCCCRbbCbb	LCCCCCCCCR			bLcCbCbbLcCb	LLCRcCbbCbb	CCCCCRbbCbb	LCCCCCCCCR
1 группа	bLcCbCbbLcCb	0	6	10	14	2 группа	LLCRcCbbCbb	6	0	10	14
	bLcCbCbbLcCb	0	6	10	14		LLCRcCbbCbb	6	0	10	14
	bbCbCbbLcCb	2	10	10	18		LLCRcCbbCbb	6	0	10	14
	bLcCbCbbLcCb	0	6	10	14		bLLCRcCbbCbb	4	2	10	14
	bLcCbCbbLcCb	0	6	10	14		bLLCRcCbbCbb	4	2	10	14
	bLLCRcCbbCbb	4	8	10	14		LLCRcCbbCbb	6	0	10	14
	bLLCRcCbbCbb	4	8	10	14		LLCRcCbbLcCb	8	4	12	14
	bLcCbCbbLcCb	0	6	10	14		LLCRcCbbLcCb	8	4	12	14
	bLcCbCbbLcCb	0	6	10	14		bLLCRcCbbCbb	4	2	10	14
	bLcCbCbbLcCb	0	6	10	14		LLCRcCbbCbb	6	0	10	14
	bLcCbCbbLcCb	0	6	10	14		LLCRcCbbCbb	6	0	10	14
	bLcCbCbbLcCb	0	6	10	14		LLCRcCbbCbb	6	0	10	14
	bLcCbCbbLcCb	0	6	10	14		LLCRcCbbCbb	6	0	10	14
	bLcCbCbbLcCb	0	6	10	14		LLCRcCbbCbb	6	0	10	14
bLcCbCbbLcCb	0	6	10	14	LLCRcCbbCbb	6	0	10	14		

Рис. 4. Расширенные лингвистические описания исследуемых кривых 1-й и 2-й группы

Группы	Описания кривых	Эталон группы и расстояние $r(\hat{T}, T_j)$				Группы	Описания кривых	Эталон группы и расстояние $r(\hat{T}, T_j)$			
		1	2	3	4			1	2	3	4
		bLcCbCbbLcCb	LLCRcCbbCbb	CCCCCRbbCbb	LCCCCCCCCR			bLcCbCbbLcCb	LLCRcCbbCbb	CCCCCRbbCbb	LCCCCCCCCR
3 группа	CCCCCRbbCbb	10	8	0	10	4 группа	LLCCCCCCCCR	18	14	12	2
	CCCCCRbbCbb	10	8	0	10		LLCCCCCCCCR	20	16	14	4
	CCCCCRbbCbb	10	8	0	10		bLCCCCCCCCR	14	14	10	0
	CCCCCRbbRbb	10	10	2	12		bLCCCCCCCCR	14	14	10	0
	bCCCCRbLcCb	10	12	2	12		bLCCCCCCCCR	14	14	10	0
	bCCCCRbbCbb	10	10	2	12		bLCCCCCCCCR	14	14	10	0
	CCCCCRbbRbb	10	10	2	12		bLCCCCCCCCR	14	14	10	0
	CCCCCRbbCbb	10	8	0	10		bLCCCCCCCCR	14	14	10	0
	CCCCCRbbCbb	10	8	0	10		bLCCCCCCCCR	14	14	10	0
	CCCCCRbbCbb	10	8	0	10		bLCCCCCCCCR	14	14	10	0
	CCCCCRbbCbb	10	8	0	10		bLCCCCCCCCR	14	14	10	0
	CCCCCRbbCbb	10	8	0	10		bLCCCCCCCCR	14	14	10	0
	CCCCCRbbCbb	10	8	0	10		bLCCCCCCCCR	14	14	10	0
	CCCCCRbbCbb	10	8	0	10		bLCCCCCCCCR	14	14	10	0
CCCCCRbbCbb	10	8	0	10	bLCCCCCCCCR	14	14	10	0		

Рис. 5. Расширенные лингвистические описания исследуемых кривых 3-й и 4-й группы

Положение символов описаний кривых на рис. 4 и 5 соответствуют естественным положениям участков на этих кривых. Каждой группе соответствует усредненная расширенная цепочка символов (эталон группы), полученных по $\varphi(T) = \sum_{j=1}^n r(T, T_j) = \min$, а справа для каждой кривой указаны расстояния $r(\hat{T}, T_j)$ от соответствующей ей расширенной цепочки до эталонов всех 4-х групп, вычисленные по формуле (1), отражающие минимальное число элементарных трансформаций, переводящих эталон в эту цепочку.

Для удобства восприятия расширенных лингвистических описаний и наглядности, на рис. 6 на кривых из разных групп отмечены символы, присвоенные каждому анализируемому участку.

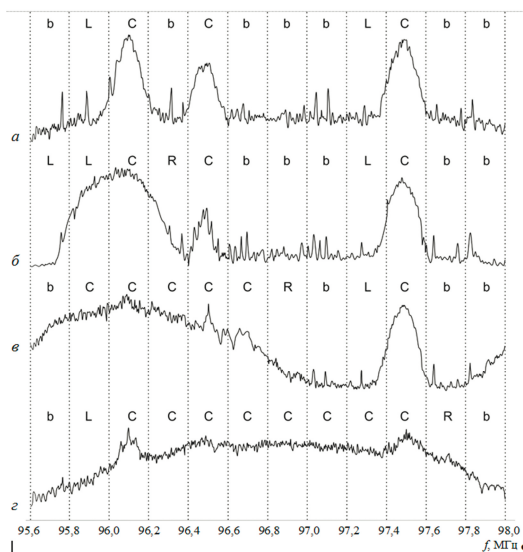


Рис. 6. Экспериментальные кривые спектрограмм радиочастот FM-диапазона и их расширенные лингвистические описания при 4-х состояниях: без подавления (а); с подавлением одного сигнала (б); 2-х сигналов (в); 3-х сигналов (г)

Полученные расширенные лингвистические описания позволяют судить о характере поведения кривых и понимать, в каком состоянии исследуемого процесса они находятся.

Заключение

По результатам первого эксперимента можно сделать вывод, что приведенные описания экспериментальных кривых, полученные методами лингвистического анализа экспериментальных данных научных исследований, представляют короткие и надежные правила для анализа спектрограмм радиочастот.

Применение эталонов позволяет безошибочно 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.

СПИСОК ЛИТЕРАТУРЫ

1. *Намиот Д. Е., Романов В. Ю.* Анализ данных для программных репозиториях // International Journal of Open Information Technologies. 2018. № 4. С. 18–23.
2. *Заруцкий С. А., Власенко Е. А.* Автоматизация анализа данных экспериментальных исследований // Электронный журнал ИВД. 2018. № 1 (48). URL: http://ivdon.ru/uploads/article/pdf/IVD_115_Zaruckij_Vlasenko.pdf_e7e45a41f4.pdf (дата обращения: 18.01.2020).
3. *Данилов В. В., Третьяков И. А., Шалаев А. В., Рушечников Я. И.* Алгоритмы идентификации переходных участков экспериментальных кривых с применением аппроксимации // Сб. науч. тр. Донец. ин-та железнодорож. транспорта. 2018. № 48. С. 19–23.
4. *Данилов В. В., Третьяков И. А., Рушечников Я. И.* Алгоритмизация присвоения символов анализируемым участкам экспериментальных кривых // Сб. науч. тр. Донец. ин-та железнодорож. транспорта. 2018. № 51. С. 15–22.
5. *Третьяков И. А., Данилов В. В.* Функции сложности для выделения и распознавания характерных участков экспериментальных кривых // Вестн. Донец. национ. ун-та. Сер. А: Естественные науки. 2017. № 2. С. 101–107.

Статья поступила в редакцию 16.04.2020

ИНФОРМАЦИЯ ОБ АВТОРАХ

Третьяков Игорь Александрович – Донецкая Народная Республика, 283001, Донецк; Донецкий национальный университет; зав. учебной лабораторией кафедры радиофизики и инфокоммуникационных технологий; es75i@mail.ru.

Данилов Владимир Васильевич – Донецкая Народная Республика, 283001, Донецк; Донецкий национальный университет; д-р техн. наук, профессор; зав. кафедрой радиофизики и инфокоммуникационных технологий; ut5iv@mail.ru.



RESEARCH OF RADIO FREQUENCY SPECTROGRAMS USING METHODS OF LINGUISTIC ANALYSIS

I. A. Tretiakov, V. V. Danilov

*Donetsk National University,
Donetsk, Donetsk People's Republic*

Abstract. The article presents the study of experimental curves of the spectral spectra of radio waves of the FM-range developed on a laboratory model. The algorithm was used to define sections, for which the complexity function took locally minimal values. The standard was determined for each

section of the curve, in which all the arithmetic mean ordinates of all areas correspond to the certain class. For a more extended linguistic description of the experimental curves, it is proposed to compile the description taking into account the location of the curve sections on the abscissa axis. The obtained extended linguistic description of the curve will reflect not only classes of simple events, but their phases as well. As a result of applying the linguistic analysis system for the analysis of spectral radiograms, it can be inferred that the experimental curves are presented in the form of short and reliable rules for the analysis of the radiogram spectrum. The use of standards allows to accurately represent each chain of characters in each group with a minimum distance to the standard. The obtained extended descriptions quite accurately describe the behavior of the curves studied.

Key words: data analysis, experimental curves, linguistic description, standard, chain of characters.

For citation: Tretiakov I. A., Danilov V. V. Research of radio frequency spectrograms using methods of linguistic analysis. *Vestnik of Astrakhan State Technical University. Series: Management, Computer Science and Informatics*. 2020;3:45-51. (In Russ.) DOI: 10.24143/2072-9502-2020-3-45-51.

REFERENCES

1. Namiot D. E., Romanov V. Iu. Analiz dannykh dlia programmnykh repozitoriev [Data analysis for software repositories]. *International Journal of Open Information Technologies*, 2018, no. 4, pp. 18-23.
2. Zarutskii S. A., Vlasenko E. A. Avtomatizatsiia analiza dannykh eksperimental'nykh issledovani [Automation of analysis of experimental research data]. *Elektronnyi zhurnal IVD*, 2018, no. 1 (48). Available at: http://ivdon.ru/uploads/article/pdf/IVD_115_Zaruckij_Vlasenko.pdf_e7e45a41f4.pdf (accessed: 18.01.2020).
3. Danilov V. V., Tret'iakov I. A., Shalaev A. V., Rushechnikov Ia. I. Algoritmy identifikatsii perekhodnykh uchastkov eksperimental'nykh krivykh s primeneniem approksimatsii [Algorithms for identifying transition sections of experimental curves using approximation]. *Sbornik nauchnykh trudov Donetskogo instituta zheleznodorozhnogo transporta*, 2018, no. 48, pp. 19-23.
4. Danilov V. V., Tret'iakov I. A., Rushechnikov Ia. I. Algoritmizatsiia prisvoeniia simvolov analiziruemym uchastkam eksperimental'nykh krivykh [Algorithmization of character assignment to analyzed sections of experimental curves]. *Sbornik nauchnykh trudov Donetskogo instituta zheleznodorozhnogo transporta*, 2018, no. 51, pp. 15-22.
5. Tret'iakov I. A., Danilov V. V. Funktsii slozhnosti dlia vydeleniia i raspoznavaniia kharakternykh uchastkov eksperimental'nykh krivykh [Complexity functions for isolating and recognizing characteristic portions of experimental curves]. *Vestnik Donetskogo natsional'nogo universiteta. Seriya A: Estestvennye nauki*, 2017, no. 2, pp. 101-107.

The article submitted to the editors 16.04.2020

INFORMATION ABOUT THE AUTHORS

Tretiakov Igor Aleksandrovich – Donetsk People's Republic, 283001, Donetsk; Donetsk National University; Head of the Laboratory at the Department of Radiophysics and Infocommunication Technologies; es75i@mail.ru.

Danilov Vladimir Vasilevich – Donetsk People's Republic, 283001, Donetsk; Donetsk National University; Doctor of Technical Sciences, Professor; Head of the Department of Radiophysics and Infocommunication Technologies; ut5iv@mail.ru.

