

Научная статья
УДК 004.056:004.93
<https://doi.org/10.24143/2072-9502-2023-1-64-70>
EDN IINRKY

Структурно-лингвистический подход к описанию спектрограмм электромагнитных излучений, создаваемых вычислительной техникой

Игорь Александрович Третьяков

*Донецкий национальный университет,
Донецк, Россия, i.tretiakov@mail.ru*

Аннотация. Рассмотрено решение научно-технической задачи в области информационной безопасности – задачи обнаружения электромагнитных излучений, создаваемых вычислительной техникой, – в рамках структурно-лингвистического подхода к анализу экспериментальных данных. Подробно рассмотрена реализация этапа присвоения сегментированным участкам спектрограмм символов некоторого алфавита, соответствующих определенным типам поведения, а также реализация этапа анализа полученных последовательностей символов. Описаны модель построения морфологической грамматики и механизм генерации текстов, мало отличающихся от правильных. Этап формирования грамматики был реализован на языке программирования Python 3.7. Выбор данного языка программирования обосновывается его кроссплатформенностью, низким порогом вхождения, а также широким спектром применения: от автоматизации математических вычислений и машинного обучения до разработки веб-приложений. Также одним из преимуществ этого языка является наличие множества качественных библиотек, в том числе используемых для текущей разработки. Для оценки эффективности и точности разработанных алгоритмов проведен статистический эксперимент. Приведены оценка вероятности верной классификации лингвистических цепочек экспериментальных кривых спектрограмм в нужные группы и общая оценка точности верного определения всех кривых в нужные группы. В результате применения структурно-лингвистического подхода к задаче обнаружения электромагнитных излучений, создаваемых вычислительной техникой, можно заключить, что полученные лингвистические описания исследуемых спектрограмм представляют собой короткие и надежные правила для их анализа и позволяют с высокой точностью в автоматизированном режиме определять отклонения в наблюдаемых спектрах от заданных эталонов.

Ключевые слова: вычислительная техника, электромагнитные излучения, структурно-лингвистический подход, спектрограмма, формирование грамматики, классификация

Для цитирования: *Третьяков И. А.* Структурно-лингвистический подход к описанию спектрограмм электромагнитных излучений, создаваемых вычислительной техникой // Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика. 2023. № 1. С. 64–70. <https://doi.org/10.24143/2072-9502-2023-1-64-70>. EDN IINRKY.

Original article

Structural-linguistic approach to the description of spectrogram of electromagnetic radiation spectrograms generated by computing technology

Igor A. Tretiakov

*Donetsk National University,
Donetsk, Russia, i.tretiakov@mail.ru*

Abstract. In this paper, the solution of an urgent scientific and technical problem in the field of information security is considered – the problem of detecting electromagnetic radiation generated by computer technology, within the framework of a structural and linguistic approach to the analysis of experimental data. The implementation of the stage of assigning segmented sections of spectrograms of symbols of some alphabet corresponding to certain types of behavior is considered in detail. As well as the implementation of the analysis stage of the received sequences of sym-

bols. A model of morphological grammar construction and a mechanism for generating texts that differ little from the correct ones are described. The grammar formation stage was implemented in the Python 3.7 programming language. The choice of this programming language is justified by its cross-platform nature, low entry threshold, as well as wide application: from automation of mathematical calculations and machine learning, to the development of web applications. Also, one of the advantages of this language is the availability of many high-quality libraries, including those used for current development. A statistical experiment was conducted to evaluate the effectiveness and accuracy of the developed algorithms. An estimate of the probability of the correct classification of linguistic chains of experimental spectrogram curves into the desired groups and a general assessment of the accuracy of the correct determination of all curves into the desired groups is given. As a result of the application of a structural-linguistic approach to the problem of detecting electromagnetic radiation generated by computer technology, it can be concluded that the obtained linguistic descriptions of the studied spectrograms represent short and reliable rules for their analysis and allow for high accuracy in automated mode to determine deviations in the observed spectra from the specified standards.

Key words: computer technology, electromagnetic radiation, structural-linguistic approach, spectrogram, grammar formation, classification

For citation: Tretiakov I. A. Structural-linguistic approach to the description of spectrogram of electromagnetic radiation spectrograms generated by computing technology. *Vestnik of Astrakhan State Technical University. Series: Management, Computer Science and Informatics*. 2023; 1:64-70. (In Russ.). <https://doi.org/10.24143/2072-9502-2023-1-64-70>. EDN IINRKY.

Введение

В настоящее время актуальной научно-технической задачей в области информационной безопасности является задача обнаружения электромагнитных излучений, создаваемых вычислительной техникой [1–5]. В данной работе рассмотрено решение этой задачи в рамках структурно-лингвистического подхода к анализу экспериментальных данных, основные идеи которого рассмотрены в [6, 7].

Структурно-лингвистический подход предполагает последовательность реализации трех основных этапов обработки спектрограмм: выделения и распознавания характерных участков (сегментация); присвоения выделенным участкам символов некоторого алфавита, соответствующих определенным типам поведения кривой (формирование грамматики); анализа полученных последовательностей символов.

Первый этап реализован и представлен авторами в [8, 9]. Данная статья является продолжением исследований и посвящена реализации второго и третьего вышеуказанных этапов. В рамках второго этапа набор присваиваемых символов представляет собой алфавит, в котором компоненты являются кодовыми обозначениями поведения кривой на каждом участке. Для формирования такого алфавита необходимо применять алгоритмы автоматической классификации, которые будут осуществлять распределение массивов векторов на классы, количество которых определяется самим алфавитом, и устанавливать критерии, по которым каждый новый вектор будет распределен в тот или иной класс, иными словами, присваивать им конкретные символы.

В качестве массива экспериментальных данных для реализации процедуры формирования грамматики послужили спектрограммы электромагнитных излучений, полученные с помощью программно-определяемой радиосистемы на базе RTL2832 и R820T в ГОУ ВПО «Донецкий национальный университет».

Программно-определяемые радиосистемы (SDR) представляют собой набор программных и специальных аппаратных средств, которые позволяют решать круг задач анализа взаимодействия радиоизлучений в широком диапазоне частот. Некоторые SDR системы могут использоваться для решения задач мониторинга спектра радиочастот [10, 11]. Данный класс устройств, благодаря возможностям программного управления, реализует функции физического уровня, что обеспечивает возможность обработки различных типов сигналов без изменения аппаратной части принимающего устройства.

На рис. 1 показан процесс выявления и регистрации электромагнитных излучений и образец получаемых спектрограмм.

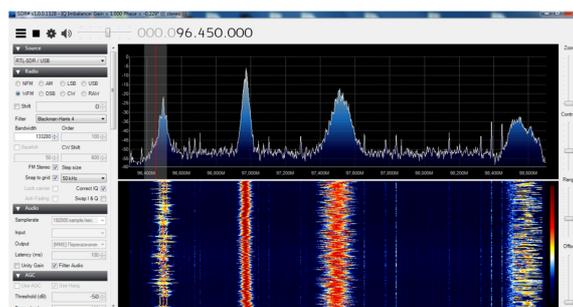


Рис. 1. Визуализация спектра радиобстановки

Fig. 1. Visualization of the radio spectrum

Формальная постановка задачи

В результате выполнения первого этапа, а именно последовательного применения алгоритмов сегментации и классификации выделенных фрагментов [8, 9], анализируемая спектрограмма визуализируется в виде ряда из символов некоторого конечного алфавита. Ряд такого вида целесообразно рассматривать как некоторый текст T на

неизвестном языке J (специальным образом адаптированным для точного описания анализируемых спектрограмм), а каждую последовательность символов текста T (цепочку) целесообразно рассматривать как некоторую фразу на данном языке.

Для описания получаемых спектрограмм зададим следующий алфавит: нарастание сигнала (передний фронт) обозначим символом L (left), высокий уровень сигнала (центральная часть) – C (central), спад сигнала (задний фронт) – R (right). Учет фоновых (неинформативных) участков в данном этапе не рассматривается. Таким образом, сформирован алфавит языка описания исследуемых спектрограмм, состоящий из трех символов $A = \{L, C, R\}$.

С такой точки зрения задачу формирования языка описания спектрограмм электромагнитных излучений, создаваемых вычислительной техникой, можно сформулировать следующим образом:

- задано некоторое количество текстов T в виде массива рядов символов;
- необходимо сформировать словарь D из отдельных лингвистических единиц (стабильных последовательностей символов);
- необходимо представить каждый из обрабатываемых текстов T в виде одного или нескольких слов этого словаря D .

Учитывая вышесказанное, пусть имеется всего один текст $T = \langle t_1, \dots, t_N \rangle$ большой длины. Каждая пара индексов (i, j) , $i \leq j$, $1 \leq i, j \leq N$ исключает из текста T некоторый сегмент последовательности, образующийся посредством удаления в T символов с индексами, меньшими i и большими j : $s(i, j) = \langle t_i, t_{i+1}, \dots, t_j \rangle$. Каждому сегменту s соответствует образ – ряд его символов, у которых удалены данные о местоположении этого сегмента (индексы) в тексте T . Очевидно, что в тексте может иметься несколько сегментов, соответствующих одному образу. Необходимо найти такие наборы образов, из которых можно составить заданный текст T . Один из таких наборов и есть искомый словарь D .

Процедура формирования грамматики

Разделение S текста T на непересекающиеся сегменты формирует некоторый словарь $D(S)$. Такое формирование словаря $D(S)$ можно рассматривать как выявление набора триггеров изменений состояний исследуемого процесса, которые отражены в символическом представлении спектрограмм в виде стабильных последовательностей символов. Например, в задаче идентификации речи в условиях помех [12] такого рода символы можно рассматривать как минимальные смысловозначительные единицы языка (фонемы). В таких услови-

ях выделение фонем в сигнале речи рассматривается не как процедура сегментации, а как процедура его структурно-лингвистического анализа.

Пусть D – словарь, в котором J – множество всех конечных последовательностей символов T из алфавита A (язык), тогда подмножество $\hat{J}(D) \subset J$ является подмножеством всех последовательностей T , представляющих собой последовательность слов из словаря D . Таким образом, $\hat{J}(D)$ – язык, детерминированный словарем D . Последовательность T считается правильной последовательностью языка $\hat{J}(D)$ тогда и только тогда, когда имеется такое ее разбиение S , что образ каждого сегмента из S является словом из D . Любой такой язык, каждая последовательность слов из D которого является правильной последовательностью, считается языком с морфологической грамматикой. Для использования такой морфологической грамматики при анализе других текстов T , т. е. других экспериментальных данных, необходимо применять процедуру преобразования символьного текста в упорядоченный набор слов или же процедуру разбиения текста T на слова из словаря D . Иными словами, необходима процедура верного распознавания этого текста на языке $\hat{J}(D)$.

Однако исследуемые спектрограммы не всегда представляются в виде правильных текстов достаточно точно. В контексте рассматриваемой задачи такая процедура осуществляется с целью автоматизации анализа экспериментальных данных спектрограмм, которые изменяют свое состояние случайно, т. е. без какой-либо определенной закономерности, в том числе и применение алгоритмов сегментации и классификации на предшествующем этапе может вносить эффект случайности в формируемые лингвистические последовательности. Исходя из этого такую морфологическую грамматику следует расценивать как модель выбора оптимальной аппроксимирующей функции для текстов, которые формируются экспериментальными данными спектрограмм.

Для полноценного ее использования процедуру генерации правильных текстов необходимо расширить некой искажающей процедурой, которая будет генерировать последовательности, мало отличающиеся от правильных. Для ее реализации следует ввести меру сходства двух случайных последовательностей из одного алфавита A , которая будет являться степенью искажения при переходе между последовательностями. В качестве процедуры искажения может применяться трансформационная грамматика, которая будет содержать некое множество элементарных трансформаций, т. е. минимальных искажений. Процедура построения трансформационной грамматики подробно описана в [6, 7]. Тогда мерой сходства между двумя последовательностями целесообразно выбрать

Таблица 1

Table 1

**Оценка вероятности верной классификации
Estimation of the probability of correct classification**

Группа	1 группа		2 группа		3 группа		4 группа	
	<i>n</i> , %	<i>A</i> , %						
1 группа	66	100	80	70	100	50	6	40
	20	90	14	60			88	30
	14	80	6	50			6	10
2 группа	20	70	60	100	6	60	100	30
	60	60	26	90	80	50		
	20	50	14	80	14	40		
3 группа	100	50	60	60	54	100	60	50
			20	50	40	90	40	40
			20	40	6	80		
4 группа	6	40	6	40	6	60	60	100
	74	30	86	30	74	50	26	90
	20	10	6	20	14	40	14	80

В табл. 1 по вертикали и горизонтали приведены наименования групп, на пересечениях указаны вероятности и точность верной классификации кривых исследуемых спектрограмм из одной группы в другую. Таким образом, 66 % кривых первой группы со 100 % точностью определены в первую группу, 20 % – с 90 % точностью, 14 % – с 80 % точностью и т. д. В качестве критерия точности

принята мера сходства (1) анализируемых цепочек с эталоном группы, где 100 % – полное совпадение. Из анализа табл. 1 следует, что минимальная точность, при которой кривая верно определена в нужную группу, составляет 80 %.

В табл. 2 приведена общая оценка точности верного определения всех кривых в нужные группы.

Таблица 2

Table 2

**Общая оценка точности верной классификации
General assessment of the accuracy of the correct classification**

Группа	1 группа, <i>A</i> , %	2 группа, <i>A</i> , %	3 группа, <i>A</i> , %	4 группа, <i>A</i> , %
1 группа	90	60	50	27
2 группа	60	90	50	30
3 группа	50	50	90	45
4 группа	27	30	45	90

В табл. 2 по вертикали и горизонтали приведены наименования групп, на пересечениях указана точность верной классификации кривых исследуемых спектрограмм из одной группы в другую. Таким образом, кривые из первой группы верно определены в первую группу с точностью 90 %, во вторую – 60 %, в третью – 50 %, в четвертую – 27 % и т. д. Из анализа табл. 2 следует, что точность верного определения кривой в нужную группу составляет 90 %.

Заключение

В результате применения структурно-лингвистического подхода к задаче обнаружения электромагнитных излучений, создаваемых вычислительной техникой, можно заключить, что полученные лингвистические описания исследуемых спектрограмм представляют собой короткие и надежные правила для их анализа и позволяют с высокой точностью в автоматизированном режиме определять отклонения в наблюдаемых спектрах от заданных эталонов.

Список источников

1. Фаустов И. С., Токарев А. Б., Сладких В. А., Козьмин В. А., Крыжко И. Б. Радиоконтроль служебных параметров сигналов Bluetooth // Системы управления, связи и безопасности. 2021. № 3. С. 135–151.

2. Ашихмин А. В., Виноградов А. Д., Рембовский А. М., Сладких В. А. Способ однопозиционного местоопределения источников радиоизлучения с использованием бортового радиопеленгатора беспилотного летательного

аппарата вертолетного типа // Системы управления, связи и безопасности. 2021. № 4. С. 40–57.

3. Третьяков И. А., Данилов В. В. Спектральный анализ радиосигналов в реальном времени на основе применения эхо-эффекта // Вестн. Астрахан. гос. техн. ун-та. Сер.: Управление, вычислительная техника и информатика. 2022. № 1. С. 53–59.

4. Третьяков И. А., Данилов В. В. Исследование спектрограмм радиочастот методами лингвистического анализа // Вестн. Астрахан. гос. техн. ун-та. Сер.: Управление, вычислительная техника и информатика. 2020. № 3. С. 26–33.

5. Рушечников Я. И., Третьяков И. А. Автоматизация процедуры обнаружения технических каналов утечки информации и побочных электромагнитных излучений // Взаимодействие вузов, научных организаций и учреждений культуры в сфере защиты информации и технологий безопасности: сб. ст. по материалам Междунар. конф., посвящ. памяти д-ра техн. наук, проф. А. А. Тарасова и д-ра техн. наук, старшего науч. сотр. О. В. Казарина (Москва, 19–20 апреля 2022 г.) / под ред. Д. А. Митюшина. М.: Изд-во РГГУ, 2022. С. 240–245.

6. Моттль В. В., Мучник И. Б. Лингвистический анализ экспериментальных кривых // ТИИЭР. 1979. Т. 69. № 5. С. 12–39.

7. Чистова Г. К., Волчихин В. И. Методы и процедуры построения лингвистической системы обнаружения и распознавания нарушителя // Вестн. МГТУ им. Н. Э. Баумана. Сер.: Приборостроение. 2004. № 3. С. 96–114.

8. Данилов В. В., Третьяков И. А., Шалаев А. В., Рушечников Я. И. Алгоритмы идентификации переходных участков экспериментальных кривых с применением аппроксимации // Сб. науч. тр. ДОНИЖТ. 2018. № 48. С. 19–23.

9. Данилов В. В., Третьяков И. А., Рушечников Я. И. Алгоритмизация присвоения символов анализируемым участкам экспериментальных кривых // Сб. науч. тр. ДОНИЖТ. 2018. № 51. С. 15–22.

10. Рушечников Я. И., Данилов В. В. Информационная технология радиомониторинга на основе программно-определяемой радиосистемы // Вестн. Донец. национ. ун-та. Сер. Г: Технические науки. 2020. № 1. С. 31–36.

11. Рушечников Я. И., Данилов В. В., Борщевский С. В. Информационная технология автоматизированной локализации источника излучения // Вестн. Донец. национ. ун-та. Сер. Г: Технические науки. 2020. № 4. С. 26–34.

12. Третьяков И. А., Кожекина Е. Н., Сыровацкий В. И. Текстонезависимая идентификация речи в условиях помех // Вестн. Донец. национ. ун-та. Сер. Г: Технические науки. 2022. № 2. С. 64–77.

References

1. Faustov I. S., Tokarev A. B., Sladkikh V. A., Kozmin V. A., Kryzhko I. B. Radiokontrol sluzhebnykh parametrov signalov Bluetooth [Radio monitoring of service parameters of Bluetooth signals]. *Sistemy upravleniia svyazi i bezopasnosti*, 2021, no. 3, pp. 135-151.

2. Ashikhmin A. V., Vinogradov A. D., Rembovskii A. M., Sladkikh V. A. Sposob odnopolozitsionnogo mestoopredeleniia istochnikov radioizlucheniia s ispolzovaniem bortovogo radiopelengatora bespilotnogo letatel'nogo apparata vertoletnogo tipa [Method of single-position location of radio emission sources using an on-board radio direction finder of an unmanned aerial vehicle of helicopter type]. *Sistemy upravleniia svyazi i bezopasnosti*, 2021, no. 4, pp. 40-57.

3. Tretiakov I. A., Danilov V. V. Spektralnyi analiz radiosignalov v realnom vremeni na osnove primeneniia ekho-effekta [Spectral analysis of radio signals in real time based on the application of the echo effect]. *Vestnik Astrakhanskogo gosudarstvennogo tekhnicheskogo universiteta. Seriya: Upravlenie vychislitelnaia tekhnika i informatika*, 2022, no. 1, pp. 53-59.

4. Tretiakov I. A., Danilov V. V., Issledovanie spektrogramm radiochastot metodami lingvisticheskogo analiza [Investigation of radio frequency spectrograms by methods of linguistic analysis]. *Vestnik Astrakhanskogo gosudarstvennogo tekhnicheskogo universiteta. Seriya: Upravlenie vychislitelnaia tekhnika i informatika*, 2020, no. 3, pp. 26-33.

5. Rushechnikov Ia. A., Tretiakov I. A. Avtomatizatsiia protsedury obnaruzheniia tekhnicheskikh kanalov utechki informatsii i pobochnykh elektromagnitnykh izlucheniia [Automation of the procedure for detecting technical channels of information leakage and side electromagnetic radiation]. *Vzaimodeistvie vuzov nauchnykh organizatsii i uchrezhdenii kultury v sfere zashchity informatsii i tekhnologii bezopasnosti: sbornik statei po materialam Mezhdunarodnoi konferentsii, posviashchennoi pamiatii doktora tekhnicheskikh nauk, professora A. A. Tarasova i doktora tekhnicheskikh*

nauk, starshego nauchnogo sotrudnika O. V. Kazarina (Moskva, 19-20 aprelia 2022 g.). Pod redaktsiei D. A. Mitushina. Moscow, Izd-vo RGGU, 2022. Pp. 240-245.

6. Mottl V. V., Muchnik I. B., Lingvisticheskii analiz eksperimentalnykh krivykh [Linguistic analysis of experimental curves]. *TIIEER*, 1979, vol. 69, no. 5, pp. 12-39.

7. Chistova G. K., Volchikhin V. I. Metody i protsedury postroeniia lingvisticheskoi sistemy obnaruzheniia i raspoznavaniia narushitel'ia [Methods and procedures for constructing a linguistic system for detecting and recognizing an intruder]. *Vestnik MGTU im N. E. Baumana. Seriya: Priboroostroenie*, 2004, no. 3, pp. 96-114.

8. Danilov V. V., Tretiakov I. A., Shalaev A. V., Rushechnikov Ia. I. Algoritmy identifikatsii perekhodnykh uchastkov eksperimentalnykh krivykh s primeneniem approksimatsii [Algorithms for identification of transitional sections of experimental curves using approximation]. *Sbornik nauchnykh trudov DONIZHT*, 2018, no. 48, pp. 19-23.

9. Danilov V. V., Tretiakov I. A., Rushechnikov Ia. I. Algoritmizatsiia prisvoeniia simvolov analiziruemykh uchastkam eksperimentalnykh krivykh [Algorithmization of the assignment of symbols to the analyzed sections of experimental curves]. *Sbornik nauchnykh trudov DONIZHT*, 2018, no. 51, pp. 15-22.

10. Rushechnikov Ia. I., Danilov V. V. Informatsionnaia tekhnologiia radiomonitoringa na osnove programmno-opredeliaemoi radiosistemy [Information technology of radio monitoring based on a software-defined radio system]. *Vestnik Donetskogo natsionalnogo universiteta. Seriya G: Tekhnicheskii nauki*, 2020, no. 1, pp. 31-36.

11. Rushechnikov Ia. I., Danilov V. V., Borshchevskii S. V. Informatsionnaia tekhnologiia avtomatizirovannoi lokalizatsii istochnika izlucheniia [Information technology of automated radiation source localization]. *Vestnik Donetskogo natsionalnogo universiteta. Seriya G: Tekhnicheskii nauki*, 2020, no. 4, pp. 26-34.

12. Tretiakov I. A., Kozhekina E. N., Syrovatskii V. I. Tekstonezavisimaa identifikatsiia rechi v usloviakh pomekh [Text-independent speech identification under interference

conditions]. *Vestnik Donetskogo natsionalnogo universiteta. Seriya G: Tekhnicheskie nauki*, 2022, no. 2, pp. 64-77.

Статья поступила в редакцию 14.12.2022; одобрена после рецензирования 27.12.2022; принята к публикации 18.01.2023
The article is submitted 14.12.2022; approved after reviewing 27.12.2022; accepted for publication 18.01.2023

Информация об авторе / Information about the author

Игорь Александрович Третьяков – кандидат технических наук; доцент кафедры радиофизики и инфокоммуникационных технологий; Донецкий национальный университет; i.tretiakov@mail.ru

Igor A. Tretiakov – Candidate of Sciences in Technology; Assistant Professor of the Department of Radiophysics and Infocommunication Technologies; Donetsk National University; i.tretiakov@mail.ru

