
**СИСТЕМНЫЙ АНАЛИЗ ПРИБОРОВ
И ИЗМЕРИТЕЛЬНЫХ МЕТОДИК**

УДК 543.51+ 681.2–5

© В. В. Манойлов, А. Г. Кузьмин, И. В. Заруцкий, Ю. А. Титов, Н. С. Самсонова, 2019

**МЕТОДЫ ОБРАБОТКИ И ИССЛЕДОВАНИЕ ВОЗМОЖНОСТЕЙ
КЛАССИФИКАЦИИ МАСС-СПЕКТРОВ ВЫДЫХАЕМЫХ ГАЗОВ**

Анализ состава выдыхаемого воздуха является одним из важных направлений неинвазивной медицины. В качестве примера использования данного метода в научном приборостроении рассматривается математическая обработка масс-спектров, полученных на квадрупольном масс-спектрометре МС7-200. Описываемые в работе методы обработки данных на основе линейного и квадратичного дискриминантного анализа позволяют разделить обрабатываемые масс-спектры выдыхаемых газов на две группы: масс-спектры выдыхаемых газов здоровых людей и масс-спектры выдыхаемых газов людей с возможными патологиями.

Кл. сл.: математические методы обработки данных, масс-спектрометры, дискриминантный анализ

ВВЕДЕНИЕ

Анализ состава выдыхаемого воздуха может использоваться в целях выявления заболеваний, предсказания реакции организма на конкретный вид лечения и мониторинга эффективности терапии. Диагностика на основе анализа выдыхаемых газов имеет ряд преимуществ по сравнению с традиционными лабораторными методами. Анализ газовой смеси безопасен для персонала, т.к. не связан с работой с химическими и биологическими жидкостями. Он является относительно дешевым, занимает немного времени и позволяет обнаруживать детектируемые вещества в минимальных концентрациях.

В качестве примера использования данного метода в аналитическом приборостроении рассматривается обработка масс-спектров, полученных на квадрупольном масс-спектрометре МС7-200 (рис. 1) с электронной ионизацией и прямым капиллярным вводом пробы при анализе наличия патологий у пациентов по спектрам выдыхаемых газов [1–5]. Забор выдыхаемого воздуха осуществлялся в специальные тестовые емкости, после чего производился анализ.

**КРАТКАЯ ИНФОРМАЦИЯ
О МАСС-СПЕКТРОМЕТРЕ МС7-200****Решаемые задачи:**

- анализ состава выдыхаемого воздуха для здоровых и, возможно, больных людей,
- экологический анализ состава воздуха в населенных пунктах и на промышленных предприятиях,
- оценка состава газов в производственных процессах.

Описание работы прибора:

Анализируемый газ при давлении атмосферы через капиллярный ввод подается в камеру ионизации источника ионов с электронным ударом. Получившиеся ионы вводятся в масс-анализатор квадрупольного типа. Получившиеся в процессе регистрации масс-спектрометрические сигналы обрабатываются с помощью специализированного программного обеспечения и сравниваются со спектрами в библиотеке стандартных масс-спектров, затем проводится идентификация отдельных компонент спектра и определение их концентрации. Капиллярная система ввода пробы в масс-спектрометр позволяет проводить анализ на расстоянии до 5 м от прибора. Для анализа используется до 10 мл пробы в мин. При некоторых анализах проба нагревается. В вакуумной системе используется диффузионный или турбомолекулярный насос.

Основные технические параметры:

- диапазон регистрируемых масс:— от 2 до 300 а. е. м.,
- разрешение по массовым числам — 1,
- скорость регистрации — до 1 масс-спектра в с.

Области применения прибора:

- медицина — анализ состава выдыхаемых газов и процессов их изменения при дыхании, диагностика патологий по наличию компонент метаболизма в выдыхаемых пробах, санитарно-эпидемиологический мониторинг;
- экология — оценка загрязнений в окружающем воздухе;
- промышленность — анализ выбросов в технологических процессах вредных веществ, таких как толуол, бензол, стирол и пр.



Рис. 1. Масс-спектрометр MS7-200

ОБРАБОТКА ДАННЫХ

При проведении массовых скрининговых профилактических осмотров с использованием масс-спектрометра MS7-200 актуальной представляется задача автоматической классификации людей по группам здоровья. В данной работе для такой классификации людей использованы методы линейного и квадратичного дискриминантного анализа масс-спектров выдыхаемых газов. Непосредственно перед классификацией в дискриминантном анализе необходимо сначала провести вычисление дискриминационных коэффициентов с помощью обучающей выборки. Обучающая выборка состоит из двух групп: первая группа — это масс-спектры выдыхаемых газов здоровых людей и вторая группа — это масс-спектры людей, имеющих патологии. После того как проведено обучение, приступают непосредственно к классификации.

Дискриминантными признаками для классификации служат значения амплитуд на определенных массах линейчатого масс-спектра (массы: 53 — изопрен, 58 — ацетон, 60 — уксус, 67 и 68 — изопрен), а также два отношения амплитуд пиков. Первым отношением является отношение амплитуды пика на массе 58 к амплитуде пика на массе 60. Вторым отношением является отношение амплитуды пика на массе 67 к амплитуде пика на массе 58.

Первым этапом обработки масс-спектров является обнаружение пиков и превращение непрерывного спектра в линейчатый. Для их обнаруже-

ния используется метод поиска на основе согласованной фильтрации [6]. В этом методе используется свертка исходного сигнала с сигналом, описывающим форму пика. Свертка вычисляется следующим образом:

$$s_1(t) = \int_0^T f_{out}(t) g(t - \tau) d\tau, \quad (1)$$

где t — независимая переменная, время, $f_{out}(t)$ — зарегистрированный масс-спектр, $g(t)$ — форма пика, T — длина масс-спектра.

В качестве формы пика в данной работе использовались две функции: первая — на основе гауссианы $g(t) = \exp\left[-(t/w)^2\right]$, где w — средняя полуширина пика в масс-спектре, и вторая — на

основе функции Эрмита $\psi(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2^n n! \sqrt{\pi}}} H_n(t)$,

где $H_n(t) = (-1)^n e^{t^2} \frac{d^n e^{-t^2}}{dt^n}$ — полином Эрмита порядка n . В данной работе $n = 2$.

В результате такой конволюции амплитуды пиков возрастают примерно в 20–25 раз, а среднее квадратичное отклонение шума примерно в 3–5 раз для гауссовых функций и 10–15 раз для функций Эрмита. Функции Эрмита дают преимущество в повышении разрешения плохо разрешенных пиков по сравнению функциями Гаусса.

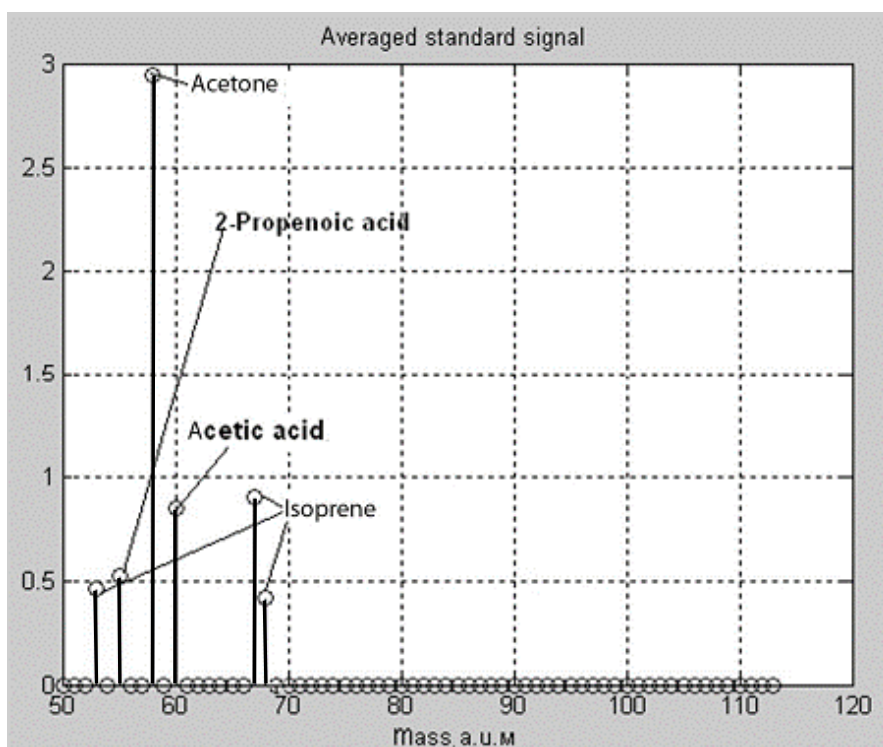


Рис. 2. Эталонный масс-спектр

Процедура свертки позволяет надежно обнаружить пики спектра и определить их положение. При вычислении интеграла (1) используется теорема Планшереля, согласно которой преобразование фурье-свертки двух сигналов равно произведению их фурье-образов. После этого сигнал $s_1(t)$ вычисляется с помощью обратного преобразования Фурье.

Применение прямого и обратного преобразования Фурье использовано в данной работе, т.к. вычисление интеграла свертки по формуле (1) требует гораздо большего объема вычислений, чем выполнение прямого и обратного преобразований Фурье с помощью алгоритмов быстрого преобразования. Полученный сигнал $s_1(t)$ сравнивается с порогом h_0 , с помощью которого обнаруживаются пики и определяются их положения. Если величина $s_1(t)$ превышает порог h_0 , то эту точку t считаем принадлежащей пику.

Следующим шагом обработки является отбор масс-спектров здоровых людей для их включения в обучающую выборку. Необходимость выполнения этого шага объясняется тем, что даже в масс-спектрах практически здоровых людей могут содержаться спектральные линии с амплитудами, характерными для масс-спектров больных людей. Выполнение этого шага осуществляется путем вы-

явления масс-спектров, содержащих наименьшее отклонение от эталонного масс-спектра, т.е. усредненного масс-спектра здоровых людей. На рис. 2 представлен эталонный масс-спектр. Наименьшее отклонение от эталонного масс-спектра вычислялось по следующим алгоритмам: евклидово расстояние (N2), "сити-блок" — сумма модулей разности амплитуд по каждой массе (N1), максимальная разность компонент. На основании проведенных вычислений была составлена обучающая выборка масс-спектров.

Результатом проведения обучения является набор коэффициентов дискриминантной функции, которые вычисляются по одному из алгоритмов многомерной статистики, рассмотренных в [6]. В качестве переменных в алгоритмах многомерной статистики использовались не исходные значения амплитуд пиков на определенных массах, а главные компоненты, полученные путем преобразования исходных данных по методу главных компонент (Principal Component Analysis — PCA).

С помощью вычисленных коэффициентов дискриминационных функций находятся границы, которые разделяют исходные масс-спектры выделяемых газов на группы: здоровых людей и людей с патологиями. На рис. 3 показаны результаты проведения дискриминантного анализа соответственно для линейного (рис. 3, а) и квадратичного методов (рис. 3, б).

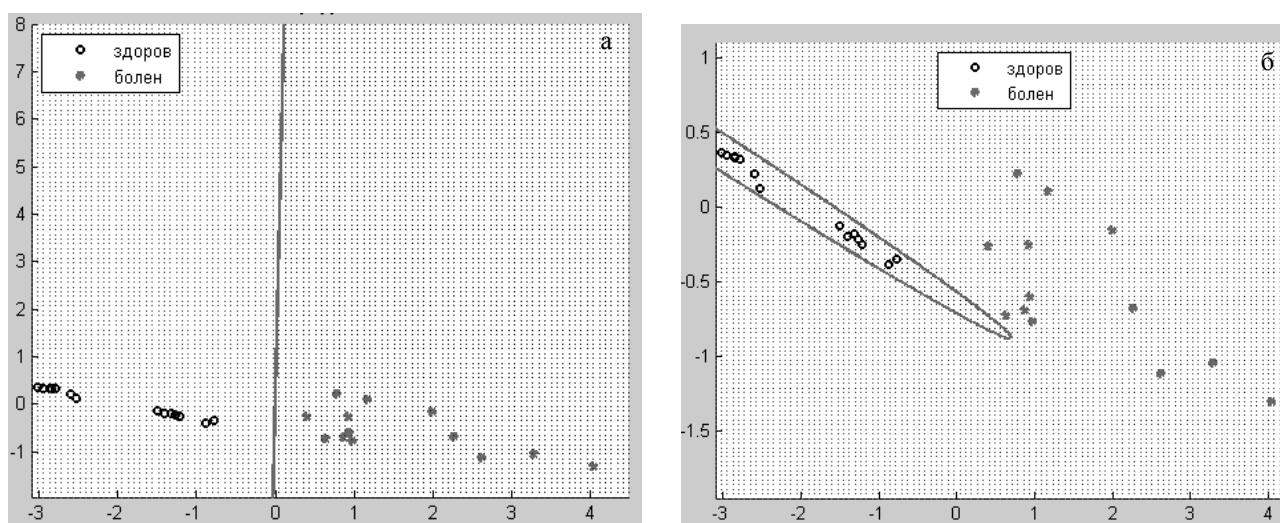


Рис. 3. Результат дискриминантного анализа.

По горизонтальной оси — значения переменных по первой главной компоненте, по вертикальной оси — значения переменных по второй компоненте.

а — результат линейного дискриминантного анализа, б — результат квадратичного дискриминантного анализа

ЗАКЛЮЧЕНИЕ

Рассмотренные алгоритмы дают возможность автоматического принятия решений о различии масс-спектров без визуального анализа информации, представленной в графическом виде, и имеют следующие преимущества: простота реализации, возможность автоматического принятия решения о принадлежности проверяемого сигнала к определенному классу.

Работа выполнена в ИАП РАН в рамках государственного задания № АААА-А16-116041310010-6.

СПИСОК ЛИТЕРАТУРЫ

1. Кузьмин А.Г., Ткаченко Е.И., Орешко Л.С., Титов Ю.А. Перспективы метода масс-спектрометрической ароматодиагностики по составу выдыхаемого воздуха // Тезисы докладов X Евразийской научной конференции "ДОНОЗОЛОГИЯ-2014", 18–19 декабря 2014 г., СПб., 2014. С. 229–231
2. Патент РФ № 94763, 27.05.2010.
3. Кузьмин А.Г., Титов Ю.А. Малогабаритные масс-спектрометры для динамических исследований состава выдыхаемого воздуха // Труды I Международной научно-практической конференции "Высокие технологии, фундаментальные и прикладные исследования в физиологии и медицине". Ч. 3. СПб., 23–26 ноября 2010 г. Изд-во СПбГПУ, 2010. С. 266–270.
4. Кузьмин А.Г., Ткаченко Е.И., Орешко Л.С., Титов Ю.А., Балабанов А.С. Метод масс-спектрометрической экспресс-диагностики по составу выдыхаемого воздуха // Медицинский академический журнал. 2016. Т. 16, № 4. С. 106–107.
5. Kuzmin A.G., Tkachenko E.I., Oreshko L.S., Titov Yu.A., Balabanov A.S. The method of medical instant diagnostics based on real-time mass-spectrometric analysis of exhaled air composition // 3rd ICMM PAN-ASIA PACIFIC CONGRESS ON MILITARY MEDICINE. Abstracts. 08.08.2016–12.08.2016. St.-Petersburg. P. 181–182.
6. Манойлов В.В., Титов Ю.А., Кузьмин А.Г., Заруцкий И.В. Методы обработки и классификации масс-спектров выдыхаемых газов с использованием дискриминантного анализа // Научное приборостроение. 2016. Т. 26, № 3. С. 50–57. URL: <http://iairas.ru/mag/2016/full3/Art7.pdf>

**Институт аналитического приборостроения РАН,
г. Санкт-Петербург** (Манойлов В.В., Кузьмин А.Г.,
Заруцкий И.В., Титов Ю.А., Самсонова Н.С.)

Контакты: Самсонова Наталья Сергеевна,
kolomna.88@mail.ru;
Манойлов Владимир Владимирович,
manoilov_vv@mail.ru

**Физико-технический институт имени А. Ф. Иоффе
РАН, г. Санкт-Петербург** (Самсонова Н.С.)

Материал поступил в редакцию 27.07.2018

METHODS OF PROCESSING AND INVESTIGATION OF THE POSSIBILITIES OF CLASSIFICATION OF MASS SPECTRA OF EXHALED GASES

V. V. Manoilov¹, A. G. Kuzmin¹, I. V. Zarutskiy¹, Yu. A. Titov¹, N. S. Samsonova^{1,2}

¹*Institute for Analytical Instrumentation of RAS, Saint-Petersburg, Russia*

²*The Ioffe Institute of RAS, Saint-Petersburg, Russia*

Analysis of exhaled air is one of the important areas of non-invasive medicine. As an example of the use of this method in scientific instrumentation, mathematical processing of mass spectra obtained on a quadrupole mass spectrometer MC7-200 is considered. The methods of data processing described on the basis of linear and quadratic discriminant analysis allow us to divide processed mass spectra of exhaled gases into two groups: mass spectra of exhaled gases of healthy people and mass spectra of exhaled gases of people with possible pathologies.

Keywords: mathematical methods of data processing, mass spectrometers, discriminant analysis

REFERENCES

1. Kuzmin A.G., Tkachenko E.I., Oreshko L.S., Titov Yu.A. [Perspective of a method of mass-spectrometry aroma diagnosis for composition of the exhaled air]. *Tezisy dokladov X Evrazijskoj nauchnoj konferencii "DONOZOLOGIYA-2014"* [Theses of reports of the X Eurasian scientific DONOZOLOGY-2014 conference]. Saint-Petersburg, 18–19 December 2014, pp. 229–231. (In Russ.).
2. Patent RF no. 94763. Prioritet 27.05.2010. (In Russ.).
3. Kuzmin A.G., Titov Yu.A. [Small-sized mass spectrometers for dynamic researches of composition of the exhaled air]. *Trudy I Mezhdunarodnoj nauchno-prakticheskoy konferencii "Vysokie tekhnologii, fundamental'nye i prikladnye issledovaniya v fiziologii i medicine". Ch. 3* [Proc. I of the international scientific and practical conference "High Technologies, Basic and Applied Researches in Physiology and Medicine". Part 3]. Saint-Petersburg, 23–26 November 2010, SPBGPU, 2010, pp. 266–270. (In Russ.).
4. Kuzmin A.G., Tkachenko E.I., Oreshko L.S., Titov Yu.A., Balabanov A.S. [Method of mass-spectrometry express diagnostics on composition of the exhaled air]. *Meditsinskiy Akademicheskij Zhurnal* [Medical Academic Journal], 2016, vol. 16, no. 4, pp. 106–107. (In Russ.).
5. Kuzmin A.G., Tkachenko E.I., Oreshko L.S., Titov Yu.A., Balabanov A.S. The method of medical instant diagnostics based on real-time mass-spectrometric analysis of exhaled air composition. *3rd ICMM PAN-ASIA PACIFIC CONGRESS ON MILITARY MEDICINE. Abstracts*. 08.08.2016–12.08.2016, St.-Petersburg, pp. 181–182.
6. Manoylov V.V., Titov Yu.A., Kuz'min A.G., Zarutskiy I.V. [Methods for data processing and classification for mass spectra of exhaled gases using discriminant analysis]. *Nauchnoe Priborostroenie* [Scientific Instrumentation], 2016, vol. 26, no. 3, pp. 50–57.
URL: <http://iairas.ru/en/mag/2016/full3/Art7.pdf> (In Russ.).

Contacts: Samsonova Natalya Sergeevna,
kolomna.88@mail.ru;
Manoylov Vladimir Vladimirovich,
manoilov_vv@mail.ru

Article received in edition 27.07.2018