

МНОГОПАРАМЕТРИЧЕСКИЙ АНАЛИЗ СОСТАВА ФАРМАЦЕВТИЧЕСКОЙ ПРОДУКЦИИ НА ОСНОВЕ ТЕРАГЕРЦОВОЙ СПЕКТРОСКОПИИ ВО ВРЕМЕННОЙ ОБЛАСТИ

Д. А. Борисевич, А. М. Гончаренко, А. В. Ляхнович,
Г. В. Синицын, М. А. Ходасевич*

УДК 535-14;543.42

Институт физики НАН Беларуси,
220072, Минск, Беларусь; e-mail: m.khodasevich@ifanbel.bas-net.by

(Поступила 12 октября 2020)

Многопараметрические методы применены к спектрам пропускания фармацевтической продукции, полученным с помощью быстрого Фурье-преобразования временных форм терагерцового электромагнитного поля. Последовательное использование метода главных компонент и линейного дискриминантного анализа позволило классифицировать два лекарственных средства, а метод независимых компонент — в смеси трех лекарственных средств восстановить спектр одного компонента и выявить характерную особенность второго.

Ключевые слова: терагерцовая спектроскопия во временной области, метод главных компонент, дискриминантный анализ, метод независимых компонент.

Multivariate methods were applied to the transmission spectra of pharmaceutical products obtained using the fast Fourier transform of the temporal forms of the terahertz electromagnetic field. Successive use of the principal component analysis and linear discriminant analysis allowed us to classify two drugs, and by the independent component analysis to recover the spectrum of the first component in a mixture of three drugs and identify the characteristic feature of the second.

Keywords: terahertz time-domain spectroscopy, principal component analysis, discriminant analysis, independent component analysis.

Введение. Уникальная особенность излучения терагерцового (ТГц) частотного диапазона в применении к спектроскопическим исследованиям — получение химической и даже структурной информации о веществах в бумажной или пластиковой упаковке. В этом случае термин “неразрушающий контроль” относится не только к объекту исследования, но и к упаковочному материалу. Поэтому практическое применение ТГц-спектроскопии постоянно расширяется и находит новые области, требующие проведения малозатратных, непродолжительных по времени измерений характеристик объектов без предварительной подготовки даже в форме извлечения из упаковки. Например, большой интерес вызывает бесконтактный и неразрушающий контроль фармацевтических препаратов, особенно при существенно возросших рисках инфекционных заболеваний. Классические аналитические методы идентификации фармацевтической продукции и ее подделок включают в себя жидкостную хроматографию [1], ядерный магнитный резонанс [2], спектроскопию ИК-диапазона [3] и комбинационного рассеяния [4], получение изображений в ближней ИК-области спектра [5]. Многочисленные исследования (см., например, [6—9]) демонстрируют большой потенциал ТГц-спектроскопии, в том числе для диагностики лекарственных препаратов. В этой области частот находятся характерные спектры многих химических веществ, применяемых в медицине и фармацевтике. Современные лекарства представляют собой сложные смеси из действующих и вспомогательных веществ

MULTIVARIATE ANALYSIS OF THE COMPOSITION OF PHARMACEUTICAL PRODUCTS BASED ON TERAHERTZ TIME-DOMAIN SPECTROSCOPY

D. A. Borisevich, A. M. Goncharenko, A. V. Lyakhnovich, G. V. Sinitsyn, M. A. Khodasevich*
(B. I. Stepanov Institute of Physics of the National Academy of Sciences of Belarus, Minsk, 220072, Belarus; e-mail: m.khodasevich@ifanbel.bas-net.by)

с высокоточной дозировкой, качественный и количественный спектральный анализ которых крайне затруднен. Требуется применение дорогостоящей контрольной аппаратуры с высокими метрологическими параметрами. В таких условиях существенный вклад в повышение точности диагностики лекарств вносят методы многопараметрической обработки результатов измерений.

Для оценки возможностей применения многопараметрических методов совместно с ТГц-спектроскопией к спектрам готовых обезболивающих фармацевтических препаратов в таблетированной форме, измеренным с помощью импульсного спектрометра ТГц-диапазона, для классификации двух лекарственных средств и определения состава модельной смеси из трех лекарственных средств применены линейный дискриминантный анализ в пространстве главных компонент (PCA-LDA) [10] и метод независимых компонент (ICA) [11]. Как правило, анализ состава фармацевтических препаратов на соответствие нормативным требованиям проводится в лабораторных условиях для специально подготовленных образцов, что позволяет оптимизировать соотношение сигнал/шум и чувствительность измерительной аппаратуры, диапазон измеряемых плотностей и частот для каждого конкретного продукта либо их набора [6—9]. В некоторых работах спектральный анализ выполняется в специфически подобранных условиях, что позволяет добиться высокой точности результатов [7]. Исследователи также уделяют внимание развитию методов контроля структурных параметров таблеток: толщины, однородности защитной оболочки [8] и пористости [9], которые влияют на растворимость и, соответственно, скорость усвоения препарата.

В настоящей работе рассматриваются готовые фармацевтические изделия, не оптимизированные для спектроскопических измерений, что делает описываемые методы более подходящими для применения в качестве промышленной аналитической технологии по сравнению с исследованиями специально подготовленных образцов.

Экспериментальное оборудование и измерения. Исследованы три лекарственных препарата белорусских производителей, часто применяемые для устранения болевого синдрома: анальгетические средства парацетамол и ацетилсалициловая кислота, а также спазмолитическое средство дроперин. Для представительности исследуемой выборки таблетки различались по году изготовления (2010—2019 гг.) и производителю (“Белмедпрепараты”, Борисовский ЗМП, “Лекфарм”). Для регистрации временных профилей импульсного ТГц-излучения использованы промышленно выпускаемый спектрометр T-Spec производства компании Ekspla UAB совместно с TeraVil Ltd (Вильнюс, Литва) [12] и разработанный в Институте физики НАН Беларуси спектрометр ТГц-диапазона “Тера-спектрометр” [13], накачиваемые фемтосекундными импульсами лазерного излучения с $\lambda = 1.03$ мкм и действующие по принципу спектроскопии во временной области. Первый относится к разряду доступных для массового применения приборов, второй является уникальным многофункциональным комплексом с настраиваемыми параметрами соотношения сигнал/шум и ширины спектрального диапазона. Временные профили излучения, прошедшего сквозь образцы, и опорного сигнала в случае отсутствия образца регистрировались в базовой конфигурации, частичная схема ТГц-тракта которой для примененных спектрометров приведена на рис. 1.

Расстояние между зеркалами 126 мм, образец располагался вне перетяжки, ближе к ПЗ1 для того, чтобы излучение максимально заполнило апертуру таблетки. Вследствие особенностей фокусировки широкополосных коротких импульсов излучения [14] перетяжка для низких частот смещается

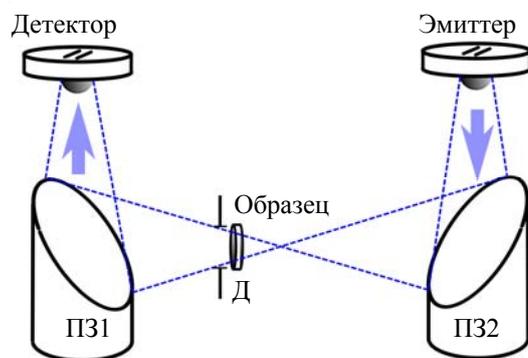


Рис. 1. Схема расположения образцов в тракте распространения ТГц-излучения: ПЗ1, ПЗ2 — внеосевые параболические зеркала, Д — диафрагма

от фокусирующего зеркала ПЗ2 дальше, чем расположена перетяжка высокочастотных составляющих излучения. Возможные блики блокировались апертурой держателя образца, отверстие которой на 1 мм меньше диаметра таблетки (12 мм для ацетилсалициловой кислоты и парацетамола и 7 мм для дротаверина). Апертура располагалась за таблеткой со стороны детектора (на рис. 1 диафрагма Д). Такой размер и расположение диафрагмы обеспечивали близкое к максимальному заполнение размера образца излучением, не приводя при этом к его аттенюации. Это позволило снизить относительную долю излучения, претерпевающего фазовые искажения на диаметральном разделительном штрихе, нанесенном на одну сторону таблетки, до уровня, который не оказывал заметного влияния на результат, а также избежать влияния неопределенности в позиционировании образца и наклоне штриха при поточных измерениях.

Длина пути распространения ТГц-излучения от эмиттера до детектора 625 мм, относительная влажность воздуха в лабораторных условиях ~45 %. Такой уровень влажности соответствует условиям хранения лекарственных средств ($\leq 60 \pm 5$ %) и типичным для медицинских учреждений (40—60 %). Эти условия не оптимальны по влиянию сильно выраженных в ТГц-диапазоне линий поглощения паров воды. Наибольшее влияние на результаты применения многопараметрических методов оказывают погрешности определения спектров в области частот >1.4 ТГц, которая лежит за пределами регистрируемых над уровнем шумов спектров исследуемых таблеток. Длительность регистрируемого временного интервала с профилем излучения 120 пс соответствует перемещению отражателя линии задержки фемтосекундных импульсов 20 мм и обеспечивает спектральное разрешение не хуже 10 ГГц.

Результаты и их обсуждение. С помощью быстрого преобразования Фурье зарегистрированных временных профилей ТГц-поля вычислены спектры прошедшего через образец излучения, нормированные на референсный спектр, который определен в отсутствие образца. Поскольку спектры пропускания образцов являются исходными данными для применения методов PCA, LDA и ICA, требующих воспроизводимости данных и мало чувствительных к регулярным искажениям спектров, дополнительная предобработка временных и спектральных зависимостей в виде устранения эффекта многолучевой интерференции вследствие переотражений в плоскопараллельном слое [9] или так называемого эффекта Кристиансена вследствие рассеяния на микроструктуре образцов не проводилась.

Для многопараметрического моделирования взяты спектры пропускания семи таблеток парацетамола и 10 таблеток ацетилсалициловой кислоты в диапазоне частот 0.4—1.37 ТГц, в которых на частотах 1.07 и 1.14 ТГц присутствуют пики двух самых сильных линий поглощения воды. Диапазоны <0.4 и >1.37 ТГц не улучшают качества идентификации препаратов и поэтому исключены из рассмотрения.

В многомерном пространстве спектральных отсчетов идентификация препаратов характеризуется отсутствием наглядности и невысокой возможностью обобщения. Для понижения размерности пространства, в котором идентификация будет наглядной, применен метод PCA. Этот метод последовательно выбирает в качестве новых переменных, называемых главными компонентами, линейные комбинации исходных переменных. Условиями выбора каждой следующей ГК являются ортогональность предыдущим и максимальная дисперсия спектральных данных исследуемых образцов, которую она описывает. Перегиб зависимости суммарной описываемой дисперсии определяет оптимальное количество ГК, равное в рассматриваемом случае двум. На оставшиеся пятнадцать ГК приходится <6.5 % дисперсии вычисленных спектров таблеток парацетамола и ацетилсалициловой кислоты.

В двумерном пространстве ГК с помощью метода LDA определена линейная функция, дискриминирующая множества таблеток двух рассматриваемых лекарственных препаратов (рис. 2). Поскольку в качестве предикторов выступают две ГК, по определению не коррелированные, предсказательная способность дискриминантного анализа максимальна. Линейная функция перпендикулярна отрезку, соединяющему центры выборок обоих лекарственных препаратов, и проходит через его центр. Найденная дискриминантная функция делит пространство двух ГК на две полуплоскости: с одной из них сопоставлена категориальная переменная, указывающая на принадлежность к множеству таблеток парацетамола, а с другой — ацетилсалициловой кислоты. Из рис. 2 видно, что исследуемые с помощью PCA-LDA образцы классифицированы правильно.

Вторая задача в процессе выполнения работы — определение состава модельной смеси из трех лекарственных средств с помощью метода ICA, предназначенного для решения проблемы слепого разделения источников. В качестве исходных препаратов использованы парацетамол, ацетилсалициловая кислота и дротаверин, спектральные зависимости оптической плотности которых представлены на рис. 3, а в частотном диапазоне 0.3—1.0 ТГц. Видно, что спектральная зависимость оптической

плотности для таблеток дрогаверина имеет ярко выраженный пик на частоте 0.52 ТГц, соответствующий содержащемуся в них моногидрату лактозы. Остальные вспомогательные вещества (повидон, тальк, стеарат кальция и картофельный крахмал) не проявляются в спектре. Более сложен и структурирован спектр для таблеток парацетамола, в состав которого кроме действующего вещества входят повидон, кроскармеллоза натрия, стеариновая кислота и картофельный крахмал. Спектр для ацетилсалициловой кислоты, в состав таблетки которой не входит никаких вспомогательных веществ, совершенно не специфичен. Тройные смеси рассматриваемых лекарственных препаратов промоделированы в соотношении 8:1:1. Спектральные зависимости оптической плотности смесей представлены на рис. 3, б.

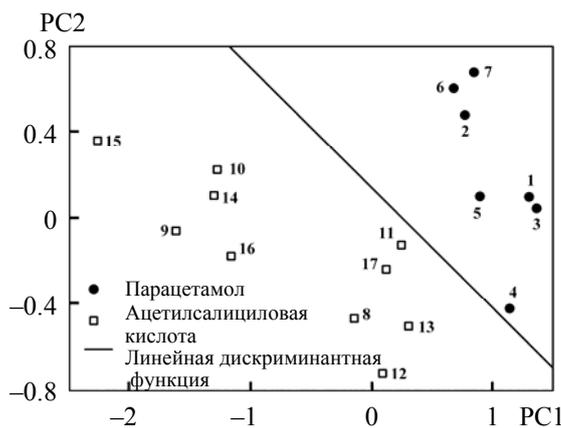


Рис. 2. Представление выборок таблеток парацетамола и ацетилсалициловой кислоты в двумерном пространстве главных компонент их ТГц-спектров пропускания и их дискриминация

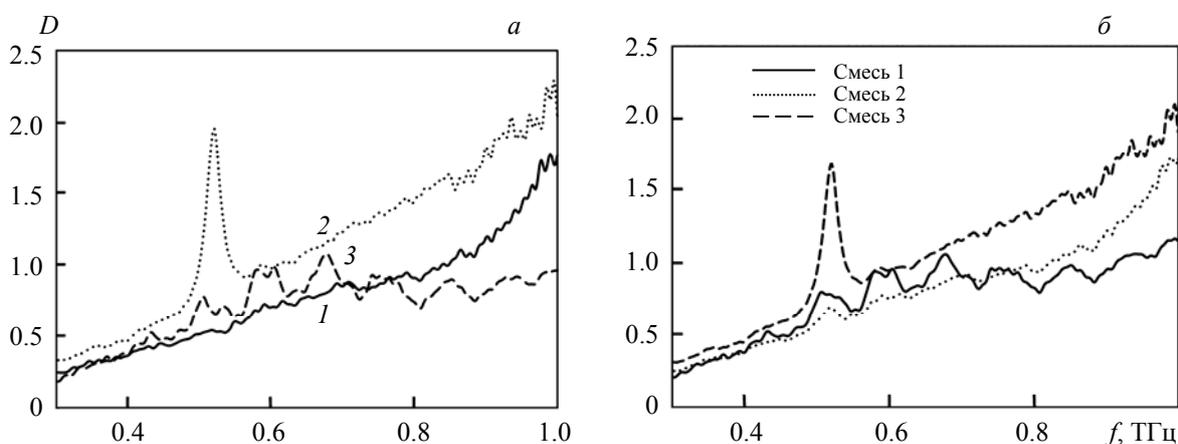


Рис. 3. Спектральные зависимости оптической плотности ацетилсалициловой кислоты (1), дрогаверина (2) и парацетамола (3) (а) и их тройных смесей в соотношении 8:1:1 (б) в ТГц-области

Указанное соотношение лекарственных препаратов в смеси взято лишь для определенности, так как в методе ИСА обработка данных включает в себя приравнивание к нулю средних значений спектров смесей и масштабирование восстановленных спектров с целью выравнивания их дисперсий. В реализации ИСА используются три спектра смесей, поскольку их оптимальное количество равно количеству лекарственных препаратов в смеси. Для успешной реализации метода в рассматриваемом случае необходимо выполнение двух условий: спектры лекарственных препаратов должны быть статистически независимы и амплитудные гистограммы спектров не должны иметь гауссов вид. Согласно центральной предельной теореме, сумма независимых случайных переменных имеет распределение, более близкое к гауссову, чем распределение любой из исходных переменных. Исходя

из этих условий, алгоритм ICA ищет разложение спектров смесей на независимые спектры, которые характеризуются либо максимальным отличием амплитудных гистограмм от гауссовой зависимости, либо минимальной величиной взаимной информации.

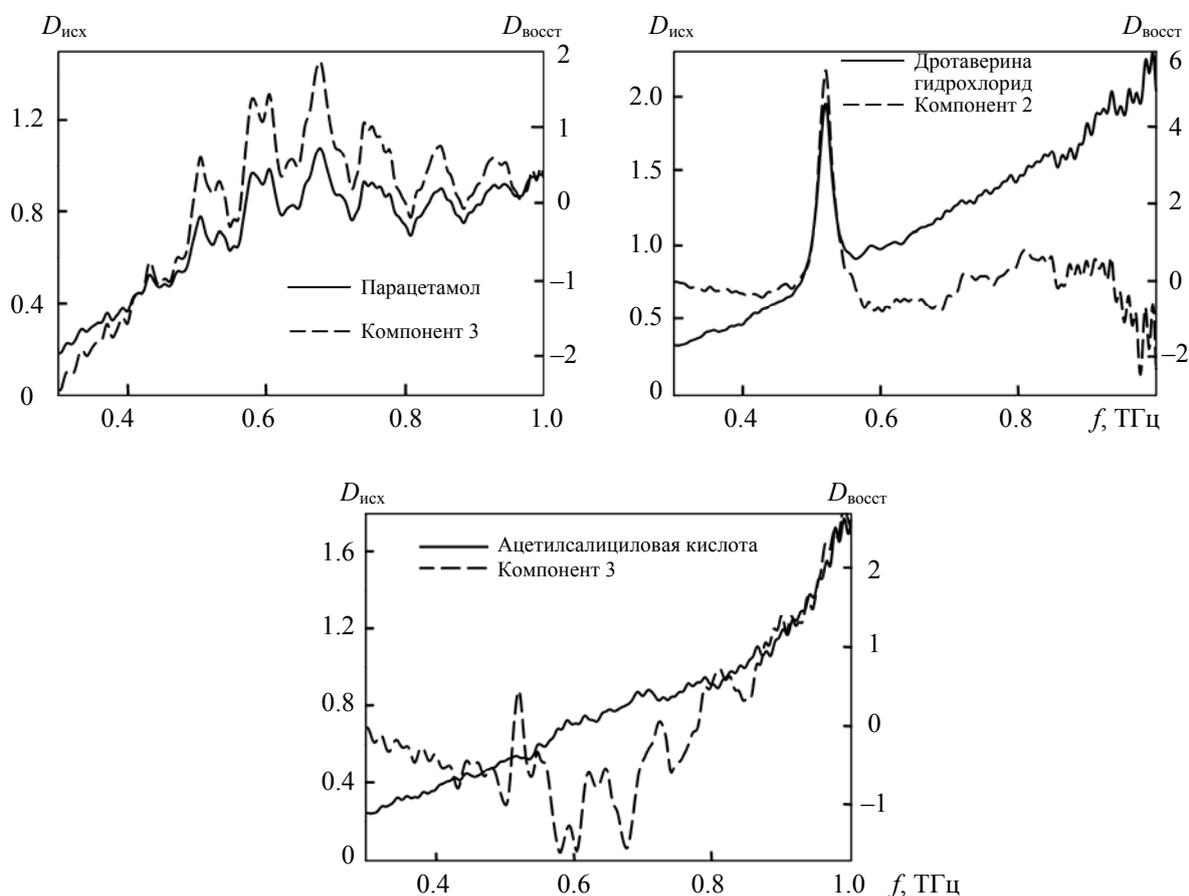


Рис. 4. Сравнение спектральных зависимостей оптической плотности исходных лекарственных средств и независимых компонентов, $D_{\text{восст}}$ — восстановленный спектр

Результаты ICA представлены на рис. 4. Видно, что имеющий множество особенностей спектр парацетамола с хорошей точностью восстанавливается методом ICA. Также этот метод позволяет выявить единичный пик в спектре дропераина. Только спектр ацетилсалициловой кислоты не удалось восстановить из спектров трех смесей. Анализ гистограмм спектров лекарственных препаратов показывает, что они имеют форму, сходную с крылом гауссова распределения, а для парацетамола и ацетилсалициловой кислоты еще и одинаковые особенности. Поэтому можно предположить, что для ICA спектры этих таблеток не являются статистически независимыми. Однако при нарушении условий применимости ICA в результате восстановления спектров трехкомпонентной смеси один компонент выявлен достоверно и еще для одного обнаружена его характерная особенность. Это позволяет сделать вывод о применимости метода ICA для обработки ТГц-спектров сложных смесей и фармацевтических препаратов.

Заключение. Показана возможность применения терагерцовой спектроскопии во временной области в сочетании с многопараметрическими методами обработки спектров в качестве промышленной аналитической технологии для проведения классификации и идентификации фармацевтической продукции. Линейный дискриминантный анализ в пространстве двух главных компонент спектров пропускания ацетилсалициловой кислоты и парацетамола в диапазоне 0.4—1.37 ТГц позволяет провести их достоверную классификацию. Метод независимых компонент в приложении к спектральным зависимостям оптической плотности трехкомпонентной смеси из парацетамола, дропераина и

ацетилсалициловой кислоты в диапазоне 0.3—1.0 ТГц позволяет однозначно идентифицировать парацетамол и выявить присутствие моногидрата лактозы в дротаверине.

- [1] **J. Yao, Y.-Q. Shi, Z.-R. Li, S.-H. Jin.** *J. Chromatogr.*, **853** (2007) 254—259
- [2] **S. Trefi, C. Routaboul, S. Hamieh, V. Gilard, M. Malet-Martino, R. Martino.** *J. Pharm. Biomed. Anal.*, **47** (2008) 103—113
- [3] **F. E. Dowell, E. B. Maghirang, F. M. Fernandez, P. N. Newton, M. D. Green.** *J. Pharm. Biomed. Anal.*, **48** (2008) 1011—1014
- [4] **Y. Roggo, K. Degardina, P. Margot.** *Talanta*, **81** (2010) 988—995
- [5] **C. Gendrin, Y. Roggo, C. Collet.** *J. Pharm. Biomed. Anal.*, **48** (2008) 533—553
- [6] **J. Qin, L. Xie, Yibin Ying.** *Anal. Chem.*, **86** (2014) 11750—11757
- [7] **N. Laman, S. Sree Harsha, D. Grischkowsky.** *Appl. Spectr.*, **62**, N 3 (2008) 319—326
- [8] **N. Odani, Sh. Mohan, E. Kato, H. Feng, Y. Li, M. Nayeem Hossain, J. K. Drennen III, Carl A. Anderson,** *Europ. J. Pharm. Biopharm.*, **145** (2019) 35—41
- [9] **P. Bawuah, D. Markl, D. Farrell, M. Evans, A. Portieri, A. Anderson, D. Goodwin, R. Lucas, J. Axel Zeitler.** *J. Infrared, Millimeter, and Terahertz Waves*, **41** (2020) 450—469
- [10] **M. Manfredi, E. Robotti, F. Quasso, E. Mazzucco, G. Calabrese, E. Marengo.** *Spectrochim. Acta A: Mol. Biomol. Spectrosc.*, **189** (2018) 427—435
- [11] **A. Tharwat.** *Applied Computing and Informatics* (2020), <https://www.emerald.com/insight/content/doi/10.1016/j.aci.2018.08.006/full/pdf?title=independent-component-analysis-an-introduction>
- [12] <http://www.teravil.lt/t-spec.php>
- [13] **А. М. Гончаренко, Г. В. Сеницын, В. Л. Малевич, А. В. Ляхнович.** *Неразрушающий контроль и диагностика*, № 2 (2015), <http://science.by/nauka/57/594/>
- [14] **В. Л. Малевич, Г. В. Сеницын, Н. Н. Розанов.** *Опт. и спектр.*, **4** (2019) 667—670