

М.А. Ходасевич¹, канд. физ-мат. наук;
А.В. Ляхнович¹; Х. Ериклиоглу²

¹(Институт физики НАН Беларуси, г. Минск)

²(Ближневосточный технический университет, г. Анкара)

ДИФФЕРЕНЦИАЦИЯ ШОКОЛАДА ПО СОСТАВУ И ПРОИЗВОДИТЕЛЮ НА ОСНОВЕ АНАЛИЗА МЕТОДОМ ГЛАВНЫХ КОМПОНЕНТ ПРЕДОБРАБОТАННЫХ СПЕКТРОВ ПРОПУСКАНИЯ ТГц ДИАПАЗОНА

В настоящее время в мире интенсивно развиваются неразрушающие и бесконтактные методы диагностики состава, качества и аутентичности продукции пищевой промышленности. Многочисленные разработки основаны на спектроскопической технике измерений. К традиционным видам УФ, видимой и ИК спектроскопии добавился и ТГц диапазон частот. Эффективность спектроскопических методов диагностики значительно повышается в сочетании с многопараметрическими методами анализа данных.

В качестве объекта анализа мы выбрали шоколад, различающийся как по производителю ("Коммунарка" (РБ), "Спартак" (РБ), "Ulker" (Турция), "Волшебница" (РФ) и "АVK" (Украина), так и по типу (горький, десертный, молочный). Спектры образцов толщиной $3,0 \pm 0,1$ мм регистрировались с использованием разработанного в Институте физики НАН Беларуси спектрометра [1] в диапазоне частот < 2 ТГц, где шоколад демонстрирует достаточную прозрачность. Факторами, затрудняющими извлечение полезной информации из измеренных спектров, являются интерференционные эффекты и наличие паров воды в измерительном тракте. Мы сознательно не принимали мер к удалению паров, в отличие от [2], моделируя условия измерения в обычных производственных либо складских помещениях.

Сложность состава исследуемых образцов существенно затрудняет детальный покомпонентный анализ. В ряде работ предложены методы диагностики качества продукции на основе ТГц изображений (безотносительно состава), либо отдельных составляющих продукта, таких как сахар либо какао-масло [3]. Отвлекаясь от детального анализа состава, мы ставили целью настоящей работы демонстрацию возможности выполнения дифференциации продукции по типу и производителю по так называемому методу «отпечатка пальца».

Применение метода главных компонент [4] к изображенным на рисунке 1, *a* спектрам образцов шоколада, обязательная предобработка которых заключалась только в центрировании на каждой длине волны, показало, что первая и вторая главные компоненты описывают

93,5% суммарной дисперсии данных, а на представленном на рисунке 1б графике счетов спектров в первую и вторую главные компоненты исследуемые образцы кластеризуются неудовлетворительно с точки зрения их принадлежности к определенным типам шоколада.

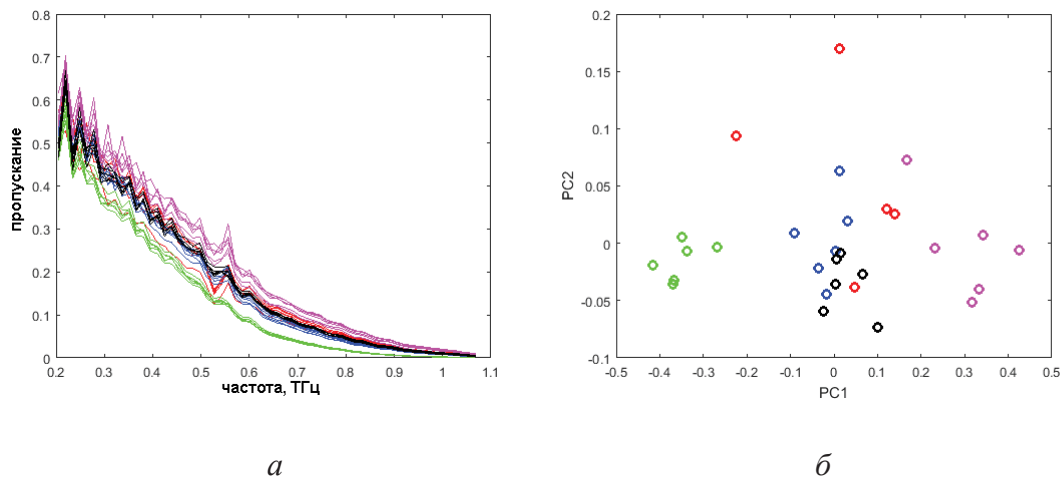


Рисунок 1 – Спектры пропускания образцов шоколада в диапазоне частот 0,2–1,1 ТГц (а) и график счетов спектров в первую и вторую главные компоненты (б) (цвета на обоих графиках соответствуют определенному типу шоколада и его производителю)

Для улучшения кластеризации в качестве дополнительной предобработки спектров был использован метод адаптивных итеративно взвешенных наименьших квадратов со штрафом (airPLS) [5]. Такой вид предобработки позволяет избавиться от шумов, вызванных наличием указанных выше факторов, затрудняющих извлечение полезной информации. В отличие от классического применения метода airPLS нами использовались именно найденные базовые линии, представленные на рисунке 2, а. В этом случае первая и вторая главные компоненты базовых линий спектров описывают 95,6% суммарной дисперсии данных. На рисунке 2, б представлен соответствующий график счетов базовых линий в первую и вторую главные компоненты.

Видно, что качество кластеризации исследованных образцов шоколада существенно улучшилось. Таким образом, нами показана возможность проведения дифференциации шоколада по типу и производителю по методу «отпечатка пальца» в пространстве главных компонент спектров пропускания в ТГц диапазоне частот с привлечением для их предобработки метода адаптивных итеративно взвешенных наименьших квадратов со штрафом. Целью дальнейшего исследования будет являться применение к счетам в первые две главные компоненты базовых линий методов кластерного или дискриминантного анализа для нахождения правил классификации образцов шоколада или дискриминирующих функций, соответственно.

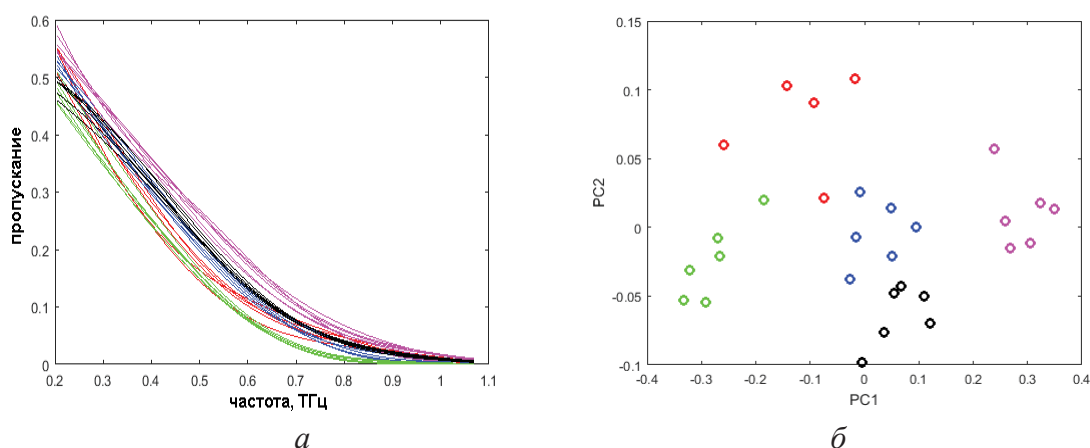


Рисунок 2 – Базовые линии спектров пропускания образцов шоколада, определенные методом airPLS (а), и график счетов базовых линий в первую и вторую главные компоненты (б)

Работа выполнена при финансовой поддержке в рамках исследовательской и инновационной программы Европейского Союза Horizon 2020 (грант № 101008228).

ЛИТЕРАТУРА

1. Терагерцовый спектрометрический комплекс с использованием фотопроводящих антенн и Yb:KYW лазера / А. М. Гончаренко [и др.] // Сборник научных трудов, IV Конгресс физиков Беларуси, 24–26 апреля, Минск, 2013. С. 82–83.
2. Oblitas J., Ruiz J. Multivariate Analysis for the Classification of Chocolate According to its Percentage of Cocoa by Using Terahertz Time-Domain Spectroscopy (THz-TDS) // Proceedings. 2021. V. 70. P. 109-1–6.
3. Weiller S., Tanabe T., Oyama Y. Terahertz Non-Contact Monitoring of Cocoa Butter in Chocolate // World Journal of Engineering and Technology. 2018. V. 6. P. 268–274.
4. Esbensen K. H., Geladi P. Principal Component Analysis: Concept, Geometrical Interpretation, Mathematical Background, Algorithms, History, Practice // Comprehensive Chemometrics / ed.: S. Brown, R. Tauler, B. Walczak. Elsevier, 2009. P. 211–226.
5. Zhang Z. M., Chen S., Liang Y. Z. Baseline correction using adaptive iteratively reweighted penalized least squares // Analyst. 2010. V. 135(5). P. 1138–1146.