

Министерство образования и науки Российской Федерации

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

Физический факультет
Кафедра общей физики (ОФ)

ДОПУСТИТЬ К ЗАЩИТЕ В ГЭК



Руководитель ООП

д-р физ.-мат. наук, профессор
факультет О. Н. Чайковская

подпись

« 15 » июня 2018 г.

Зав. кафедрой ОФ

д-р физ.-мат. наук, профессор
В.П. Демкин

подпись

« 15 » июня 2018 г.

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА БАКАЛАВРА

СОВЕРШЕНСТВОВАНИЕ МЕТОДА ИЗМЕРЕНИЯ СПЕКТРОВ ПОГЛОЩЕНИЯ СЛЮНЫ ЧЕЛОАЕКА НА ТЕРАГЕРЦОВОМ СПЕКТРОМЕТРЕ T-SPEC

по основной образовательной программе подготовки бакалавров
направление подготовки 03.03.02 - Физика

Кудряшов Кирилл Андреевич

Руководитель ВКР канд. физ.-
мат. наук

А.В. Борисов

подпись

« 15 » июня 2018 г.

Автор работы студент группы

№ 05466

Кудряшов К.А. Кудряшов

подпись

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	3
1 Инструменты и способы получения данных	7
1.1 Экспериментальная установка.....	7
1.2 Программное обеспечение	9
2 Метод главных компонент	12
2.1 Общий принцип МГК.....	12
2.2 Понижение размерности	14
2.3 Матрица счетов.....	15
2.4 Матрица нагрузок	17
3 Диагностика плоского лишая на основе анализа данных терагерцовой спектроскопии слюны человека	18
3.1 Анализ данных терагерцовой спектроскопии с помощью статистических методов.....	18
3.2 Совершенствование метода измерения спектров слюны человека	23
ЗАКЛЮЧЕНИЕ	31
СПИСОК ИСТОЧНИКОВ И ЛИТЕРАТУРЫ.....	32

ВВЕДЕНИЕ

В данной работе рассматривается задача совершенствования метода измерения спектров поглощения слюны человека. Для получения спектров слюны используется терагерцовый спектрометр: T-spec. Терагерцовый (ТГц) диапазон заключен между инфракрасной и микроволновой частью электромагнитного спектра: 100 ГГц - 10 Трц. ТГц-волны обладают рядом характерных особенностей: низкая ионизирующая способность, сильно поглощаются водой, высокая проникающая способность.

Медицинская диагностика [1] является одной из наиболее перспективных областей применения ТГц-спектроскопии. Концентрация воды в тканях человеческого организма высока, поэтому ТГц-волны не могут проникать глубоко в тело человека. Следовательно, проводить диагностику по проходящему через организм излучению принципиально невозможно. Тем не менее, существует возможность использования изображающих ТГц-систем и ТГц-спектроскопии для диагностики заболеваний эпителиальных тканей организма по отраженному от среды излучению. В настоящее время существует ряд методов [2] неинвазивной диагностики заболеваний эпителиальных тканей:

Магнитно-резонансная томография не используется для изучения онкологических поражений по причине того, что разрешение систем магнитно-резонансной томографии составляет 1 мм по всем трём направлениям. С помощью данного метода можно получить полезную информацию о поражении организма, лишь при условии того, что поражение находится на глубине, не превышающей 15 мм под поверхностью кожи [3].

Магнитно-резонансная микроскопия используется для получения трёхмерных изображений структуры онкологических поражений кожи. Такие изображения имеют латеральные (вдоль поверхности) разрешения 78 мкм и разрешение по глубине 19 мкм. Предельная глубина, на которой магнитно-резонансная микроскопия визуализацию структуры составляет 800 мкм [4] Информации о составе химических тканей магнитно-резонансная

микроскопия получить не позволяет. Использование магнитно-резонансной томографии и магнитно-резонансной микроскопии для диагностики заболеваний эпителиальных тканей нецелесообразно по причине дороговизны и громоздкости оборудования.

Цифровые камеры высокого разрешения [5] и различные виды кожной микроскопии [6] позволяют анализировать лишь поверхностные характеристики поражений.

Высокочастотные ультразвуковые исследования (20 ... 100 МГц) позволяют визуализировать трёхмерную структуру онкологического поражения с разрешением в 200 мкм по латеральным координатам и в 100 мкм по глубине [7]. Возможная глубина высокочастотных ультразвуковых исследований составляет 7 мм. Отраженный от объекта звуковой сигнал не содержит информации о химических свойствах тканей, поэтому возможности данного метода исследований ограничены [8].

Рентгеновские изображающие системы в целях диагностики заболеваний кожи не используются, по причине того, что рентгеновское излучение вредно для человеческого организма. Кроме того с помощью рентгеновских лучей не всегда можно обнаружить онкологические новообразования в эпителиальных тканях.

Конфокальная отражающая микроскопия обеспечивает изучение строения кожи на глубине до 100 ... 200 мкм с латеральным разрешением до 2 мкм и разрешением по глубине до 3 ... 5 мкм [9]. Высокое разрешение позволяет исследовать строение клеток, визуализировать движение крови в реальном времени и масштабе. Недостатком конфокальной отражающей микроскопии является небольшая глубина диагностики.

Оптическая когерентная томография – метод диагностики, который использует оптическое излучение. Этот метод обеспечивает получение изображения внутренней структуры кожи на глубине до 1 ... 1,5 мм при этом теоретический предел разрешения по глубине составляет 5 ... 15 мкм [10]. Важно отметить, что сильное рассеяние видимого и инфракрасного

излучения в биологических тканях снижает качество изображений, уменьшает разрешение по глубине. С помощью оптической когерентной томографии нельзя получить информацию о химическом составе тканей.

Лазерное доплеровское исследование приповерхностного кровотока кожи даёт возможность получать информацию на глубине до 200 мкм. Лазерные доплеровские исследования часто используются для диагностики аллергических заболеваний. Глубина диагностики данного метода недостаточна для эффективного исследования новообразований кожи.

Таким образом показано, что ни один из перечисленных методов не позволяет однозначно установить вид ракового поражения кожи, найти все очаги роста новообразования.

Эффективным средством ранней неинвазивной диагностики злокачественных поражений могут стать терагерцовые (ТГц) изображающие системы и ТГц-спектроскопия. ТГц излучение [11] не обладает вредными для биообъектов факторами, подобно вышеупомянутому рентгену и гамма-квантам.

Терагерцовый диапазон удобен для спектроскопических исследований, так как в нём лежат сильные линии поглощения многих молекул, линии поглощения которых в других диапазонах очень слабы (например, металлоорганические молекулы). Электромагнитные волны терагерцового диапазона поглощаются металлами, но проникают через пластмассы, бумагу, сухую древесину. Терагерцовые частоты примерно соответствуют колебательным частотам молекул. Отраженные, поглощенные в среде прошедшие сквозь неё терагерцовые волны несут в себе богатейшую информацию об объекте. Однако волны этого диапазона трудно [12] генерировать и детектировать. Всё это определяет широкий спектр применения этого излучения. Терагерцовое излучение испытывает значительно меньшее Релеевское рассеяние, нежели излучение инфракрасного и видимого диапазонов. Все вещества, атомы которых находятся в возбужденном состоянии, излучают и поглощают световые

волны, энергия которых определенным образом распределена по длинам волн. Данная работа основана на законе поглощения света. Если пропускать белый свет сквозь холодный, неизлучающий газ, то на фоне непрерывного спектра источника появляются темные линии. Газ поглощает наиболее интенсивно свет как раз тех длин волн, которые он испускает в сильно нагретом состоянии. Темные линии на фоне непрерывного спектра - это линии поглощения, образующие в совокупности спектр поглощения.

1 Инструменты и способы получения данных

1.1 Экспериментальная установка

В работе использовался терагерцовый спектрометр EKSPLA T-Spec. Он устроен следующим образом:

Оптическая схема установки для ТГц спектроскопии с разрешением по времени (THz-TDS) (Рисунок 1.1) Фотопроводящая антенна, освещаемая ультракороткими лазерными импульсами, используется для генерации и регистрации ТГц излучения. Лазер накачки должен генерировать импульсы на длине волны 1050 ± 40 нм, с длительностью 10 – 150 фс и с частотой следования порядка 30 – 100 МГц, а также иметь выходную мощность не более 100 мВт. Для более эффективного сбора и фокусировки ТГц излучения подложка в виде линзы, изготовленная из высоко резистивного кремния, прикрепляется к задней стороне каждой антенны.

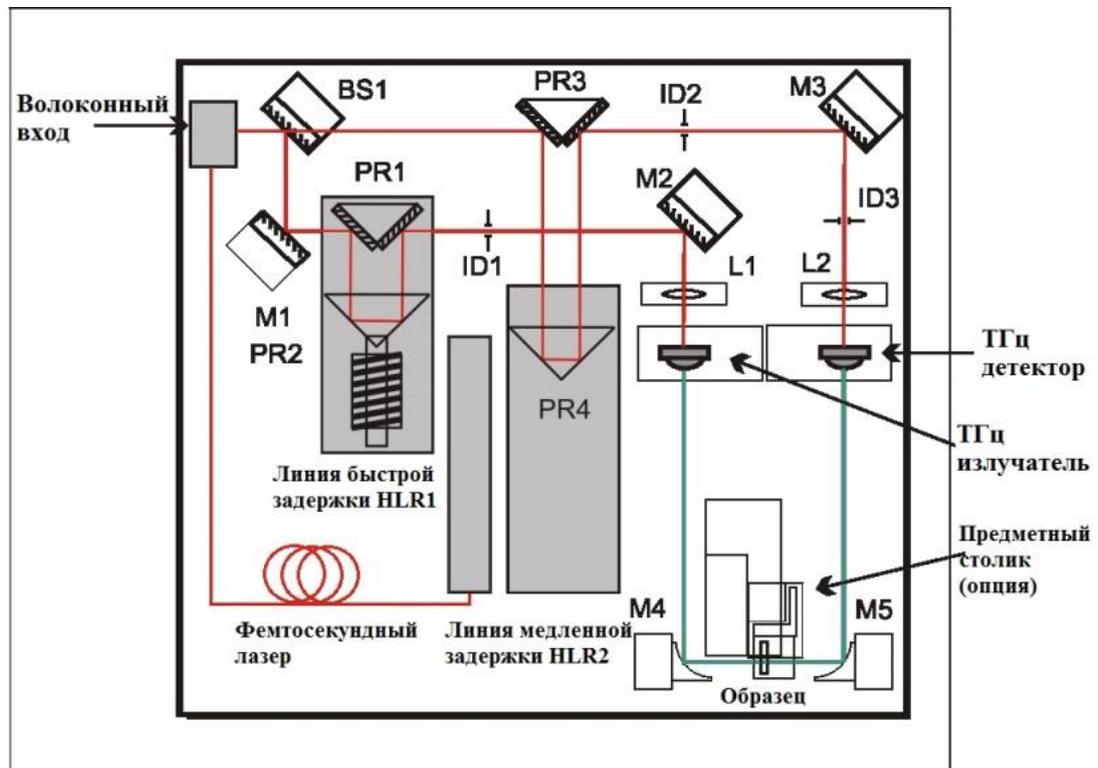


Рисунок 1.1 Оптическая схема прибора

Волоконный вход - это источник ТГц излучения. Луч накачки, попадая на диафрагменный делитель пучка (зеркало) BS1, разделяется на два

пучка. Коэффициент деления связан соотношением 55:45 и не зависит от поляризации лазерного излучения. Зеркала M1 и M2 используются, чтобы направлять луч на излучатель через линию быстрой задержки. Данная линия задержки построена на основе полого уголкового отражателя HLR1,двигающегося с частотой 10 Гц, закрепленного на быстро колеблющейся линейной обмотке, и двух призм с углом 90° (PR1), неподвижно закрепленных напротив уголкового отражателя. После линии быстрой задержки луч накачки направляется на ТГц излучатель с помощью зеркала M2. Линза L1 фокусирует луч накачки в зазор на фотопроводящей антенне.

Зеркала M1 и M2 используются, чтобы направлять луч на излучатель через линию быстрой задержки. Данная линия задержки построена на основе полого уголкового отражателя HLR1,двигающегося с частотой 10 Гц, закрепленного на быстро колеблющейся линейной обмотке, и двух призм с углом 90° (PR1), неподвижно закрепленных напротив уголкового отражателя. После линии быстрой задержки луч накачки направляется на ТГц излучатель с помощью зеркала M2. Линза L1 фокусирует луч накачки в зазор на фотопроводящей антенне.

Вторая часть луча накачки направляется на линию медленной задержки, состоящей из пологого уголкового отражателя или призмы PR4 на шаговом двигателе и неподвижно закрепленной призмы PR3 или зеркал. Далее луч направляется на антенну детектора зеркалом M3. Линза L2 фокусирует луч накачки в зазор на фотопроводящей антенне.

PR4 – это уголкового отражателя. Уголкового отражателя отражает свет в том направлении, в котором он пришел (свойство уголкового отражателя). Это снижает конструкционные требования к установке такой призмы. Перемещение уголкового отражателя приводит к изменению области работы спектра. Для нулевой разности хода призма PR4 ставится на определённый угол. Необходимое в дальнейшем Фурье преобразование всегда выполняется относительно нулевой разности хода. При нулевой разности хода (когда расстояние между зеркалами в интерферометре Майкельсона одинаковы),

будет максимальный интерференционный сигнал. При движении призмы PR4 выравниваются длины путей, пройденных до детектора по плечу 2 и длины путей, пройденных по плечу 1 до детектора (Рисунок 1.2). Луч на два в интерферометре Майкельсона делится для получения автокорреляционной функции (чтобы 2 луча интерферировали). Один луч с постоянной фазой, а у второго луча (где движется зеркало) фаза меняется. Лучи приходят с двух сторон на детектор. Детектор регистрирует максимальный сигнал. И в точке детектора они суммируются.

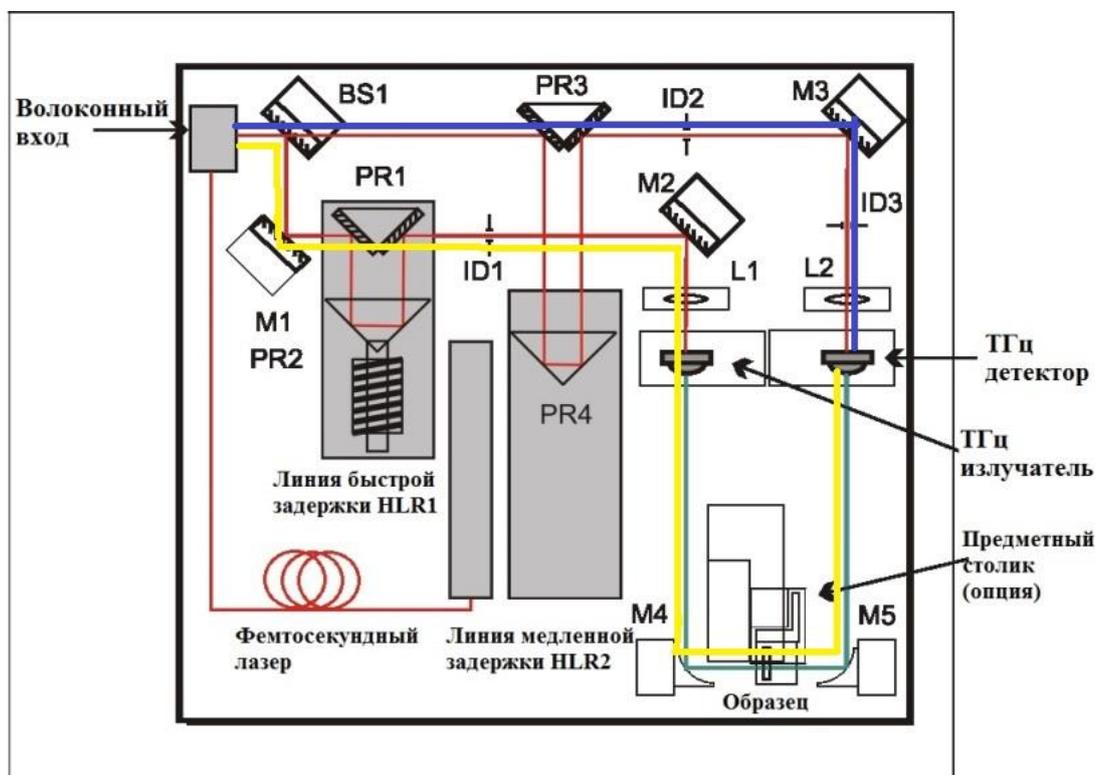


Рисунок 1.2 - Оптическая схема прибора. Синим цветом обозначено первое плечо – путь, пройденный от BS1 до детектора. Жёлтым цветом обозначено второе плечо

1.2 Программное обеспечение

Для снятия данных использовалось программное обеспечение THz Spectrometer 2D. Оно предназначено для регистрации ТГц сигнала, его поглощения и пропускания (10 раз/с). Также оно служит для превращения временного сигнала в частотный и получения спектра частот этого сигнала. При оцифровке сигнала получаем массив данных, где в одной строке находится время, а в другой значение интенсивности сигнала в момент

времени. После выполнения над этим массивом преобразования Фурье [13] получаем другой массив данных, в котором вместо интенсивности сигнала по времени, будет интенсивность сигнала для каждой из частот.

Быстрое преобразование Фурье:

$$C_k = \frac{1}{T} \int_0^T f(t) \cdot e^{-ik\omega t} dt \quad (1),$$

где ω основная частота, k волновое число, $f(t)$ - действительная функция, заданная на временном интервале. C_k - комплексная функция, заданная на частотном интервале.

Поскольку быстрое преобразование Фурье это в первую очередь алгоритм компьютерного вычисления, то задача переводится в дискретную форму. Для этого выбираются дискретные моменты времени $t_n = n \cdot \Delta t$, где Δt период дискретизации, затем выбираются дискретные значения функции в эти моменты времени $x_n = f(n \cdot \Delta t)$ так, что на полном периоде функции оказывается $N \cdot \Delta t = T$, где T - период функции. Подставим это в формулу (1): интеграл превратится в сумму и dt превратится в Δt .

Интеграл превратился в обычную функцию:

$$C_k = \frac{1}{N} \sum_{n=0}^{N-1} x_n \cdot e^{\frac{-2i\pi nk}{N}}, \quad (2)$$

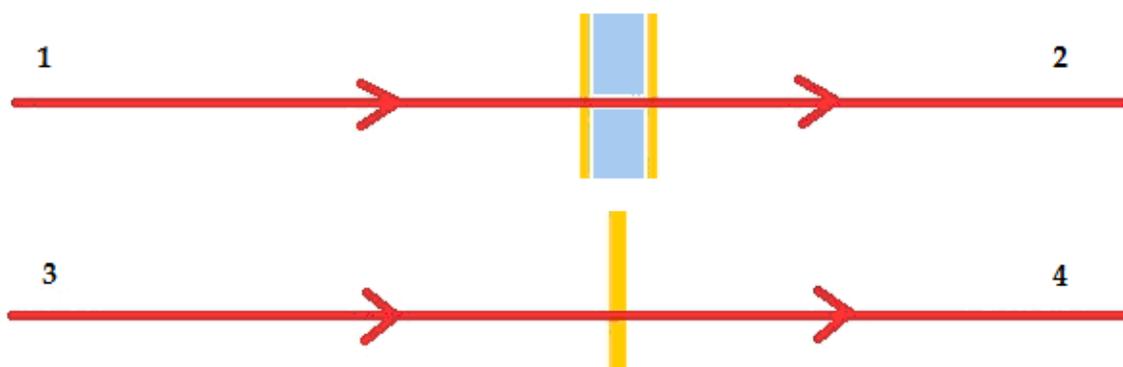
где $\frac{1}{N}$ - масштабный коэффициент, на относительную величину коэффициента для разных гармоник он не влияет. Исследуется соотношение интенсивностей рассматриваемого сигнала и референтного сигнала, для чего применяется закон Бугера.

Закон Бугера – Ламберта - Бера (закон Бугера) - физический закон, определяющий ослабление параллельного монохроматического пучка света при распространении его в поглощающей среде.

$$I(l) = I_0 e^{-k(v)l}, \quad (3)$$

где $I(l)$ - интенсивность прошедшая слой l , l – слой поглощающего вещества, I_0 – интенсивность на входе в вещество.

Показатель поглощения $k(\nu)$ — коэффициент, характеризующий поглощение средой и зависящий от длины волны λ поглощаемого света. Эта зависимость называется спектром поглощения вещества. Величина ν/k имеет размерность длины и показывает характерное расстояние, на котором интенсивность падает в e раз. Ход луча через образец слюны в кювете и пустую кювету (Рисунок 1.3)



Обозначения:

- 1) $I_{0об}$ – сигнал на входе в кювету с образцом
- 2) $I_{об} = I_{0об} e^{(-k_1 l)}$ – сигнал на выходе из кюветы с образцом
- 3) $I_{0кв}$ – сигнал на входе в пустую кювету
- 4) $I_{кв} = I_{0кв} e^{(-k_2 l)}$ – сигнал на выходе из пустой кюветы

Рисунок 1.3 Отношение сигналов $I_{0об} = I_{0кв}$

$$\text{Пропускание: } \frac{I_{об}}{I_{кв}} = e^{-(k_1 - k_2)l} \quad (4)$$

$$\text{Поглощение: } -\ln \frac{I_{об}}{I_{кв}} = (k_1 - k_2)l \quad (5)$$

Закон Бугера справедлив для широкого класса объектов: жидких, твердых и газообразных сред, а также смесей различных веществ, причем в очень широких пределах изменения интенсивности излучения.

2 Метод главных компонент

2.1 Общий принцип МГК

Для сравнения полученных спектров используется метод главных компонент (МГК). Метод главных компонент [14] (Principal Component Analysis) – базовый подход, для решения разнообразных задач, он применяется к данным, записанным в виде матрицы X – прямоугольной таблицы чисел размерностью I строк и J столбцов.

$$\begin{aligned} X &= x_{ij}, \text{ где} \\ i &= 1, \dots, I \\ j &= 1, \dots, J \end{aligned} \quad (6)$$

Данные зачастую содержат в себе нежелательную составляющую, эту составляющую называют шумом. Шум независимо от своей природы не содержит искомой информации. Шум проявляет себя через корреляционные связи между переменными. Погрешности в данных могут привести к появлению не систематических, а случайных связей между переменными. Понятие эффективного ранга и скрытых, латентных переменных, число которых равно этому рангу, является важнейшим понятием в МГК.

Интуитивно–понятная геометрическая интерпретация метода главных компонент (Рисунок 2.1). Пусть имеются только две переменные x_1 и x_2 . Каждой строке исходной таблицы соответствует точка на плоскости с соответствующими координатами. Проведем через них прямую, так, чтобы вдоль нее происходило максимальное изменение данных. Эта прямая будет названа первой главной компонентой – PC1. Спроецируем исходные точки на ось первой главной компоненты. Теперь можно предположить, что все экспериментальные точки должны лежать на данной оси, но какие–то силы отклонили их от такого положения. Тогда все отклонения от новой оси можно считать ненужной информацией или шумом. Проверить, важна ли эта часть данных, можно поступив с этими остатками так же, как с исходными данными – найти в них ось максимальных изменений. Она будет названа второй главной компонентой – PC2. Такой алгоритм необходимо выполнять

до тех пор, пока от шума не останется случайный хаотический набор величин. В таком случае можно будет считать, что пренебрежение искомыми данными исключено.

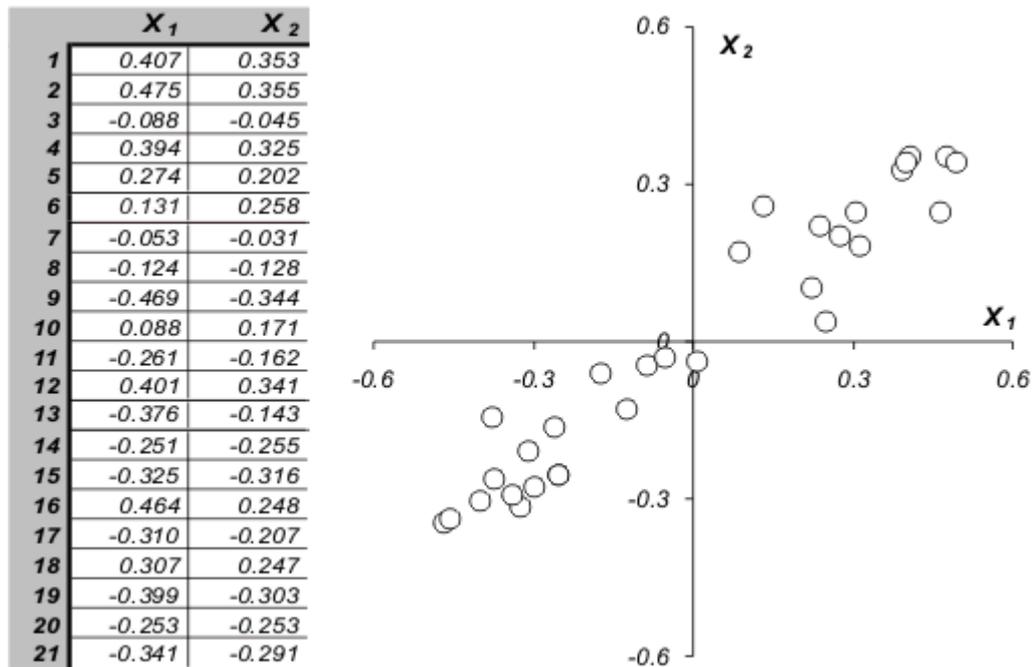


Рисунок 2.1 Графическое представление двумерных данных

Процесс выделения главных компонент в общем случае:

1. Начало координат переносится в центр облака данных – нулевая главная компонента (PC0)
2. Выбирается направление максимального изменения данных – первая главная компонента (PC1)
3. При условии, что данные описаны не полностью, выбирается еще одно перпендикулярное к первому направлению (PC2), которое описывает оставшееся изменение в данных.

В результате размерность становится значительно меньше. Таким образом, возможно, упростить данные на порядки: от 1000 переменных до 2. При этом все переменные учитываются. Найденные главные компоненты дают искомые переменные, которые управляют устройством данных.

2.2 Понижение размерности

Метод главных компонент используется для существенного понижения размерности данных (Рисунок 2.2.). Исходная матрица X заменяется новыми матрицами T и P , размерность которых, A , меньше, числа переменных (столбцов) J у исходной матрицы X

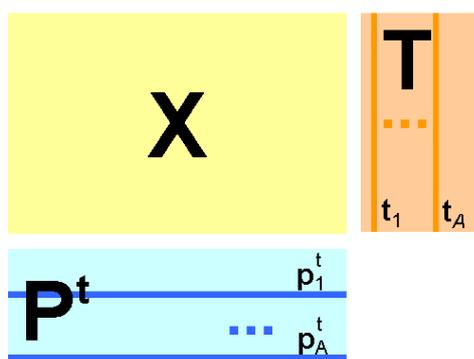


Рисунок 2.2 Декомпозиция матрицы X

Вторая размерность – число образцов (строк) I сохраняется. При условии, что правильно выполнена декомпозиция (размерность A выбрана верно), матрица T несет в себе столько же информации, сколько было изначально, в матрице X . При этом матрица T меньше и проще, чем X .

Пусть имеется матрица переменных X размерностью $(I \times J)$, где I – число образцов (строк), а J – это число независимых переменных (столбцов), число которых ($J \gg 1$). В методе главных компонент используются новые, формальные переменные t_a ($a=1, \dots, A$), которые являются линейной комбинацией исходных переменных x_j ($j=1, \dots, J$)

$$t_a = p_{a1}x_1 + \dots + p_{aj}x_j \quad (7)$$

С помощью этих новых переменных матрица X разлагается в произведение двух матриц T и P (Рисунок 2.3) –

$$X = TP + E = \sum_{a=1}^A t_a p_a^t \quad (8)$$

Матрица T называется матрицей счетов (scores). Ее размерность $(I \times A)$.

Матрица P называется матрицей нагрузок (loadings). Ее размерность ($J \times A$).

E – это матрица остатков, размерностью ($I \times J$).

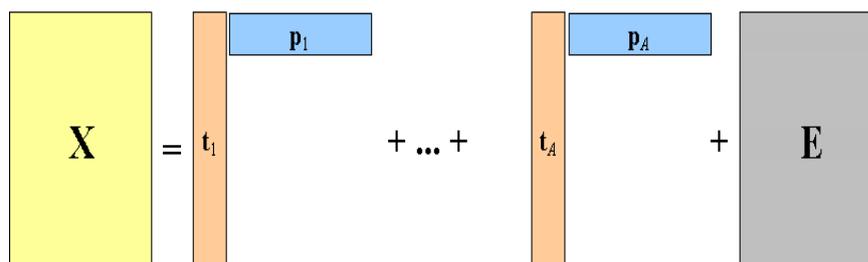


Рисунок 2.3 Разложение по главным компонентам

Новые переменные t_a – главные компоненты (Principal Components), от которых берёт своё название метод. Число столбцов – t_a в матрице T , и p_a в матрице P , равно A – число главных компонент (PC). Эта величина заведомо меньше числа переменных J и числа образцов I .

Важным свойством PCA является ортогональность (независимость) главных компонент. Вследствие чего матрица счетов T не перестраивается при увеличении числа компонент, а к ней прибавляется еще один столбец – соответствующий новому направлению. Тоже происходит и с матрицей нагрузок P .

2.3 Матрица счетов

Матрица счетов T дает проекции исходных образцов (J – мерных векторов x_1, \dots, x_I) на A -мерное подпространство главных. Строки t_1, \dots, t_I матрицы T – это координаты образцов в новой системе координат. Столбцы t_1, \dots, t_A матрицы T – ортогональны и представляют проекции всех образцов на одну новую координатную ось (Рисунок 2.4). При исследовании данных данным методом, особое внимание уделяется графикам счетов. Они несут информацию об устройстве данных. На графике счетов каждый образец изображается в координатах (t_i, t_j) , например – (t_1, t_2) , обозначаемых PC1 и PC2. Близость двух точек означает их положительную корреляцию. Точки, расположенные под прямым углом, являются некоррелированными, а те, которые расположены диаметрально противоположно – отрицательно коррелированными.

$$\begin{aligned}
 T &= t_{ja}, \text{ где} \\
 i &= 1, \dots, I \\
 a &= 1, \dots, A
 \end{aligned}
 \tag{9}$$

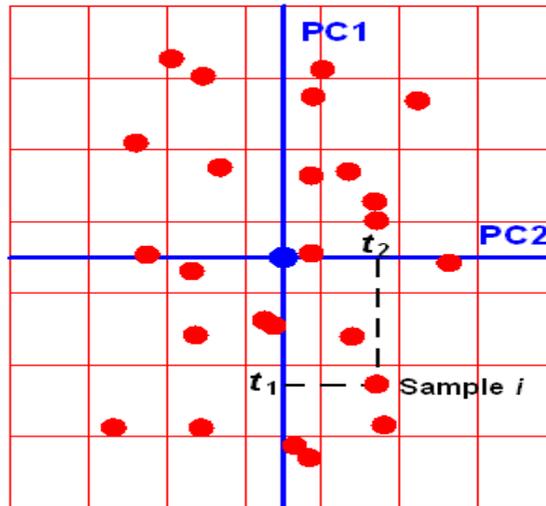


Рисунок 2.4 График счетов

Подробнее об извлечении из графиков счетов полезной информации, будет рассказано в примере.

Для матрицы счетов имеют место следующие соотношения –

$$T^t T = \Lambda = \text{diag} \{ \lambda_1, \dots, \lambda_A \}, \tag{10}$$

где величины $\lambda_1 \geq \dots \geq \lambda_A \geq 0$ – собственные значения, которые характеризуют важность каждой компоненты

$$\lambda_a = \sum_{i=1}^I t_{ia}, a = 1, \dots, A \tag{11}$$

Нулевое собственное значение λ_0 определяется как сумма всех собственных значений.

$$\lambda_0 = \sum_{a=1}^A \lambda_a = \text{sp}(X^t X) = \sum_{i=1}^I \sum_{j=1}^J x_{ij}^2 \tag{12}$$

2.4 Матрица нагрузок

Матрица нагрузок P – это матрица перехода из исходного J -мерного пространства переменных x_1, \dots, x_J в A -мерное пространство главных компонент. Каждая строка матрицы P состоит из коэффициентов, связывающих переменные t и x_1 . Например, a -я строка – это проекция всех переменных x_1, x_J на a -ю ось главных компонент. Каждый столбец P – это проекция соответствующей переменной x_j на новую систему координат.

$$\begin{aligned} p &= p_{ai}, \text{ где} \\ a &= 1, \dots, A \\ j &= 1, \dots, J \end{aligned} \quad (13)$$

График нагрузок (Рисунок 2.5) применяется для исследования роли переменных. На этом графике каждая переменная x_j отображается точкой в координатах (p_i, p_j) , например (p_1, p_2) . Его анализ подобно графику счетов показывает связанные и независимые переменные. Совместное исследование парных графиков счетов и нагрузок даёт много полезной информации о данных.

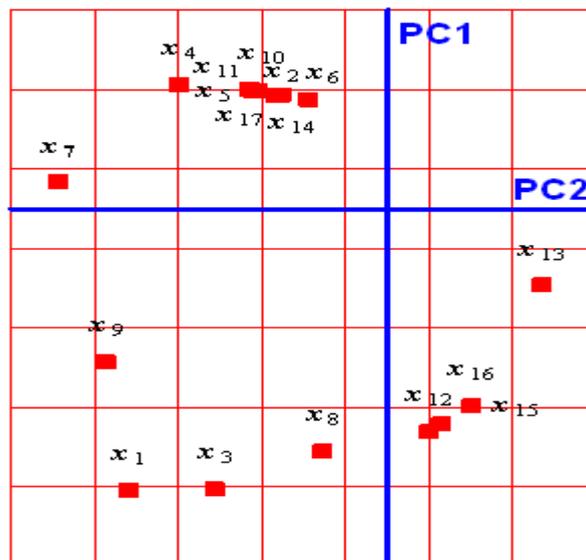


Рисунок 2.5 График нагрузок

В методе главных компонент нагрузки – это ортогональные нормированные вектора, т.е. $P^t P = I$

3 Диагностика плоского лишая на основе анализа данных терагерцовой спектроскопии слюны человека

3.1 Анализ данных терагерцовой спектроскопии с помощью статистических методов

В данном разделе подробно рассматривается обработка результатов. Предметом анализа были пробы слюны двух групп людей: группа здоровых и группа больных. Исследуемое заболевание – плоский лишай ротовой полости. Анализ осуществляется следующим образом:

При оцифровке сигнала данными заполняется массив, в одной строке которого находится время, а в другой значение интенсивности сигнала в момент времени. Путём преобразования Фурье временной сигнал превращается в частотный. Теперь вместо интенсивности сигнала по времени от всего сигнала, будет интенсивность сигнала для каждой из частот. Для классификации используются отношение интенсивности рассматриваемого сигнала и интенсивности сигнала взятого за референтный (поглощение). Выбор спектров для исследования происходил при помощи пространственного распределения по интенсивности сигнала (Рисунок 3.1), чем ярче цвет, тем больше величина интенсивности соответственно.

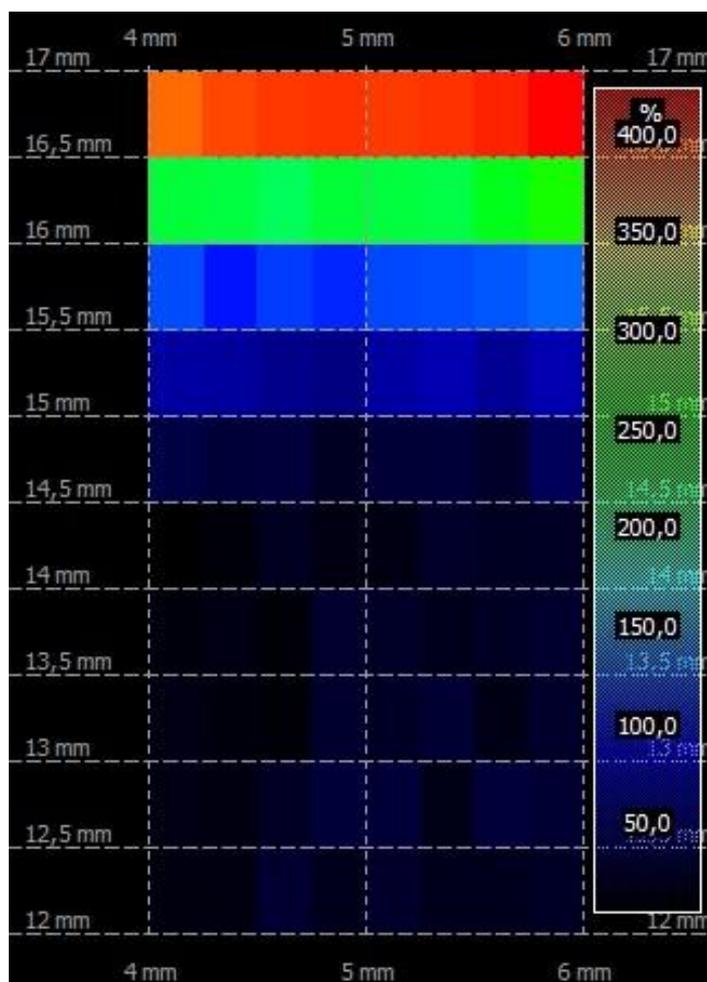


Рисунок 3.1 Пространственное распределение сигнала по интенсивности

Выбраны спектры с определённой интенсивностью сигнала, остальные в вычислениях не использовались. Для каждой пробы таких спектров получается определённое количество. Все отобранные таким образом спектры были помещены в массивы: массив спектров здоровых людей и массив спектров людей, которые имеют исследуемое заболевание. Из этих спектров заполняется матрица. Из этих массивов были посчитаны два средних спектра (Рисунок 3.2): средний спектр для больных и средний спектр для здоровых.

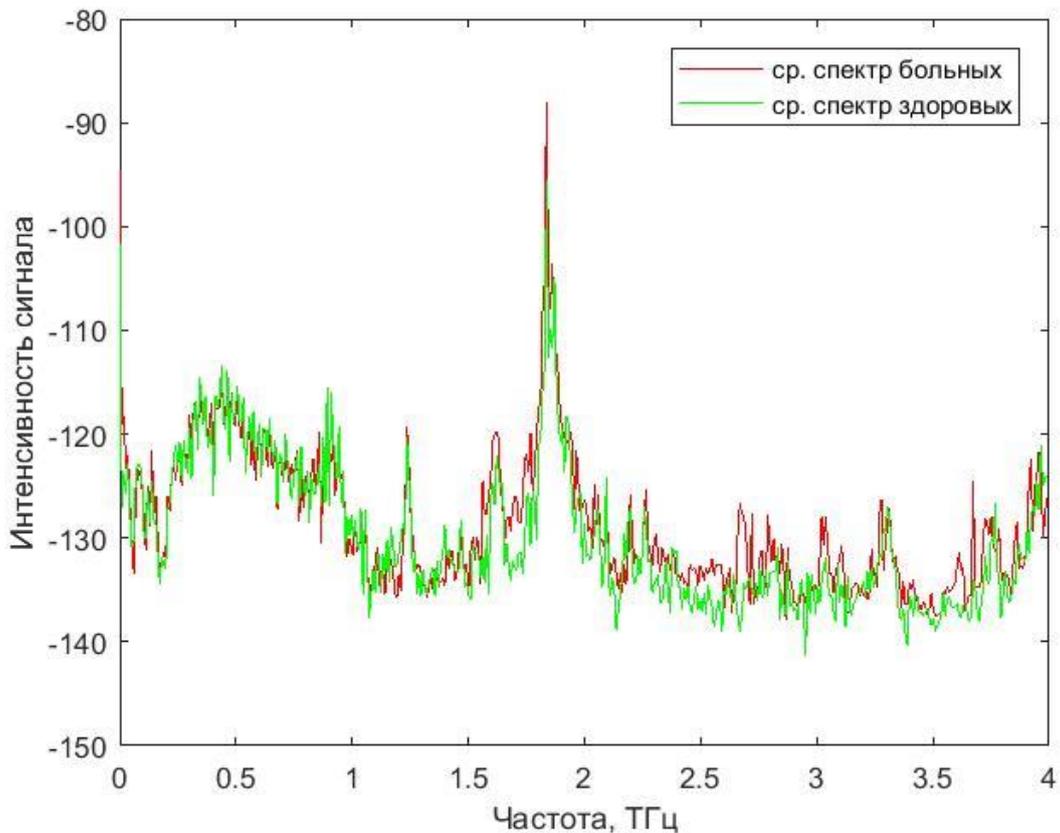


Рисунок 3.2 – График средних спектров больных и здоровых

Полученные данные обрабатываются путём сравнения результатов исследования данных двух групп. Построенные средние спектры служат для визуального наблюдения различий между усреднёнными результатами пациентов имеющих заболевание и пациентами, у которых упомянутое заболевание отсутствует. Для того чтобы сделать эту разницу более заметной были построены графики дисперсий (Рисунок 3.3) для средних значений спектров каждой группы.

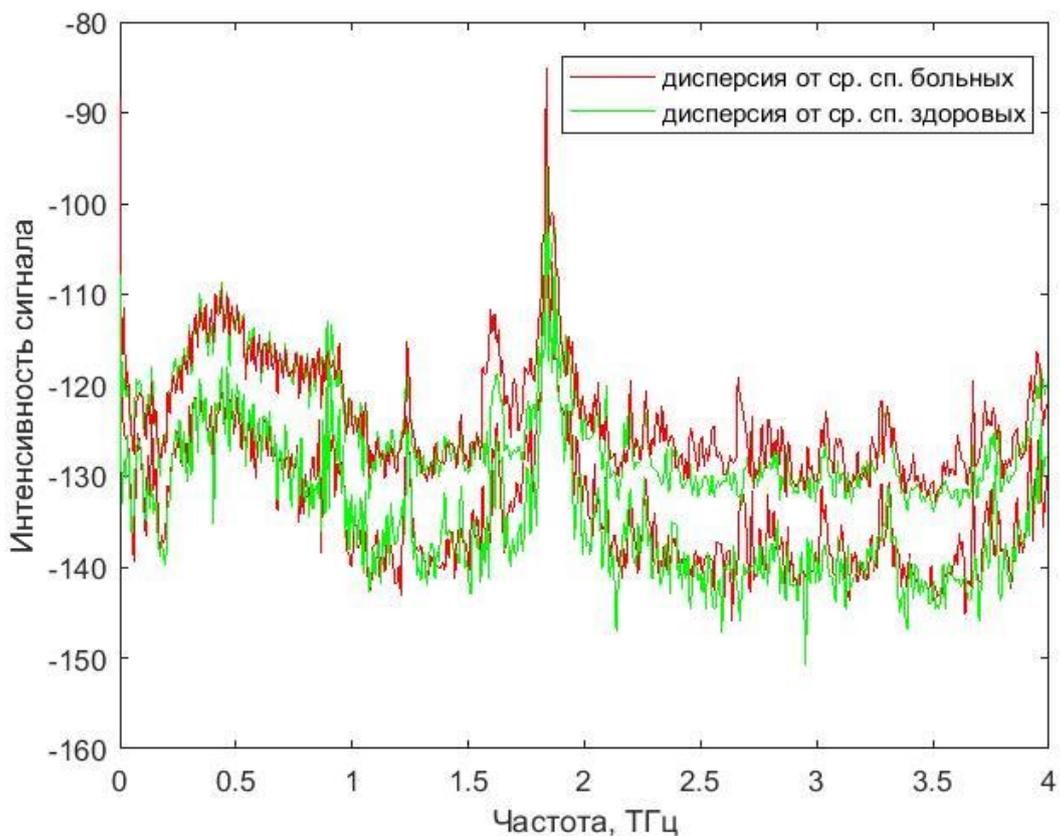


Рисунок 3.3 – График дисперсий больных и здоровых

После построения дисперсии, на определённых частотах наблюдается существенная разница между средними спектрами исследуемых групп, на других частотах такая разница отсутствует.

Для поиска более явных различий использовались иные методы анализа [14], в частности, метод главных компонент. В вышеупомянутых матрицах каждый столбец соответствует спектру. Данной матрице ставится в соответствие другая матрица, в которой столько же столбцов, но число строк уменьшено. Теперь каждому спектру соответствует набор чисел (главных компонент). Их вклад в характеристику спектра уменьшается с номером главной компоненты. Построены главные компоненты с целью уменьшить пространство признаков и максимально снизить влияние нежелательных составляющих (случайных отклонений) на результаты сравнения (Рисунок 3.4).

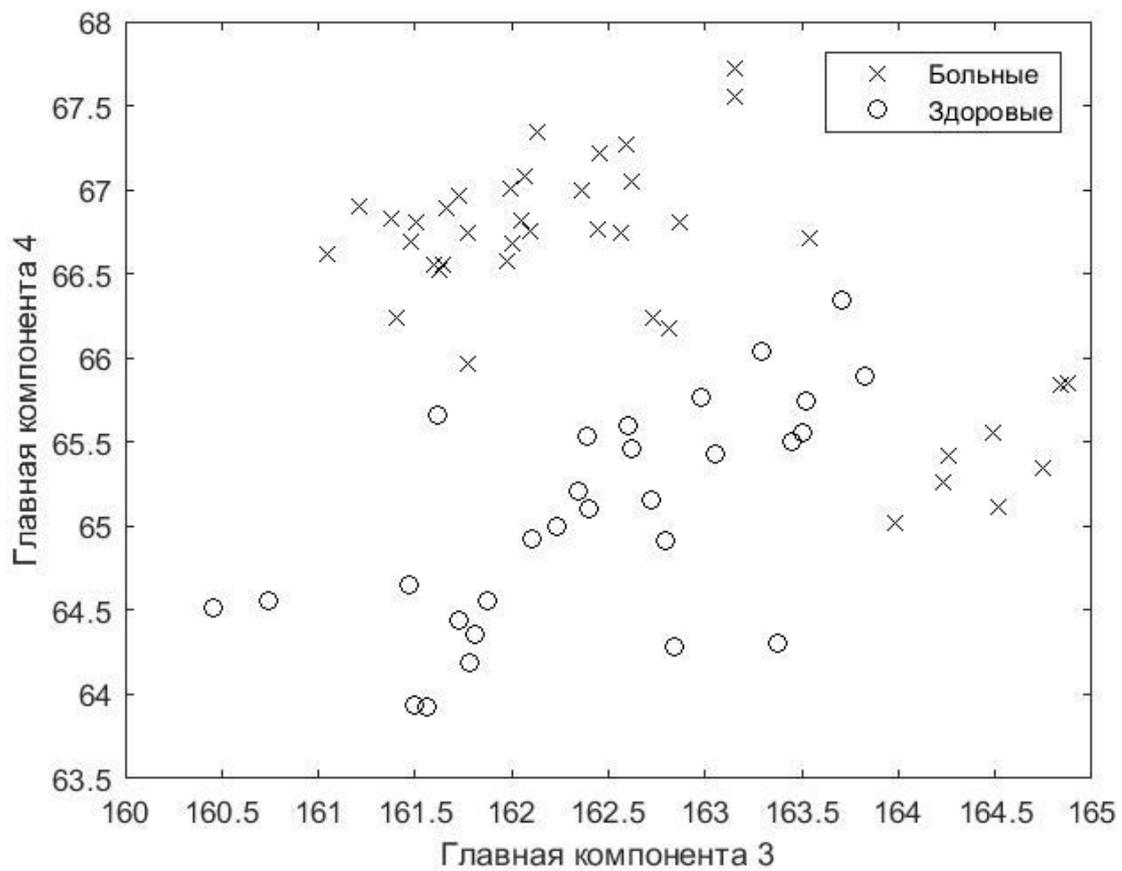


Рисунок 3.4(а) – Распределение данных по главным компонентам

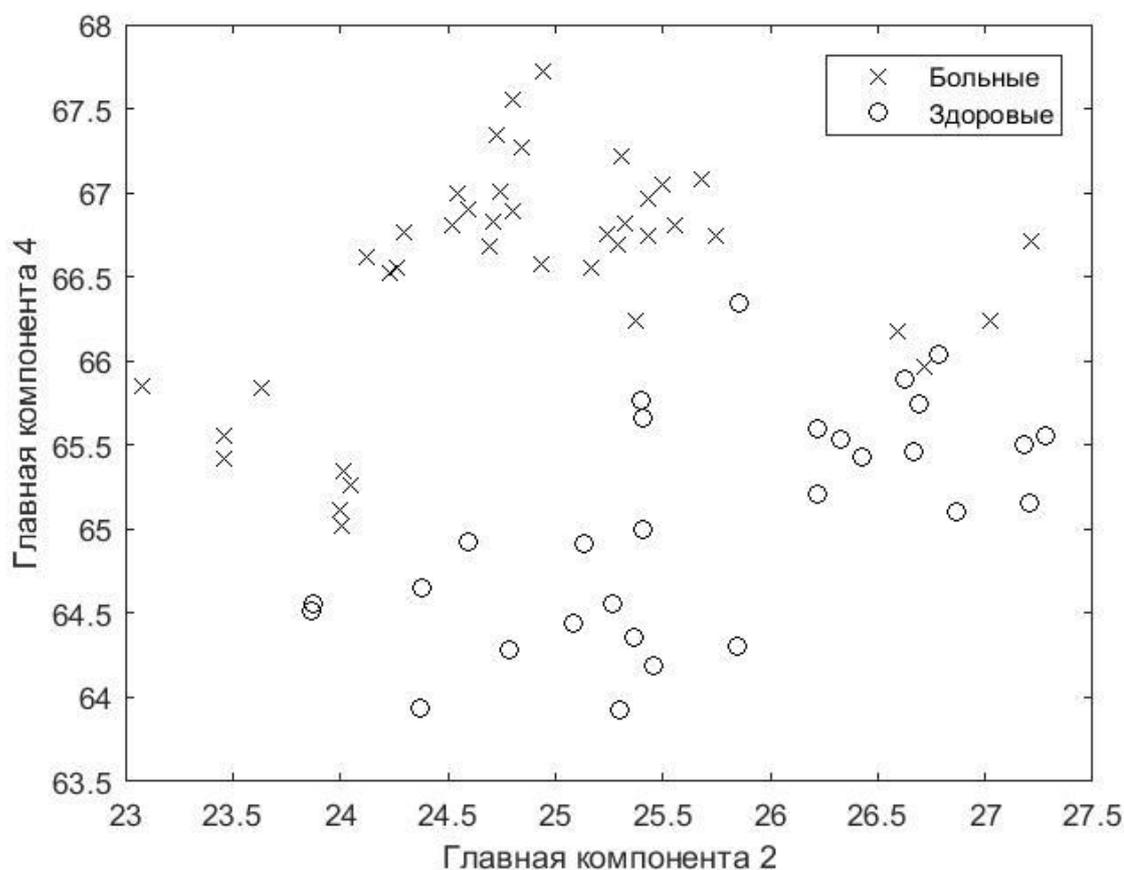


Рисунок 3.4(б) – Распределение данных по главным компонентам

3.2 Совершенствование метода измерения спектров слюны человека

Следующим шагом стала задача повышения точности эксперимента. С этой целью были сняты пробы шести образцов слюны человека. Это было сделано при помощи того же терагерцового спектрометра T-Spec. Держателем для слюны выступала кювета из другого материала – фторопласта, который обладает большей прозрачностью для терагерцового излучения, чем пластик. Сначала сигнал был пропущен через пустую кювету с целью получения сигнала, который был взят за референтный. После чего, тот же сигнал был пропущен через кювету со слюной внутри и зафиксирован. Данное действие было проделано для шести образцов. Референтный сигнал переснимался каждый раз заново при включении прибора. В результате было получено 90 значений сигнала (15 значений для каждого из образцов). Каждое значение представляет собой отношение интенсивности сигнала от

времени, которое путём преобразования Фурье было превращено в отношения интенсивности к частоте. Преобразование Фурье было реализовано программно.

Из полученных значений была заполнена матрица, в которой первым столбцом выступает координата от частоты, вторым столбцом выступает величина интенсивности сигнала или поглощение. Из 15 сигналов, которые соответствуют одной пробе получено одно среднее значение. Данная операция была проведена для каждой пробы. На основании чего были построены графики (Рисунок 3.5).

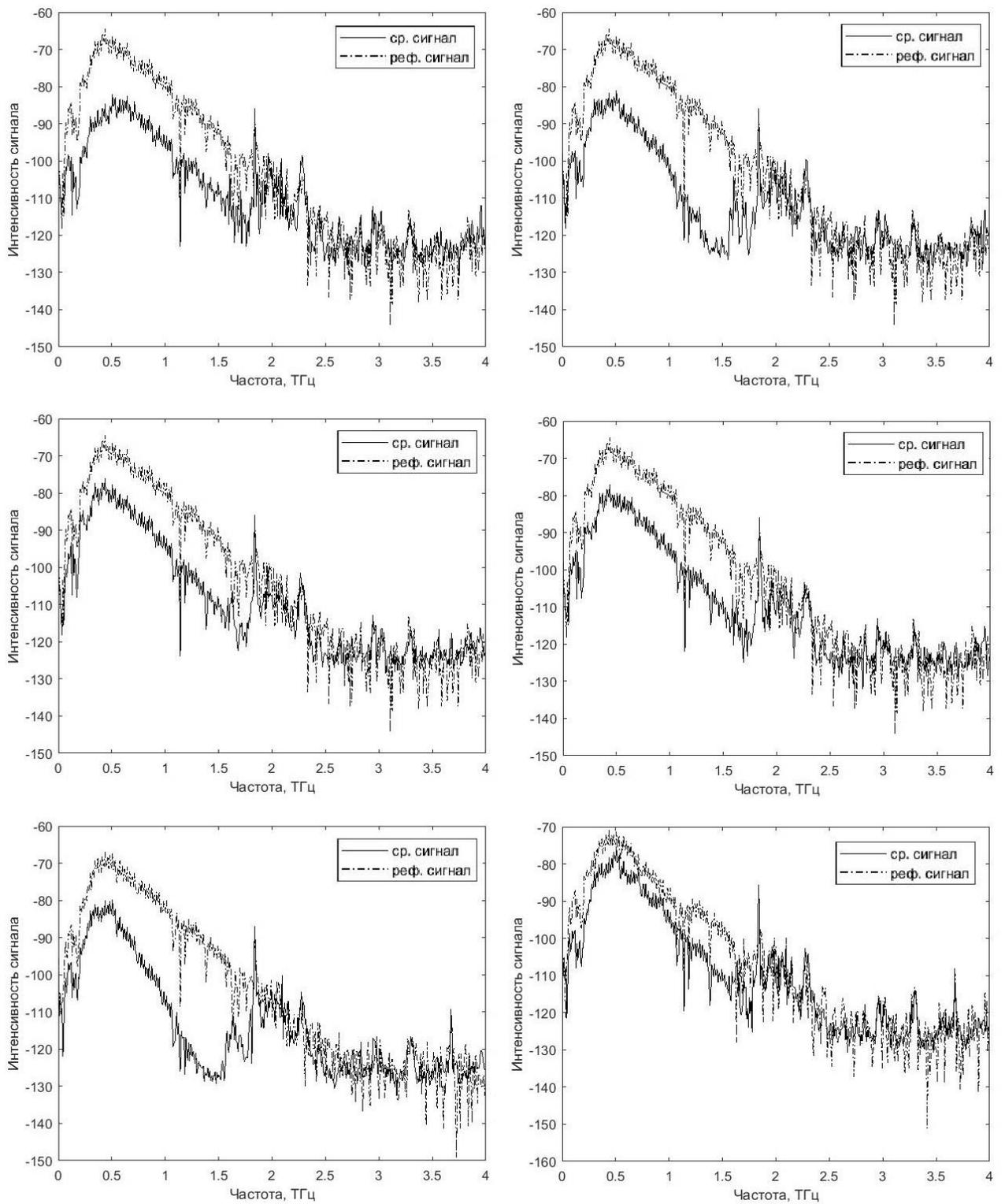


Рисунок 3.5 – Средние значения спектров слюны каждой пробы и референтный сигнал соответствующий данной пробе

Так же построены графики с использованием одномерной медианной фильтрации (Рисунок 3. 6). Медианой числовой последовательности x_1, x_2, \dots, x_n при нечетном n является средний по значению член ряда, получающегося, если упорядочить эту последовательность по возрастанию

или убыванию. Медианный фильтр [15] представляет собой оконный фильтр, который последовательно скользит по массиву сигнала и на каждом шаге возвращается один из элементов, попавших в окно (апертуру) фильтра. Выходной сигнал y_k скользящего медианного фильтра шириной $2n+1$ для текущего отсчета k формируется из входного временного ряда $\dots, x_{k-1}, x_k, x_{k+1}, \dots$ в соответствии с формулой: $y_k = \text{med}(x_{k-n}, x_{k-n+1}, \dots, x_{k-1}, x_k, x_{k+1}, \dots, x_{k+n-1}, x_{k+n})$, где $\text{med}(x_1, \dots, x_m, \dots, x_{2n+1}) = x_{n+1}$; x_m – элементы вариационного ряда, т. е. ранжированные в порядке возрастания значений: $x_1 = \min(x_1, x_2, \dots, x_{2n+1}) \leq x(2) \leq x(3) \leq \dots \leq x_{2n+1} = \max(x_1, x_2, \dots, x_{2n+1})$. Таким образом, медианная фильтрация осуществляет замену значений отсчетов в центре апертуры медианным значением исходных отсчетов внутри апертуры фильтра.

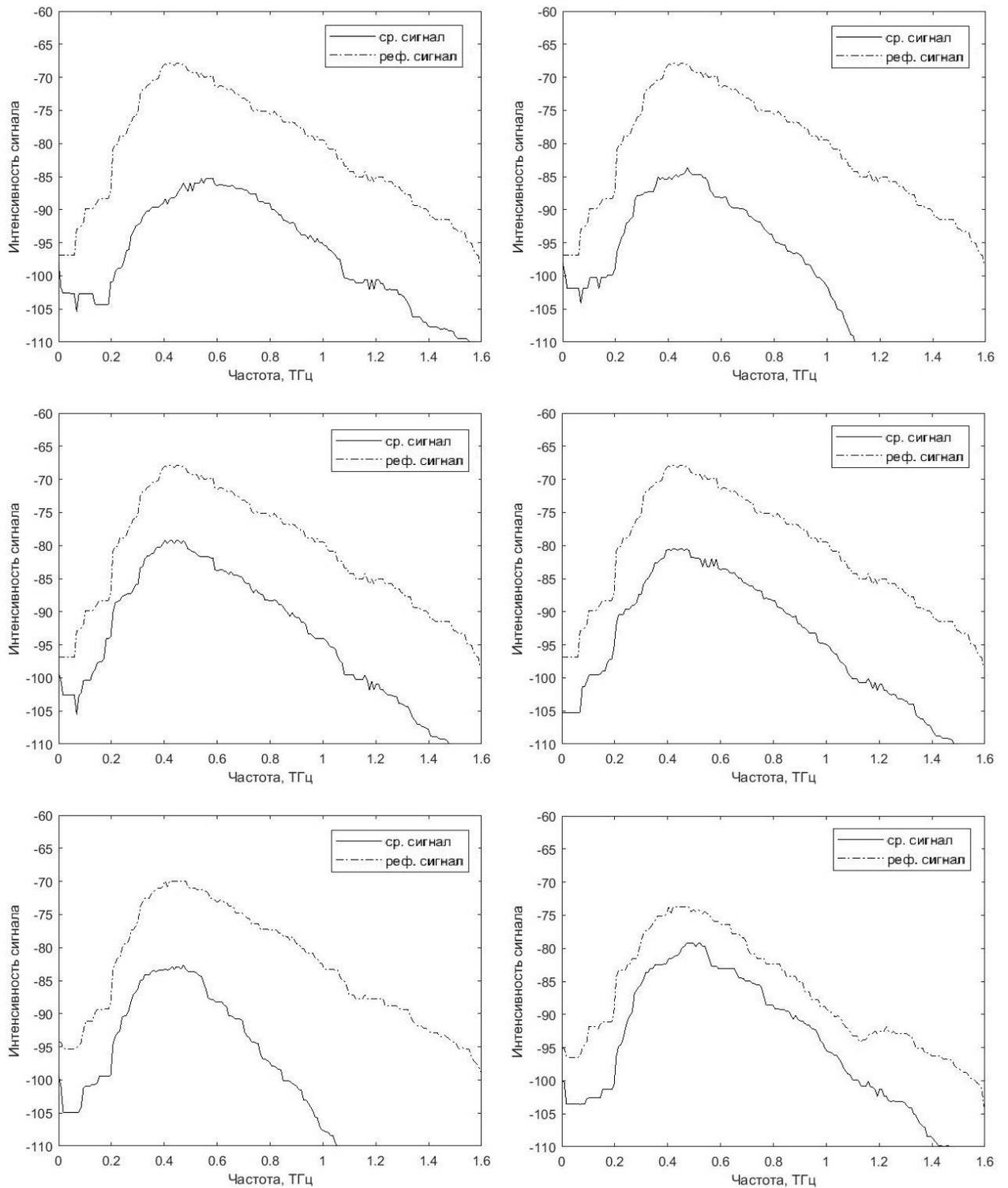


Рисунок 3.6 – Средние значения спектров слюны каждой пробы и референтный сигнал соответствующий данной пробе (с использованием одномерной медианной фильтрации, $n=15$)

Внимание было сосредоточено на сравнении спектров поглощения данных проб. Как показано на графике (Рисунок 3.7.), их значения разделились на 3 группы. Спектры 1 группы близки друг к другу и имеют относительно ровный и характерный наклон, это даёт основание полагать,

что эксперимент при снятии данных проб был проведён правильно. Спектры 2 и 3 групп имеют существенное отклонение от 1 группы. Что вызвано дисперсией сигнала из-за частичного прохождения его через пузыри воздуха и пластиковую прокладку. Первая группа была отдельно выделена с использованием одномерной медианной фильтрации (Рисунок 3.8)

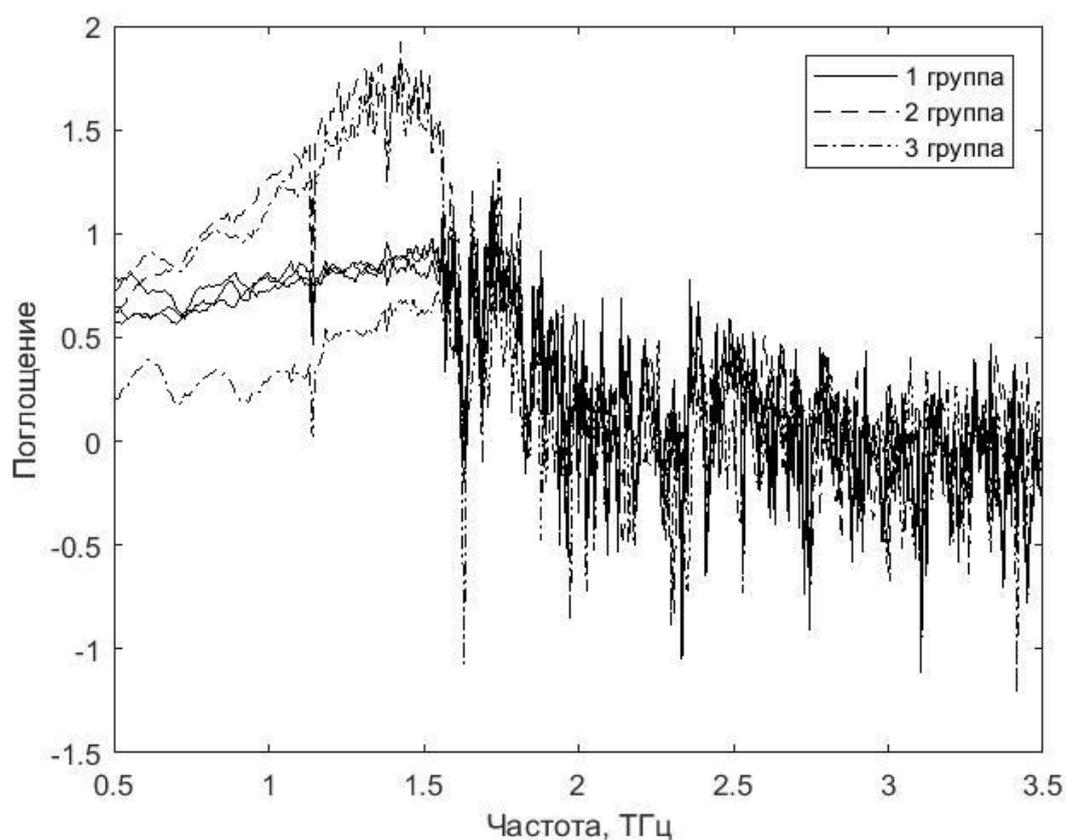


Рисунок 3.7 – График средних спектров поглощения

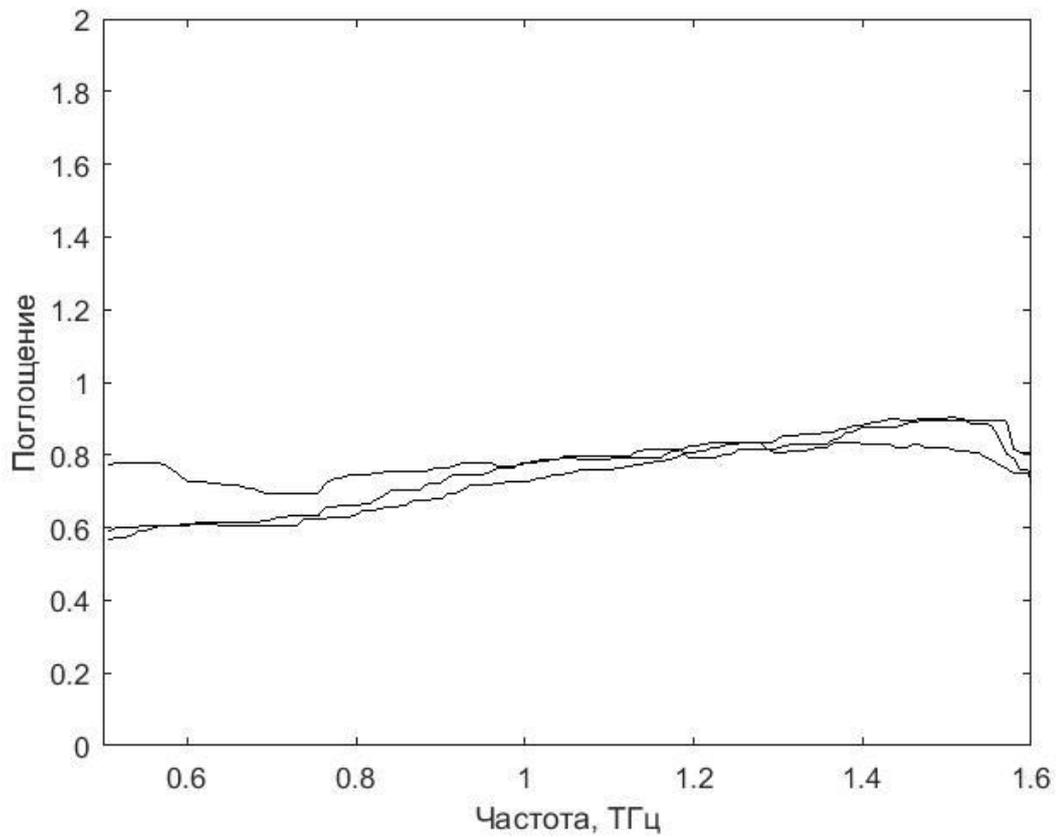


Рисунок 3.8 – Спектры поглощения первой группы с одномерной медианной фильтрацией (с использованием одномерной медианной фильтрации, $n=15$)

Дисперсия данных спектров (Рисунок 3.9) не превышает допустимых пределов.

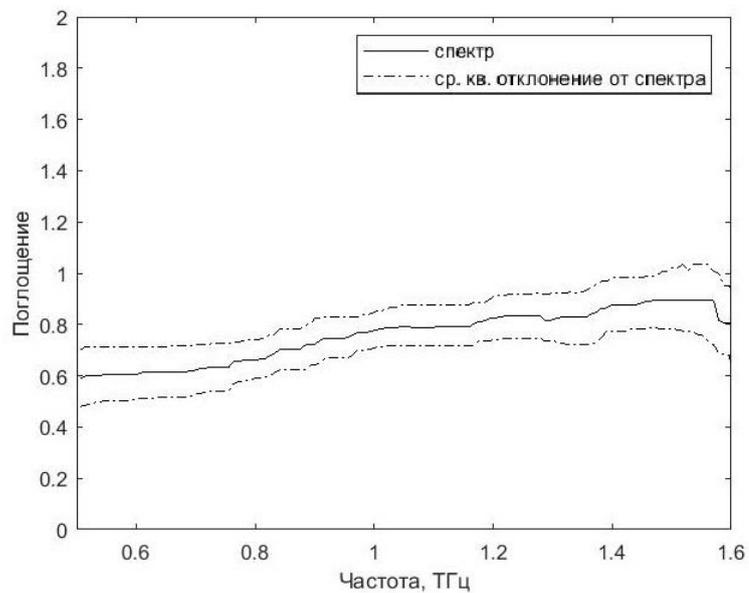
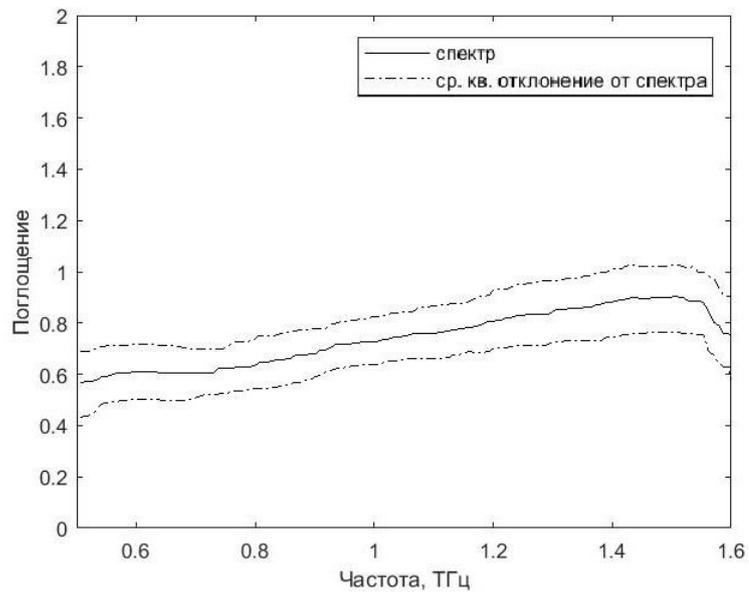
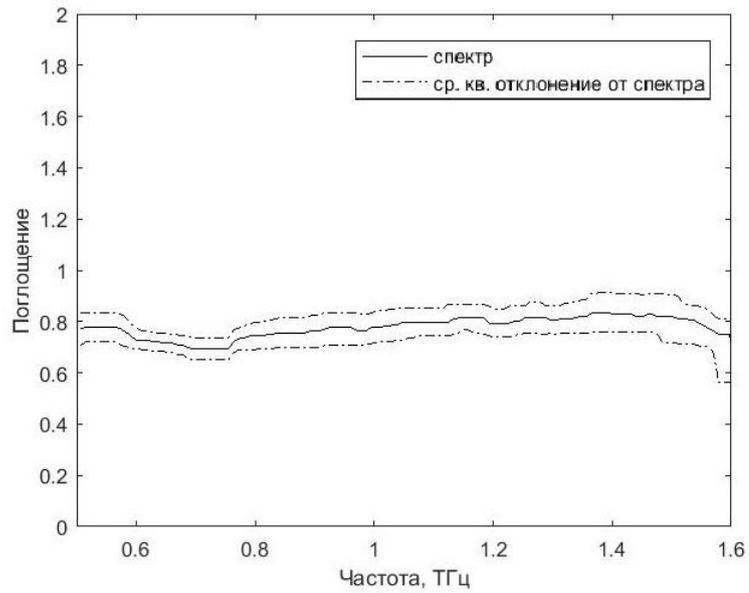


Рисунок 3.9. – дисперсия относительно средних спектров первой группы (с использованием одномерной медианной фильтрации, $n=15$)

ЗАКЛЮЧЕНИЕ

В ходе работы был проведён ряд экспериментов на человеческой слюне. Целью опытов было классифицирование спектров слюны больных и здоровых людей. Данная классификация была получена, но имела недостатки, в частности неоднозначное разделение данных. По этой причине было сделано уточнение процедуры получения спектров поглощения.

Для решения поставленной задачи были рассмотрены спектры полученные с использованием более точных инструментов измерения: кюветы из более прозрачного для терагерцового излучения материала – фторопласт. В ходе эксперимента результаты разделились на три группы. На основании чего была подтверждено, что различного рода ошибки при измерении спектров поглощения слюны человека существенным образом влияют на полученные результаты. Спектры первой группы были определены, как наиболее точные и пригодные для исследований. Спектры второй и третьей группы путём анализа были определены, как спектры полученные в результате совершения методических ошибок. Сделан вывод, что причиной ошибок служило частичное прохождение терагерцового пучка через пузыри воздуха и пластиковую прокладку, что способствовало дисперсии терагерцового пучка и, как следствие из этого, искажению полученных спектров поглощения.

Полученные в результате исследования знания позволят сделать метод измерения спектров поглощения слюны человека на терагерцовом спектрометре T-Spec более точным. Что позволит повысить точность разделения больных и здоровых.

СПИСОК ИСТОЧНИКОВ И ЛИТЕРАТУРЫ

1. Yury Kistenev^a, Alexey Borisov^a, Alexander Shapovalov^a, Olga Baydik^b, Maria Titarenko^b Diagnostics of oral lichen planus based on analysis of volatile organic compounds in saliva // ^aTomsk State University, Tomsk, Russia, 36 Lenina ave., Tomsk, 634050 Russia ^bSiberian State Medical University, Tomsk, Russia, 2 Moskovskiy tract, Tomsk, 634050, Russia.
2. К. И. Зайцев, В. Е. Карасик, Е. В. Новицкая, В. И. Рыжий, И. Н. Фокина Экспериментальное исследование возможности диагностики рака кожи с помощью терагерцовой спектроскопии // Вестник МГТУ им. Н.Э. Баумана. Сер. «Приборостроение». 2012 г.
3. H e i n r i t z H., B e n z e l W., H o f f m a n n K., I r o H. Imaging superficial skin tumors of the ENT area. High frequency ultrasound in comparison with computer-ized tomography and magnetic resonance tomography // HNO. 43:6–11. 1995.
4. S o n g H. K., W e h r l i F. W., M a J. In vivo MR microscopy of the human skin // MRM. 37:185–191. 1997.
5. A s p r e s N., E g e r t o n I. B., L i m A. C. Imaging the skin // Australas J. Derma-tol. 44(1): 19–27. 2003.
6. Cohen D. E., Sanguenza O. P., Fass E. In vivo cutaneous surface microscopy: revised nomenclature // Int. J. Dermatol. 32(4): 257–258. 1993.
7. Bleve M., Capra P., Pavonetto F. Ultrasound and 3D Skin Imaging: Methods to Evaluate Efficacy of Striae Distensae Treatment // Dermatology Research and Practice. 2012. Vol. 2012. Article ID 673706. 10 p.
8. Hoffmann K., Happe M., Schuller S. Ranking of 20 MHz sonography of malignant melanoma and pigmented lesions in routine diagnosis // Ultraschall Med. 20:104-109. 1999.
9. Branzan A. L., Landthaler M., Szeimies R.–M.. In vivo confocal scanning laser microscopy in dermatology // Lasers Med Sci. 22(2):73–82. (2007).

10. Mogensen M., Thrane L., Jorgensen T. M. OCT imaging of skin cancer and other dermatological diseases // J. Biophotonics. 2(6–7):442–451. (2009).
11. Богачев И. В. Ватульян А. О. Дударев В. В., Об одном методе идентификации свойств многослойных мягких биологических тканей // Российский журнал биомеханики Выпуск № 3 / 2013 г.
12. А.С. Сырнаева В.В. Черкасов // Проблемы спектроскопии излучения терагерцового диапазона. СГГА, Новосибирск 2009 г.
13. Снеддон И. Преобразования Фурье // М. Изд-во иностранной литературы 1955г.
14. Померанцев А.Л., Метод главных компонент [Электронный ресурс] // Российское хемометрическое общество - [Б. м., б. г.]. – URL: <http://rcs.chph.ras.ru/>(дата обращения 25.5.2017).
15. Медианная фильтрация [Электронный ресурс] // Национальная библиотека им. Н. Э. Баумана [Б. м., б. г.]. – URL: http://ru.bmstu.wiki/Медианная_фильтрация(дата обращения 17.3.2018).

Отчет о проверке на заимствования №1

Автор: Печерицын Алексей Анатольевич pecher@phys.tsu.ru / ID: 449043

Проверяющий: Печерицын Алексей Анатольевич (pecher@phys.tsu.ru / ID: 449043)

Отчет предоставлен сервисом «Антиплагиат»- <http://www.antiplagiat.ru>

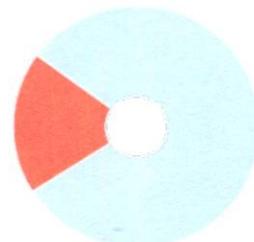
ИНФОРМАЦИЯ О ДОКУМЕНТЕ

№ документа: 19
Начало загрузки: 12.06.2018 07:27:40
Длительность загрузки: 00:00:01
Имя исходного файла: Kudryashov-dipl
Размер текста: 1879 кБ
Символов в тексте: 32742
Слов в тексте: 4074
Число предложений: 304

ИНФОРМАЦИЯ ОБ ОТЧЕТЕ

Последний готовый отчет (ред.)
Начало проверки: 12.06.2018 07:27:41
Длительность проверки: 00:00:03
Комментарии: не указано
Модули поиска:

ЗАИМСТВОВАНИЯ **ЦИТИРОВАНИЯ** **ОРИГИНАЛЬНОСТЬ**
19,08% **0%** **80,92%**



Заимствования — доля всех найденных текстовых пересечений, за исключением тех, которые система относит к цитированиям, по отношению к общему объему документа.
Цитирования — доля текстовых пересечений, которые не являются заимствованиями, но система посчитала их как таковыми корректным, по отношению к общему объему документа. Сюда относятся оформление по ГОСТу цитаты, общепотребительные выражения, фрагменты текста, найденные в источниках из коллекции нормативной правовой документации.
Текстовое пересечение — фрагмент текста проверяемого документа, совпадающий или почти совпадающий с фрагментом текста источника.
Источник — документ, проиндексированный в системе и одобренный в модуле поиска, по которому проводится проверка.
Оригинальность — доля фрагментов текста проверяемого документа, не обнаруженных ни в одном источнике, по которым шла проверка, по отношению к общему объему документа.
Заимствования, цитирования и оригинальность являются отдельными показателями и в сумме дают 100%, что соответствует всему тексту проверяемого документа.
Обращаем Ваше внимание, что система находит текстовые пересечения проверяемого документа с проиндексированными в системе текстовыми источниками. При этом система является инструментом выявления фактов нарушения оригинальности и достоверности заимствований или цитирований, а также заимствования фрагментов проверяемого документа, оставшихся в компетенции проверяющего.

№	Доля в отчете	Доля в тексте	Источник	Ссылка	Актуален на	Модуль поиска	Блоков в отчете	Блоков в тексте
[01]	8,42%	10,95%	Метод Главных Компоне...	http://bourabai.kz	25 Мар 2016	Модуль поиска Интернет	35	45
[02]	0,89%	4,47%	Конференция	http://studfiles.ru	27 Июл 2016	Модуль поиска Интернет	4	22
[03]	0,29%	3,05%	fRTVYt2NYV.pdf	https://disser.spbu.ru	29 Июл 2017	Модуль поиска Интернет	1	13

Еще источников: 12

Еще заимствований: 9,49%

Отв. за проверку на «Антиплагиат»
ст. преп. Печерицын А.А. 
12.06.2018