ESCUELA TÉCNICA SUPERIOR DE INGENIEROS INDUSTRIALES Y DE TELECOMUNICACIÓN

UNIVERSIDAD DE CANTABRIA



Proyecto Fin de Carrera

Identificación de la composición de tejidos biológicos a partir de medidas de reflectancia difusa

(Identification of biological tissue composition through diffuse reflectance measurements)

Para acceder al Titulo de

INGENIERO DE TELECOMUNICACIÓN

Autor: Eusebio Real Peña

Julio – 2012



E.T.S. DE INGENIEROS INDUSTRIALES Y DE TELECOMUNICACION

INGENIERÍA DE TELECOMUNICACIÓN

CALIFICACIÓN DEL PROYECTO FIN DE CARRERA

Realizado por: Eusebio Real Peña **Director del PFC:** Olga M. Conde Portilla

- **Título:** "Identificación de la composición de tejidos biológicos a partir de medidas de reflectancia difusa"
- Title: "Identification of biological tissue composition through diffuse reflectance measurements "

Presentado a examen el día:

para acceder al Título de

INGENIERO DE TELECOMUNICACIÓN

<u>Composición del Tribunal:</u> Presidente (Apellidos, Nombre): Cobo García, Adolfo Secretario (Apellidos, Nombre): Conde Portilla, Olga María Vocal (Apellidos, Nombre): Casanueva López, Alicia

Este Tribunal ha resuelto otorgar la calificación de:

Fdo.: El Presidente

Fdo.: El Secretario

Fdo.: El Vocal

Fdo.: El Director del PFC (sólo si es distinto del Secretario)

V° B° del Subdirector

Proyecto Fin de Carrera Nº (a asignar por Secretaría)

Palabras clave

Reflectancia difusa, espectroscopia visible, análisis de componentes principales, principal component analysis, PCA, parámetro de absorción, parámetro de esparcimiento

Agradecimientos

Agradezco enormemente la gran ayuda prestada por Alma y Jose durante la realización de éste trabajo. Da gusto trabajar con buenos compañeros que te sacan de un apuro.

Agradezco a Olga su dedicación y sobre todo la energía que aporta.

También agradezco su ayuda a Laura, Warren y Pablo por todas las manos que me han echado en éste trabajo y desde hace años.

Contenido

Capítulo 1 Introducción y estructura del proyecto	1
1.1 Introducción	1
1.2 Estructura del proyecto	5
Capítulo 2 Interacción luz materia	6
2.1 Tipos de interacción	6
2.2 Interacción luz-tejido biológico	9
2.3 Reflectancia Difusa en tejidos biológicos	11
Capítulo 3 Materiales y métodos	14
3.1 Materiales	14
3.2 Métodos	24
Capítulo 4 Análisis espectral de la reflectancia	
4.1 Introducción	
4.2 Herramientas utilizadas	
4.3 PCA: Análisis de Componentes Principales	
4.4 Medidas del diámetro del spot	
Capítulo 5 Identificación y cuantificación de componentes en tejidos	50
5.1 Primer análisis: análisis visual de los espectros	50
5.2 Identificación de componentes mediante análisis espectral de la imagen	52
5.3 Cuantificación de composición mediante el análisis de disoluciones	64
Capítulo 6 Conclusiones y líneas futuras	71
6.1 Conclusiones	71
6.2 Líneas futuras	72
Bibliografía	74

Introducción y estructura del proyecto

1.1 Introducción

La luz visible es una radiación electromagnética cuya frecuencia está comprendida entre los $7,5 \cdot 10^{14}$ y $4 \cdot 10^{14}$ Hertzios. En el vacío esta relación se traduce al rango de longitudes de onda que va desde 400 hasta 750 nanómetros. El concepto 'luz' en realidad abarca todo el espectro electromagnético, aunque en óptica es más normal trabajar en el rango espectral que abarca las radiaciones ultravioleta (100 a 400nm), visible (400 a 750nm) e infrarrojo (750nm a 1mm). La longitud de onda de la luz es su parámetro más importante puesto que condiciona su interacción con la materia ya que está íntimamente relacionada con su energía. Es determinante en la refracción, penetración en tejidos, absorción, etc.

La luz es una parte esencial de la vida puesto que participa en numerosos procesos biológicos. Su simple colisión con determinadas moléculas puede producir alteraciones en uno de sus enlaces que cambien por completo las propiedades del material, incluso a longitudes de onda de baja energía. Este efecto se llama (foto)isomerización y un ejemplo típico es el 'daño' producido por la luz UV-B (280 a 315nm) en la epidermis. Aun así, cuando se trata de aplicar técnicas ópticas para la caracterización o terapia de un tejido biológico, se hace patente en casos como el ejemplo anterior que hay que ser cuidadoso en los métodos y parámetros a utilizar. Por el efecto dañino visto en el ejemplo anterior, el rango UV-B no es una opción.

Pero ¿y qué longitudes de onda utilizar? Cada región del espectro se debe a distintas formas de interacción de la radiación con el tejido, en general a cómo absorben los materiales la energía recibida de los fotones radiados. Las radiaciones de tipo rayos X son perjudiciales porque sus fotones llevan una gran cantidad de energía que les permite interactuar y modificar la estructura molecular. La forma de interacción de la luz visible no es dañina porque la energía de sus fotones es baja y sólo les permite excitar electrones para acceder a un estado energético superior. En el caso de los infrarrojos, su energía es aún menor y hace 'vibrar' a las moléculas. Cualquier cuerpo caliente emite radiación en el rango espectral infrarrojo.

- 5	Spectroscopic	Techniques for	Biological and	d Biomedical	Applications
	1 1		0		11

Spectral Region*	Wavelength* (cm)	Energy* (kcal mol ⁻¹)	Techniques	Properties
γ-Ray	10-11	3×10^{8}	Mössbauer	Nucleus properties
X-ray	10-8	3×10^{5}	X-ray diffraction/scattering	Molecular structure
UV	10-5	3×10^{2}	UV absorption	Electronic states
Visible	6×10 ⁻⁵	5×10^{3}	Visible absorption	Electronic states
Infrared	10-3	$3 imes 10^{0}$	Luminescence IR absorption	Electronic states Molecular vibrations
Microwave	10 ⁻¹ 10 ⁰	3×10^{-2} 10^{-3}	Microwave Electron paramagnetic resonance	Rotations of molecules Nuclear spin
Radio-frequency	10	3×10^{-4}	Nuclear magnetic resonance	Nuclear spin

*Approximate values

Figura 1.1 Resumen de técnicas espectroscópicas, rango espectral y forma de interactuar con la materia [1].

Todas las técnicas que se basan en el análisis del espectro, óptico o no, para la caracterización de materiales son llamadas *técnicas espectroscópicas*. Se excita el material con una fuente de radiación y se mide la respuesta espectral, que estará determinada por cómo interactúa ese material con esa radiación. Dentro de la variedad de técnicas cada una tiene sus ventajas e inconvenientes. Las técnicas ópticas son las que menor penetración tienen (<3mm) pero son también las que tienen una mayor resolución sin ser perjudiciales para el organismo, como lo son los rayos X.

En las técnicas ópticas de tratamiento el rango espectral depende del objetivo y el límite lo marca el riesgo que se quiera/pueda tomar. Se utilizan por ejemplo técnicas de fotoablación que destruyen el tejido para eliminar elementos no deseados. En las técnicas de diagnóstico el rango está más limitado. En general, interesa que estas técnicas pasen lo más desapercibidas posibles para el organismo y es precisamente por eso que se opta por las técnicas ópticas.

En los tejidos, el rango de longitudes de onda utilizadas está limitado por la absorción de sus componentes, principalmente el agua y la hemoglobina presente en la sangre. La zona en la que estos componentes tienen una absorción menor se le llama *ventana terapéutica* o diagnóstica. Abarca desde los 600nm hasta los 1400nm, desde el color naranja hasta la región de infrarrojo cercano.



Figura 1.2 Ventana terapéutica (zona entre las líneas punteadas) y espectro de los principales elementos absorbentes en tejidos [1].

La región de estudio óptima es la región visible-infrarrojo. Pero ésta no es la única consideración a tener en cuenta. Hay distintos tipos de técnicas para caracterizar distintos parámetros del tejido y cada una de ellas tiene su objetivo. Se puede medir la cantidad de luz absorbida por la muestra, reflejada, disipada, re-emitida (fluorescencia, fosforescencia, esparcimiento o *scattering* de Raman) o, ya no cantidades sino otras propiedades de la luz como polarización, dicroísmo, tiempos entre eventos, etc.

La técnica que se aplica en éste trabajo es el análisis de la reflectancia difusa. Ésta técnica se sirve de la luz esparcida y absorbida en la superficie, y justo debajo de ésta, para la caracterización espectral. No es una técnica nueva pero aún se está viendo las posibilidades que ofrece y de qué forma obtener los parámetros del tejido. Se ha utilizado con éxito para medidas de la calidad de productos alimentarios, como frutas [2],[3] o productos cárnicos [4]. En tejidos se ha aplicado para la detección de tejido cancerígeno en zonas de expuestas y de fácil acceso como son el tejido epitelial que recubre la boca [5],[6] o la piel en distintas partes del cuerpo y con distintos objetivos como caracterización [7],[8],[9],[10], o detección de distintas lesiones como angiomas (crecimientos anómalos de las arterias en una zona determinada) [12], o incluso diabetes [10].

En la mayor parte de estudios la luz es transmitida mediante fibras que entran en contacto con el material, tanto en transmisión como en recepción puesto que es relativamente sencillo juntar ambas fibras en un dispositivo y ponerlas en contacto con la piel o el tejido bucal. Sin embargo, hay regiones en las que sería más cómodo trabajar sin contacto o donde incluso sería perjudicial para el organismo, como es por ejemplo el caso de los aneurismas de aorta en los que la pared arterial está debilitada y el contacto podría producir una fisura; u otro tipo de heridas y abiertas en las que el contacto sería molesto. En el caso de los productos alimentarios es más sencillo utilizar sistemas no guiados y sin contacto puesto que facilitan la medida en cadena de toda la producción.

En el presente proyecto se ha utilizado un sistema básico sin contacto que oscila entre ambos planteamientos. Las muestras a analizar serán muestras ex vivo (tejido muerto y fuera del organismo). La elección de la fuente viene determinada, como se mencionó anteriormente, por la ventana terapéutica. Puesto que se quiere hacer un estudio del espectro se utiliza una fuente blanca de rango espectral ancho, de modo que se puede caracterizar el tejido en una zona amplia de longitudes de onda. En general, el tipo de fuentes utilizadas en reflectancia difusa son lámparas halógenas que abarcan el rango visible e infrarrojo. En algunos casos se utilizan fuentes láser, pero éstas solo permiten analizar una longitud de onda determinada y por tanto sólo un componente en la muestra.

El sistema de captación es la parte de la que hay menos datos de otros estudios. Puesto que en la mayor parte de trabajos se hacen medidas con contacto, éstas suelen ser con fibra óptica o en ocasiones con esferas integradoras que permiten captar prácticamente toda la reflexión, aunque el resultado de usar unas u otras es parecido. La segunda opción no es posible si se quiere trabajar sin contacto, por lo que se ha elegido trabajar con fibra. La captación con fibra se hace en general utilizando una sonda que suele llevar varios haces, pero aquí se ha sustituido por una fibra de gran diámetro de forma que tenga un gran cono de captación y permita obtener una gran parte de las reflexiones producidas.

Por último, el dispositivo de medida debe ser un espectrómetro que sirva para captar los picos característicos de los componentes a medir. La absorción del agua limita el espectro de trabajo de 200 a 1400nm. En esta zona se pueden medir los espectros de los constituyentes principales de los tejidos como son la sangre y la grasa (y el agua con una gran absorción). La sangre limita la ventana para medir cualquier otro componente más allá de los 600nm, y es en la zona de 500 a 600nm donde su espectro es más característico. Por tanto, el área de análisis

de 200 a 1100nm (que debido a la sensibilidad del espectrómetro se reduce a 450-750nm) es útil para detectar y medir la cantidad de sangre y se ha comprobado durante éste trabajo que también se puede detectar grasa. Si se aumenta el espectro medible hasta 1600nm, la grasa y el agua cuentan con picos de absorción más altos que facilitan la medida de concentración de agua y grasa [11], pero habría que incluir otro espectrómetro que cubra ese rango espectral.



Figura 1.3 Espectros en absorción de hemoglobina oxigenada (HbO2) desoxigenada (Hb) grasa y agua [13].

La motivación para tratar de identificar la presencia tanto de sangre como de grasa en los tejidos se expone a continuación. La cantidad de sangre es en muchos casos un indicador del estado del tejido. Un tejido cancerígeno se caracteriza por un crecimiento anómalo (tumor) que necesita de una mayor cantidad de sangre que lo alimente. Esto provoca un aumento del número de vasos sanguíneos y por tanto de la cantidad de sangre en la región. Por otro lado, la cantidad de grasa también es característica del estado de los tejidos. En el caso de las arterias, se forman tapones o trombos compuestos principalmente de grasa. Y en ocasiones se confunden quistes de grasa con tumores.

En cuanto al análisis de las medidas, generalmente se estudia el espectro en las longitudes de onda en las que se conoce la absorción de un componente determinado y se estudia la variación en ese punto como indicador de la medida. Elegir un único punto hace al sistema de detección poco fiable frente al ruido o pequeñas variaciones y, es más, no se sabe hasta qué punto puede haber una pequeña cantidad de el compuesto buscado o en realidad no se encuentra presente y la medida se debe a la contribución espectral de otro elemento en esa longitud de onda. Cuanto mayor sea el número de puntos a analizar, menor es el efecto de ese ruido y las desviaciones en una longitud de onda. El problema de analizar una gran zona del espectro es que hay que trabajar con muchos datos y es difícil diferenciar. Es aquí donde entra el algoritmo PCA (*Principal Component Analysis*). Este algoritmo extrae las componentes principales del espectro medido, obtenidas como método de diferenciación entre zonas. Si se aplica éste algoritmo a las medidas obtenidas de un tejido, en cada zona tendrá distinta cantidad de sangre, grasa, etc. y por tanto el algoritmo obtiene como componentes principales los espectros de los componentes principales. Así las componentes principales del espectro son también las componentes principales que lo diferencian entre regiones.

Es común aplicar PCA para disminuir el número de datos a tener en cuenta para hacer medidas y caracterizaciones. Lo que se pretende en éste trabajo no es disminuir el número de componentes sino ver cuáles son las principales, identificar a partir de éstas los principales elementos presentes en el tejido y a través del mismo algoritmo medir la cantidad de cada uno de los componentes identificados.

Para el análisis de datos y la toma de medidas se ha utilizado el programa MATLAB[®]. Además de ofrecer numerosas funciones y posibilidades de tratamiento de datos, permite controlar dispositivos conectados al ordenador a través de distintos interfaces. Para éste proyecto se ha desarrollado un ejecutable que permite controlar la captura y almacenamiento de datos del espectrómetro desde MATLAB. Se ha añadido también un motor que permite desplazar la muestra en un eje para tomar varias medidas en distintas zonas. Éste motor está también controlado desde MATLAB de forma que se pueda sincronizar el movimiento del motor con la captura mediante el espectrómetro y el almacenamiento de los datos. Los datos almacenados se pueden tratar sincronizados con las capturas o a posteriori.

1.2 Estructura del proyecto

Para alcanzar los objetivos planteados durante la introducción, se estructura el proyecto en cinco capítulos que se introducen a continuación.

- El **capítulo 2**, es una introducción a la interacción de la luz con los tejidos biológicos y el caso especial de la reflectancia difusa. Es importante conocer el tipo de tejido y sus peculiaridades para poderlo caracterizar.
- El **capítulo 3** enumera y explica las muestras a medir, los componentes y los distintos montajes que se han utilizado para tomar las medidas así como sus objetivos.
- En el capítulo 4 se explica el diagrama de flujo que se ha seguido para el análisis de los datos. También se explican los principios del algoritmo PCA y cómo interpretar sus resultados.
- En el **capítulo 5** se analizan las medidas tomadas y se explican los resultados obtenidos.
- Finalmente, en el **capítulo 6** se resume los resultados y conclusiones obtenidas en éste trabajo y se plantean mejoras y líneas a seguir para continuar con el tema tratado.

Interacción luz-materia

La interacción luz-materia es un mecanismo básico que se utiliza día a día para obtener información acerca del entorno. Con un simple vistazo se puede obtener información de color, forma, o intensidad. Sin embargo, la luz percibida (y la que no) lleva mucha más información de la que se puede analizar a simple vista. Al entrar la luz en contacto con cualquier material se verán modificadas varias de sus propiedades como color, ángulo, polarización, intensidad, etc. Por ejemplo, se puede apreciar fácilmente que un metal incandescente está caliente, puesto que emite energía en forma de luz. También se sabe que un tejido negro absorbe más luz que uno blanco, puesto que la cantidad de luz reflejada es menor. O se sabe que una superficie está caliente cuando la luz se refracta a poca distancia de su superficie, como sucede con los llamados espejismos. Se hace evidente que poder medir y cuantificar esos datos ayudaría a conocer mejor las propiedades del material. Es en este punto donde entran las técnicas ópticas: utilizan la luz como elemento interrogante y aprovechan los distintos tipos de interacciones para obtener los parámetros ópticos que caracterizan al material en cuestión. Una vez medidos los parámetros del material se puede conocer datos como su composición y su estructura atómica.

2.1 Tipos de interacción

Cuando la luz se encuentra con un material, su trayectoria se ve afectada. En general, parte de la luz es reflejada, parte transmitida y parte absorbida.

La reflexión en la capa superficial del material, viene determinada principalmente por el ángulo de incidencia y se puede caracterizar a priori con óptica geométrica. Sin embargo, una vez que la luz penetra en el material se producen interacciones con las partículas que lo componen y que dan lugar a distintas reflexiones y absorciones. Para poder modelar la interacción en este caso, sería necesario conocer la composición exacta del material y aún así habría un elevado número de interacciones a analizar. Existen métodos probabilísticos basados en el método de Monte Carlo que simulan el comportamiento de la luz en un material compuesto por partículas conocidas.

Se enumeran a continuación los tipos de interacción a tener en cuenta en el proyecto.

Reflexión

La reflexión es uno de los fenómenos ópticos básicos. Cuando un rayo de luz colisiona con una superficie, este rayo se ve reflejado y cambia su dirección. Es lo que nos permite vernos en un espejo por ejemplo. La reflexión de la luz se produce en la interfaz entre materiales y es debida a la diferencia de índices de refracción entre los medios. Viene dada por la siguiente relación:

$$I_R = I_o R \tag{2.1}$$

Siendo I_o la intensidad incidente, I_R la intensidad reflejada, $R = \left(\frac{n1-n2}{n1+n2}\right)^2$ el coeficiente de reflexión y n1, n2 los índices de refracción del medio inicial y reflectivo respectivamente.

Se pueden distinguir tres tipos de reflexión: especular, compuesta y difusa. Los dos primeros casos se producen cuando el rayo incidente y el rayo reflejado forman el mismo ángulo con respecto a la perpendicular a la superficie. Si la superficie es lisa será reflexión especular y los rayos reflejados paralelos entre sí. Si la superficie es rugosa, será reflexión compuesta y los rayos reflejados no serán paralelos pero se mantiene la máxima potencia reflejada en el mismo ángulo que cuando era especular. La reflexión difusa es debida a dos efectos. El primero de ellos es la luz que se refleja en la superficie del material pero el ángulo de la reflexión no es el especular. Es por tanto toda la luz que no se refleja en la dirección especular. El otro efecto es el que se da por ejemplo en los tejidos, cuando la luz entra en el tejido, interacciona con éste y sale de él. Si se analiza la reflexión en un tejido se tendrá reflexión difusa de los dos tipos, pero en otros materiales la luz apenas penetra y por lo tanto no interacciona con el material, tan sólo se ve reflejada.



Figura 2.1 Reflexión especular y difusa.

Al producirse una reflexión, la mayor parte de la potencia es reflejada en la dirección especular. Sin embargo, esta luz se refleja sin interactuar con el material y por tanto no lleva información relevante a cerca de su composición. Es la reflexión difusa la que penetra en el material y por tanto la que interactúa con éste. Así que para medir las características del material es necesario considerar la reflexión difusa y evitar la reflexión especular, que actuaría a modo de ruido de gran potencia.

Absorción

Al igual que la reflexión, la absorción es uno de los fenómenos ópticos más importantes. La luz reflejada es la que acaba llegando al ojo, pero si ésta tiene distintos colores se debe a que, el material que la refleja, absorbe muchas de las longitudes de onda de la luz y sólo refleja unos pocos.

La absorción consiste en la asimilación de la energía lumínica por parte del medio. Para que esta asimilación se produzca, el fotón debe llevar la energía necesaria (en general, ni más ni menos) para que los átomos del material puedan transportar un electrón a un nivel energético superior. Puesto que los fotones a cada longitud de onda llevan una energía determinada, cada material sólo absorberá la luz a unas longitudes de onda dadas. Esto da una curva de absorción característica para cada uno.

La absorción de un medio viene descrita por la ley de Beer-Lambert:

$$I(d) = I_0 e^{-\mu_a d}$$
 (2.2)

Siendo I_0 la intensidad inicial, *d* la distancia recorrida y μ_a es el coeficiente de absorción por unidad de longitud en $[cm^{-1}]$.

• Esparcimiento o Scattering

Cuando los fotones colisionan con las partículas del material, estos se ven esparcidos en todas direcciones. Esto es el llamado esparcimiento o *scattering*. El *scattering* se produce en cualquier medio que esté formado por partículas: en el tejido biológico son las células e incluso los orgánulos que las componen, por ejemplo cuando se pega la mano a una linterna y se ve como el tejido desprende una luz roja. También el efecto por el que el cielo es azul: la luz de bajas longitudes de onda (azul) es más esparcida debido al tamaño de las partículas en la atmósfera. Es así que dependiendo del tipo de medio y del tamaño de las partículas, se dan distintos tipos de *scattering*. Al contrario de lo que sucedía con la absorción, no importa el material sino su forma, estructura y composición.

El caso más sencillo de estudiar se da cuando el tamaño de las partículas es varios órdenes de magnitud mayor que la longitud de onda de la luz incidente. En este caso se aplica la óptica geométrica, puesto que lo que se tiene es un caso de reflexión más que de esparcimiento. Cuando las partículas son mucho menores que la longitud de onda, se produce *scattering* de Rayleigh o elástico, lo que significa que no se pierde energía durante el esparcimiento. Este es el tipo de *scattering* que da el color azul al cielo. El caso intermedio se denomina régimen de Mie, aunque la teoría de *scattering* de Mie es aplicable para cualquier relación longitud de onda–tamaño de partícula, siempre y cuando las partículas del material sean esféricas o de una geometría similar.

El esparcimiento se caracteriza mediante el coeficiente de *scattering*, μ_s y el factor de anisotropía de scattering g.

- El factor de *scattering* indica la cantidad de luz esparcida. Se define como $\mu_s = \sigma_s \rho \ [cm^{-1}]$ siendo ρ la densidad de partículas y $\sigma_s = \frac{P_{scatt}}{I_0}$ la sección de scattering, definida como potencia esparcida respecto a la intensidad inicial. Resulta más intuitiva la longitud de *scattering* $l_s = \frac{1}{\mu_s} [cm]$, que indica la distancia media que recorrerá la luz en el medio entre dos esparcimientos sucesivos, *mean free-path*.
- El factor de anisotropía g es la media del coseno del ángulo de esparcimiento y por tanto indica la dirección de la luz esparcida. El scattering de Rayleigh se caracteriza por su radiación isotrópica (g=0) mientras que en el régimen de Mie, la luz esparcida es solo es desviada ligeramente de su trayectoria inicial y sigue siendo altamente direccional (g=1). El caso g=-1 indica que la luz esparcida se dirige hacia atrás, hacia la fuente.



Figura 2.2 Ángulo de scattering [1]

Aparte de los tipos de *scattering* ya mencionados, se puede producir esparcimiento de tipo inelástico (de Raman o Brillouin). Este tipo de esparcimientos conllevan pérdida de energía y re-emisión a diferentes longitudes de onda, sin embargo su ocurrencia es mucho menor (varios órdenes de magnitud) y por tanto no son relevantes ni medibles con la instrumentación utilizada.

2.2 Interacción luz-tejido biológico

Los tejidos biológicos son tejidos muy heterogéneos y por tanto altamente dispersivos (medio turbio). Este tipo de tejidos está compuesto por una gran variedad de células, membranas y fibras con distintas formas, tamaños y propiedades ópticas. Por ejemplo, los glóbulos rojos de la sangre, tienen forma plana y su tamaño es de unas 7 micras; su principal componente es la hemoglobina, que presenta un color rojo y alta absorción en longitudes de onda visibles. El colágeno está formado por cadenas alargadas de aminoácidos de un diámetro de unos 10 nanómetros. Y las propias células, con tamaños nanométricos. Son comunes además, estructuras multicapa como es el caso de las paredes arteriales, la piel o el epitelio y las capas musculares que recubren los órganos.

Debido a la variedad de tamaños y formas de los componentes del tejido, éste presenta distintos tipos de reflexiones y esparcimientos. Trabajando con luz visible y parte del infrarrojo cercano (380nm-1000nm), las partículas más pequeñas (nm) producen un esparcimiento de Rayleigh. Las células, cuyo tamaño es del orden de los 500 nanómetros hasta cientos de micras de tamaño, producen un esparcimiento en la región de Mie. Las células más grandes son aptas para el estudio con esta teoría puesto que su forma es, en general, aproximadamente esférica.

En la siguiente figura se muestra la diferencia en el coeficiente de *scattering* cuando el tamaño de una partícula se clasifica en una u otra región. Para partículas pequeñas (en relación con la longitud de onda utilizada), se cumple la relación de Rayleigh, que tiene un factor $1/\lambda^4$ (dispersa más las longitudes de onda cortas). Cuando las partículas tienen un tamaño similar a la longitud de onda (región de Mie), ya no hay tal relación y prima más el tamaño de la partícula (la pendiente de la curva va disminuyendo a medida que aumenta el tamaño de la partícula).



Figura 2.3 Coeficiente de esparcimiento, scattering, en función del tamaño de las partículas del medio [1].



Figura 2.4 (A) Célula y orgánulos. (B) *Scattering* de Rayleigh producido por los orgánulos. El ángulo indica la desviación de la luz incidente. [1]

Puesto que el *scattering* depende fuertemente del tamaño, forma y estructura de los componentes del tejido, resulta útil para hacer valoraciones en este sentido. Midiendo el esparcimiento, se pueden detectar variaciones en la forma o tamaño de las componentes del tejido, que pueden indicar posibles anomalías.

En cuanto a la absorción, ésta viene determinada principalmente por los átomos y moléculas que conforman el tejido, más que por su tamaño. El principal componente de todo tejido vivo es el agua y de hecho es el principal absorbente. La región espectral en la que su absorción disminuye da paso a la llamada 'ventana terapéutica' (600 a 1400nm), puesto que fuera de este rango la radiación se ve fuertemente absorbida. La región de mínima absorción del agua se extiende más allá de esta ventana, desde unos 200nm (UV cercano) hasta unos 1400nm (IR cercano), pero entra en juego también la absorción de la sangre.



Diagnostic window

Figura 2.3 Coeficiente de absorción de distintos tejidos biológicos. Se indica la 'ventana terapéutica' (*Diagnostic window*) [14].

La segunda componente más abundante e importante es la sangre. Su espectro es dominante en la ventana terapéutica, de hecho la limita a la región de 600 a 1400nm, y dificulta la medida de cualquier otro tipo de tejido. La absorción de la sangre viene determinada por la hemoglobina que, oxigenada o no, es el principal componente de los glóbulos rojos. La concentración de sangre es un indicador importante del estado de los tejidos, sin embargo, debido a su alta absorción impide ver el espectro del resto de componentes. Este es, de hecho, el principal impedimento que se encuentran las técnicas ópticas en el análisis de tejidos en vivo.

La penetración de los fotones en el tejido es difícil de medir, sin embargo se han realizado modelos que describen con gran exactitud la morfología del tejido. Con estos modelos y mediante simulaciones aplicando el método de Monte Carlo se han hecho estimaciones tanto de la profundidad de penetración como de los parámetros ópticos del tejido [15], [16], [17]. El método de Monte Carlo es un método numérico que permite resolver problemas físicos y matemáticos mediante la simulación de variables aleatorias. Se utiliza cuando los problemas tienen difícil solución por métodos exclusivamente analíticos o numéricos y que dependen de factores aleatorios, como es el caso de la distribución de las partículas en un tejido biológico. Estos modelos requieren conocer la composición del tejido con mucha exactitud y por tanto no sirven para el diagnóstico si lo que se pretende es medir el estado y propiedades de los componentes.



Figura 2.4 Profundidad de penetración simulada con Monte Carlo para un medio con μ_s =5mm⁻¹, μ_a =0.4mm⁻¹, g=0.8. Fibras de diámetro de núcleo 0.2mm(A) y 0.4mm(B). La línea punteada indica la penetración efectiva [18].

2.3 Reflectancia Difusa en tejidos biológicos

Cuando la radiación lumínica entra en contacto con un tejido colisiona en múltiples ocasiones con las partículas y elementos que lo conforman. Estas colisiones diseminan la radiación en todas direcciones, de forma que, tras múltiples esparcimientos, buena parte de ella acaba siendo absorbida. Sin embargo, puesto que la luz se ve rebotada en todas direcciones, una parte de ella sale del tejido y por tanto puede ser medida y analizada. La luz que consigue escapar del tejido es muy difusa y no sería útil para formar una imagen, pero, puesto que viene determinada por el tamaño, forma y distribución de las partículas (*scattering*) y su absorción, es muy útil para caracterizar estos valores en la muestra.

La reflectancia difusa es un fenómeno que se produce en las primeras capas tras la superficie. Al adentrarse en el tejido, los fotones son esparcidos y absorbidos. Cuanto más se

adentre un fotón en el tejido, más probabilidades tiene de ser absorbido. Por tanto, es más difícil que consiga rebotar con dirección a la superficie y no ser asimilado por el medio que lo rodea. Las inversas de los coeficientes de *scattering* y absorción, denominadas respectivamente longitudes de *scattering* y absorción $l_s = \frac{1}{\mu_s}$ y $l_a = \frac{1}{\mu_a}$, representan la distancia media entre dos eventos de *scattering* o de absorción. La caracterización por reflectancia difusa tiene como objetivo medir estos dos parámetros y, a través de ellos, identificar la composición de las muestras, la concentración de sus componentes y su tamaño.

A la hora de trabajar con tejidos biológicos, el espectro a utilizar está limitado a la ventana terapéutica. Es, por tanto, lo más común trabajar con fuentes de luz halógena o cualquier otra que proporcione luz blanca de amplio espectro. Este tipo de fuentes abarca el rango espectral desde el visible hasta los infrarrojos, siendo ideal para el estudio en tejidos biológicos. El problema que tienen las fuentes halógenas es la elevada temperatura que alcanzan lo que les hace inviables para trabajar con tejidos vivos. Para este tipo de aplicaciones se suelen utilizar diodos LED de alto brillo que combinan alta radiación con condiciones de emisión de luz fría pero limitando su emisión a una región espectral más limitada. Es importante que la luz a utilizar sea de amplio espectro para poder estudiar el comportamiento espectral en un rango amplio. En ocasiones la aplicación solamente requiere analizar una longitud de onda. En estos casos se utilizará un laser o diodo de espectro estrecho.

La reflectancia difusa tiene como objetivo la medida de la luz retro-esparcida por el material. Es vital para estas medidas evitar las reflexiones directas producidas en la superficie de la muestra puesto que tienen una alta potencia y no interaccionan con el tejido, es decir, no aportan información sobre el mismo. Los montajes típicos utilizan una fibra en transmisión y otra en recepción (o varias) de un diámetro de núcleo de unas 200 micras. Estas fibras están en contacto directo con la muestra, de forma que la luz recibida es necesariamente proveniente del tejido. En función de la distancia entre fibras se puede medir la propagación a distintas profundidades y en medios multicapa se podría llegar a medir el grosor de estas [19].



Figura 2.5 Esquema básico de un montaje para la medida de reflexión difusa [1].

En algunas aplicaciones es preferible que no haya contacto entre el aparato de medida y la muestra. Es el caso de aplicaciones en las que se requiera movimiento del sistema de captación para tomar múltiples muestras. O situaciones en las que no sea posible el contacto como por ejemplo, en el tema de tejidos en vivo, el caso de las aneurismas de aorta. En un

aneurisma es muy importante conocer el estado de la pared arterial, pero el contacto del sistema de medida con la superficie arterial deteriorada podría provocar la ruptura de esta, lo cual sería fatal para el paciente. En estos casos se puede realizar un montaje que no necesite de contacto. Se ilumina y se capta la luz ya sea con las fibras ópticas al aire o con un sistema focalizador formado por lentes o esferas. En este caso hay que orientar la parte de transmisión y de recepción de forma que no haya reflexión especular, esto es, formando un ángulo distinto con la superficie. Las fibras a utilizar deben ser de un diámetro mayor que en el caso en el que hay contacto puesto que están a una distancia mayor de la muestra.

La luz retro esparcida por un medio turbio, como son este tipo de tejidos, pierde las características de onda que tenía al salir de la fuente como son fase y coherencia (cada fotón captado tendrá una fase distinta y aleatoria). A la salida del medio, solo es posible medir la intensidad, caracterizada por el coeficiente de atenuación del material ($\mu_t = \mu_a + \mu_s$).

$$R(d) = I_0(1 - R^2)e^{-\mu_t d}$$
(2.3)

Siendo *R(d)* la reflectancia medida a tras haber recorrido la luz una distancia *d*. Con esta técnica sólo se puede obtener la atenuación, no se pueden obtener por separado el coeficiente de absorción y el de *scattering*, ni el factor de anisotropía (puesto que no se mide al ángulo de la luz reflejada). Para medir esos parámetros se utilizan sistemas con esferas integradoras para captar toda la luz esparcida y ecuaciones semi-empíricas contrastadas por simulaciones de Monte Carlo [20].

Al corregir la señal en reflexión para eliminar la dependencia espectral de la fuente se divide la señal capturada entre I_0 , por tanto se tiene

$$\frac{R(d)}{I_0} = (1 - R^2)e^{-\mu_t d}$$
(2.4)

Para obtener el coeficiente de atenuación del medio, se aplica el logaritmo a la expresión anterior:

$$-\ln\left(\frac{R(d)}{I_0}\right) = -\ln\left((1-R^2)e^{-\mu_t d}\right) = -\ln(1-R^2) + \mu_t d$$
(2.5)

Por lo tanto,

$$\mu_t = -\frac{1}{d} \cdot \ln\left(\frac{R(d)}{I_0}\right) + offset$$
(2.6)

Si se mantienen las muestras a una distancia constante de la fuente y la fibra receptora, lo único que cambia es el espectro medido en reflectancia, pudiendo ver así las variaciones entre materiales o zonas de estos.

Los datos que se obtienen, como se ha mencionado, sólo indican el coeficiente de atenuación. La obtención del resto de coeficientes es más costosa y sería necesaria para caracterizar el tejido, pero eso no es lo que se busca en este trabajo. Lo que se pretende obtener, es un método sencillo para distinguir regiones dentro de una muestra, en base a diferencias en las concentraciones de los componentes. Con el coeficiente de atenuación se pueden medir esas diferencias y por tanto diferenciar zonas.

Si a la señal medida R(d) no se le aplica el logaritmo, se trabaja entonces con la señal de reflexión y no con el coeficiente de atenuación. Aun así, la respuesta sigue conteniendo gran parte de la información. Se ha comprobado que, en efecto, el espectro en reflexión normalizado se aproxima mucho a la atenuación, pero los mejores resultados se obtienen siguiendo la ley de Beer-Lambert, es decir, trabajando con -log(R) en lugar de R.

Materiales y métodos

El presente trabajo se centra principalmente en la espectroscopia de reflectancia difusa y el análisis con PCA de los datos obtenidos. Todas las medidas han sido realizadas en el rango de luz visible, con una fuente de luz halógena y un espectrómetro que abarca este rango espectral. El análisis y captura de los datos se ha hecho con MATLAB ya que facilita y agiliza la sincronización de equipos y el análisis.

Las medidas iniciales se hicieron con un sistema fijo que consistía en una fuente de luz y una fibra receptora. Más adelante se hicieron distintas configuraciones optimizando la posición de los elementos y añadiendo un sistema de posicionamiento X-Y manual. Finalmente, se añadió un motor que permite el posicionamiento automático y la sincronización de capturas y movimiento con MATLAB.

Aparte de las medidas en reflectancia, se han tomado medidas en absorción de disoluciones de tintes para comprobar el funcionamiento del algoritmo PCA.

3.1 Materiales

3.1.1 Muestras

3.1.1.1 Tejidos biológicos

Para las medidas de reflectancia difusa se han utilizado muestras de distintos tejidos biológicos procedentes de animales (vaca y cerdo) y adquiridas en un comercio de alimentación. Las muestras se han elegido por su composición en cuanto a tipos de células más características de cada tejido.

- **Hígado de vaca**. Se ha elegido éste tejido debido a que el hígado suele contener sangre en abundancia. Esto permite ver el espectro de la hemoglobina y distinguir entre regiones con más o menos sangre de forma visual.



Figura 3.1 Muestra de hígado.

 Carne de ternera de una zona de tejido muscular. Las células musculares necesitan sangre para recibir los nutrientes necesarios, pero la cantidad de ésta no es tan abundante como en el hígado por ejemplo. Cabría esperar ver diferencias en los espectros de ambos tejidos sobre todo en cuanto a los picos de absorción en la zona característica de la hemoglobina.



Figura 3.2 Muestra de músculo.

- **Panceta de cerdo**. La carne de cerdo tiene un alto contenido en grasa, pero claramente la panceta, que se obtiene de la capa externa de las zonas más grasas del animal, está repleta. Las diferencias entre zonas de grasa y zonas con tejido muscular sirven para ver las diferencias entre los espectros de una y otra zonas.



Figura 3.3 Muestra de panceta.

 Carne picada. Está formada por una masa heterogénea de grasa y carne. Se pretende comprobar si se detectan las diferencias entre las partes con grasa y con carne y el funcionamiento del algoritmo y el elemento de captación en una muestra con tantas diferencias.



Figura 3.4 Muestra de carne picada.

- **Carne de cerdo adobada**. Se ha elegido éste tipo de carne puesto que el tratamiento en adobo hace que la carne pierda las características de 'frescura' del tejido. La principal diferencia con el resto de tejidos es la presencia de grasa y agua. De hecho, en los espectros no se aprecia presencia de hemoglobina.
- Tejido de la mano. Puesto que el sistema utiliza luz visible y no es agresivo en absoluto, se tomaron medidas en distintas zonas de la mano y el antebrazo. La principal diferencia de éste tejido con los anteriores es, evidentemente, que es un tejido vivo mientras que el resto están muertos. Sin embargo, los espectros obtenidos son muy similares y contienen los mismos picos de absorción característicos de la hemoglobina oxigenada.

3.1.1.2 Disoluciones de tintes textiles

Para poder medir la concentración de cada componente y saber que la medida es correcta hay que tomar medidas de muestras con concentraciones conocidas. Puesto que la cantidad de componentes en los tejidos biológicos es desconocida, es necesario crear muestras artificiales de composición conocida, *phantoms*. Se pueden crear *phantoms* de distintos materiales, incluso con tejidos vivos y sangre [21], [22], [23], [24]. Sin embargo, es un proceso delicado y hay que contar con ciertos materiales especiales para su preparación. Es por eso que en primer lugar se decidió hacer unos *phantoms* no de sangre y elementos biológicos sino con tintes de uso textil para estudiar el comportamiento del parámetro de absorción.

Los tintes utilizados son tintes en polvo con el nombre *Remazol®*, del fabricante *Dystar*. Se presenta en polvo y por tanto ha de ser disuelto en agua para caracterizarlo. Se han utilizado tres tintes de distinto color (rojo, azul y amarillo) y se han diluido cada uno en agua en concentraciones de 0,01g tinte/2ml agua. Con las disoluciones resultantes se han obtenido mezclas con distintas concentraciones de cada tinte. Puesto que todos los tintes se han diluido con la misma cantidad de agua, la proporción entre las cantidades de cada tinte es la misma que las proporciones entre las disoluciones. Por tanto, si se toman 0,5ml de cada una disolución, habrá un tercio de cada tinte en la mezcla resultante y la cantidad de agua se puede obviar.



Figura 3.5 Disoluciones en agua de los tintes utilizados.

Las disoluciones obtenidas no tienen partículas en suspensión por lo que el coeficiente de atenuación es mayoritariamente debido a la absorción y en menor medida debido al

esparcimiento. Es por eso que la mejor forma de caracterizar las mezclas es en absorción, lo que se hace con un sistema de iluminación-captación diferente al de reflectancia difusa. Las medidas en absorción consisten en iluminar la muestra y recoger la luz que la atraviesa. Para esto se vierten las disoluciones en las cubetas que se muestran en la figura superior y se mide su espectro con la ayuda del soporte para cubetas que se menciona en el apartado siguiente.

3.1.2 Componentes ópticos

Se enumera la lista de componentes utilizados para los diversos montajes. Posteriormente se describe cada uno de ellos y su utilización en el proyecto.

- 1. Espectrómetro
- 2. Posicionadores
- 3. Motor y controladora
- 4. Fuentes de luz
- 5. Fibra óptica
- 6. Soporte para cubetas
- 7. Material de referencia
- 8. Material de calibración

3.1.2.1 Espectrómetro

Un espectrómetro es un instrumento que permite medir el espectro óptico de la luz que recibe en un rango determinado. Su principal componente es una rejilla que descompone la luz incidente en las longitudes de onda que la componen, de forma que con un sensor CCD se puede medir el nivel de intensidad para cada una de ellas y obtener el espectro. A continuación se describe el funcionamiento del aparato de forma más detallada.



Figura 3.6 Esquema del espectrómetro utilizado [imagen del fabricante].

En la figura superior se muestra un esquema general de un espectrómetro, en este caso el utilizado durante el proyecto. El punto **1** es la entrada de luz. Está preparado para un conector d fibra con rosca, de tipo SMA 905. A continuación (**2**) se encuentra el *slit*: un cristal con una abertura alargada que deja pasar solo una línea de luz. El punto **3** indica la posición de los filtros si los hubiera. El elemento **4** es un espejo colimador cuya función es enviar los rayos de

luz paralelos entre sí (colimados) hasta el siguiente elemento. Este espejo tiene el tamaño exacto para una fibra de entrada de apertura numérica 0.22 (ángulo de aceptancia de 12.7⁹). El elemento **5** es la rejilla o red de difracción. Este elemento es un cristal que tiene en su superficie hendiduras a modo de prismas sucesivos. La densidad de hendiduras determina la resolución espectral de la rejilla y del espectrómetro. Como ya se ha mencionado, su función es dividir la luz en las longitudes de onda que la componen. El elemento **6** es un espejo para focalizar la luz en el receptor **8**. Finalmente, el elemento **7** lleva todo el sistema de captura y control de los datos.

El espectrómetro utilizado es el HR2000CG-UV-NIR de *Ocean Optics* [25]. Tiene un rango espectral de 200 a 1100nm con una resolución de 1nm. Utiliza un sensor CCD de silicio de 2048 pixeles (IL511B de *SONY* [26]). La sensibilidad espectral del sensor limita el uso real del espectrómetro para esta aplicación al rango 450-750nm para no superar los 3dB de caída de potencia.



Figura 3.7 Sensibilidad espectral del sensor CCD del espectrómetro.

En cuanto al eje de intensidad, tiene un rango de hasta 16.000 cuentas y un tiempo de integración (captura) desde 1ms hasta 6 segundos. El numero de cuentas guarda relación con la intensidad, tomando 75 fotones/cuenta a 400nm y 41 fotones/cuenta a 600nm.

El espectrómetro viene con un software propio llamado *Spectrasuite*. Este software permite ajustar todos los parámetros de la captura, hacer correcciones en transmitancia y absorbancia, hacer varias capturas simultáneas y alguna utilidad adicional. Lo que no permite es la sincronización con MATLAB o cualquier otro programa y por eso ha sido necesario crear un programa propio. Se ha escrito en C++ un código utiliza las librerías de *Ocean Optics* y que permite comunicarse con el dispositivo a través de la API (*Application Programing Interface*) de éste. Con este código se obtiene un *ejecutable.exe* que puede ser llamado desde MATLAB y recibir parámetros de entrada, en este caso el tiempo de integración en microsegundos y el nombre del archivo de salida.

La fibra óptica a la entrada del espectrómetro debe ser la proporcionada por el fabricante o en su defecto otra fibra con apertura numérica de 0.22. Debajo de la **Figura 3.6** se explica el motivo (elemento 4 del esquema).

3.1.2.2 Posicionadores

El posicionador manual utilizado junto con el motor y en el caso no automatizado es el M-423 series de *Newport*. Tiene una precisión de 0.02mm y un desplazamiento máximo de 46mm.



Figura 3.8 Posicionador manual [imagen del fabricante].

3.1.2.3 Motor y controladora

Se ha utilizado un motor y controladora del fabricante PI (*Physik Instrumente*). La controladora es la *MercuryC-862*. Se comunica con el ordenador mediante la interfaz RS-232 (puerto serie) y con el motor con unos conectores propios que le permiten ordenar la ejecución de un movimiento y ver la posición. Permite conectar otros controladores en serie para manejar hasta 16 dispositivos. Tiene un conjunto de comandos propios [27] y la capacidad de almacenar macros (conjuntos de parámetros que se ejecutan juntos de una sola vez). Puesto que se conecta por puerto serie, los comandos se le indican en formato ASCII, al igual que se lee la información del estado que pueda proporcionar.

El motor es del mismo fabricante pero viene sin etiquetado, por lo que debe ser caracterizado. El desplazamiento no se indica por distancia sino por cuentas, por tanto hay que hacer una traducción para saber la equivalencia. Para hacer esta traducción se le indica que se mueva un gran número de cuentas, próximo al rango para disminuir el efecto del error. Se mide el desplazamiento en milímetros en cuatro ocasiones con dos elemento de medida de gran precisión: un calibre y un micrómetro. Se encuentra una equivalencia de 17.050 cuentas/mm con una diferencia entre medidas de 10 cuentas/mm (0,058%). Además, el controlador tiene un software que permite medir el error de desplazamiento en número de cuentas. Los mayores errores encontrados son en grandes desplazamientos y se ha medido un error máximo de 10 cuentas que traducido a milímetros da una cota de 0.00058mm. En el proyecto se hacen medidas de hasta 0.01mm por lo que el error mencionado (que sólo se produce en grandes desplazamientos, de unas 200.000 cuentas) es poco relevante.

3.1.2.4 Fuentes de luz

Se han utilizado tres fuentes de luz, todas ellas halógenas (lámparas con filamento de tungsteno y un gas del grupo químico de los halógenos) por lo que el rango espectral va desde el visible hasta el infrarrojo cercano (>750nm). Con éste tipo de fuentes hay que tener en cuenta que conllevan un tiempo de calentamiento para la estabilización del nivel espectral que suele ser de unos 20 a 30 minutos. Hay que esperar ese tiempo desde que se enciende hasta que se toman las medidas para que no haya fluctuaciones entre las medidas.

- La fuente DH-2000 de *Ocean Optics* [28] ha sido utilizada en las medidas de absorción. Aparte de la fuente halógena (de tungsteno) incluye una fuente de deuterio (no utilizada) que sirve de complemento espectral para abarcar la zona ultravioleta



(<400nm). Se especifica el tiempo de calentamiento de 40min para la parte de deuterio y 20 min para la halógena.

Figura 3.9 Izquierda espectro de la fuente de Ocean Optics. Derecha imagen de la fuente [28].

- La fuente utilizada para medir la reflectancia de los tejidos es una la fuente PL-950 del fabricante *Fiber Lite [29]* que utiliza una lámpara halógena de cuarzo. Lleva acoplada una fibra con un cabezal que distribuye la luz en forma de línea. Como su nombre indica, sirve para iluminar una franja de la muestra de forma uniforme y es muy útil para orientar la fibra receptora controlando el ángulo de iluminación.



Figura 3.10 Fotografía de la fuente con el elemento que produce iluminación en una línea.



Figura 3.11 Esquema del montaje realizado con la fuente de línea y la fibra receptora en ángulo.

 La última fuente, utilizada en el montaje para medir la altura del receptor, es un foco halógeno de gran potencia. Se ha elegido este tipo de fuente porque ilumina una gran área de forma uniforme. Debido a su alta potencia (>500W) desprende mucho calor por lo que no es viable para iluminar tejidos biológicos. Con estos últimos se emplea la fuente de línea.



Figura 3.12 Fotografía de la fuente de iluminación distribuida.

3.1.2.5 Fibra óptica

Las fibras utilizadas son fibras de 600 micras de diámetro de núcleo del tipo UV/SR-VIS de *Ocean Optics* [30].Están diseñadas para trabajar en la región visible y tiene alta atenuación hasta los 400nm y desde los 850nm, por lo tanto tiene el mismo rango de trabajo que el espectrómetro. Se ha elegido fibras de gran tamaño para facilitar la captación de luz. En los montajes se ha capturado la luz con la fibra desnuda apuntando a la muestra. Para conseguir una focalización más fina se puede colocar algún elemento entre la fibra y el material a estudiar, como una lente biconvexa o una lente de tipo GRIN (focalización gradual).



Figura 3.13 Atenuación de las fibras ópticas utilizadas.

La fibra de entrada al espectrómetro debe tener una apertura numérica de 0.22 $(\theta_{máx} = \pm 12.7^{\circ})$ para que el dispositivo mida correctamente (**Apartado 3.1.2.1**). La apertura numérica de una fibra (*A.N.*) guarda relación con el ángulo de aceptancia máximo del núcleo $(\theta_{máx})$ y éste a su vez depende de los índices de refracción del núcleo (*n1*) y la cubierta (*n2*) de la fibra.

$$A.N. = n \cdot sen\theta_{max} = \sqrt{n_1^2 - n_2^2}$$
(3.1)

El ángulo $\theta_{máx}$ se define como el ángulo máximo respecto al eje de la fibra para el que ésta es capaz de captar luz. *n* es el índice de refracción del medio de entrada a la fibra, en general aire por lo que n=1.

3.1.2.6 Soporte para cubetas

Como su nombre indica, se trata de un soporte preparado para colocar cubetas con disoluciones y medir su absorción. Tiene dos conectores para conectar las fibras a los extremos cada uno con una lente colimadora. Las cubetas son de plástico y permiten trabajar entre 200 y 900nm.



Figura 3.14 CU-V de Ocean Optics [imagen del fabricante].

3.1.2.7 Material de referencia

El elemento utilizado como referencia es una pastilla de Spectralon[®] SRS-99-020 [31]. Spectralon[®] es una resina termo-plástica basada en un fluoro-polímero que produce más cantidad de reflectancia difusa que cualquier otro material en el rango UV-Vis-NIR del espectro. Por ello, es un material producido por el fabricante *Labsphere* que está diseñado para estandarizar la respuesta en reflectancia difusa de los materiales. Su reflectancia es superior al 99% en el rango 400-1500 nm, y mayor al 95% al rango 250-2500 nm. Su respuesta en plana en el rango de 250-2500nm con variaciones de ±4%. Su respuesta es también altamente lambertiana, esto es, la reflexión en el material es isotrópica y la intensidad reflejada depende del ángulo de observación, es decir, es máxima en la perpendicular a la superficie y decae de forma proporcional al coseno del ángulo.



Figura 3.15 (Izquierda) Respuesta espectral del material Spectralon para la gama de productos. (Derecha) Formato comercial del material [imagen del fabricante].

3.1.2.8 Material de calibración

Para comprobar la calibración del espectrómetro se ha utilizado una pastilla del material WCS-MC-020 (Multi-Component Wavelength Calibration Standard) de *Labsphere* [32]. Este material es un compuesto que combina óxidos de tres tierras raras (óxido de holmio, óxido de erbio y óxido de disprosio) para producir picos con características de absorción intensas y resolubles y así poder servir como patrón de referencia espectral para calibrar espectrofotómetros de reflectancia.



Figura 3.16 Formato comercial del material WCS de Labsphere [imagen del fabricante].

3.2.1Calibración espectral del espectrómetro

El primer paso antes de realizar cualquier medida es comprobar la calibración del espectrómetro. Para esto hace falta contar con un espectro conocido y medible, a ser posible con picos muy marcados. Para esto se emplea un material de *Labsphere*, WCS (*Wavelength Calibration Standard* ver **figura 3.16**) fabricado para este fin. La ventaja de éste material es que su espectro tiene múltiples picos a determinadas longitudes de onda. Tras hacer una medida del material y tras la corrección mencionada en el apartado anterior, se da por buena la calibración del aparato.





Figura 3.17 (Arriba) espectro del material WCS facilitado por el fabricante y longitudes de onda de los picos. (Abajo) espectro medido con los picos. Los picos medidos están dentro del rango que indica el fabricante [13].

3.2.2 Procedimiento de medida

Al medir los espectros fruto de la interacción con materia, estos vienen siempre marcados por el espectro de la fuente. Puesto que lo que se quiere caracterizar es la muestra y no la fuente, es preciso hacer una corrección del espectro para eliminar la dependencia del espectro de la muestra del espectro de la fuente. Esta corrección se debe hacer en cada medida que se toma. Antes de empezar a tomar medidas es necesario establecer el *tiempo de integración* del espectrómetro, esto es, el tiempo que permanece capturando intensidad para cada medida, de forma que se aproveche el rango de amplitudes disponible pero sin saturar el equipo. Según se indica en la descripción de este equipo, el tiempo de integración no ha de ser exacto pero es conveniente ajustarlo para aprovechar al máximo la cuantificación del espectro y sin saturar el receptor. Para esto, antes de hacer las medidas se mide con la pastilla del material de referencia Spectralon[®] y se ajusta el tiempo de integración hasta que los valores de intensidad estén cerca del máximo de cuentas del equipo. Cualquier otra muestra que se mida será de menor intensidad que la referencia y, por tanto, no saturará el receptor.

En los primeros montajes, las medidas se han hecho con el sistema quieto y medidas simples: se toma la referencia, se ajusta el tiempo de integración, se mide el ruido de oscuridad y a continuación se hacen las medidas. A continuación se introdujo un sistema de posicionamiento en dos ejes. Esto permite hacer una 'fotografía' o tomar una imagen del espectro del material haciendo desplazamientos de la muestra frente al punto de medida. Pero lleva su tiempo desplazar la muestra en uno o dos ejes, tomar una medida y guardar su orden dentro de la matriz de datos, volver desplazar la muestra y así sucesivamente. Es por eso que se decidió agilizar el proceso e incorporar un motor al montaje para automatizar las medidas en uno de los ejes (lo ideal habría sido hacerlo en dos ejes pero uno de los motores no funcionaba correctamente).

Para hacer el proceso automático, hay que sincronizar los desplazamientos del motor con la toma de muestras del espectrómetro y el tiempo de integración que este necesita para tomar cada medida. Para esto se centraliza el proceso desde MATLAB. Se controla el motor mediante el puerto serie del ordenador y una controladora y el espectrómetro se controla mediante el puerto usb con una aplicación de elaboración propia.



Figura 3.18 Esquema de uno de los montajes realizados.

3.2.2.1 Corrección de los espectros en reflexión/absorción

Al hacer una medida espectral utilizando una fuente de luz, el espectro medido es principalmente el espectro original de la fuente con las modificaciones producidas en la muestra. Hay que sustraer la característica de la fuente para tener una caracterización del material que no esté condicionado a esa fuente en concreto. Por tanto, para eliminar la característica espectral de la fuente de luz se ha de corregir el espectro medido. Lo mismo sucede con el ruido de oscuridad del dispositivo de medida. El ruido de oscuridad es un ruido blanco y por tanto tiene un espectro plano que añade un offset a la señal. La corrección del espectro consiste en tomar una medida de la respuesta espectral de la fuente y una medida sin entrada de luz en el dispositivo receptor. De esta forma se tienen los dos espectros y se pueden sustraer de la señal a medir.



Figura 3.19 (Izquierda) Espectro de la fuente de luz halógena medido en reflectancia sobre el Spectralon[®]. (Derecha) Ruido de offset medido con la fibra óptica tapada.

En un sistema diseñado para caracterizar espectros de absorción la medida del espectro de la fuente se hace dejando libre el espacio entre la fuente y el receptor. En un sistema de medida de reflectancia difusa, la forma de medir el espectro de la fuente de luz debe ser también en reflectancia. El material de referencia sobre el que se mida la reflexión de la fuente debe presentar una respuesta espectral lo más plana posible. Típicamente se ha utilizado la alúmina, pero hay otros materiales con una respuesta más plana, como es el sulfato de bario (BaSO₄), oxido de magnesio (MgO) o Spectralon[®] (Se detallan en el **apartado 3.1.2.7** sus características).

Una vez obtenido el espectro del ruido de oscuridad $N(\lambda)$ y el espectro de la fuente $B(\lambda)$ (ya sea en reflectancia o en absorción), se realizan las medidas del espectro si $S(\lambda)$ se corrigen de acuerdo a la siguiente ecuación:

$$S(\lambda)_{corregido} = \frac{S(\lambda)_{medido} - N(\lambda)}{B(\lambda) - N(\lambda)}$$
(3.2)



Figura 3.20 Espectro en reflexión de una muestra de tejido. (Izquierda) Espectro capturado. Las líneas de puntos indican la región del espectro recortada. (Derecha) Espectro corregido.

Con esta corrección se elimina la dependencia de la fuente y otros elementos del sistema y se elimina el ruido, pero el rango espectral sigue dependiendo del espectrómetro. Debido a la sensibilidad de éste, el espectro queda ruidoso en las longitudes de onda de los extremos. En esa zona el espectrómetro no capta la intensidad suficiente debido a su baja sensibilidad espectral. Es por eso que la señal para esas longitudes de onda es en su mayor parte ruido de oscuridad y al hacer la corrección ésta zona queda muy distorsionada. Teniendo esto en cuenta, el rango espectral de las medidas queda limitado a un rango menor del nominal, que finalmente está 450 a 850nm.

Hay que mencionar que aunque se haga una corrección del espectro, los tubos fluorescentes producen unos picos en el espectro que, como se ve en la **Figura 3.20** y **3.21**, no se pueden eliminar por completo. Por tanto es necesario aislar el sistema de este tipo de iluminación. Una opción es cubrir el banco de medida con un material que no deje pasar la luz visible, por ejemplo una mampara de cartón o cartulina, o apagar la luz del ambiente si se puede.

3.2.3 Bancos de medidas

Los bancos experimentales se centran en tres áreas:

- Medidas en una posición fijas. El objetivo de estos montajes es medir los espectros de distintos materiales y disoluciones para ver su respuesta espectral y los componentes que lo forman.
- Medidas con desplazamiento en dos ejes. Consiste en tomar varias muestras a lo largo y ancho del tejido para después procesar los datos y hallar sus constituyentes principales, así como hacer un mapa de las regiones según la concentración de sus componentes.

3.2.3.1 Medidas en una posición fija

Se ha utilizado dos bancos de medidas: el primero en reflectancia difusa y el segundo en absorción con diferentes versiones en el primero de ellos.

3.2.3.1.1 Reflectancia difusa

Este montaje es la toma de contacto con el método de medida. Sirve de entrenamiento para establecer los rangos de funcionamiento del sistema y la posición óptima de los elementos: hay que determinar la posición de la fibra receptora y la fuente, la una respecto a la otra y respecto a la muestra. Y también para establecer el procedimiento de medida.

En el primer montaje se utilizo una fibra en recepción y otra en transmisión. El problema de éste montaje es que la luz que llega al material y es captada por la fibra en recepción no es suficiente para caracterizar el material con detalle. Se optó entonces por la fuente de línea que es capaz de proporcionar más potencia y en un área mayor de la muestra.



Figura 3.21 Esquema del montaje con recepción y transmisión en fibra. La zona de iluminación debe ser mayor que la de captura.

- Elemento de captación. Para la captura se utiliza una fibra apuntando a la muestra. El ángulo que forma con la superficie del tejido debe variar en concordancia con el ángulo de iluminación para evitar captar la reflexión especular.
- Fuente de luz. Se ilumina con la fuente de línea puesto que permite aplicar suficiente potencia y de forma homogénea a la vez que permite 'apuntar' a la zona de captura (spot de la fibra de captura) sin necesidad de gran precisión.
- Distancia a la muestra. La distancia de la fuente a la muestra se toma de forma que la región de iluminación sea amplia y cubra sobradamente la zona de captación. La distancia de la fibra de captura a la muestra se ajusto de forma experimental. Variando ésta distancia desde 0mm (contacto con el tejido) hasta 3cm se comprobó que la altura donde se captaban con más detalle los espectros es a aproximadamente 1cm de la muestra. A distancias inferiores es más fácil que se ensucie la fibra con la muestra y es difícil iluminar.
- Orientación de los elementos. Puesto que hay que evitar la reflexión especular, la orientación de la fuente de iluminación y la fibra de captura debe estar en un ángulo adecuado. La mejor forma de conseguirlo es colocando uno de los elementos en la perpendicular a la superficie de la muestra y el otro elemento formando un ángulo menor.



Figura 3.22 Posiciones de la fuente de línea y la fibra de captura. A la derecha la fibra se encuentra perpendicular a la muestra y la fuente de iluminación formando un ángulo entre 30 y 60º. A la izquierda se colocan de forma inversa.

Una vez hallada la mejor posición para los elementos del montaje, se puede establecer un procedimiento de medida. Éste procedimiento es bastante sencillo pero es necesario mantenerlo:

- en primer lugar se ha de encender la fuente puesto que ha de calentar para que se estabilice su emisión
- a continuación se tapa la fibra y se toma una medida que servirá para hacer la corrección del ruido de oscuridad del espectrómetro
- después de media hora la respuesta de la fuente ya es estable
- se orienta la luz hacia la zona de muestreo y se coloca la pastilla de material de referencia
- se ajusta la altura a 1cm aproximadamente de su superficie. En este paso habrá que ajustar el tiempo de integración al valor más alto sin saturar y se toma una medida que será el espectro de referencia para la corrección de la fuente
- a continuación se pueden colocar las muestras y se puede proceder a tomar las medidas.

Como se ha mencionado anteriormente, la iluminación de los tubos fluorescentes no se puede corregir completamente porque contiene unos picos demasiado pronunciados. Por eso es necesario tomar las medidas con la iluminación ambiental apagada.



Figura 3.23 Esquema del montaje para medidas estáticas de reflectancia difusa.

3.2.3.1.2 Absorción

El sistema para la medida de absorción es más simple que el anterior puesto que el tema de iluminación y captación queda cubierto al utilizar el soporte para cubetas. Como su nombre indica, su función consiste en sostener las cubetas con las disoluciones entre la fibra de emisión y la de captura. Tiene una tapa que permite aislarlo de la luz ambiental mientras se toma la medida. Se conecta la fuente halógena-deuterio, con salida de fibra óptica al soporte para cubetas y este al espectrómetro. Puesto que las cubetas tienen dos orientaciones posibles hay que mantener la orientación elegida. Para hacer la corrección del espectro se toma en primer lugar una muestra con una cubeta rellena con la misma agua utilizada para disolver los tintes y se ajusta el tiempo de integración en el espectrómetro. A continuación se tapa la fuente con el obturador y se toma otra muestra que sirve para eliminar el ruido del sistema.



Figura 3.24 Esquema del montaje para medidas de absorción. Detalle del contenedor de cubetas.

El procedimiento de medida es similar al caso anterior. La fuente a utilizar también es halógena por lo que tiene un tiempo de calentamiento también de media hora. La captura del espectro de referencia tiene ahora como objetivo no sólo eliminar el espectro de la fuente del espectro final, sino también el espectro del plástico de la cubeta y del disolvente de la mezcla a ser posible (en este caso agua).

- en primer lugar se ha de encender la fuente para que caliente y se estabilice su emisión
- a continuación se cierra la salida de luz de la fuente con el obturador, se cubre el porta cubetas con su tapa y se toma una medida que servirá para hacer la corrección del ruido de oscuridad del espectrómetro
- después de media hora la respuesta de la fuente ya es estable
- se llena una cubeta idéntica a las que se utilizarán con las muestras y se llena de agua
- se ajusta el tiempo de integración al máximo rango de amplitud sin saturar y se captura el espectro de referencia en absorción

En todas las capturas hay que tapar el porta cubetas antes de tomar la medida para aislar el sistema de la luz ambiente y evitar diferencias de niveles de de intensidad entre medidas.

3.2.3.2 Medidas con desplazamiento en dos ejes

Estas medidas se han hecho para hacer una imagen espectral del tejido: se recorre el tejido en dos ejes y se obtiene el espectro en cada punto. Al final se tiene una matriz con los espectros en cada punto que se puede comparar con una fotografía del tejido. El resultado es similar a la descomposición de una imagen hiperespectral. En un sistema hiperespectral se toman varias imágenes a distintas longitudes de onda. Aquí se toman medidas del espectro en varios puntos de forma que al final se tiene el espectro en cada punto de igual forma que en las anteriores.



Figura 3.25 Imagen y espectro en cada punto superpuesto.

3.2.3.2.1 Sistema manual

El primer montaje se realizó con dos posicionadores manuales que permiten el desplazamiento en dos ejes. Manteniendo la fuente de luz y la fibra de recepción fija, se varía la posición de la muestra y se van almacenando las medidas. El desplazamiento de la muestra se debe calcular para evitar que las áreas asociadas al cono de aceptancia de la fibra detectora se solapen. Por tanto, debe ser mayor que la zona de captura de la fibra (mayor que el radio del spot de la fibra) para la altura a la que se encuentra.

Para ver bien la zona a la que apunta la fibra, se coloca otra en el mismo eje de manera que ambas apuntan al mismo punto cuando la altura es de 1 cm. Esta comprobación se hace para el primer punto y el resto ya no es necesario puesto que el desplazamiento será constante y mayor que la zona de captura.


Figura 3.26 (Derecha) Montaje con la fibra para apuntar encendida. (Izquierda) Medida de una muestra con el montaje descrito.

3.2.3.2.2 Sistema automatizado

Puesto que hacer las medidas con el sistema manual es un trabajo tedioso y lento, se acopló un motor al montaje (en principio dos motores pero uno de ellos no funcionaba correctamente y hubo que desplazar un eje manualmente). La diferencia que hay con este montaje respecto al anterior es que la muestra se desplaza automáticamente y las capturas se sincronizan con el movimiento, por tanto sólo hay que hacer cada ciclo un desplazamiento en uno de los ejes. El resto de consideraciones son las mismas. Se muestra un esquema del montaje en la siguiente figura.



Figura 3.27 Esquema del montaje incluido el motor. El resto de montajes son simplificaciones de éste.

En todos los montajes anteriores las capturas del espectro se realizan desde el programa *Spectrasuite* proporcionado por *Ocean Optics*. Los espectros capturados se almacenan y se tratan posteriormente con MATLAB. El software indicado tiene incorporada algunas funciones para corregir los espectros, calcular la transmitancia o absorbancia y alguna más. Sin embargo para el análisis resulta más cómodo almacenar los espectros sin procesar y a posteriori realizar cualquier cambio. En éste banco de medidas, se añade el motor con su controladora para desplazar la muestra. El motor también tiene un programa propio que permite controlarlo desde el ordenador introduciendo los comandos de control. Por tanto, ambos elementos, espectrómetro y motor, tienen programas propios que permiten controlarlos desde el ordenador. Aun así el control sigue siendo manual.

Para automatizar el proceso de medida hay que controlar ambos dispositivos de forma automática y sincronizar sus actuaciones para que las medidas se tomen en el punto indicado y no haya desplazamientos mientras se está capturando. Para esto se utiliza MATLAB, como ya se ha comentado anteriormente.

En primer lugar, el espectrómetro se controla a través del puerto USB del ordenador. Para pasar y recibir datos se ha programado un pequeño ejecutable que pasa como parámetros el tiempo de integración y el nombre del archivo de salida. En primer lugar se comprueba si hay algún espectrómetro conectado y si no es así toma la medida indicada y se guarda el archivo con las medidas en el fichero indicado.

Para programar éste ejecutable se han utilizado los drivers que proporciona el fabricante. Estos drivers, de nombre *Omnidriver* xx55, están creados en *java* de forma que sirven para cualquier sistema operativo. Aunque esté basado en java, se puede programar en distintos lenguajes (*C*, *C++*, *C#*, *Pascal*, *Delphi*, *LabVIEW*, *Visual Basic*) haciéndoles comunicar con la plataforma *java* a través de librerías. En este caso, por familiaridad con el lenguaje, se ha escogido *C++*. Es necesario un entorno de programación para éste lenguaje, como es *Visual C++* de Windows, la librería del espectrómetro para el sistema operativo de Windows (*.dll*) y las cabeceras para C++ (*.h*) para enlazar con código del programa y crear el *ejecutable.exe*. También es necesario tener instalado el *JDK (Java Development Kit)* que incorpora las librerías de java necesarias para poder desarrollar aplicaciones y que funcione correctamente la plataforma *java*.

Una vez creado el ejecutable, se le llama desde MATLAB y se le indican los parámetros de entrada.



Figura 3.28 Esquema de comunicación de MATLAB con el espectrómetro.

La comunicación con el motor es mucho más sencilla que en el caso anterior. El motor va conectado a una controladora que es quien recibe las órdenes y controla el estado. La controladora se conecta al ordenador a través del puerto serie o interfaz RS-232. Tiene una lista de comandos para mover el motor, leer el estado de éste o establecer parámetros como velocidad, punto de origen, etc. Estos comandos están explicados en el manual xx16 y puesto que comunica a través del puerto serie, se envían desde MATLAB como caracteres alfanuméricos. Antes de establecer la comunicación se debe abrir el puerto serie en el ordenador y al terminar se debe cerrar para no dejarlo bloqueado.



Figura 3.29 Esquema de comunicación de MATLAB con el motor.

Una vez descrito el funcionamiento de los dos elementos se pasa a describir la toma de medidas:

- Al igual que en los casos anteriores, se ha de encender la fuente halógena, la fuente de línea en éste caso, treinta minutos antes de empezar a tomar las medidas para que se estabilice su respuesta espectral.
- Una vez pasado éste tiempo se coloca la pastilla de referencia en reflectancia a 1 cm de la fibra de captura y se ajusta el tiempo de integración para aprovechar el rango de amplitudes del espectrómetro sin saturarlo.
- A continuación se tapa la fibra y se toma una medida para corregir el ruido de oscuridad.
- Se coloca la muestra y la fibra a 1 cm de ésta.
- En último lugar, se ejecuta el código de MATLAB que automatiza la medida. El código para las medidas debe seguir en cualquier caso el siguiente orden.



Figura 3.30 Diagrama de flujo seguido para tomar medidas con el sistema automatizado.

Para tomar las medidas de forma correcta se debe esperar el tiempo de integración desde que se da la orden al espectrómetro de tomar una medida antes de continuar ejecutando el código. Lo mismo sucede con el motor, cuando se le manda moverse se debe esperar a que haya completado su movimiento para proseguir con la ejecución. Si no se hace así puede ocurrir que comience una medida cuando el motor aún está en movimiento o que el motor se mueva cuando la medida aún no ha concluido.

Una vez se han completado las medidas se tienen todos los datos almacenados y se puede proceder al análisis de los espectros obtenidos.

Análisis espectral de la reflectancia

4.1 Introducción

En este capítulo se describirán los diferentes métodos aplicados sobre las medidas ópticas de reflectancia difusa. Básicamente se va a realizar:

- un estudio *cualitativo* intentando descubrir de forma ciega la identidad de los materiales existentes en el tejido
- un estudio *cuantitativo* para medir la concentración en la que se encuentran esos componentes.

Se partirá como ya se ha comentado de medidas de reflectancia difusa realizadas con el espectrómetro. Estas medidas son adquiridas y posteriormente analizadas por una plataforma controlada de forma centralizada con Matlab[®]. Puesto que el análisis de la reflectancia difusa es una parte importante del presente trabajo, se explica en este capítulo el flujo de trabajo y las herramientas o funciones utilizadas para ello. La **figura 4.1** recoge de forma gráfica este flujo de trabajo:



Figura 4.1 Esquema de flujo de éste trabajo.

La reflectancia difusa es capturada por el espectrómetro. Los datos obtenidos son los espectros de las muestras a medir. Estos espectros están condicionados por el espectro de la fuente de iluminación, así como por la respuesta espectral de las fibras ópticas y del sensor del espectrómetro. Para eliminar su dependencia espectral se corrige el espectro en reflexión, esto es, se elimina el espectro debido a todos los elementos del montaje. El espectro lleva información acerca del tejido, su coeficiente de atenuación, por medio de una exponencial negativa (ley de Beer-Lambert), por tanto para extraer esa información se debe aplicar el logaritmo neperiano y aplicar ese cambio de signo. Este paso puede crear problemas con la aplicación posterior de PCA puesto que no admite entradas de valores infinitos ya que este algoritmo se basa en la maximización de la varianza y un valor infinito estropearía los resultados. Por eso se debe realizar un paso anterior para comprobar que el espectro no toma valores nulos ($log(0)=\infty$) y eliminar aquéllos picos de ruido que se producen sobre todo en las longitudes de onda de los márgenes espectrales debidos principalmente a la baja respuesta espectral del CCD del espectrómetro en estas zonas.

Una vez se preprocesan los espectros, se puede aplicar el algoritmo PCA. De éste algoritmo se obtienen dos tipos de resultados: los *coeficientes*, de los que se intentará interpretar e identificar los constituyentes del tejido (análisis cualitativo) y los *scores*, a partir de los que se cuantificará la concentración de estos (análisis cuantitativo). Finalmente, se comprobará que tanto la reconstrucción de las concentraciones, como la composición del material son las obtenidas calculando el coeficiente de correlación de Pearson de éstas con las concentraciones conocidas y los espectros conocidos respectivamente.

4.2 Herramientas utilizadas

Se describen a continuación las principales herramientas usadas para el análisis de los datos. Puesto que se trabaja con MATLAB, estas herramientas se traduce en funciones implementadas para dicho entorno.

4.2.1 Obtención de las componentes principales (PCA)

Este algoritmo es la base del presente proyecto para la identificación de componentes en un tejido. Se explica de forma más extendida en el **apartado 4.3**. En cuanto a su implementación en MATLAB, se realiza mediante la función *princomp* [39].

4.2.2 Correlación

Del análisis de los datos se obtiene bien una componente principal (PC) de la que se quiere comprobar si se parece a los espectros conocidos, o bien una estimación de la concentración de un compuesto de la que se necesita comprobar si coincide con la concentración real. Para ver cuánto se parece una figura a otra se utiliza el coeficiente de correlación de Pearson (r). Este coeficiente mide la correlación lineal entre las señales de entrada en un rango independiente de las amplitudes de entrada siendo 0 para señales completamente distintas y 1 cuando la señal coincide. En general, se considera que dos señales 'se parecen' para r>0,95. El coeficiente de Pearson entre dos señales x e y viene definido por [40]:

$$r = \frac{Cov(x,y)}{\sigma_x \sigma_y}$$
(4.1)

Siendo Cov(x, y) la covarianza entre las señales x e y, σ_x , σ_y la desviación estándar de x e y respectivamente.

Es importante también el valor p (p-value). Este valor revela el nivel de confianza sobre el análisis realizado, se define como la probabilidad de encontrar un vector que tenga una

correlación tan alta como r, suponiendo que en realidad las señales no están correladas. Lo que se consigue es ver si r es significativo, es decir, comprobar que no es pura casualidad obtener ese valor. Típicamente se considera un valor de correlación apto si p<0,05.

La implementación de esta comprobación se hace a través de la función de Matlab corrcoef [40].

4.2.3 Ajuste de curvas

Para el ajuste de curvas se ha utilizado dos funciones distintas: fit y polyfit.

- La función *fit* [41] sirve para ajustar los valores obtenidos por la función. Se le indica la variable independiente y las dependientes, la función a la que se pretende ajustar. Como resultado devuelve los valores de las constantes para el ajuste y una estructura con diversos datos sobre la bondad del ajuste, entre ellos el error cuadrático medio o RMSE (*Root Mean Square Error*).
- La función *polyfit* [42] sirve para hacer ajustes polinómicos junto con *polyval* [42]. Un ajuste polinómico de grado *n* es un ajuste del tipo:

$$p(x) = p_1 x^n + p_2 x^{n-1} + \dots + p_n x + p_{n+1}$$
(4.2)

El uso que se le da es para ajustar los valores obtenidos de concentraciones por tanto el grado es 1. La función busca los coeficientes p_i que consiguen un mejor ajuste en base al criterio del mínimo error cuadrático.

La reconstrucción de la recta de ajuste con este polinomio se hace con el comando *polyval* [43].

La comprobación de los ajustes se hace en ambos casos con el RMSE (*Root Mean Square Error*). Este estimador viene a expresar la media del error cuadrático cometido en todos los puntos del ajuste y está expresado en las mismas unidades que los valores. Si x es la señal original, \hat{x} la ajustada y n el número de puntos a ajustar, se define el RMSE como:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \hat{x}_i)^2}{n}}$$
(4.3)

4.3 PCA: Análisis de Componentes Principales

PCA es un método matemático utilizado para reducir el número de variables en un sistema de elementos correlados. Este método efectúa una transformación de ejes, de forma que en el sistema transformado, los vectores sean ortogonales y por tanto incorrelados. Al eliminar la correlación de los vectores iniciales se reduce la redundancia en los datos y por tanto el número de variables que aportan información al sistema. PCA ordena el nuevo sistema en orden de varianza, por lo que se pueden tomar sólo los primeros y prescindir del resto sin perder información. Se puede entender como un método de compresión. En ocasiones, como es el caso de este proyecto, se puede reducir de mil componentes a tan solo un par de ellas. La figura 3.2 muestra un ejemplo de la aplicación de PCA. A la izquierda sistema original. La varianza es significativa en ambos ejes. A la derecha el sistema tras la transformación con PCA. En el eje vertical se ha reducido la varianza considerablemente y por tanto esta variable es prescindible puesto que apenas aporta información.



Figura 4.2 Ejemplo sencillo de aplicación de PCA [33].

Éste método es llamado de análisis multi-variable puesto que trabaja con varios vectores típicamente observaciones o medidas (espectros en esta aplicación) que comparten un eje o referencia (rango espectral o longitudes de onda). Es además un método ciego, puesto que sólo recibe como entrada las observaciones. Sin necesidad de información adicional acerca de los datos del sistema, obtiene las componentes principales en base a sus momentos estadísticos.

Típicamente, PCA se utiliza para eliminar datos redundantes en un sistema y tomar solamente los que contienen más variabilidad para caracterizar a la fuente de datos con estas componentes [35], [36], [37]. En este trabajo lo que se pretende conseguir del algoritmo es identificar los constituyentes principales de un tejido en base a su espectro en reflectancia y medir la cantidad de los constituyentes en distintas regiones de la muestra.

Una forma sencilla de comprender el funcionamiento de PCA es verlo como un detector de diferencias. Si se le indican M vectores, el método va a buscar la forma más sencilla de diferenciar entre ellos, esto es, va a buscar las diferencias más significativas.

4.3.1 Descripción del algoritmo PCA

La entrada del algoritmo es la matriz de observaciones X, formada por vectores columna que contienen las medidas en cada caso. Si se miden N puntos en cada medida y se realiza M medidas, X será una matriz de dimensiones NxM.

$$X = \begin{bmatrix} x_1(\lambda) \\ x_2(\lambda) \\ \vdots \\ \vdots \\ x_M(\lambda) \end{bmatrix}^T \quad con \ \lambda = 1, \dots, N \tag{4.4}$$

El símbolo T significa traspuesta. Para este proyecto, se miden M espectros en distintas muestras, cada uno formado por N longitudes de onda. Puesto que se quiere clasificar los espectros, la dimensión a reducir es el número de longitudes de onda.

A la matriz de entrada se le aplican transformaciones lineales, de modo que la matriz resultante, Y, se construye de la siguiente forma:

$$y_i(\lambda) = \sum_{j=1}^M x_j(\lambda) \cdot W_{ji} \quad i = 1, \dots, N$$
(4.5)

Siendo W la matriz de coeficientes para la transformación. Por construcción se cumple:

$$Y_{NxM} = X_{NxM} \cdot W_{MxM} \tag{4.6}$$

Para conseguir la matriz de coeficientes, el algoritmo PCA consta de los siguientes pasos [33], [34]:

1. Sustraer la media de cada medida (*M*), Puesto que el algoritmo se basa en encontrar la máxima varianza. La media de cada fila se define como sigue:

$$\bar{x}_m = \frac{\sum_{i=1}^N x_m(i)}{N}$$
 (4.7)

Ahora *X* será la matriz inicial con media por filas nula.

2. Calcular la matriz de covarianza

$$C_{M \chi M} = X^T X \tag{4.8}$$

 X^T es la matriz X traspuesta. La matriz de covarianza es simétrica respecto a la diagonal principal, que es unitaria.

3. Calcular los autovalores y autovectores de la matriz de covarianza.

Se tiene la matriz A y se quiere transformar en la matriz B. Para esto, habrá una matriz V no nula tal que VA = VB. V es la matriz de autovectors de A y B es la matriz de autovalores de A asociado a V.

Puesto que A tiene un número de vectores propios igual su dimensión (es lo que se quiere conseguir) A es diagonalizable a través de V: $VAV^{-1} = B$ siendo B una matriz cuyos elementos de la diagonal principal son los autovalores y el resto son ceros. La matriz V contiene en sus filas los autovectores. Siendo la matriz V de autovectores ortogonales o perpendiculares, esto se traduce en que $V^T = V^{-1}$.

En el caso que estamos tratando, se calculan los autovalores y autovectores de la matriz de covarianza para construir la matriz de coeficientes de la siguiente forma: $C = VDV^{-1}$

4. Ordenar las componentes. El último paso de PCA para reducir el número de variables es colocar los autovectores según el orden de los autovalores en orden descendente. Así, los primeros vectores serán los más relevantes. Una vez colocados, la matriz de coeficientes buscada es $W_{MXM} = V_{MXM}$. La varianza acumulada será la que proporcionan los autovalores correspondientes a cada autovector.

Una vez organizados los vectores por tamaño, se tiene:

$$C = X^T X = V D V^{-1}$$

La matriz de scores es $Y = X \cdot W$ y W es V, así que despejando de la ecuación anterior:

$$X^T X V = V D$$
$$X V = (X^T)^{-1} V D$$

Al final se tiene $Y_{NxM} = X_{NxM} \cdot W_{MxM}$ los scores $Y = (X^T)^{-1} VD$. La expresión anterior es la transformación completa. Para tomar solo las componentes principales, se tomará W_{pxM} y por tanto se obtiene Y_{pxN} . Un buen método para delimitar el número de componentes p a tomar es tener en cuenta la varianza acumulada, esto es, el tamaño de los autovalores respecto al total.

Una consideración muy importante a tener en cuenta es que este algoritmo no conserva el signo de los vectores de entrada a la salida. Esto es debido a que lo único en lo que se basa el método es en encontrar la máxima varianza entre vectores pero en ningún momento se considera el signo de los valores. Este hecho se traduce en que algunos de los resultados puedan salir desfasados 180º, 'dados la vuelta'.

4.3.2 PCA y el logaritmo

Los datos de entrada al algoritmo PCA serán los diferentes espectros medidos. Una vez que los espectros están corregidos y se ha eliminado la influencia de la fuente de luz y el espectrómetro (capítulo 4) se tiene el espectro completo de la muestra. Siendo estrictos, a este espectro medido se le debe aplicar el logaritmo neperiano y cambiarle el signo, puesto que no se trata de medir un espectro sino la característica en atenuación de la muestra. Sin embargo, aplicar el logaritmo a señales muy pequeñas produce como resultado señales muy altas. Si se tiene una respuesta espectral cercana a cero, como sucede por ejemplo con el espectro de un colorante rojo en longitudes de onda bajas, el logaritmo de éste tenderá a infinito. Es fácil comprender que la atenuación en ese punto no es infinita sino que es muy alta, pero el algoritmo de PCA no tolera entradas no numéricas como son $\pm\infty$. Además, las fluctuaciones producidas por el ruido en estas zonas se ven magnificadas al aplicar el logaritmo.

Este hecho produce que en ocasiones no se pueda aplicar el logaritmo. Aun así, en algunos casos se puede hacer una aproximación que permite aplicar PCA y obtener resultados satisfactorios.

Los espectros que se manejan son siempre de amplitud menor que la unidad puesto que han sido previamente corregidos en reflexión, esto es, normalizados frente al espectro de la fuente. Cuando se aplica el logaritmo a una función x que toma valores cercanos a cero, el teorema de Taylor dice que se puede aproximar el logaritmo por una función polinómica. El desarrollo en serie de Taylor de una función es el siguiente [35]:

$$\lim_{\substack{\Delta x \to 0 \\ x \approx a}} f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots$$
(4.9)

Siendo la función f(x) el logaritmo del espectro en reflectancia corregido (amplitud <1) en algunas zonas donde la señal no varía demasiado se tiene:

$$\lim_{\substack{\Delta R(\lambda) \to 0 \\ R(\lambda) \approx a}} \ln(R(\lambda)) \approx \ln(a) + \frac{1}{a \cdot 1!} (R(\lambda) - a) = \frac{1}{a} R(\lambda) + \ln(a) - 1$$
(4.10)

Lo que resulta en una función lineal con la reflectancia, por lo que la información que aporta el coeficiente de atenuación se puede llegar a aproximar por el espectro antes de aplicar el logaritmo y de esta forma se puede aplicar el algoritmo de PCA. Como se comentó en el **apartado 2**, el espectro en reflexión antes de aplicar el logaritmo, $R(\lambda)$, lleva gran parte de la información que contiene su coeficiente de atenuación. Sin embargo, la aproximación anterior solo es correcta cuando los valores medidos de $R(\lambda)$ son de una amplitud similar. Es por eso que en algunos casos se puede hacer la aproximación y en otros no.

En general resulta más conveniente 'limpiar' la señal eliminando los ceros del espectro. En algún caso esto se puede hacer cortando la región del espectro donde sus valores son próximos a cero. Pero esto conlleva una pérdida importante de información sólo asumible en algunos casos. Otra opción es limitar la amplitud del espectro antes de aplicar el logaritmo en función de la energía medida. Se calcula el histograma, que muestra el número de ocurrencias que hay en la medida para las distintas amplitudes. Por ejemplo, en la **figura 4.3** hay entre 150 y 200 ocurrencias (puntos espectrales) con una amplitud de entre 0.1 y 0.15. Se pueden tomar distintas consideraciones para establecer el límite por el que recortar, generalmente para mantener un porcentaje de energía (un número de valores respecto del total) como son los cuartiles que dividen la medida en zonas que abarcan el 25%, 50%, 75% de energía.



Figura 4.3 Histograma de una de las medidas espectrales.

En las medidas tomadas, se ve que efectivamente entre amplitudes de 0.1 y 0.15 hay una alta ocurrencia y por tanto ésta región hay que considerarla. Así que el límite a establecer será, en este caso, 0.1. El máximo también está limitado y se podría tomar como límite 0.5 o 0.55 para conservar más ocurrencias, pero en realidad no es necesario porque el valor máximo va a ser menor que 1 (al corregir el espectro, el máximo sólo lo alcanza el material de corrección) y éste valor no da ningún problema en PCA al aplicar el logaritmo.

4.3.3 Interpretación de PCA

En origen, el algoritmo toma como parámetro de entrada únicamente una matriz de observaciones X y, de forma ciega, hace un tratamiento de los datos para obtener aquellos más relevantes. Como resultado, el algoritmo devuelve tres resultados: la matriz con las componentes principales Y (típicamente llamados *SCORES*), la matriz de coeficientes W y los autovalores de la covarianza de X (normalmente llamados *latent*).

En este trabajo la matriz X está formada por vectores fila de medidas compuestos por los valores para las distintas longitudes de onda. Al aplicar el algoritmo, los resultados obtenidos y

su interpretación son los que se explican a continuación con un ejemplo en el que sólo son relevantes dos componentes.

En los *SCORES* aparecen las componentes principales obtenidas con el algoritmo. Esto significa que, de todas las longitudes de onda, sólo se habrá quedado con unas pocas y además habrán sido transformadas al nuevo espacio ortogonal. En el ejemplo, a la entrada hay tres medidas con 2048 longitudes de onda, a la salida puede haber, como en el ejemplo, sólo dos longitudes de onda representativas (componentes principales) que son suficientes para diferenciar los tres vectores de medidas.



Figura 4.4 Dibujo de los primeros SCORES. Sólo son relevantes dos componentes. Los SCORES diferencian entre regiones del vector de entrada a modo de mapa.

La matriz de coeficientes es el operador, o matriz de transformación, por el que se transforma de un espacio a otro. Si se han identificado tan sólo dos componentes principales, con las dos primeras columnas de ésta matriz se puede transformar otras medidas al nuevo espacio. El resto de vectores suelen ser más ruidosos o nulos. La **figura 4.5** muestra los autovectores para el caso de la **figura 4.4**, se puede comprobar cómo los correspondientes a las componentes mayores a la segunda no aportan información espectral significativa.



Figura 4.5 Primeras columnas (longitudes de onda) de la matriz de coeficientes. Son suficientes las dos primeras para realizar una transformación sin perder información.

Y el vector de autovalores, *latent*, sirve para medir la varianza que aporta cada nueva componente al sistema. Es el mejor método para ver cuántas componentes es necesario tomar para conseguir mantener la cantidad deseada de información. Siguiendo con el ejemplo

anterior, se puede comprobar cómo a partir de la segunda PC la varianza acumulada llega a un punto de saturación.



Figura 4.6 Vector latent. Indica la varianza acumulada al incrementar el número de componentes a considerar.

4.4 Medidas del diámetro del spot

Para saber con exactitud el tamaño del área de iluminación/captación de la fibra, se puede conocer sabiendo la distancia de la muestra a la terminación de la fibra. Si se sabe con exactitud la distancia y la apertura numérica de la fibra, es sencillo calcular el diámetro del spot. Pero si el sistema focalizador no es una fibra o no se conoce la distancia, es más difícil conocer ese diámetro. Lo que se pretende en este apartado es medir el diámetro del spot con métodos de reflectancia difusa y de forma automática e independiente del sistema de captación. El tener una medida precisa de esta magnitud permitirá automatizar el sistema de captación de reflectancia moviendo la muestra de tejido en la magnitud adecuada a la colocación del sistema en cada momento.



Figura 4.7 Esquema del spot de una fibra óptica.

El diámetro del spot se calcula como:

$$\emptyset' = \emptyset + 2b = \emptyset + 2d \cdot tan\theta_{max}$$

Siendo θ_{max} el ángulo de aceptancia dado por el fabricante, de 12,7º, Ø el diámetro de la fibra, 600µm, y d la distancia que es la variable independiente. Se tiene entonces la ecuación lineal que define el ancho del spot con la distancia:

$$\phi' = 0.6 + 0.4507 \cdot d \text{ [mm]}$$

Como resultado de las medidas, se debería obtener una recta con los mismos parámetros que la anterior, lo cual indicaría que se está midiendo correctamente el diámetro.

Con el montaje diseñado se pretende medir el ancho del spot en base a la cantidad de potencia captada mientras se desplaza la muestra. Se coloca un elemento que refleje bien la luz como es una placa de alúmina encima de un material que no sea buen reflector, como una cartulina negra. Cuando el área que captura la fibra se encuentre en la zona de alúmina, la potencia captada será máxima. Al desplazarse la muestra, llega un momento en el que el área ya no sólo comprende la región de alúmina y empieza a adentrarse en la región oscura, por lo que la potencia recibida será menor. Cuando el spot de captación esté ubicado sobre la zona oscura, la potencia capturada será mínima.



Figura 4.8 Esquema del área de captura al desplazarse la muestra. Izquierda posiciones del spot. Derecha intensidad medida al desplazarse el spot. La medida del diámetro se toma entre las líneas verticales.

Se toman medidas para alturas de 2 a 10 mm y se recorre un eje con el motor en pequeños incrementos. Para ajustar la altura se utiliza un poste con una abrazadera y unos 'tacos' cuyos grosores son las alturas necesarias (2 a 10mm). Como iluminación se utiliza un foco halógeno de gran potencia de forma que ilumine la muestra de manera uniforme. Para que la relación del diámetro del spot con la altura sea lineal, la fibra debe colocarse en la perpendicular a la superficie de la muestra. De otra manera, la luz incidiría en ángulo y el spot no sería redondo sino elíptico y por tanto la potencia capturada no se incrementaría de forma lineal con la altura.



Figura 4.9 Fotografía del montaje. Al lado de este se encuentran los 'tacos' de distintas alturas para ajustar la distancia fibra-muestra.

Las medidas se realizan cada cierta distancia (0.1 mm, 0.05 mm ó 0.01 mm dependiendo del día) a lo largo de la muestra y para alturas que van de 2 a 10 mm. La idea original es medir el diámetro del spot como la distancia comprendida en la zona donde la potencia de luz captada por la fibra varía en el rango desde el 95% al 5%. Pero éste método resulto ser poco preciso porque ligeras fluctuaciones en los puntos medidos provocan una gran diferencia en la medida del diámetro. Se producen incongruencias como el hecho de que para alturas que el spot debería ser menor, queda mayor que el anterior.



Figura 4.10 Valores medidos con el desplazamiento del motor. En rojo límites 95%-5% para la medida del diámetro del spot.

Debido a este hecho se decide cambiar el método de medida por un ajuste por la función gaussiana. Para cada altura, se hace un ajuste de la respuesta obtenida por una distribución Normal, tomando como radio del spot 2σ (94% de potencia). Los diámetros medidos son similares a los obtenidos con el otro método pero más precisos y congruentes.



Figura 4.11 Ajuste de los puntos medidos (verde) por una gaussiana (rojo) para una distancia fibra-muestras de 8mm.

Tras obtener los diámetros con la gaussiana se ajusta su distribución por una recta, que resulta la siguiente:



 $\phi'_{ajuste} = 0.4677 + 0.1095 \cdot d \text{ [mm]}$

Figura 4.12 Diámetro del spot real (negro) y estimado (rojo) para uno de las medidas.

Se tomaron medidas en días distintos con el mismo montaje y los datos del ajuste fueron:

Día	КО	К1
1	1.1554	0.1870
2	0.8956	0.2057
3	0.9354	0.2191
4	1.0705	0.1889
5	1.0387	0.2188

$$\emptyset'_{ajuste1} = k0 + k1 \cdot d$$



Figura 4.13 Diámetros de spot obtenidos en distintos días.

Las medidas del diámetro del spot parecen no ser directas. Se comprueba a continuación si la medida no es directamente el ancho del spot pero pueda ser proporcional. Para esto modifica la pendiente multiplicando los diámetros medidos por distintos factores y se toma aquel que de una máxima correlación con la recta de diámetro del spot real.

Día	КО	K1	Factor
1	2.7730	0.4489	2.4
2	1.9702	0.4525	2.2
3	1.9644	0.4601	2.1
4	2.5691	0.4534	2.4
5	2.1813	0.4595	2.1



Figura 4.14 Diámetros medidos en uno de los casos tras aplicarle el factor.

Aplicando el factor indicado, se ha igualado la pendiente de la curva que representa el diámetro del spot al de la pendiente del spot, pero esto ha descuadrado el valor de la constante independiente.

Como conclusión de las medidas y análisis hechos en éste apartado, se determina que la medida realizada en reflectancia difusa y la ecuación teórica para el cálculo del diámetro del spot no coinciden.

El diámetro calculado a partir de las medidas mantiene la dependencia lineal con la distancia fibra-muestra, pero el spot que permite captar información a una distancia tan grande (>1mm) parece ser en realidad menor de lo que indica la óptica geométrica.

Según la óptica geométrica, el área de captura es todo el spot debido al ángulo de aceptancia máxima de la fibra, pero según las medidas realizadas la zona de captura es menor. Esto se debe a que la distribución de luz en la fibra no es uniforme, ni viajando por su interior ni a su entrada o salida. La distribución de energía en la fibra cae al incrementar el radio, de forma que una mayor parte de la potencia viaja por el núcleo y al acercarse a los bordes disminuye de forma exponencial. Esta energía distribuida en los bordes del spot apenas aporta energía al cómputo global y de hecho las medidas del spot efectivo, el área que realmente sirve para la captura, es bastante menor al calculado.



Figura 4.15 Esquema del spot de iluminación y la distancia entre pasos. (Izquierda) considerando el planteamiento de óptica geometría. (Derecha) según las medidas.

El estudio de éstas medidas tiene el objetivo de determinar el desplazamiento Δd necesario para capturar espectros en puntos sucesivos de una muestra sin solapar información. Según la óptica geométrica la distancia entre capturas adyacentes debe ser al menos el diámetro del spot, Ø. Pero en realidad, la distancia debe ser menor puesto que el spot efectivo, la parte de éste que aporta energía a la medida, es menor. Si se muestrea aplicando desplazamientos de igual valor al diámetro real se quedaría una zona de la muestra sin analizar. Si el desplazamiento se aplica en función del diámetro medido se puede ajustar el salto al diámetro efectivo del spot de captación y se hace una focalización más precisa. Así se evita dejar zonas sin caracterizar y se aprovecha el "efecto focalizador" que permite medir áreas de muestra de un tamaño menor. Con éste método se puede determinar cuál es ese radio de captura efectivo, puesto que a simple vista no se puede saber que parte del spot aporta una potencia significativa.

Identificación y cuantificación de componentes en tejidos

Las medidas realizadas se clasifican en tres objetivos: el análisis de los espectros de forma visual, el análisis de los espectros a través del algoritmo PCA y la medida de la concentración de componentes aplicando PCA.

5.1 Primer análisis: análisis visual de los espectros

Las primeras medidas corresponden a espectros de tejidos y están tomadas con el sistema de medida fijo. Se analiza y compara carne de cerdo adobada (tejido muerto), presumiblemente sin sangre y con más grasa y tejido de la mano (tejido vivo), con más sangre. El objetivo de este análisis es proporcionar una idea global y cualitativa de los datos que se van a medir y sus características principales.



Figura 5.1 Espectros de tejido muerto (izquierda) y vivo (derecha) en comparación con tejido graso .

Se comparan los espectros de grasa, medidos en partes de la muestra donde solo hay grasa, con el espectro del tejido muerto de la zona de carne. A simple vista se puede apreciar que desde los 600nm en adelante la forma y niveles son muy parecidos. Sin embargo en la zona inferior del espectro, los valores de grasa son muy inferiores a los de las otras muestras, en torno a un 20%. A simple vista, la zona de 450-500nm es una región útil para detectar la grasa del tejido.



Figura 5.2 Espectro de tejido vivo y muerto comparados.

En la figura superior se ha contrapuesto el espectro de las muestras de tejido vivo y de tejido muerto. Los espectros son muy parecidos pero se aprecian dos máximos en el rango 500-600nm en el espectro del tejido vivo. Estos máximos representan la principal diferencia entre los tipos de tejidos. Para analizarlos mejor, se aplica el logaritmo neperiano al espectro del tejido vivo y se le cambia el signo, según lo visto en el capítulo 4, con lo que se obtiene la atenuación del tejido μ_t .



Figura 5.3 Izquierda espectro medido en el tejido vivo. Derecha espectros de hemoglobina y agua [1].

Comparando el espectro medido con los espectros de hemoglobina (oxigenada y desoxigenada), parece que los picos vistos en la zona de 500 a 600nm se corresponden con picos en absorción de la hemoglobina oxigenada. Esto explica que en el tejido vivo se midan estos picos y en el muerto (que está adobado y no contiene sangre) no estén presentes. Aparte de los picos mencionados, también parece haber correlación entre el espectro medido y el de hemoglobina desoxigenada y agua en la región de 700 a 800nm.



Figura 5.4 Espectro de tejido vivo tomado en distintas zonas de este

Puesto que la región del espectro que se corresponde con la sangre está identificada, se toman distintas medidas en distintas zonas del tejido vivo (mano y antebrazo). En la imagen superior se aprecian diferencias en los niveles de esos picos y por lo tanto en los niveles de sangre. Por lo tanto, parece posible medir la cantidad de sangre presente en una región analizando los niveles de reflectancia en esta región del espectro.

5.2 Identificación de componentes mediante análisis espectral de la imagen

Después de los resultados del apartado anterior, parece factible diferenciar entre zonas de un mismo tejido en base a alguno de sus componentes, como sangre, grasa o agua. Se modifica el montaje y se añaden los posicionadores para poder desplazar las muestras y tomar imágenes espectrales de los tejidos. Esto es, analizar el espectro en distintos puntos para luego hacer una comparación y compararlo con una imagen del tejido. Las muestras a medir son hígado de ternera (alto contenido en sangre), tejido graso (carne de cerdo con un alto contenido en grasa), músculo (células musculares) y picado de carne (muy heterogénea con partes de carne y de grasa). En la imagen inferior se muestra la media de los espectros capturados en estos tejidos. Se puede ver que en el hígado y el musculo tienen un mayor nivel de sangre que se hace patente en los picos antes mencionados. En los otros tejidos la cantidad es considerablemente menor.



Figura 5.5 Media de los espectros tomados para cada tipo de tejido.

5.2.1 Análisis en el tejido de hígado

A continuación se explica el procedimiento y el análisis con las muestras de hígado, siendo análogo para el resto de tejidos.

En primer lugar se toman las medidas y se analizan los espectros. En la figura 5.6 se ve la imagen de la que proceden las medidas, a la derecha y a la izquierda el espectro para los puntos marcados que son los puntos de cruce de las líneas. En algunos de ellos se perciben los picos de absorción de la sangre y en otros no. Esto se debe a la presencia o no de sangre en esa zona y es visible en la imagen. La zona más blanca tiene sangre en su superficie y la zona oscura es más bien tejido, con menos sangre. Esas diferencias se aprecian en los espectros medidos en cada zona.



Figura 5.6 Imagen del tejido (derecha) y su espectro en los puntos señalados (izquierda). Los espectros han sido corregidos en reflexión y se encuentran en el rango de 500 a 800nm.

Puesto que a simple vista se puede decidir más o menos en que zonas hay sangre y en cuales no, aplicando el algoritmo de PCA se simplificará el número de componentes a analizar, esto es, se simplificará el número de longitudes de onda a tener en cuenta. Al aplicar PCA se obtienen tres datos distintos. El primer dato es un vector que indica la varianza acumulada, de

forma que estableciendo un valor razonable, como pueda ser el 99%, se puede decidir el número de componentes necesarias y despreciar el resto. En el caso de las muestras de hígado, el número de componentes necesarias según este criterio es 3.



Figura 5.7 Varianza acumulada (normalizada) en función del número de componentes tenidas en cuenta. Sólo se cuantifica para las primeras 10 componentes porque el valor rápidamente llega al 100%.

A la hora de identificar las componentes, la varianza acumulada es un buen método para descartar algunas, pero no dice cuál de ellas se corresponde con el espectro de los elementos constituyentes del tejido. Para hacer esta distinción se ha de trabajar con la matriz de coeficientes.

La matriz de coeficientes contiene los vectores que se utilizan para hacer las transformaciones entre espacios vectoriales asociados a la traslación y rotación vinculada al análisis PCA. Estos vectores se obtienen a partir del tratamiento que hace PCA de todas las medidas hechas. El algoritmo PCA funciona, a grandes rasgos, buscando un patrón que le permita distinguir entre todos los vectores de la matriz de entrada. Es por ello que el espectro de un constituyente puede influir en la diferenciación y esta influencia se hará patente en las columnas de la matriz de coeficientes.

En el caso de los tejidos biológicos y disoluciones, los patrones para diferenciar regiones proceden de los espectros de los materiales constituyentes. Si un elemento, como la sangre, aporta una componente espectral muy característico al espectro global de la muestra, la diferencia de niveles de sangre se traducirá en la elección de esta componente espectral. Recordando del **Apartado 5.1** como la presencia de sangre se medía en la región de 500 a 600nm, puesto que era la zona que más diferenciaba los tejidos (el resto del espectro era prácticamente el mismo) PCA selecciona el espectro en esta zona (aportado por la sangre) como una Componente Principal.



Figura 5.8 Columnas de la matriz de coeficientes (llamadas PC) frente a los espectros conocidos [13]. Según la evolución de la varianza se necesitan 3 PCs para cubrir el 99% de varianza. Si se toman más, como son la 4 y la 5, se demuestra que podrían ser despreciables por ser ya muy ruidosas.

Para comprobar cuál es el constituyente del tejido que determina cada componente principal, se comprueba la correlación de sus espectros con las componentes principales obtenidas con PCA. Esto es, se correlan las columnas de la matriz de coeficientes con los espectros de algunos constituyentes del tejido biológico como son la hemoglobina desoxigenada (*Hb*), hemoglobina oxigenada (*HbO2*), agua y grasa. Hay que mencionar que PCA presenta incertidumbre del signo, es decir, no lo tiene en cuenta en el cálculo y sólo se encarga de encontrar la variación máxima en los datos. Es por eso que alguna componente puede salir invertida y la correlación será entonces negativa. Por ello, el coeficiente de correlación se ha de tomar en valor absoluto.



Figura 5.9 Componente nº2 del análisis (PC2) y espectro de hemoglobina oxigenada. Derecha correlación de las 5 primeras componentes con los espectros conocidos. Los valores de correlación más altos se muestran en blanco. Los cuadros negros reflejan una correlación baja y un nivel de confianza (p-value) mayor de 0,05.

Como resultado se obtiene que la segunda componente obtenida del análisis (PC2) ofrece un coeficiente de correlación con el espectro de hemoglobina oxigenada de 0,968. Esto indica que la PC2, utilizada para diferenciar los vectores de medidas, es obtenida casi por completo de la hemoglobina oxigenada (sangre) que hay en el tejido. Esto significa que de todas las medidas, la cantidad de hemoglobina oxigenada es un componente que ayuda a diferenciar entre ellos.

El resto de correlaciones de PCs con los espectros conocidos son inferiores. La hemoglobina desoxigenada también da un valor alto (0,923) porque el espectro de *Hb* y *HbO2* son muy parecidos, pero sólo el segundo tiene esos dos picos característicos. Para comprobar la validez de las correlaciones se ha establecido un nivel de confianza <0.05 tal y como se explicaba en el **Capítulo 4**.

El último resultado del análisis son los llamados *Scores*. Estos valores forman la matriz de salida tras la transformación:

$$score_{NxM} = entrada_{NxM} \cdot coeficientes_{MxM}$$
 (5.1)

El score n-ésimo se construye multiplicando la fila n-ésima de la matriz de entrada (espectro medido) por la columna n-ésima de coeficientes (*PCn*). Así, si una PC se corresponde con la hemoglobina oxigenada, el score que se ha construido con esa PC, reflejaría y sería un indicador de la la cantidad de PC que tiene cada punto del score, esto es, la cantidad de hemoglobina.

En este caso y en el resto, el score de la primera componente tiene una correlación muy alta no con determinadas componentes sino con el espectro en general. Si se dibuja este score se puede apreciar a simple vista la correlación con los espectros medidos.



Figura 5.10 Izquierda mapa o score de la PC1. Derecha espectros medidos en el hígado. En la zona resaltada se aprecia la presencia de sangre.

En la **figura 5.10**, en los score se marca según la cantidad con un color más fuerte o más claro (si la componente queda positive o negativa será blanco o negro). Ésta diferencia de intensidad se comprueba que coincide con los espectros en los puntos donde se aprecian mejor los picos de absorción de la hemoglobina oxigenada.

5.2.2 Análisis en todos los tejidos

Para el resto de tejidos se aplica el mismo método. Puesto que se pretende detectar más componentes se aplican también enventanados de las regiones características de los espectros conocidos para comprobar la correlación con las componentes medidas. Los resultados obtenidos se muestran a continuación.

5.2.2.1 Tejido muscular

Según el criterio de la varianza, son necesarias 3 componentes.

La correlación más alta es la de la segunda componente con la hemoglobina oxigenada con un valor absoluto de 0,978 (es negativo por la ambigüedad de signo del algoritmo). Sin embargo, a la vista de las imágenes de la figura superior, se observa un pico de absorción entre 700 y 800nm que coincide con el espectro de absorción de la grasa. Detectado esto, se "enventana" esa región y se calcula la correlación. Por "enventanado" se entiende recortar las dos señales a comparar para tomar solamente una parte del espectro. Las ventanas se aplican en las regiones del espectro donde se detectan picos que pueden ser característicos de alguno de los materiales conocidos, así se compara solamente esa región concreta de la componente principal con el espectro conocido y toda la componente, que puede estar formada por la composición de respuestas espectrales de distintos elementos en el tejido.



Figura 5.11 Cuatro primeras PC (columnas de la matriz de coeficientes) y correlación de las cinco primeras con los espectros conocidos.

El resultado de esta correlación dice que ese pico de atenuación no se debe a la grasa sino a la hemoglobina desoxigenada, que tiene un pico en esa región que no se ve en el espectro completo.



Figura 5.12 Correlación de la componente 1 con el espectro de grasa (izquierda) y hemoglobina desoxigenada (derecha). La correlación con la hemoglobina es mucho mayor.

El pico en 760nm no coincide con la grasa. Pero en la región que va desde 800 nm en adelante, el espectro de hemoglobina es plano y la grasa sí tiene singularidades. Se enventana esta región y, en efecto, la PC4 da una correlación de 0,957 con el espectro de grasa. El problema es que en esa región el sensor del espectrómetro tiene poca sensibilidad y el espectro medido es muy ruidoso.



Figura 5.13 Izquierda correlación de las componentes con los espectros conocidos tras el enventanado. Derecha superposición del espectro de grasa y la PC4 en la región espectral enventanada.

Finalmente se muestran los scores, la imagen del tejido y los espectros y se ve como todas ellas correlan visualmente, lo mismo que se ha obtenido con los datos.





Figura 5.14Análisis espectral del tejido muscular: (a) mapa de scores para la PC1_x; (b)tejido muscular y rejilla de muestreo; (c) espectros medidos.

En la figura 5.14 se muestra los scores obtenidos con PCA (A), el tejido analizado (B) y los espectros medidos en los puntos indicados del tejido (C). El espectro en C se debe a los componentes en B. Se aprecia diferencias sobre todo en cuanto al contenido en hemoglobina. En A se muestra el mapa obtenido con una de las componentes de PCA al recibir como entrada los espectros en C. La componente en cuestión es la número 1, que si bien no es debida a la hemoglobina, se debe a la forma completa de los espectros. En las tres figuras hay más cantidad de hemoglobina en la esquina superior derecha.

5.2.2.2 Tejido graso

En este tejido la clasificación está mucho más clara puesto que hay zonas con mucha grasa y zonas con sangre. Según el criterio de la varianza, sólo es necesaria una componente para hacer distinciones. Puesto que hay regiones con grasa y regiones sin ella, la distinción hecha por PCA se basa en la hemoglobina, igual que en casos anteriores, aunque las zonas con hemoglobina y la cantidad de esta son mínimas. La correlación es de 0,928 entre la PC1 y la hemoglobina desoxigenada. Pero este factor de correlación es más bajo que en los casos anteriores. En efecto, la diferenciación no es tanto por la hemoglobina sino por el nivel de absorción en esta zona. Analizando la **Figura 5.1** y **5.2** de este capítulo, en el tejido graso la atenuación es menor en la misma zona en la que se hace patente la hemoglobina. En este caso la diferenciación se hace en función de la absorción en esa zona, no en función de los picos de hemoglobina.



Figura 5.15 Tejido graso: (a) espectros medidos; (b) muestra de tejido graso. Derecha espectros obtenidos de la imagen de la derecha.

En la figura superior se aprecia como en la 3ª fila (correspondiente a la grasa del tejido) la atenuación es menor en las longitudes de onda bajas que en las filas de su alrededor. Además en los puntos centrales de ésta fila el nivel es aún mayor, debido a que el área de captación comprende una región mayor de grasa.

Si se enventanan las componentes, como en el tejido anterior, se ve como la PC2 correla en la misma zona (>800nm) con el espectro de grasa con un factor de 0,95.



Figura 5.16 Correlación de las componentes del análisis con PCA y los espectros conocidos en el rango 500 a 900 nm.



Figura 5.17 Correlación de las componentes del análisis con PCA y los espectros conocidos en el rango de 800 a 900nm.

Finalmente, se muestra el mapa de scores, donde éste vuelve a correlar con la imagen como grasa-no grasa.



Figura 5.18 Imagen de origen de las capturas y scores de PC1 para los puntos señalados.

Aunque la PC2 se deba al espectro de la grasa, la reconstrucción con scores para comparación de forma visual con la muestra se identifica mejor con la PC1. Esto es porque, en ocasiones, según los casos vistos, la primera componente no identifica los componentes del tejido sino el espectro completo. Es por eso que, siendo espectro visible, la imagen se parece más a simple vista al mapa con estos scores.

5.2.2.3 Tejido mixto

En éste tejido se necesitan 3 componentes para conservar el 99% de varianza. La correlación de las principales PCs con los espectros conocidos dan valores muy bajos, pero al igual que en los dos casos anteriores, se observan partes del espectro que sí que correlan visualmente con los espectros conocidos. Se aprecia en las PC2 y PC3 el pico de absorción del agua hacia los 670nm y los picos de hemoglobina entre 500 y 600 nm.



Figura 5.19 (a) componente PC2 y (b) componente PC3 del tejido mixto.

Se enventana en primer lugar la región de 500 a 600nm, donde se encuentran los picos de la hemoglobina. Aunque se ve la influencia de ésta, la PC3 en esa zona es debida principalmente a la absorción de la grasa y el agua con correlaciones de 0,953 y 0,940 respectivamente (el resto de componentes tienen una correlación menor).



Figura 5.20 PC3 en el rango 500 a 600nm comparada con el espectro de grasa (izquierda) y agua (derecha).

Haciendo un enventanado en la otra zona señalada, de 600 a 700nm se obtiene que las componentes correlan más con la hemoglobina desoxigenada (0,928) que con el agua (0.846). Todas las componentes tienen una alta correlación con los espectros de hemoglobina y agua pero ninguna de ellas se corresponde completamente a uno de estos componentes.



Figura 5.21 Correlación de las componentes con los espectros conocidos en la ventana de 600 a 690nm.

Por último, se enventana la región de 800 a 900nm en la que la grasa era característica. En esta región la correlación más alta se da entre la componente PC5 y el espectro de grasa (r=0,910).



Figura 5.22 Correlación de las PC con espectros conocidos en la región de 800 a 900 nm. A la derecha la componente con la correlación más alta, que se da con la grasa.

Analizando las distintas ventanas espectrales se comprueba que varios componentes tienen una correlación alta con las componentes obtenidas, pero estas componentes están formadas a lo largo de todo su espectro por composiciones de los espectros de los distintos tipos de constituyentes. Los valores más altos de correlación obtenidos se dan entre el agua y la grasa con una de las PC para la región de 500 a 600 nm, donde también se ve la influencia del espectro de la hemoglobina oxigenada (**Figura 5.19**).

Al igual que sucede con las PC, los scores no aclaran mucho. Si parece que en efecto, las zonas con más grasa son las que se marcan en los scores de un color más claro, como parece indicar el enventanado en 500 a 600nm.



Figura 5.23 Scores e imagen de la muestra. Las zonas más claras de los primeros parecen indicar la presencia de grasa en la muestra.

Como resumen de las medidas y resultados de este apartado, se reseña que se puede determinar la presencia de uno o más constituyentes en los tejidos biológicos analizados. La información contenida en las componentes PCA permite extraer el significado físico y químico de los compuestos. En ocasiones hay que hacer un enventanado para comparar la correlación de las PC obtenidas con el espectro de constituyentes conocidos. Este enventanado se hace en las zonas en las que los espectros de constituyentes presentan singularidades en el espectro, picos de absorción. Los constituyentes para la comparación son algunos de los que se sabe que conforman los tejidos biológicos y se ha comprobado que en efecto, se puede encontrar más de un constituyente con este enventanado.

5.3 Cuantificación de composición mediante el análisis de disoluciones

En este apartado se analizan los datos tomados de la medida en absorción de distintas soluciones. Tras ver los resultados del apartado anterior, que determina que se puede detectar la presencia de determinados constituyentes en un tejido, se pretende poder medir la concentración de estos. Para hacer esta comprobación es necesario saber la concentración de constituyente para compararla con el resultado obtenido del análisis con PCA. La concentración quedará determinada en los scores obtenidos del análisis, no de forma absoluta pero si de forma relativa.

5.3.1 Concentración de un solo constituyente

En primer lugar se prepararon varias disoluciones de tinta roja en agua con distintos niveles de agua. Para trabajar con el coeficiente de atenuación se aplica el logaritmo neperiano a los espectros y se le cambia de signo. Este hecho obliga a recortar el espectro en los puntos que tome valores nulos puesto que $log(0) = \infty$ y el algoritmo PCA no admite valores infinitos.



Figura 5.24 (a) Espectros de transmisión y (b) absorbancia de las disoluciones de tinte rojo en agua con distintas concentraciones. A la izquierda una vez aplicado –log(R) y recortados los valores nulos. Las leyendas indican la concentración de la tinta roja en la disolución.

Se aplica PCA a una matriz formada por el los vectores que contienen distintas concentraciones de rojo. El algoritmo, como se ha visto en los casos anteriores, obtiene las distintas componentes (PC) que multiplica por el vector de entrada para hacer la transformación. Según el criterio de la varianza acumulada, para un 99% sólo es necesaria una PC. Esta PC coincide con el espectro del tinte rojo de nivel más alto y presenta un coeficiente de correlación de 0,999. En la siguiente tabla se indica el coeficiente de correlación de la PC1 con el espectro de cada disolución.

concentración	0.333	0.5	0.6	0.667	1
Coef. de correlación con PC1	0.989741	0.995331	0.997198	0.997896	0.999161



Figura 5.25 PCs resultado de aplicar PCA a la matriz que contiene distintas concentraciones de tinte rojo.

Si se mira a los scores que da el algoritmo, se ve en el mapa (Figura 5.26) que, una vez normalizados, muestran un escalado relacionado con la concentración (decreciente por filas) de tinte rojo. Para hacer un buen ajuste no es suficiente con normalizar respecto al valor más alto puesto que se basaría todo el ajuste en un único punto. Es más exacto hacer un ajuste polinómico de primer orden, puesto que la concentración es lineal.

$$p(x) = p_1 x + p_2 \tag{5.2}$$

 p_1 será el coeficiente que ofrezca un mejor ajuste según el criterio del mínimo error cuadrático. Para comprobar si el ajuste es bueno, se utiliza el coeficiente de correlación (r), el factor de mérito (r^2) por si el ajuste fuera negativo y el error cuadrático medio (RMSE), que viene a representar la media de las desviaciones entre la curva real y la ajustada (ver **capítulo 3**).

Se toman estos valores y se representan normalizados frente a la concentración conocida de tinta roja en agua. La concentración reconstruida con PCA coincide con la concentración real con un coeficiente de correlación de 0,98682.



Figura 5.26 (a) Scores obtenidos con PCA, las concentraciones están ordenadas por filas de la mayor a la menor; (b)modelo de regresión del ajuste, concentración estimada frente a concentración real.

Datos del ajuste:

Concentración Real Reconstrucción 0.3333 0.3283 0.5000 0.4712 0.6000 0.6071 0.6667 0.7306 1.0000 0.9628

Coeficiente de correlación r=0.986824; Factor de mérito del ajuste r²=0.973822; Error cuadrático medio del ajuste RMSE=0.035710

Por tanto se puede afirmar que la reconstrucción es correcta y, como se sugería en apartados anteriores, que en los scores va incluida información de la cantidad del constituyente.

5.3.2 Concentración de varios constituyentes

Puesto que en caso de un solo constituyente se consigue reconstruir la concentración de este en cada vector, se procede a hacer lo mismo pero con varios tintes. Se preparan disoluciones de tres tintes textiles distintos: rojo, azul y amarillo. Se introduce al algoritmo PCA los vectores de mezcla, corregidos, aplicando –log(espectro) y eliminando la región de ceros. En la siguiente tabla se muestran las concentraciones de los tres tintes en las mezclas.

Tinte\Mezcla	1	2	3	4	5	6	7	8	9	10
Rojo	0.3333	0.2857	0.2500	0.2222	0.1818	0.1538	0.1333	0.2353	0.3158	0.3810
Azul	0.3333	0.4286	0.5000	0.5556	0.4545	0.3846	0.3333	0.2941	0.2632	0.2381
Amarillo	0.3333	0.2857	0.2500	0.2222	0.3636	0.4615	0.5333	0.4706	0.4211	0.3810



Figura 5.27 Espectro de los tres tintes antes de la mezcla.



Figura 5.28 Espectro de las distintas mezcla. La leyenda indica la concentración del tinte rojo en la mezcla. Izquierda espectro. Derecha absorbancia (*-log(espectro*)).

Se aplica el algoritmo y se tienen los resultados. En primer lugar, sólo se necesita una componente. Esta componente tiene un coeficiente de correlación con el tinte azul de 0,9994, por tanto, presumiblemente viene determinada por la presencia de éste tinte en la mezcla. Como sucedía en el caso anterior y como se sugería en los experimentos con tejidos vivos, si esta PC es muy parecida al tinte azul, vendrá determinada por la concentración de este tinte en la mezcla.



Figura 5.29 Análisis de tintes textiles, PCs obtenidas del análisis.

Para comprobar si en efecto la PC1 determina la concentración en la mezcla del tinte azul, se calcula la correlación de los scores con la concentración del tinte azul en la mezcla.


Figura 5.30 Scores obtenidos con PCA. A la derecha se dibujan frente a la concentración real.

Datos del ajuste para la concentración de tinte azul:

Concentración Real Reconstrucción 0.3333 0.3380 0.4286 0.4217 0.5000 0.5008 0.5556 0.5484 0.4545 0.4583 0.3846 0.3965 0.3333 0.3401 0.2941 0.2762 0.2632 0.2664 0.2381 0.2390 Coeficiente de correlación r=0.99667; Factor de mérito del ajuste $r^2=0.99346$;

Factor de mérito del ajuste r²=0.99346; Error cuadrático medio del ajuste RMSE=0.008102

La reconstrucción para cualquier otro tinto con esta u otra componente da un valor muy inferior.

A modo de resumen, se puede medir la concentración de al menos un componente en una disolución. Los espectros de estas mezclas no tienen singularidades como sucedía con los tejidos biológicos, por tanto es más complicado que el algoritmo encuentre diferencias entre los distintos constituyentes. En el caso de estas mezclas de tintes, el espectro viene dominado por la respuesta (absorción) del tinte azul y por lo tanto oculta la característica espectral del resto.

5.3.3 Concentración de varios constituyentes de forma no ciega

En los casos anteriores, los datos de entrada eran los vectores de mezcla para ver si era capaz de detectar los constituyentes. Y solo se consiguió detectar uno de ellos. Ahora se va a forzar al algoritmo a que detecte los componentes metiéndole a la entrada los espectros de tinte rojo, azul y amarillo. Para hacer esto no se puede aplicar el logaritmo puesto que las diferencias entre colores se deben a las transiciones por 0 y por 1 y como ya se ha comentado, el logaritmo de cero toma un valor infinito y el algoritmo no funciona.

Las PC obtenidas son una composición de los espectros de los tres tintes. Al igual que sucedía con las muestras de tejidos, hay unas zonas que se corresponden con un constituyente y otras zonas que se corresponden con otro. Para detectar cuales son cada una se hacen

enventanados y se calcula la correlación de la PCn y el espectro de cada tinte en esa ventana (W). Sólo se consideran los que den un coeficiente lo bastante alto (r >0,97).



Figura 5.31 Enventanados que dan una correlación alta. Reactivo 1, 2 y 3 se corresponden con el tinte rojo, azul y Amarillo respectivamente.

Si en los casos anteriores las PC daban la concentración, ahora se quiere comprobar si solo un trozo de las PC, el trozo enventanado, también es capaz de darla. Pero en el caso anterior se tenía los coeficientes completos y la información de la concentración estaba en las mezclas. Ahora se tiene un enventanado, entonces lo que se hace es transformar las mezclas al nuevo sistema enventanado.

Si PCA aplica la siguiente relación:

$SCORE = mezclas \cdot coefficientes$

Para pasar las mezclas al nuevo espacio sólo hay que hacer la misma operación:

 $mezcla transformada = mezcla \cdot coefficientes_W$

Una vez hecha esta transformación, se mira el vector *mezcla transformada*, que son los nuevos scores. Pero sólo ha conseguido reconstruir la concentración del tinte azul.



Concentración componente 2 teórica reconstrucción 0.3333 0.3466 0.4286 0.4368 0.5000 0.4991 0.5556 0.5352 0.4545 0.4596 0.3846 0.3985 0.3333 0.3384 0.2941 0.2720 0.2632 0.2657 0.2381 0.2333 Coeficiente de correlación r=-0.992762;

Factor de mérito del ajuste $r^2=0.985577$; Error cuadrático medio del ajuste RMSE=0.011929

Y solo funciona para el tinte azul debido a que en el espectro de mezcla no aparecen signos del resto de tintes.

La conclusión que se obtiene es que el enventanado de una región de las PC funciona también para reconstruir la concentración, pero si en la mezcla no se encuentra característica del espectro a reconstruir, por mucho que la haya en la entrada de PCA, no la va a poder reconstruir. Al igual que sucedía en las muestras biológicas, que sólo detectaba ciertos constituyentes, aquí sólo detecta el tinte azul y por lo tanto es el único que puede reconstruir.

Se hace patente entonces que para detectar más de un constituyente, estos deben tener singularidades en su espectro y estas singularidades no deben estar en la zona de absorción de otro componente porque no se verán en la mezcla.

En el caso de la hemoglobina, su absorción dificulta ver el espectro de otros elementos del tejido biológico. En el caso de los tintes sucede lo mismo, el tinte azul no permite ver el espectro de los otros dos tintes (rojo y amarillo) puesto que sus singularidades espectrales se encuentran en la zona de máxima absorción del tinte azul. Sin embargo, se ha comprobado que se puede medir la concentración (la relación entre las concentraciones de una u otra zona) cuando la presencia de éste componente ha sido detectado en la matriz de coeficientes. Y además el enventanado es capaz de sacar la concentración siempre que, como se ha mencionado, en la mezcla se encuentre el espectro del constituyente deseado.

Conclusiones y líneas futuras

6.1 Conclusiones

El objetivo del proyecto es caracterizar tejidos biológicos mediante métodos de reflexión difusa y mediante un sistema sencillo y sobre todo, sin contacto de ningún tipo del sistema con la muestra. Con este fin, se ha ido avanzando en función de distintos resultados.

- En primer lugar se ha diseñado un sistema sencillo que permite medir la reflexión difusa en los tejidos. Se ha comprobado que es posible medir el espectro de una muestra con este sistema con suficiente sensibilidad para la caracterización y diferenciación de distintos tejidos biológicos. Se ha optimizado la posición de cada componente y finalmente se ha automatizado la adquisición de la medida y su análisis mediante un control centralizado en MATLAB. Esto ha permitido tener de forma precisa el espectro en cada punto de la muestra y hacer un mapa espectral de ésta.
- Una vez obtenidos los datos, se ha trabajado con un algoritmo, PCA, que permite distinguir las componentes principales en una matriz sin saber a priori cuales son (método ciego). Se ha conseguido obtener con éste algoritmo los constituyentes principales que forman el tejido (sangre y grasa) a través de su espectro y se han establecido las condiciones necesarias para su identificación. Para que esto sea posible, es necesario que los componentes tengan alguna singularidad (picos en el espectro) y en una región donde no quede oculta por la absorción de otros componentes. En base a estas singularidades, se identifica la sustancia a la que puede pertenecer y se compara con su espectro en absorción para ver si coinciden. Si es así, se ha identificado un componente. En los tejidos biológicos, en general, la mayor parte de compuestos son conocidos y con ellos sus espectros en absorción.
- El paso anterior consigue la identificación de las componentes principales, es una medida cualitativa. Una vez conseguido esto llega el momento de hacer medidas cuantitativas. El algoritmo obtiene los componentes principales en base a las diferencias en los espectros, y estas diferencias entre dos zonas con la misma composición, se basan en la concentración de los componentes que la forman. A partir de este razonamiento, y con las componentes principales obtenidas, se ha conseguido reconstruir la concentración de tinte en una mezcla de forma diferencial, esto es, se puede saber la diferencia de concentración de una muestra a otra en términos relativos. Este resultado debería ser extrapolable al caso de los tejidos.
- Finalmente, con la intención de mejorar el sistema de medida, se pretendía medir el ancho del spot de captación de luz con técnicas de reflexión difusa. De esta forma, se pretendía estimar (que no calcular) el área de medida de una forma valida con distintos sistemas de focalización. Para esto se utilizó el montaje de los apartados anteriores con alguna modificación.

El resultado de estas medidas concluye que el cálculo teórico del diámetro no se corresponde con las medidas realizadas. Puesto que la medida es la tomada, el cálculo teórico según la óptica geométrica no es aplicable puesto que la distribución de energía que captura la fibra no es uniforme con el radio. Lo que indican las medidas tomadas es que el área de captación es en realidad menor, es decir, se capta luz en un área más pequeña de lo calculado. Las medidas conservan su linealidad con la altura y parece un buen método para medir el área de captación real con cualquier sistema de captación de spot circular.

La ventaja de éste método es que permite medir el diámetro del spot de captura con distintos montajes y elementos focalizadores, sin necesidad de mantener unas condiciones determinadas, como distancia a la muestra, ángulo de inclinación, etc. puesto que la medida se basa en la energía recibida.

6.2 Líneas futuras

Los resultados obtenidos indican que si se puede determinar la composición de una muestra mediante la medida de la reflexión difusa y sin necesidad de contacto y que además se puede medir la concentración de los componentes en diferencias muestras de forma relativa.

- Como futuras líneas, en primer lugar habría que aplicar el método de medida de la concentración a muestras formadas por componentes de tejido biológico (sangre y grasa en principio). Puesto que es necesario conocer la concentración para comprobar que se consigue una reconstrucción correcta, hay que hacer disoluciones de estos componentes en concentraciones conocidas (llamados phantoms).
- Una vez comprobado lo anterior, habría que hacer de ésta medida de concentraciones algo no sólo comparativo, sino un método válido para determinar la concentración real de, por ejemplo, sangre en un tejido. Para esto fijando el método y las condiciones (sistema de medida fijo), se podría almacenar una medida del espectro con concentración conocida, que serviría para medir la concentración real. Otra opción es tener una muestra del espectro y medirla en el mismo momento que la muestra. De esta forma las condiciones de medida serán las mismas y el sistema no tiene que ser tan robusto. El problema de esto es la 'caducidad' de los tejidos biológicos y sus componentes.
- Si se consigue tomar medidas absolutas, el siguiente paso será medir y aislar los coeficientes de absorción y esparcimiento del material, μ_a y μ_s.
- Para la detección de las componentes, se podía hacer de forma automática, detectando lo picos en las componentes obtenidas y compararlos con los espectros conocidos ya almacenados.
- El sistema de medida se ha demostrado que funciona. Aun así se puede mejorar añadiéndole algún elemento focalizador (como lentes). Esto permitiría capturar en posiciones más cercanas y de forma más precisa ya que el área de captación actual es relativamente grande, unos 5mm de diámetro y no permite ver el espectro de regiones más pequeñas que este. Y en cuanto al montaje, el rango espectral de medida está dominado por el espectro de absorción de la hemoglobina. Si se extiende el rango de medida hasta la zona de 1600nm, donde tanto la grasa como

el agua tienen unos picos de absorción mayores, se podrán detectar mejor éstas componentes y medir sus concentraciones.

Todas estas modificaciones planteadas son para aplicaciones en tejido biológico, pero, como se vio en el caso de los tintes, son aplicables a otro tipo de muestras. Puesto que el montaje es sencillo, sería fácil de extender a otras áreas con modificaciones del tipo rango espectral, tamaño de las muestras etc.

Bibliografía

[1] Tuan Vo-Dhin [et al.]. 2003. Biomedical Photonics HANDBOOK. Oak Ridge, Tennesse: CRC

[2] Zhang Haihong, Zhang Shujuan, Wang Fenghua [et al]. 2010. "Research on the Firmness of Persimmon by Near_Infrared Diffuse Reflectance Spectroscopy". En *World Automation Congress*. 19-23 Septiembre. pp. 455-459.

[3] Ishizawa Hiroaki, Takeuchi Masahiko, Nishimatsu [et al]. 2002. "Diffuse Reflectance Near-Infrared Spectral Image measurement for the Field Monitoring of Agricultural Products". En *Tecnology Conference*. Anchorage, AK, USA, 21-23 Mayo 2002. pp. 3-6.

[4] Bo Cai, Huacai Chen, Yongjun Zhang [et al]. 2010. "Non-destructive Determination of the quality components in fresh pork meat using near diffuse reflectance spectroscopy". En *Symposium on Photonics and Optoelectronic.* ISBN 978-1-4244-4964-4.

[5] Bard M.P.L., Amelink A., Noordhoek Hegt V. [et al] 2005. "Measurement of Hypoxia-related Parameters in Bronchial Mucosa by Use of Optical Spectroscopy". En *Am J Respir Crit Care Med.* Vol. 171. pp 1178-1184. Disponible en: DOI: 10.1164/rccm.200501-046OC

[6] Jayanthi J.L., Nisha G.U., Manju S. [et al] 2011. "Diffuse reflectance spectroscopy: diagnostic accuracy of a non-invasive screening technique for early detection of malignant changes in the oral cavity". En *BMJ open accesible medical research*. Disponible en: DOI: 10.1136/bmjopen-2011-000071

[7] Zonios G., Dimou A. 2006. "Modeling diffuse reflectance from semi-infinite turbid media: application to the study of skin optical properties". En *Optics Express* Vol 14. No. 19. pp 8661-8674.

[8] Zonios G., Bykowski J., Kollias N. 2001. "Skin Melanin, Hemoglobin, and Light Scattering Properties can be Quantitatively Assessed *In Vivo* Using Diffuse Reflectance Spectroscopy". En *Journal of Investigative Dermatology*. Vol. 117, No. 6. pp 1452-1457.

[9] Arimoto H., Egawa M., Yamada Y. 2005. "Depth profile of diffuse reflectance near-infrared spectroscopy for measurament of water content in skin". En *Skin Research and Technology*. Vol. 11. pp. 27-35.

[10] Suresh Anand B.S., Sujatha N., Narayanamurthy V.B. [et al] 2011. "Clinical applications of diffuse reflectance spectroscopy in normal and prediabetic subjects-A pilot study". En *Defense Science Research Conference and Expo.* 3-5 Agosto 2011. pp 1-4.

[11] Nachabé R., Hendriks B.H.W., Van der Voort M. [et al] 2010. "Estimation of biological chromophores using diffuse optical spectroscopy: benefit of extending the UV-VIS wavelenght range to include 1000 to 1600 nm". En *Optics Express.* Vol. 18 No. 24 pp. 1432-1442.

[12] Prince S., Malarvizhi S. 2009. "Analysis of Diffuse Reflectance spectra of various skin conditions by Principal Component method". En *Biomedical and Pharmaceutical Engineering, International Conference.* 2-5 Diciembre. pp 1-4.

[13] Spectra [Sitio web]. *Oregon Medical Laser Center*. Disponible en: <u>http://omlc.ogi.edu/spectra/</u>

[14] Tuchin Valery. 2007. *Tissue Optics*. 2ª Edición. Bellingham, Washington, USA: SPIE Press.

[15] Bender J., Vishwanath K., Moore L. 2009. "A Robust Monte Carlo Model for the Extraction of Biological Absorption and Scattering in Vivo". *IIEEE Transactions on biomedical engineering.* Vol 56. No. 4. pp. 960-968.

[16] Flock S.T., Patterson M.S., Wilson B.C. 1989. "Monte Carlo Modeling of Light Propagation in Highly Scattering Tissues-I: Model Predictions and Comparison with Diffusion Theory". *IIEE Transactions on biomedical engineering*. Vol 36. No. 12. pp. 1162-1168.

[17] Flock S.T., Patterson M.S., Wilson B.C. 1989." Monte Carlo Modeling of Light Propagation in Highly Scattering Tissues-11: Comparison with Measurements in Phantoms". *IIEE Transactions on biomedical engineering.* Vol 36. No. 12. pp. 1169-1173.

[18] Kanick S.C., Robinson D. J., Sterenborg H.J. [et al] 2009. "Monte Carlo analysis of single fiber reflectance spectroscopy: photon path length and sampling depth". *Physics in medicine and biology*. Vol. 54, pp. 6991-7008. Disponible en: doi:10.1088/0031-9155/54/22/016

[19] Hattery D., Hattery B., Chernomordik V. [et al]. 2004. "Differential oblique angle spectroscopy of the oral epithelium". *Journal of Biomedical Optics.* Vol 9, No. 5. pp. 951-960.

[20] Van Gemert M. J. C., Welch A.J. 1995. *Optical-Thermal Response of Laser- Irradiated Tissue.* 2ª ed. Dordrecht, Heidelberg, London, New York: Springer. Disponible en: DOI 10.1007/978-90-481-8831-4

[21] Wagniéres G., Shangguan C., Zellweger M. [et al]. 1997. "An optical phantom with tissuelike properties in the visible for use in PDT and fluorescence spectroscopy". En *Physics in Medicine and Biology*. Vol 42, pp. 1415-1426.

[22] Wang J., Davis S.C., Srinivasan S. [et al]. 2008. "Spectral tomography with diffuse nearinfrared light: inclusión of broadband frequency domain spectral data". En *Journal of Biomed. Optics*. Vol. 13, No. 4. Disponible en: doi:10.1117/1.2952006I

[23] M. De Grand A., Lomnes S.J., Lee D.S. [et al]. 2006. "Tissue-Like Phantoms for Near Infrared Fluorescence Imaging System Assessment and the Training of Surgeons". En *Journal of Biomed. Optics.* Vol. 11, No. 1.

[24] Wang J., Jiang S., Pogue B.W. [et al]. 2008. "Enhanced NIR Spectral Reconstruction with Ti: Sapphire Laser based FD system". Dartmouth College, Hanover.

[25]	Espectrómetro	HR2000	CG-UV-NIR	de	Ocean	Optics	[Especificaciones]
http://www.oceanoptics.com/products/hr2000cg.asp							
[26]	Sensor	CCD	IL511B	de	2	SONY	[Especificaciones]
http://www.oceanoptics.com/technical/ILX511B(E).pdf							

[27] Motor y controladora de *Physikinstrumente*[Especificaciones]

http://www.physikinstrumente.com/en/products/prdetail.php?sortnr=900600

[28] Fuente de luz de *Oceanoptics* [Especificaciones] http://www.oceanoptics.com/technical/dh2000cal.pdf

[29] Fuente de luz de *Fiber Lite* [Especificaciones] http://www.dolan-jenner.com/Manuals/DC950_manual.pdf

[30]FibrasópticasdeOceanoptics[Especificaciones]http://www.oceanoptics.com/catalog/OceanOpticsFibersProbes.pdf

[31] Material de referencia de *Labsphere* [Especificaciones] <u>http://www.labsphere.com/uploads/datasheets/diffuse-reflectance-standards-product-sheet.pdf</u>

[32] Material de calibración de *Labsphere* [Especificaciones] <u>http://www.labsphere.com/uploads/datasheets/wavelength-standards-product-sheet.pdf</u>

[33] Semmlow John L. 2004. *Biosignal and Biomedical Image Processing. MATLAB-Based Applications.* CRC Press.

[34] Smith Lindsey I. 2002. A tutorial on Principal Components Analysis. Cornell University.[Consulta:28dejuniode2012].Disponibleen:www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf

[35] Abellanas, Pedro. 1965. *"Elementos de matemática"*. 11ª Edición. Madrid.

[36] Palmer G.M., Zhu C., Breslin T.M. [et al] 2003. "Comparison of Multiexcitation Fluorescence and Diffuse Reflectance Spectroscopy for the Diagnosis of Breast Cancer". En *IIEE Transactions on biomedical engineering.* Vol. 50. No. 11, pp. 1233-1242.

[37] Jayanth J.L., Nisha G.U., Manju S., [et al.]. 2011. "Diffuse reflectance spectroscopy: diagnostic accuracy of a non-invasive secreening technique for early detection of malignante changes in the oral cavity". En *BMJ Open*. Disponible en: doi: 10.1136/bmjopen-2011-000071.

[38] Skala M.C., Palmer G.M., Vrotsos K.M. [et al] 2007. "Comparison of a physical model and principal component analysis for the diagnosis of epithelial neoplasias in vivo using diffuse reflectance spectroscopy". En *Opt Express*. Vol 15, No. 12. pp: 7863-7875. Disponible en: doi:10.1364/OE.15.007863.

[39] Principal component analysis (PCA) on data - MATLAB. En *The Mathworks, Inc.* [Manual] Disponible en: www.mathworks.es/help/toolbox/stats/princomp.html

[39] Principal component analysis (PCA) on data - MATLAB. En *The Mathworks, Inc.* [Manual] Disponible en: <u>www.mathworks.es/help/toolbox/stats/princomp.html</u>

[40] "Covariance matrix – MATLAB". En *The Mathworks Inc.* [Manual] Disponible en: www.mathworks.es/help/techdoc/ref/cov.html

[41] "Fit curve or surface to data – MATLAB". En *The Mathworks Inc.* [Manual] Disponible en: www.mathworks.es/help/toolbox/curvefit/fit.html

[42] "Polynomial curve fitting – MATLAB". En *The Mathworks Inc.* [Manual] Disponible en: www.mathworks.es/help/techdoc/ref/polyfit.html

[43] "Polynomial evaluation – MATLAB". En *The Mathworks Inc.* [Manual] Disponible en: www.mathworks.es/help/techdoc/ref/polyval.html